Structural characterization of human spliceosome activation by cryo-EM

Dissertation

for the award of the degree
"Doctor of Philosophy" (Ph.D.)
of the Georg-August-Universität Göttingen

within the doctoral program
International Max Planck Research School for Molecular Biology
of the Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences (GGNB)

submitted by

Cole Townsend

from Ada, Oklahoma, USA

Göttingen 2021

Thesis Committee

Prof. Dr. Holger Stark
Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry

Prof. Dr. Henning Urlaub
Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry

Dr. Alexander Stein
Membrane Protein Biochemistry, Max Planck Institute for Biophysical Chemistry

Members of the Examination Board

1st Referee: Prof. Dr. Holger Stark
Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry

2nd Referee: Prof. Dr. Henning Urlaub
Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry

Dr. Alexander Stein
Membrane Protein Biochemistry, Max Planck Institute for Biophysical Chemistry

Prof. Dr. Reinhard Lührmann
Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry

Prof. Dr. Kai Tittmann
Department of Molecular Enzymology, Georg-August-Universität Göttingen

Dr. Alex Faesen
Biochemistry of Signal Dynamics, Max Planck Institute for Biophysical Chemistry

Date of oral examination: December 8th, 2021

Affidavit

I hereby declare that this dissertation with the title "Structural characterization of human spliceosome activation by cryo-EM" has been written independently and with no other aids or sources than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

Cole Townsend

Abstract

Eukaryotic genes are transcribed as precursor mRNA (pre-mRNA), in which coding regions (exons) are interrupted by non-coding regions (introns). Introns are excised and exons are ligated together in a two-step process termed "splicing" to produce mature mRNA. Both steps of pre-mRNA splicing are catalyzed by RNA within a complex molecular machine consisting of 5 small nuclear ribonucleoproteins (U1, U2, U4/U6.U5 snRNPs) and over 150 proteins: the spliceosome. For each intron to be spliced, the spliceosome is assembled *de novo* on its pre-mRNA substrate in a stepwise manner catalyzed by DExD/H-box ATPases, which remodel RNA-RNA and RNA-protein interactions of each complex. Spliceosome assembly begins with consecutive association of U1 and U2 snRNPs with the pre-mRNA, followed by the integration of the U4/U6.U5 tri-snRNP and subsequent release of U1 snRNP to form a pre-catalytic spliceosome, or B complex. The B complex lacks a catalytic center, and must therefore be extensively remodeled in a process called activation, to form an activated complex (B^{act} complex). Spliceosome activation constitutes the largest flux in the composition of the spliceosome, with over 30 proteins being dissociated and more than 25 being integrated to form the activated complex. The activation phase is catalyzed by the DExD/H-box ATPase BRR2, which unwinds the base-pairing between U4/U6 snRNAs, leading to the dissociation of U4 snRNP and numerous proteins, and allowing for the reorganization of U6 snRNA to form intramolecular base-pairing interactions as well as intermolecular base-pairs with U2 snRNA. The resulting U2/U6 RNA-RNA network results in a triple-helix of RNA that coordinates two divalent metal cations (Mg^{2+}) which are involved in splicing catalysis. While structural and biochemical insights have been gleaned about both the pre-catalytic and activated states of the spliceosome, it is unknown whether structurally and compositionally distinct intermediates during the activation phase may exist. Moreover, the role of proteins in facilitating the formation of the RNA-based catalytic center at the core of the spliceosome is unclear. Using a previously identified small molecule chemical inhibitor of pre-mRNA splicing, we isolated spliceosomes stalled at intermediate stages of activation. By employing single particle cryo-EM and image classifications, we identified two novel and distinct states of the spliceosome following the release of U4 snRNP but prior to the formation of an activated complex, which we termed pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$. The pre-B^{act} structures offer new insights into the massive exchange of proteins during activation as well as the role of these proteins in guiding the formation of the RNA-base catalytic network formed by base-pairing interactions between U2/U6 snRNAs.

Acknowledgment

It has been a privilege to carry out my doctoral work in the Department of Structural Dynamics over the past three and a half years and I would like to convey my thanks to the people who made it a joy to come to the lab every day.

First, I thank Holger Stark for providing an excellent environment in which to carry out scientific research. From the unparalleled microscope facilities to the freedom to pursue many fascinating projects, I am truly happy to have been in such a creative and exciting lab for the past few years. I am grateful for his guidance, future-oriented thinking, and availability to discuss scientific topics at length.

I am extremely thankful to Reinhard Lührmann for allowing me the chance to work on cutting-edge topics in spliceosome structural biology. From the lectures he held during my master's curriculum, I became fascinated by the spliceosome and the possibility of studying its assembly with cryo-EM. I feel fortunate to have had countless inspiring meetings with him, during which I learned something new at each one.

I would also like to sincerely thank the members of my Thesis Advisory Committee, Henning Urlaub and Alex Stein, for their support and advice. It was always very helpful to receive feedback about ongoing projects, and to see aspects of the work from different perspectives. To Henning Urlaub I am also particularly thankful for the mass spectrometry analyses he masterfully oversees for many spliceosome projects, including the one detailed in this thesis.

In addition, I would like to extend my gratitude to members of Reinhard Lührmann's lab, the Department of Cellular Biochemistry, for their longstanding efforts to characterize difficult samples that extend our knowledge of the spliceosome. In particular, I am grateful to Dmitry Agafonov for many long and entertaining discussions about technical and theoretical aspects of spliceosome biochemistry. Furthermore, I thank Dmitry Agafonov, Majety Leelaram, and Olex Dybkov for establishing the biochemical protocol to purify and analyze the samples described in this work. Additionally, my sincere thanks go to Cindy Will and Berthold Kastner for their collaboration and guidance in completing the manuscript related to this work, as well as for deep discussions that helped me view problems from new vantage points. I am also thankful to Klaus Hartmuth, for his enthusiasm about science in general and about RNA in particular.

Next, I thank Karl Bertram for supervising my master's thesis work, including theoretical and practical aspects of single particle cryo-EM. In addition, I am thankful to him for cryo-EM data collection of the sample detailed within in this work, and for entrusting the project to me early on in my master's thesis. It was an exciting time and I appreciate his always-upbeat attitude.

I would like to express my thanks to David Haselbach, who took the time to introduce me to cryo-EM during my lab rotation in the Department of Structural Dynamics. I was always impressed by how he took care to explain things with enthusiasm, all while driving forward his own projects and having time left over to prepare delicious cakes that the whole lab frequently enjoyed.

Very importantly, I thank my fellow officemates and friends, Zhenwei Zhang and Ka Man Yip, for contributing to a superb "feng shui" that included a constant flow of stimulating conversations about topics ranging from the day-to-day to the bizarre, as well as to a fun and energetic ecosystem in which we exchanged ideas about the finer points of cryo-EM on a daily basis. It has been an outstanding time sharing our fishtank office!

My gratitude also goes out to the original members of the (in)famous Office 113: Fabian Henneberg, Karl Bertram, Kashish Singh, and Lukas Schulte. Being a visitor in their very entertaining office during my lab rotation and master's thesis was a humorous and welcoming way to be introduced to the lab, providing me with a collection of stories I remember fondly.

I sincerely thank Dietmar Riedel for his constant willingness to share anecdotes and solutions to practical problems learned from his extensive experience in electron microscopy. It was a great time working with him, whether at the CM200, Talos, or Krios 3. Additionally, I thank Erik Schliep for practical advice in cryo-EM sample preparation and for many stimulating discussions, as well as Niels Fischer for ensuring smooth running of the microscopes.

Finally, I thank my family for their steadfast support, and especially my parents for traveling across the Atlantic multiple times to visit me. I would also like to thank my wife, Grace, for her constant companionship and encouragement. In summary, I am thankful to the many people who have enriched and made possible my time in Göttingen.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Central Dogma of Molecular Biology

In all domains of life, genetic information is encoded in DNA, which forms the set of molecular instructions required for an organism to function. DNA is transcribed into RNA, which can then be translated into proteins that carry out specific roles in the cell. Gene expression was long thought to be a unidirectional flow of information that could be described as DNA –> RNA –> protein. This basic schema describing the conversion of DNA instructions into functional proteins was formulated in the mid-20th century and formed what Francis Crick termed the "Central Dogma of Molecular Biology" [1]. Towards the latter half of the 20th century, it was discovered that large segments of transcribed RNA served no clear functional role in eukaryotes. Genetic and biochemical studies indicated that much more RNA was present than was strictly required for the synthesis of the encoded proteins. Surprisingly, it was found that in many organisms, large segments of noncoding RNA are removed from the newly transcribed precursor messenger RNA (pre-mRNA), before mature mRNA is translated into protein by the ribosome. Genes bearing this characteristic were termed "split genes," to describe how coding regions are initially separated from one another by noncoding regions [2]. Coding regions were termed "exons" and remain in the RNA template to form mRNA, whereas noncoding regions were termed "introns," and are excised prior to the formation of the mature mRNA [3]. The molecular process by which exons are ligated and introns are removed was termed "splicing" [4, 5].

1.2 Pre-mRNA splicing

All nuclear pre-mRNAs have characteristic sequence features: a 5′ and 3′ splice site (ss) and a branch site (BS) (Fig. 1.1). The 5′ss and 3′ss separate the intron from the 5′ and 3′ exons and consist of a conserved sequence of GU and AG in >95% of introns [6–8]. In a minor class of introns in higher eukaryotes, but not in simpler organisms such as *S. cerevisiae*, the 5′ and 3′ splice sites are demarcated by AU and AC [9, 10]. In all types of pre-mRNAs, an adenosine nucleotide at the branch site (BS-A) is located within the intron and is the nucleophile in the first step of splicing (Fig. 1.1). Nuclear pre-mRNA splicing is carried out in two S_N2-type transesterification reactions: Step 1, or branching; and Step 2, or exon ligation (Fig. 1.1). In Step 1, a 2'OH group of the BS-A carries out an S_N2 nucleophilic attack at the 5′ end of the intron, looping the intron into an "intron lariat" [11–14]. In Step 2, the 3′OH of the terminal 3′ nt of the 5′ exon carries out an S_N2 attack the 5′ end of the 3′ exon, ligating the two exons while leading to the dissociation of the intron [15, 16]. The ligated exons form a messenger RNA (mRNA) that can exported to the cytoplasm to be translated into polypeptides by the ribosome [17].

After a gene is transcribed into pre-mRNA, it can be cleaved into numerous potential mRNA products[18, 19]. For example, a single pre-mRNA sequence can be processed to generate thousands of protein isoforms in some cases; for example, one gene in *Drosophila melanogaster* can code for up to 38,000 mRNAs [20, 21]. This phenomenon is known as "alternative splicing" and explains how relatively simple genomes can give rise to complex proteomes in eukaryotes [22]. In humans, over 90% of genes containing more than one exon are alternatively spliced [23]. Alternative splicing in lower eukaryotes (e.g., *S. cerevisiae*) has been documented but does not occur to the same extent as in higher eukaryotes [24–27]. Thus, alternative splicing is a means by which cells in higher eukaryotes such as humans can produce hundreds of thousands of unique isoforms while containing a genome of approximately 20,000 protein-coding genes [28]. The principal patterns of alternative splicing are intron retention, exon skipping, variable splice site usage, and mutual exclusion of exons (Fig. 1.2) [22].

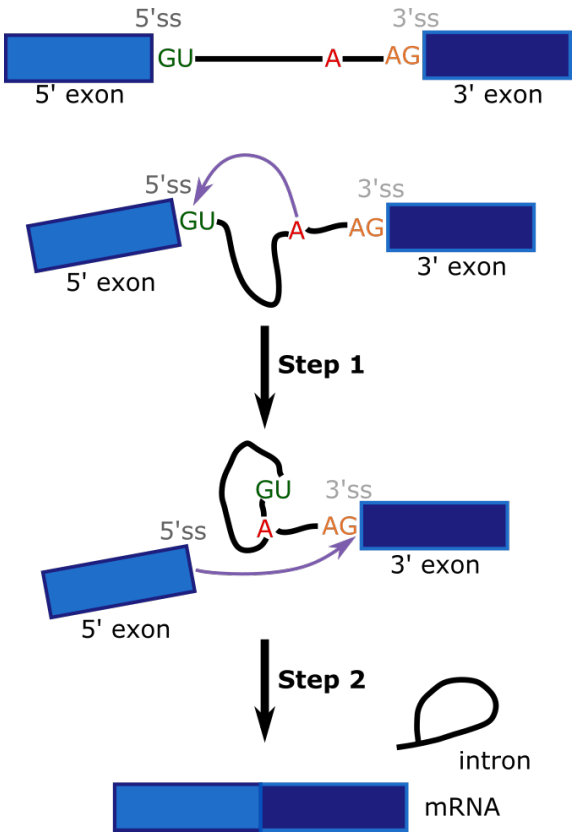

Fig. 1.1: Pre-mRNA splicing reactions. The top panel depicts the basic structure of a nuclear pre-mRNA, where the 5′ and 3′ exons are depicted as colored rectangles while the intron is shown as a solid line. Constitutive sequence elements at the 5′ss (GU) and 3′ss (AG) are colored green and orange. The BS-A nucleotide is colored red. In Step 1 of splicing, the 2'OH of the BS-A attacks the 5′ss in an S_N2-type reaction, resulting in the formation of an intron lariat. In Step 2 of splicing, the 3′ end of the 5′ exon attacks the 3′ss, ligating the two exons and allowing for the removal of the intron lariat.

Intron retention occurs when an intron is not excised from the pre-mRNA and thus becomes part of the final mRNA template. Exon skipping is the excision of exons from the pre-mRNA that could otherwise be translated into protein. Variable splice site usage is the recognition of a non-consensus intron-exon boundaries. Mutual exclusion of exons is removal of certain exons and the retention of corresponding ones, and vice versa. Alternative splicing patterns are regulated by multiple mechanisms, including binding of proteins to the pre-mRNA which negatively or positively effect the expression of the gene on which they act [22]. Alternative splicing is regulated in a tissue-specific manner, facilitating the development of highly differentiated cell types in multicellular organisms [30]. Aberrant alternative splicing is observed in many diseases, including several cancers, and can result from mutations within the splicing machinery or the gene to be expressed [31].

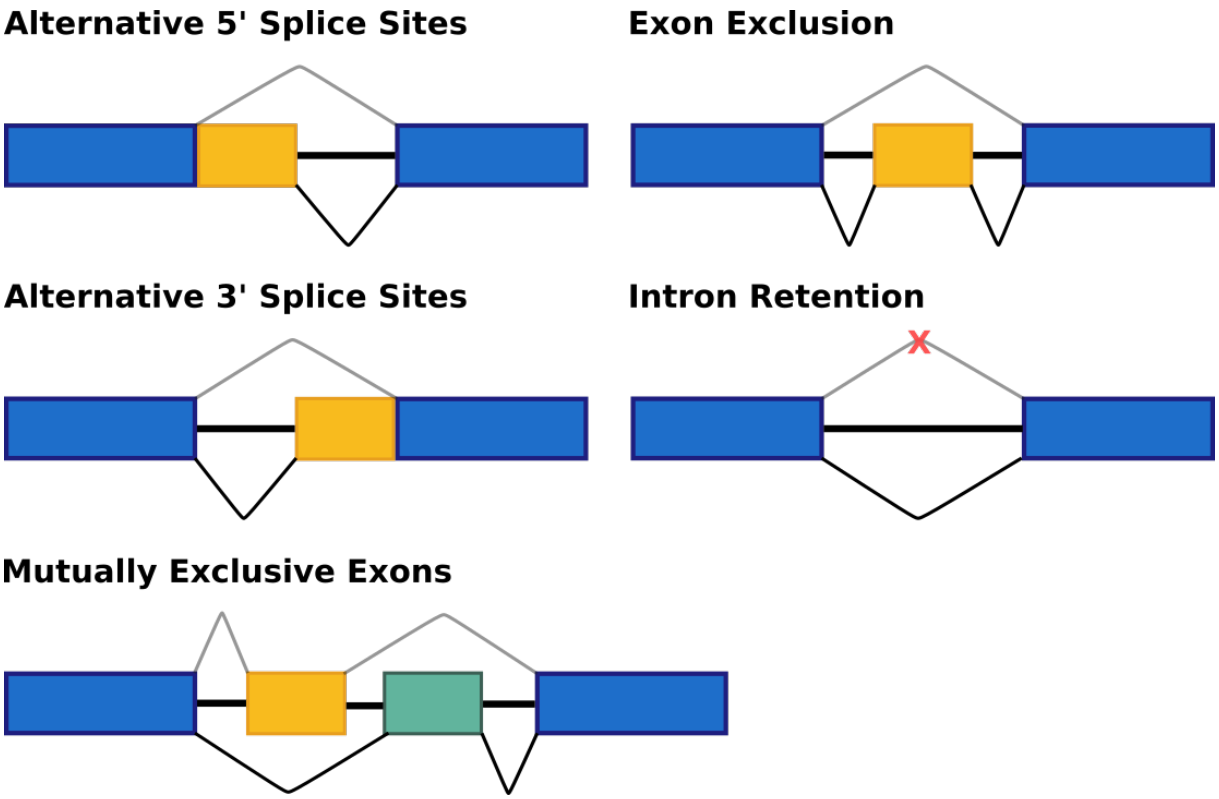

Fig. 1.2: Alternative splicing patterns. Exons are depicted as rectangles and introns as thin black lines. Blue rectangles indicate constitutively spliced exons whereas yellow or teal rectangles demonstrate exons that are included in the mRNA as a result of alternative splicing. Bent gray or black lines indicate the region that is excised. Adapted from [29].

Pre-mRNA splicing is heavily regulated by serine-arginine (SR) rich proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) that influence where splicing machinery associates to the pre-mRNA [32, 33]. SR proteins typically enhance while hnRNPs generally repress splicing of a specific transcript, although both classes of proteins can bind to the pre-mRNA at splicing enhancers or splicing silencers [34, 35]. Introns have been shown to be coated with hnRNPs, but whether hnRNPs always bind in a sequence-specific manner or as preformed complexes is not well understood [36–38].

Sequence elements of the pre-mRNA called enhancers and silencers provide binding sites for proteins that positively or negatively regulate splicing by influencing the recognition of splice sites and the assembly of the splicing machinery at specific positions [39]. An enhancer or silencer is found within the intron or the exon, and is thus termed an "intronic splicing enhancer/silencer" (ISE/ISS) or "exonic splicing enhancer/silencer" (ESE/ESS) [23].

1.3 Spliceosome

Both steps of pre-mRNA splicing are catalyzed by RNA within a multi-megadalton (up to ca. 4 MDa, with it longest axis spanning approximately 400 Å) ribonucleoprotein (RNP) complex comprised of five small nuclear RNAs (snRNAs) and over 150 proteins, the spliceosome. The spliceosome is a massive molecular machine consisting of 5 small nuclear RNPs (snRNPS): U1, U2, U4, U5, and U6 (Fig. 1.3). In addition, some higher eukaryotes have a minor spliceosome in which U1, U2, U4, and U6 snRNPs are replaced for U11, U12, U4atac, and U6atac snRNPs [9, 10]. All spliceosomal snRNPs contain several shared characteristics: uridine-rich small nuclear RNA (U snRNA), a heptameric beta-propeller Sm core, and associated proteins [40].

Biogenesis of spliceosomal snRNPs begins with the transcription of snRNA by RNA polymerase II (Pol II) for all snRNPs except U6 snRNA, which is transcribed by RNA polymerase III [42]. Pol II-transcribed snRNAs are modified by the addition of a 5′ 7-methylguanosine (m^7G) cap and 3′ polyadenylation before being exported to the cytoplasm through the nuclear pore complex [43, 44]. In contrast, U6 and U6atac snRNAs are not exported to the cytoplasm, but are instead assembled into functional snRNPs entirely within the nucleus [45, 46].

Pol II-transcribed snRNPs associate with a heptameric complex consisting of B/B', D1, D2, D3, E, F, G Sm proteins [47, 48]. This complex is called an Sm core and forms a doughnut-shaped beta-propeller that binds to snRNAs at specific RNA sequences consisting of PuAU4-6GPu [49]. Sm-bound snRNAs are subsequently bound by additional proteins and their m^7G cap is hypermethylated to form 2,2,7-tri-methyl guanosine (m^3G) and the 3′ end of each snRNA is shortened by ribonucleases [50–52]. The Sm core and m^3G cap together form a nuclear localization signal which is recognized by protein factors that aid in the import of the particle into the nucleus [53, 54]. Additional snRNP proteins are imported separately into the nucleus and join the Sm core-bound snRNA to form a mature snRNP [55, 56]. U6 and U6atac snRNAs are not bound by Sm proteins,

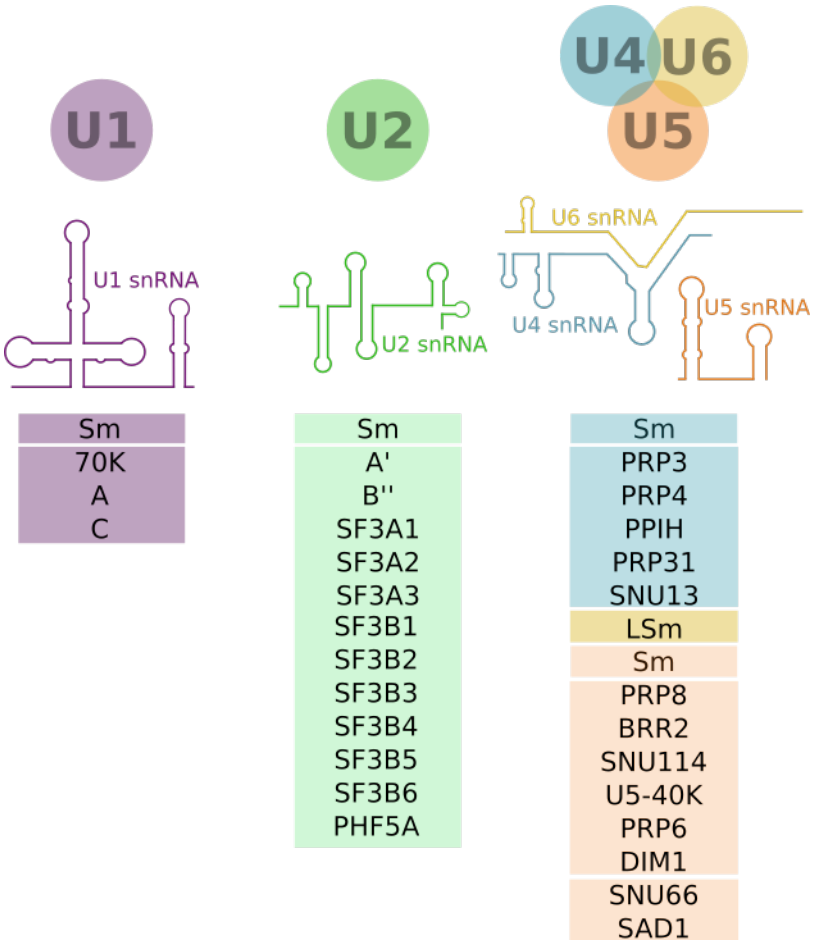

Fig. 1.3: Composition of spliceosomal snRNPs. The protein and snRNA inventory of spliceosomal snRNPs is depicted. Small nuclear ribonucleoproteins (snRNPS) consist of one or more small nuclear RNA (snRNA) bound by numerous proteins. U1 and U2 snRNPs exist as individual particles and associate to the spliceosome in a stepwise manner, yet the U4, U5, and U6 snRNPs are packaged into a triple-snRNP (U4/U6.U5 tri-snRNP) in which U4 and U6 snRNAs are base-paired and associated to U5 snRNP via protein-protein interactions. U2 snRNP contains the SF3a and SF3b complexes, which account for the majority of the molecular weight of the particle. Each snRNA is bound to an Sm core, with the exception of U6 snRNA, which binds a similar heptameric, doughnut-shaped LSm core. U1 snRNP is the smallest snRNP of the spliceosome and is responsible for recognition of the 5′ss to form the E complex. U2 snRNP recognizes and the BS-A and base-pairs with the pre-mRNA to form the U2/BS helix, in which the BS-A is flipped out. U2 and U6 snRNPs form the catalytic RNA at the core of the spliceosome, while U5 contributes to proper positioning of the pre-mRNA (5′ss). The snRNPs shown here are joined by many additional proteins during the assembly and function of the spliceosome, but snRNPs form the core structure of the complex. Adapted from [41].

but instead associate with Sm-like proteins that also form a heptameric ring structure: LSm2-8 (or "LSm proteins") [57].

Spliceosomal snRNAs undergo extensive post-transcriptional modifications, including pseudouridylation and methylation [58, 59]. In addition to its constitutive 5′ tri-methylguanosine cap, U2 snRNA contains at least 13 pseudouridines and 10 methyl groups [60]. In U2 snRNA, a pseudouridine near the branch site was shown to affect positioning of the BS-A

for Step 1 of splicing [61]. All snRNAs of the major spliceosome have been reported to undergo pseudouridylation and methylation [59].

Pre-mRNAs are also post-transcriptionally modified, which influences gene expression by altering the usage of splice sites. RNA modifications have been documented to control pre-mRNA splicing by altering how the spliceosome recognizes splice sites [62]. The enzyme double-stranded RNA-specific editase 1 (ADAR2), which converts adenosine to inosine, was shown to convert the sequence AA (adenosine-adenosine) to AI (adenosine-inosine) within the intronic region of its own pre-mRNA [62, 63]. AI can be substituted for the conserved AG dinucleotide of the 3′ss, and therefore caused the newly substituted inosine to demarcate an alternative 3′ss [62, 64]. This results in alternative splicing of the ADAR2 pre-mRNA, generating multiple isoforms of the enzyme [62, 65]. It was recently demonstrated that pre-mRNA is chemically modified *in vivo* in eukaryotes; adenosine methylation (m^6A) of the AG dinucleotide at the 3′ss prevents binding of the protein U2AF35, which is involved in recognition of the 3′ss and early assembly of the spliceosome [66]. Thus modification of the pre-mRNA can alter splice site choice, thereby regulating alternative splicing.

Assembly of the spliceosome and splicing of some pre-mRNAs occurs during transcription, and is thus a co-transcriptional process [67]. The spliceosome begins to assemble and splice pre-mRNAs as they are synthesized; that is, before they are released from Pol II and the rest of the transcription machinery [68]. U1 snRNP recognizes the 5′ss on the pre-mRNA, followed by binding of U2AF, U2 snRNP, and other splicing factors that identify the BS and the 3′ss [69, 70]. U1 snRNP also has been shown to play a role in regulating transcription by preventing premature transcription termination in a process termed "telescripting", in which U1 bound to the pre-mRNA blocks cleavage and polyadenylation at cryptic sites [71, 72]. Early insights into the structural basis of U1 snRNP's interactions with Pol II have recently been reported, outlining the molecular architecture of portions of the co-transcriptional machinery [73]. Transcription factors are also known to associate with the spliceosome, suggesting that parts of the transcription apparatus may also regulate constitutive and alternative splicing [74–76].

Splicing is also heavily regulated by post-translational modifications of core spliceosomal proteins or proteins that bind to enhancer/silencer regions in the pre-mRNA. Post-translational modifications of spliceosomal proteins regulate spliceosome function at all stages of assembly and function [77]. Protein phosphorylation is perhaps the best understood post-translation modification in the spliceosome, and it can promote or repress splicing, while also influencing which splice sites are utilized (i.e., regulation of alternative splicing) [78, 79]. The best documented class of post-translational modifications in the spliceosome are those of SR proteins [80], although the SF3B1 protein of the spliceosomal U2 snRNP is also known to undergo phosphorylation [81]. Acetylation has been proposed to be a regulator of pre-mRNA splicing, although understanding of the role of this modification is still in its early stages [82, 83]. Finally, ubiquitination was demonstrated to play a role in spliceosome assembly by binding to PRP8 (the core scaffold protein of the spliceosome) and preventing the premature dissociation of U4 snRNP [84].

1.3.1 Assembly

For each round of splicing, the spliceosome is assembled anew on its pre-mRNA substrate (Fig. 1.4), a process entailing major rearrangements in the molecular architecture and composition of the complex. The functional cycle of the spliceosome can be divided into four main phases: assembly, activation, splicing, and disassembly (Fig. 1.4). During the assembly phase, U1 snRNP base-pairs to the 5′ss to form the E complex (where "E" stands for "Early") [85, 86]. Base-pairing of U2 snRNA with a region of the intron containing the branch site (BS) forms the pre-spliceosome, or "A complex" [87, 88]. The interaction between U2 snRNA and the pre-mRNA forms a helix known as the U2/BS helix, which remains intact for both steps of splicing. The U4/U6.U5 tri-snRNP is integrated into the A complex to form the pre-B complex, which contains all 5 snRNPs of the spliceosome [89]. In the pre-B complex, a connection is established between the U2 snRNP and the U4/U6.U5 tri-snRNP via base-pairing of the 5′ end of U2 snRNA and the 3′ end of U6 snRNA, which forms a duplex, U2/U6 helix II. This short helix is bound by the U6 LSm

proteins and anchors the highly flexible U2 snRNP to the body of the tri-snRNP. Once established, helix II remains intact throughout the remainder of the splicing cycle.

At this stage the 5′ss of the intron, which is still base-paired to U1 snRNA, must be handed off to the U5 and U6 snRNAs, allowing for the formation new base-pairing interactions that anchor the 5′ss in place for Step 1 of splicing (Figs. 1.1 and 1.5) [90]. An invariant sequence of ACAGA(GA) in U6 snRNA base-pairs to the region immediately downstream of the 5′ss, forming the U6/5′ss helix. The hand-off of the 5′ss results in the dissociation of U1 snRNP and the formation of the B complex, or precatalytic spliceosome (Fig. 1.4).

1.3.2 Activation

The B complex, while containing all of the snRNPs and pre-mRNA elements needed for splicing to occur, does not yet contain an RNA-based catalytic center (Fig. 1.5). For this to occur, U4 snRNP must be dissociated from the complex, allowing for large-scale structural rearrangements to take place (Fig. 1.5 and 1.6).

U4 and U6 are extensively base-paired in the B complex and this interaction must be therefore disrupted for an activated complex to be formed. An ATP-dependent RNA helicase, BRR2, is involved in mediating the dissociation of U4 from U6 snRNA, freeing U6 snRNA to base-pair with itself and with U2 snRNA to form an internal stem-loop (ISL) and two short helices that are essential for pre-mRNA splicing (Figs. 1.7, 1.8, 1.9) [91–93].

This transition from a precatalytic to an activated spliceosome is termed the "activation" phase and constitutes the largest flux in protein and RNA composition of the entire splicing cycle, with approximately 60 proteins being exchanged in humans (32 proteins dissociated[1], 28 incorporated) [94, 95]. Many of the proteins that are exchanged are not part of snRNPs, but are rather individual proteins or pre-organized subcomplexes that are dissociated from or integrated to the spliceosome. Activation results in the B^{act}

[1] U4 Sm and U6 LSm complexes contain seven individual polypeptides each

complex – containing U2, U5, and U6 snRNPs, all of which remain in the spliceosome until splicing is complete – which contains a fully formed catalytic center (Figs. 1.7, 1.8, 1.9). However, the B^{act} is not catalytically active, and must be further remodelled to liberate the BS-A from the U2 snRNP-associated SF3a and SF3b proteins (Fig. 1.4) [96]. Structural rearrangements lead to the formation of a catalytically active complex, "B*," which is poised to carry out Step 1 of splicing (Fig. 1.4) [96, 97].

Rearrangements in base-pairing during activation also bring the BS-A and 5′ss progressively into much closer spatial proximity *en route* to intermolecular distances at which splicing can be catalyzed. In the B complex, the BS-A is separated from the 5′ss by approximately 150 Å [98]. In contrast, this distance is shortened to about 50 Å in the B^{act} complex [99], although the BS-A remains sequestered within the SF3B1 protein of the U2 snRNP and is therefore inhibited. The distance between the BS-A and 5′′ss is subsequently shortened to approximately 4 Å in the catalytically activated B* complex (as documented in *S. cerevisiae*, which has a highly similar catalytic center to that of humans) immediately prior to Step 1 of splicing [100].

1.3.3 Splicing

Step 1 of splicing results in the C complex, in which an intron loop or "lariat" is formed (Figs. 1.1 and 1.4) [101]. Exon 1 and 2 are positioned into closer proximity to prepare for exon ligation, or Step 2 of splicing, in a complex termed C*, which is catalytically active (Figs. 1.1 and 1.4) [15, 102].

1.3.4 Disassembly

Completion of Step 2 of splicing results in a post-catalytic spliceosome, or P complex, in which Exon 1 and 2 have been ligated, but the splicing machinery and intron are still present [103]. The P complex is dissociated into a messenger RNA particle (mRNP) and a complex comprised of U2, U5 and U6 snRNPs plus the intron lariat, which together

form an intron lariat spliceosome (ILS) [104]. The ILS is subsequently disassembled and recycled whereas the mRNA can be exported to the cytoplasm to be translated into protein by the ribosome.

1.3.5 Remodeling of the spliceosome by ATP-dependent DExD/H-box RNA helicases

Transitions throughout the splicing cycle are driven by eight ATP-dependent RNA helicases known as DExD/H-box ATPases [106], named after a conserved sequence of Asp-Glu-x-Asp/His ("D-E-x-D/H") residues [107]. Several DExD/H-box helicases utilize the energy from ATP hydrolysis to disrupt RNA basepairing (Fig. 1.5) [108]. DExD/H-box helicases have both remodeling and proofreading functions in pre-mRNA splicing and are required for assembly and disassembly of the spliceosome [109].

Two DEAD-box helicases, UAP56 and PRP5, are involved in the integration of the U2 snRNP onto the pre-mRNA, at which step a branch-site stem loop (BSL) of U2 snRNA is unwound to allow for base-pairing with the BS of the intron, forming the A complex [110–112].

Following the binding of the U4/U6.U5 tri-snRNP to form the pre-B complex, the DEAD-box helicase PRP28 facilitates dissociation of base-pairing between U1 snRNA and the 5′ss, permitting the destabilization of U1 snRNP as well as the transfer of the 5′ss to loop I of U5 snRNA [90, 113, 114]. This results in the pre-catalytic (B complex), which still does not yet contain a catalytic center.

The DExD/H-box helicase BRR2 is involved in unwinding the U4/U6 duplex, resulting in the dissociation of U4 snRNP and many associated proteins, while freeing the U6 snRNA to form the internal stem-loop (ISL) as well as helix Ia and helix Ib with U2 snRNA, resulting in a mature catalytic center of the activated spliceosome, or B^{act} (Fig. 1.6 and 1.7) [91, 115]. PRP2 is involved in proofreading the assembly of the RNP complex,

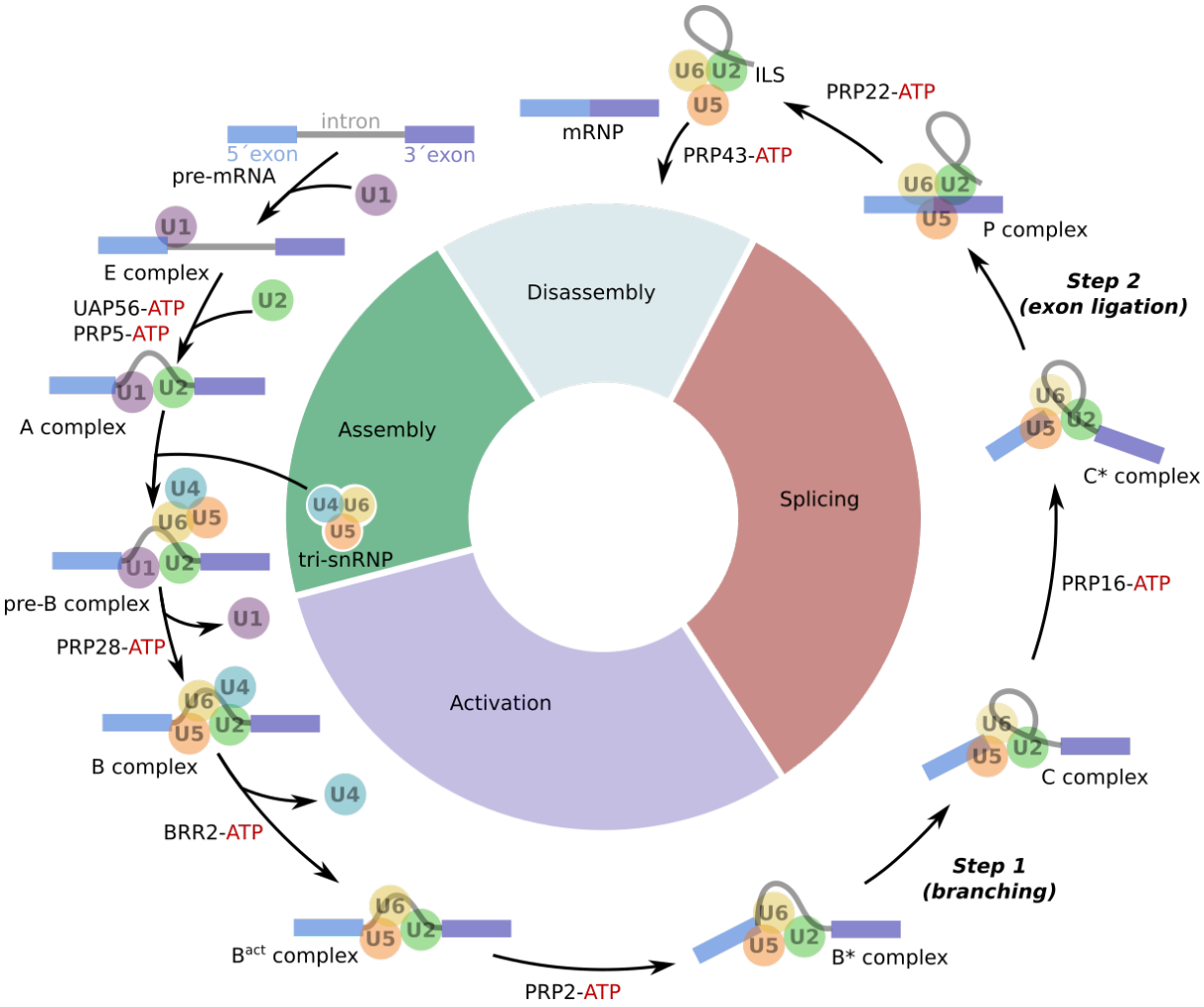

Fig. 1.4: Splicing cycle. The 4 phases of the spliceosome's cycle – assembly, activation, splicing, and disassembly – are shown by the colored wheel. ATP-dependent steps involving DExD/H-box helicases are indicated. ATP hydrolysis provides the "driving force" that catalyzes the forward progression of the splicing cycle. Some DExD/H-box helicases exert their function in the spliceosome by rearranging protein-RNA interactions instead of unwinding RNA helices. During the assembly phase, U1 snRNP recognizes the 5′ss, and U1 snRNA base-pairs to the pre-mRNA to form the E complex. Two DEAD-box helicases, UAP56 and PRP5, facilitate the binding of U2 snRNP to the pre-mRNA, where U2 snRNA recognizes and base-pairs with the region surrounding the BS-A to form the A complex, sequestering the BS-A within the HEAT repeats of the SF3B1 protein of the U2 snRNP. In the final stage of assembly, the U4/U6.U5 tri-snRNP is loosely docked to the A complex to generate the pre-B complex, containing all 5 spliceosomal snRNPs. Unwinding of base-pairing between the U1/5′ss interaction (assisted by the DEAD-box helicase PRP28) allows for the release of U1 snRNP and the transfer of the 5′ss to U5 snRNA. At this point, the invariant ACAGA(GA) sequence of U6 snRNA base-pairs to the intron downstream of the 5′ss, forming the U6/5′ss helix, resulting in the B complex, or pre-catalytic spliceosome. The activation phase begins with the unwinding of the U4/U6 snRNA duplex (mediated by DExD/H-box helicase BRR2), triggering the removal of U4 snRNP as well as a massive change in the RNA and protein architecture of the spliceosome to generate a B^{act} complex, or activated spliceosome, which contains a fully formed, RNA-based catalytic center. Activity of the DEAH-box helicase PRP2 is involved in the removal of the SF3a and SF3b complexes of the U2 snRNP, freeing the BS-A to carry out Step 1 (branching) of splicing in the B* complex. The resulting C complex is catalytically activated following the action of the DEAH-box helicase PRP16, generating the C* complex, which carries out Step 2 (exon ligation) of splicing. The postcatalytic spliceosome (P complex), containing the ligated exons as well as the intron lariat, is dissociated following the action of DEAH-box helicase PRP22, freeing the mRNA and associated proteins (mRNP) and intron lariat spliceosome (ILS). The ILS is broken down into individual snRNPs and the intron lariat in a process mediated by PRP43 (a DEAH-box helicase), completing the disassembly phase. The mRNA is exported to the cytoplasm to be translated by the ribosome. Adapted from [41, 105].

leading to rearrangements that bring about a catalytically activated spliceosome (B*) which carries out Step 1 of splicing, resulting in the C complex [116].

Following Step 1 of splicing, PRP16 is involved facilitating dissociation of the proteins YJU2 and CWC25, allowing for structural rearrangements of the reactants within the catalytic center [117, 118]. Following the action of PRP16, the catalytically activated Step 2 spliceosome (C* complex) is formed, poised to carry out exon ligation [15, 102].

PRP22 is implicated in dissociating the postcatalytic spliceosome (P complex) into an mRNP and intron lariat spliceosome (ILS), preparing the former to be exported from the nucleus and the latter to be further processed [119, 120]. Finally, PRP43 catalyzes the disassembly of the ILS, allowing U2/U6 and U5 snRNPs and associated components to be recycled [121]. Thus the major rearrangements in the RNA-RNA and RNA-protein interactions of the spliceosome are driven by DExD/H-box helicases, which use the energy of ATP hydrolysis to drive the splicing cycle forward, albeit through mechanisms that remain largely unclear.

Nearly all DExD/H-box proteins involved in splicing associate transiently with the spliceosome, although the U5 snRNP-associated helicase, BRR2, remains bound to the core spliceosomal machinery throughout the entire splicing cycle [122]. Moreover, BRR2 is repositioned throughout the entire splicing cycle [122]. Prior to activation, BRR2 docks onto U4 snRNA (see also 1.4.2). As BRR2 is responsible for the activation step, it must be tightly regulated to prevent premature unwinding of the U4/U6 duplex, involving many protein-protein interactions that either physically separate BRR2 from premature loading onto its substrate or that block the RNA binding channel of the complex [123–126]. Disruption of U4/U6 base-pairing triggers the most drastic change in RNA-RNA network of the spliceosome, paving the way for activation to occur (Fig. 1.5).

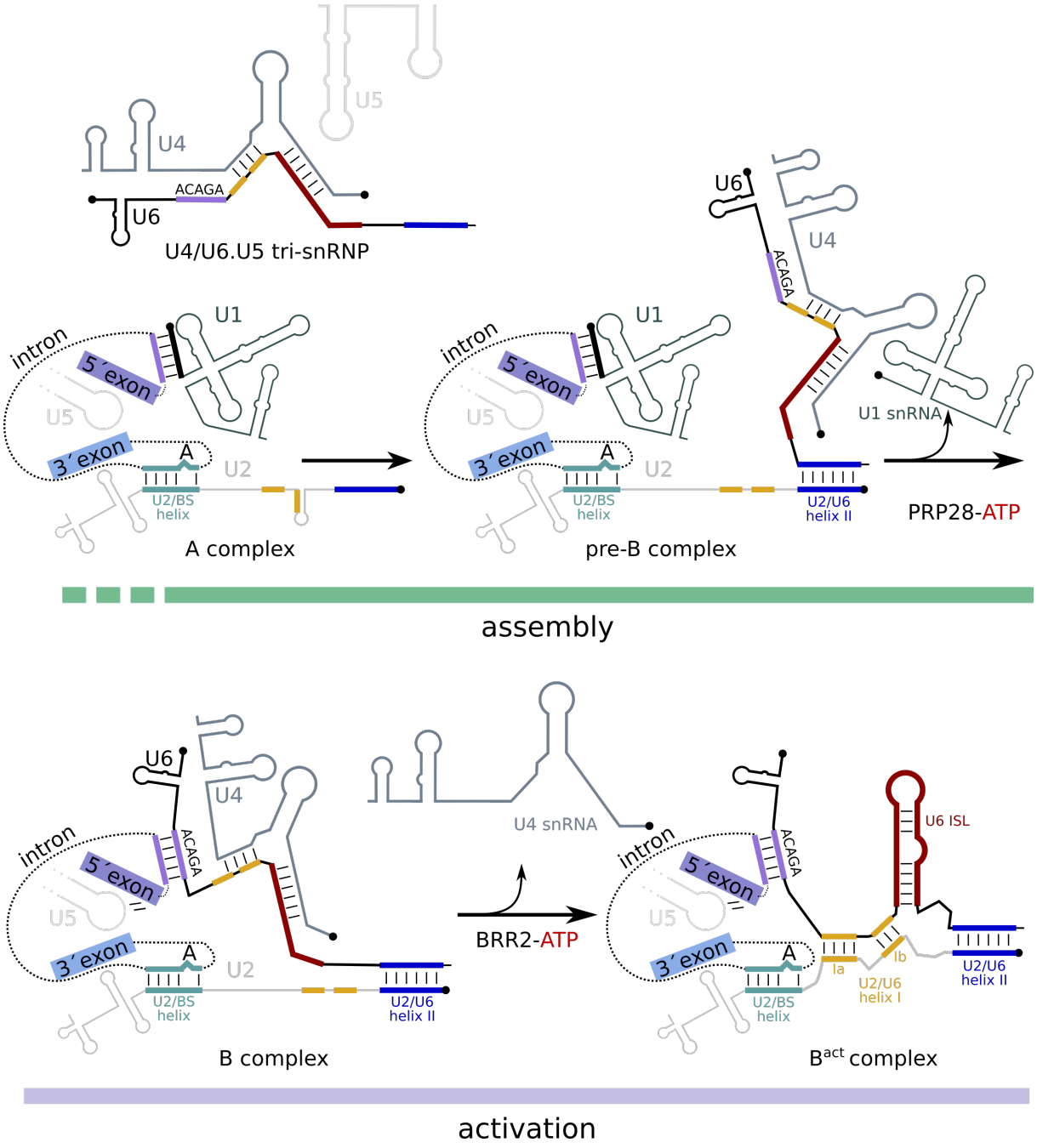

Fig. 1.5: RNA remodeling events during assembly and activation. The major events in the remodeling of the RNA-RNA interactions during assembly (upper panel) and activation (lower panel) are depicted. In the upper panel, the stippled green line indicates the continuation of assembly as the E complex, containing only the U1 snRNP, is not included. Base-pairing interactions that are either disrupted or formed are indicated as colored lines. The U2/BS helix, established in the A complex, is colored teal. The BS-A is labeled and depicted as a bulged nt in the U2/BS helix. U2/U6 helix II, first formed by the integration of the tri-snRNP to the A complex to form the pre-B complex, is dark blue. The sequences of the intron and the U6 snRNA that base-pair to form the U6/5′ss helix are indicated in light purple. New base-pairing interactions formed upon activation, which comprise the catalytic center, are indicated in dark yellow (U2/U6 helix I, consisting of helix Ia and Ib) and crimson (U6 ISL). Adapted from [41, 113].

1.4 Spliceosome activation

Catalytic transesterification steps 1 and 2 of pre-mRNA splicing are catalyzed by RNA (Figs. 1.1, 1.8, 1.9) [127]. More specifically, the catalytic center of the spliceosome is constructed from the base-pairing and tertiary interactions between U2 and U6 snRNAs. However, these interactions are not found in the B complex, as the U6 nucleotides relevant for formation of the catalytic center (A50-U74) are base-paired to U4 snRNA. Thus, the spliceosome's catalytic center is not a preformed entity, and is instead formed for the first time during activation of the spliceosome on its pre-mRNA substrate.

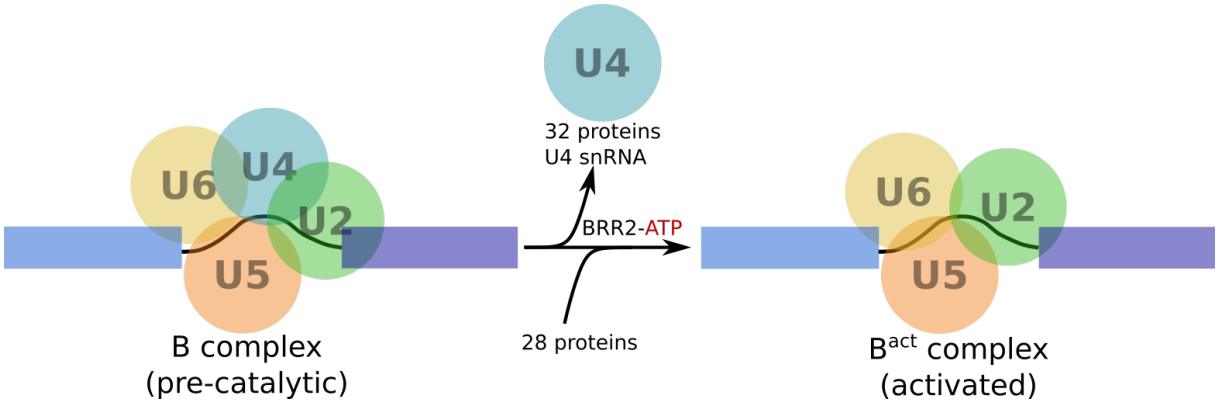

Fig. 1.6: Remodeling of the spliceosome during activation. The snRNP inventory of the B complex (pre-catalytic spliceosome) and B^{act} complex (activated spliceosome) are shown as colored circles assembled on the pre-mRNA substrate, with exons represented as rectangles and the intron depicted as a black line. BRR2 uses the energy of ATP hydrolysis to catalyze spliceosome activation, resulting in the dissociation of U4 snRNP, including approximately 32 proteins and U4 snRNA. Approximately 28 new proteins are subsequently integrated into the spliceosome, facilitating structural rearrangements and the formation of the catalytic center in B^{act}.

In order for activation to proceed, the U4/U6 duplex must be unwound by BRR2, allowing for the release of U4 snRNA and multiple proteins (Fig. 1.6). This frees U6 snRNA to establish new intramolecular base-pairing interactions as well as intermolecular base-pairing interactions with U2 snRNA. The result of this reorganization of RNA-RNA interactions is the formation of three conserved secondary structures that are the key components of the catalytic center: the U6 internal stem-loop (ISL), U2/U6 helix Ia, and U2/U6 helix Ib (Fig. 1.7) [91, 115]. The U6 nucleotides A53, G54 and C55 of helix Ib comprise the invariant sequence AGC, known as the catalytic triad. Together with the U6/5′ss, these secondary structures are positioned in a compact arrangement such that a catalytically

essential triple helix is formed (Fig. 1.7) [128]. The first two strands of the triple helix are contributed by the Watson-Crick base-pairing of U2/U6 helix Ib, consisting of the catalytic triad (U6 nts A53, G54, C55) and three nts of U2 (G20, C21, and U22) (Fig. 1.8). The third and final strand of the triple helix is contributed by the U6 nts G46, A47, and U74, which engage in Hoogsteen interactions with U6 nts G54, A53, and C55 (i.e., the catalytic triad), respectively [129]. U6 nts G46 and A47 form the 3′ end of the conserved ACAGA(GA) sequence of the U6/5′ss helix, while U74 is bulged out of the U6 ISL. U2/U6 helix Ia connects these secondary structures to allow for the triple helix to be formed, bringing the U6/5′ss, helix Ib, and the ISL into close proximity. Nucleotides of the triple helix coordinate divalent metal ions that stabilize reaction intermediates in Steps 1 and 2 of pre-mRNA splicing (Fig. 1.9) [128, 130].

Another type of intron – known as Group II – is excised by a two-step mechanism that is very similar to that found in the spliceosome [132]. Group II introns have not been detected in the human genome, but are found in the genomes of bacteria, archaea, and eukaryotic organelles (mitochondria and chloroplasts) of several eukaryotic genomes [133]. In addition, Group II introns and nuclear pre-mRNAs exhibit similar sequences at the 5′ and 3′ termini of their introns: GUGYC and AY in Group II, GU and AG in nuclear pre-mRNA introns (Fig. 1.8) [134]. Group II introns fold into catalytically active RNA structures that facilitate their own removal [106]. A Group II intron forms six conserved secondary structure domains: D1-D6 [135, 136]. At the core of a Group II intron lies an RNA triple-helix comprised of tertiary interactions between three of these domains: D2, D3, and D5 (Fig. 1.8) [132]. The nucleotides of D2 and D3 involved in this interaction are referred to as "J2/J3," as they are situated at the junction ("J") between these two domains [137]. D5 forms an internal stem-loop structure existing in all Group II introns [138].

As in the spliceosome, the Group II catalytic center contains an internal stem-loop (D5) and an invariant AGC catalytic triad. These structures interact with an upstream GA sequence (i.e., analogous to the U6 nts G46 and A47 of the ACAGA(GA) sequence in the spliceosome) to form a catalytically essential triple helix of RNA, which coordinates

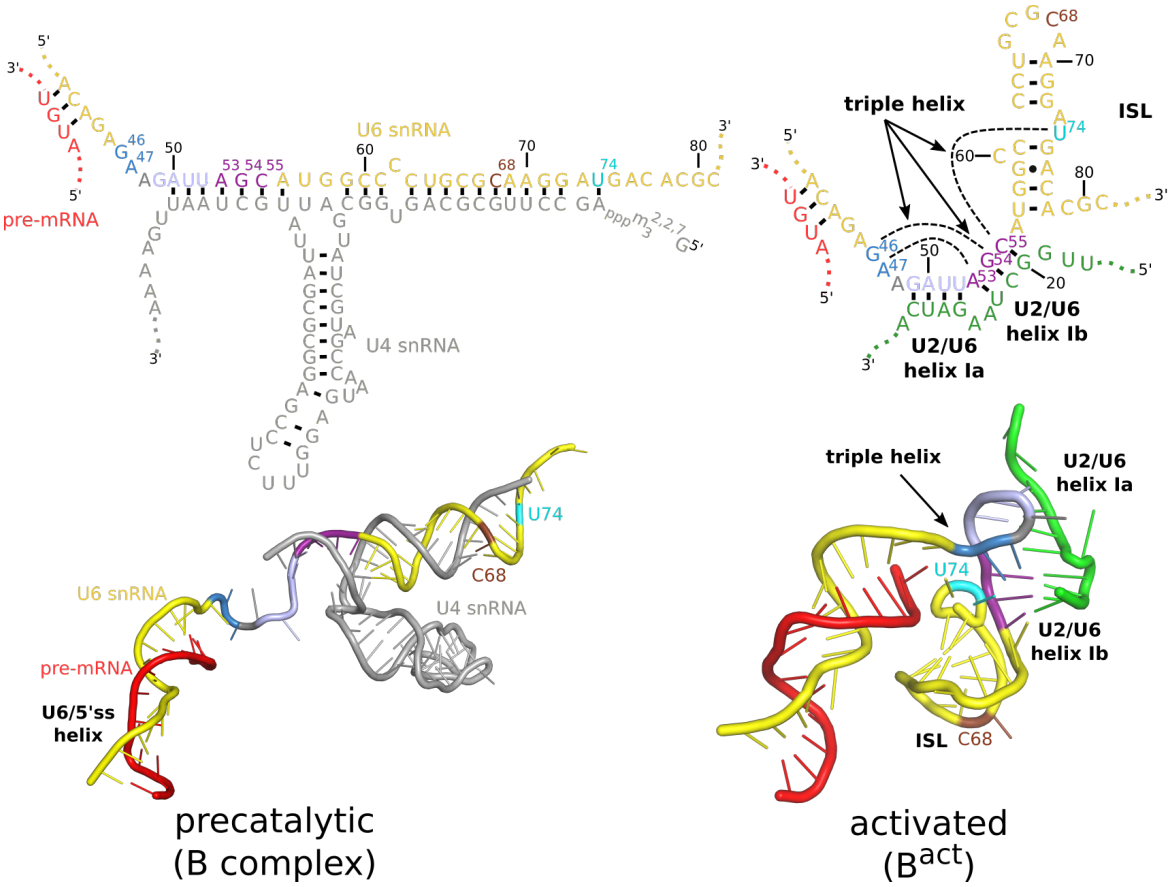

Fig. 1.7: Folding of catalytic RNA in the spliceosome. The U4/U6 and U6/5′ss RNA structures of the B complex are depicted as a 2D diagram (top-left panel) and 3D structure (bottom-left panel). U4 snRNA is depicted in gray, pre-mRNA in red, and U6 snRNA in yellow. U6 nucleotides that later contribute to the catalytic center are indicated in blue (G46, A47, which are the 3′ terminal nts of the U6 ACAGA(GA) sequence), purple (U6 nts of U2/U6 helix Ib; also called the catalytic triad: A53, G54, C55), and U74 (cyan). U6 nts (G49, A50, U51, U52) that later base-pair with U2 snRNA to form helix Ia are indicated in lavender. With the exception of G46 and A47, U6 nts that participate in catalytic center formation are base-paired to U4 snRNA. C68 is indicated in brown for orientation purposes. Following BRR2 unwinding of U4/U6 snRNA, U4 snRNA is released during activation. The mature catalytic center of the B^{act} is depicted as a 2D diagram (top-right panel) as well as a 3D structure (bottom-right panel). U2 snRNA is colored green. The nucleotides that participate in Hoogsteen interactions to form the triple helix are indicated by stippled lines in the 2D diagram (top-right panel) and by an arrow in the 3D depiction (bottom-right panel). Model for B^{act}: PDB 6FF4. Adapted from [131].

divalent metal cations required for splicing catalysis [128, 141, 142]. The 3D architecture of the RNA at the catalytic center in Group II introns is thus highly similar to that of the spliceosome (Fig. 1.9) [128]. Taken together, these lines of evidence suggest that Group II introns and nuclear pre-mRNAs may share a common evolutionary lineage, resulting in the conservation of the RNA core of the spliceosome [143].

Group II intron splicing occurs in an autocatalytic manner that does not require proteins [142]. Instead, a network of RNA-RNA interactions not found in the spliceosome stabilizes

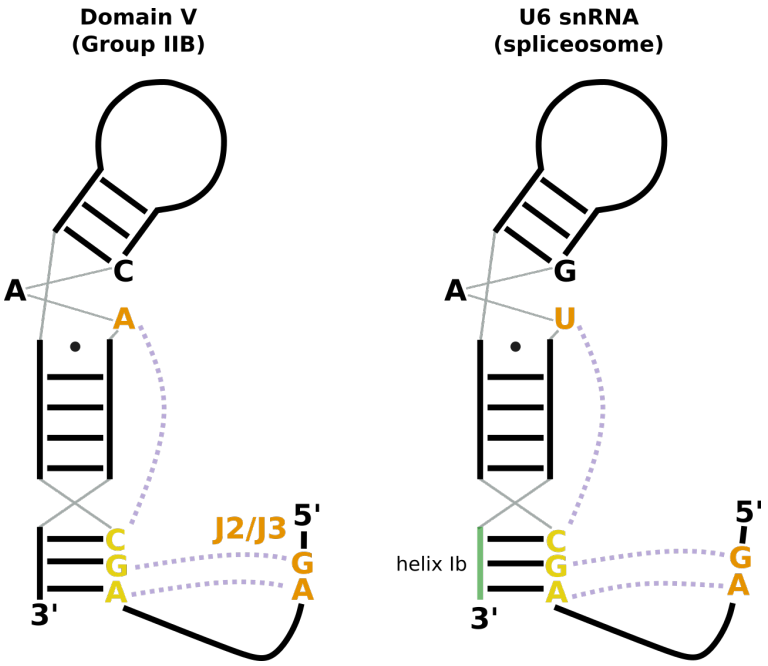

Fig. 1.8: Comparison of triple helix in Group IIB introns and spliceosomal snRNA. Nucleotides (orange) that form Hoogsteen interactions (indicated by stippled lavender line) with the ACG catalytic triad (yellow) are depicted. In Group IIB introns, an adenosine (A) nucleotide of Domain V is present instead of a uracil (U) as in the spliceosome. U2 nts of U2/U6 helix Ib (green) are indicated. The interactions of these nucleotides form a triple helix in Group IIB introns and spliceosomal snRNA. Adapted from [128, 136, 139, 140]

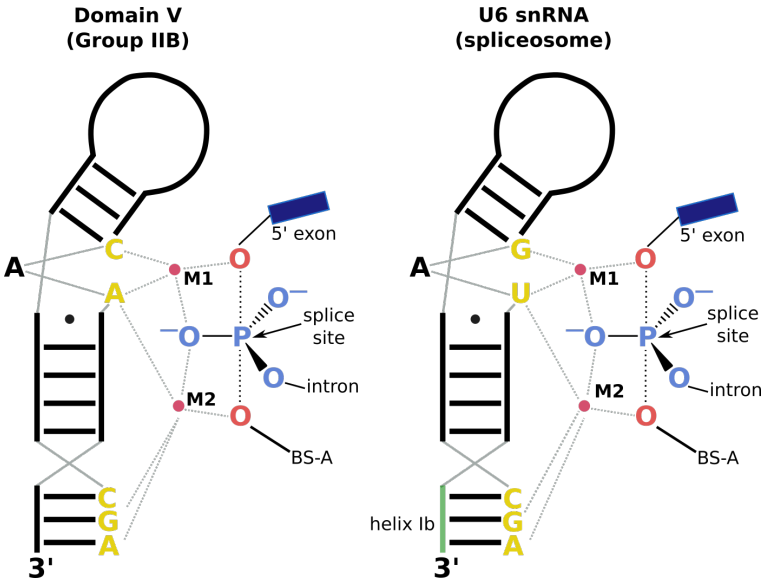

Fig. 1.9: Coordination of divalent metal ions by RNA in the catalytic center of Group IIB introns and the spliceosome. Divalent metal cations (magnesium, in the spliceosome) within the catalytic center are designated M1 and M2 (pink) and coordination thereof is indicated with stippled grey lines. Nucleotides of Domain V (Group IIB) and U6 ISL (spliceosome) that coordinate M1 and M2 are indicated in yellow. The phosphate group of the 5′ss is designated with a black arrow ("splice site"). Breakage of the covalent bond between the phosphate at the 5′ss nt and the 5′ exon, and consequent formation of a covalent bond between the 2'OH of the BS-A and the phosphate at the 5′ss nt of the 5′ end of the intron is indicated with a stippled black line. Adapted from [128, 136, 139, 140]

the Group II catalytic center [144]. In stark contrast to the spliceosome, the spliceosome's catalytic center is completely surrounded by numerous proteins. The largest of these is the most conserved and largest single component of the spliceosome by molecular weight, PRP8 [145, 146].

Structural studies of spliceosome complexes by cryo-EM

Spliceosome structure and function is highly conserved throughout eukaryotes [147], with almost all *S. cerevisiae* splicing components having homologs in humans [41]. In contrast to those of yeast, spliceosomes of higher eukaryotes such as humans are estimated to contain as many as 80 additional splicing factors, many of which are required for regulation of splicing in more complex genomes [41].

Single particle cryo electron microscopy (cryo-EM) has emerged as a standard method for reconstructing the 3D structures of dynamic macromolecular complexes, such as the spliceosome. Compared to X-ray crystallography or nuclear magnetic resonance (NMR), a major advantage of cryo-EM is that different conformations and species within a single biochemical purification can be extensively sorted through image classification techniques. Moreover, cryo-EM does not require crystallization, therefore removing a time-intensive search for crystallization conditions.

In cryo-EM, biological samples are vitrified within a thin (typically <500 Å) layer of buffer and are imaged within an electron microscope operated at liquid nitrogen temperature (-196°C) [148, 149]. Images in cryo-EM are formed as a result of the interaction of the electron beam with the sample and are collected on a direct electron detector [150]. The resulting image, called a micrograph, contains projections of the macromolecules within the sample. Due to optical imperfections (i.e., aberrations) in electron microscopes, micrographs are not perfect representations of the projected macromolecules, but are systematically blurred. This blurring is described by a point spread function (PSF), which quantitatively evaluates how distortions in the optical system influence the quality of the resulting micrograph [151]. Micrographs are recorded in real space, but determination of their PSF and corresponding ideal image is more easily achieved in Fourier space, in which

the PSF is described by the Contrast Transfer Function (CTF) [152]. The CTF takes into account the parameters of the optical system, such as spherical aberration, defocus, and electron wavelength [153]. The CTF describes the phases and amplitudes recovered in a micrograph as a function of spatial frequency (i.e., resolution). CTF correction is a method in which CTF parameters for every micrograph are estimated and applied to recover an approximation of a perfect image. Thus, it is crucial that CTF parameters are as accurate as possible to avoid introduction of artefacts in the data processing pipeline. CTF correction is typically carried out following motion correction and dose-weighting (see below).

Irradiation of thin vitreous films by the electron beam in cryo-EM results in beam-induced motion and radiation damage of the biological sample. Motion and radiation damage both limit the amount of information in a micrograph and thus reduce image quality [154, 155]. To mitigate the motion in an image, the thin film of vitreous buffer can be suspended in grids with small holes and/or on top of a thin layer of amorphous carbon film [149, 156]. Moreover, direct electron detectors with fast frame rates and are used to collect "movies" of the sample over a few seconds, allowing the individual frames to be aligned to one another to account for intra- and interframe motions during acquisition (i.e., "motion correction") [157].

Biological samples are sensitive to radiation damage, limiting the "dose" or fluence that can be used to record a micrograph (number of electrons per unit area per unit time). Radiation damage is unavoidable, but can be mitigated by low temperatures, which are thought to limit the diffusion of free radicals produced by ionizing radiation [152, 158]. High frame rates of most modern direct electron detectors used to collect micrographs also allow for the fractionation of the dose across many individual frames. Initial frames are exposed to the least radiation and therefore contain the highest resolution information [158]. Thus, weights can be assigned that take into account the relative contribution of high and low resolution signals within subframes of a micrograph, with earlier frames having higher weights for high resolution information than later frames. This approach is known as "dose weighting" [159, 160].

As a result of low electron fluence required in cryo-EM to mitigate radiation damage, micrographs have very low signal-to-noise ratios (SNR), complicating direct interpretation of single projection images of macromolecules. To overcome this low SNR, thousands of single particle images in similar orientations are averaged. Practically, this is achieved by first identifying individual individual macromolecules (i.e., "particles") within each micrograph and subsequently cropping them out as individual images. Particle images are subjected to reference-free alignment and classification, in which particles with similar orientations and properties are sorted iteratively into groups or "classes" [161].

Well aligned 2D classes often reveal ordered domain structures and often finer secondary structure features of the sample. In ideal cases, the random orientation of the sample in the vitreous buffer leads to many different orientations of particles, thus resulting in distinct "views" of the macromolecular complex. Poorly defined classes often represent damaged particles or debris that are excluded from further processing. Particle images belonging to well defined 2D classes are retained for structure determination.

The next step of image processing is to determine how 2D projection images of a macromolecule are related to the 3D object from which they are derived. Once particle images are annotated with these angular assignments (Euler angles), they are used to generate a 3D model *ab initio* from the data [162, 163]. Starting models generated in this manner are free from model bias and therefore do not introduce parameters which could lead to spurious results from an external reference. The *ab initio* starting model is then used as a reference to carry out 3D classifications, in which the 3D volume is reprojected and matched to single particle images ("projection matching") belonging to distinct classes [164]. Angular assignments of particle images are iteratively refined to result in higher resolution reconstructions, as the angular assignment becomes progressively more accurate.

3D classification is a powerful tool for sorting out conformational differences in biochemical samples in which multiple distinct macromolecular complexes may exist. In some cases, 3D classification is also used to focus only on smaller regions of a large complex that may

be flexible in relation to a large, well aligned portion of a macromolecular complex. If left unclassified, such regions are often poorly resolved due to structural heterogeneity in the overall reconstruction of the complex. Several approaches have been adopted to address this challenge, including *in silico* subtraction of stable regions directly from particle images followed by classification of the region of interest [165]. One of these approaches, known as multi-body refinement, involves signal subtraction and 3D refinement (see below) of two or more rigid bodies within a macromolecular assembly [166], allowing simultaneous refinement of several regions of a single complex. In addition, structural heterogeneity in the sample has been described by 3D principal component analysis (PCA), yielding energy landscapes that provide quantitative representations of distinct sub-populations of a macromolecular complex [99, 167].

Following computational sorting to obtain more homogeneous populations of 3D reconstructions, a map is refined against a 3D reference. All particle images contributing to a map chosen for refinement are split into 2 random subsets and independently reconstructed. The resolution of an EM map is estimated by comparing the two subsets that are divided during map refinement [168]. The overall resolution is estimated to be the spatial resolution at which the correlation of the maps is 0.143 [169]. This method is called Fourier Shell Correlation (FSC) and is the primary method of estimating resolution of EM maps [170]. FSC estimations report a single resolution value, but resolution in EM maps is often highly variable. Better aligned or more stable portions of a macromolecular complex are generally better resolved, while poorly aligned or flexible regions are less well resolved. Local resolution estimates differ substantially from global resolution estimates in many cases [171]. In some cases, the local resolution can be improved with focused classifications or multi-body refinements.

After reconstructions of the macromolecular complexes are determined and classified using 3D classification and refinement, molecular models of the complexes can be built. It is through model building and interpretation of the resulting structures that biological insights from the cryo-EM reconstructions begin to emerge. In most cases, model building begins by consulting the biochemical composition of a sample, which is either known in ad-

vance (e.g., for *in vitro* reconstitutions) or determined experimentally (e.g., endogenously purified complexes). Once the biochemical composition of the sample is well characterized, this information can be used to begin building a model of the 3D reconstructions from single particle cryo-EM. Model building is carried out iteratively, often beginning by docking the largest subunits into well defined density elements followed by docking of progressively smaller units. Experimentally determined or predicted structures of individual components often differ in conformation from those within a macromolecular complex, necessitating adjustments of inter-domain or flexible regions of reference structures within the 3D reconstruction [172]. For large assemblies containing many (e.g., $\geq$10) proteins, such as the spliceosome, it is helpful to have cross-linking mass spectrometry data to guide modeling, as protein-protein and protein-RNA crosslinks assist in determining which components are likely to interact within a macromolecular complex [173, 174]. In practice, model building is often a very time consuming step of structure determination, especially in parts of the map where regions of multiple components converge. After all components have been identified and docked within the density maps, the model is subjected to a refinement (or fitting) step to minimize geometric errors and atomic clashes such that the model is accommodated within the density, but is not overfitted [175, 176].

Cryo-EM is the method of choice for structural biology studies of spliceosomes, as it allows for classification of conformational and compositional heterogeneity of highly complex macromolecules. Structures for many individual components of the spliceosome have previously been determined by X-ray crystallography or NMR, including the pioneering achievements of structures for U1 snRNP and the core spliceosomal proteins PRP8 and BRR2 [177–181]. However, crystal structures of functional splicing complexes proved elusive due to the dynamic nature of the spliceosome. The "Resolution Revolution", resulting from major advancements in software and hardware used for single particle cryo-EM, has made it possible to determine the molecular architecture of highly dynamic complexes to sub-nanometer resolution [182], and it is reasonable to expect that further improvements will lead to resolutions better than 2 Å even for large assemblies[183, 184]. Aided by a wealth of previous biochemical studies as well as protein-protein cross-linking

mass spectrometry (XL-MS) data, cryo-EM structures for all major yeast and human intermediates (with the exception of the human E, A, and B* complexes) have been determined in the last 6 years, uncovering deep insights into assembly, activation, and function of the spliceosome [122, 185, 186].

1.4.1 Architecture of PRP8 and overview of its dynamics

PRP8 is a 220 kDa, multidomain protein that forms the central scaffold of the spliceosome, is an integral component of U5 snRNP, and joins the complex as part of the U4/U6.U5 tri-snRNP during conversion of the pre-spliceosomal A complex to the pre-catalytic B complex (Fig. 1.4) [187]. PRP8 consists of a large N-terminal domain (NTD) connected to a helical bundle (HB) via a NTD Linker (NTDL) region, a reverse transcriptase-like/endonuclease-like (RT/En) domain, an RNaseH-like (RH) domain, and a Jab1 domain connected by short linker regions (Fig. 1.10) [122]. PRP8^{RH} contains an RNase H-like active site but is not known to unwind RNA duplexes [188]. The RT/En domain contains several smaller domains, including the Thumb/X, a Switch-Loop (SWL), Linker, α-finger [122]. This multidomain organization affords the spliceosome a massive amount of flexibility in organizing the molecular architecture of the complex.

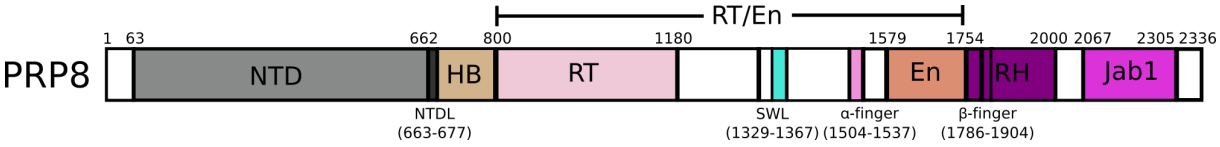

Fig. 1.10: Domain architecture of PRP8. The multidomain architecture of PRP8 is depicted. PRP8$^{RT/EN}$ is indicated, containing the PRP8^{SWL} and PRP8$^{\alpha\text{-finger}}$. PRP8^{NTD}, PRP8^{NTDL}, PRP8^{HB}, and PRP8$^{RT/En}$ surround the active site, while the C-terminal PRP8^{RH} and PRP8^{Jab1} play regulatory roles.

Early crosslinking and structural investigations determined that pre-mRNA as well as spliceosomal snRNAs associate with PRP8 and that several important secondary structures of U2 and U6 snRNAs, later found to be the catalytic center, were likely situated within a large cavity of PRP8 [145, 180, 189]. This large cavity is formed by the arrangement of the NTD, RT/En and the HB domains of PRP8, which form a compact region in which the active site U2/U6 is situated [180].

The molecular architecture of PRP8 is very comparable to the 3D structure of Group II introns, suggesting that the environment surrounding the catalytic RNA structures of the spliceosome and Group II introns is conserved, despite the fact that one compartment is composed of RNA and the other of protein and RNA, respectively [180]. This finding suggested that the protein "mold" of the spliceosome may have emerged throughout evolution to replace the RNA-only scaffold of Group II introns or a similar common ancestor [134]. However, PRP8 alone is not likely to be sufficient for the formation and stabilization of the active site, which is surrounded by additional proteins in the activated complex (see 1.4.5).

Several domains of PRP8 are highly dynamic throughout the splicing cycle, regulating the progression of spliceosome assembly and activation. The C-terminal region of PRP8 is involved in regulating the DExD/H-box helicase BRR2 to prevent premature unwinding of the U4/U6 snRNA duplex, which could lead to premature activation [124, 190, 191]. The RNase H domain was found to have multiple regulatory functions, including regulation of BRR2 activity as well as control of splicing catalysis, and has been reported to toggle between different conformations to regulate splicing fidelity [188, 192–194]. Cryo-EM structures documented the dynamic nature of PRP8, showing that the RNase H and Jab domains are highly mobile and occupy distinctly different positions in nearly every stage of the splicing cycle [195–199]. The switch-loop (SWL) domain of PRP8 ($PRP8^{SWL}$) is repositioned during spliceosome activation, and is returned to its original position after splicing is complete [200].

The overall conformation of PRP8 gradually transitions from an open to a closed conformation as the spliceosome is assembled (Fig. 1.11). In the tri-snRNP as well as the pre-B complex, PRP8 is in an open conformation, with the En domain separated from the NTD domain. Following the formation of the pre-catalytic B complex, the RT/En domain pivots on the NTD such that the En repositioned closer to NTD, yet remains in an overall open conformation. These movements are accompanied by slight rotation of the HB domain, which is connected to the NTD via the NTDL.

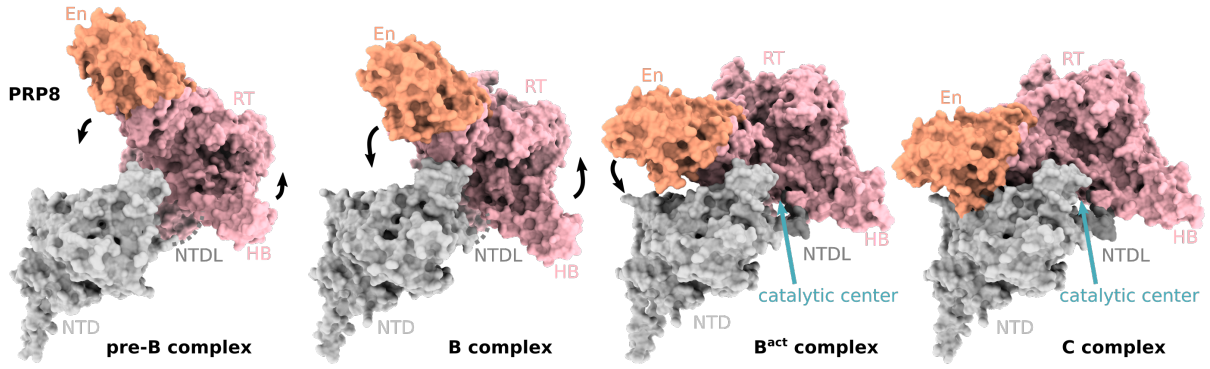

Fig. 1.11: PRP8 conformational change. PRP8^{NTD} (light grey), PRP8^{HB} (light pink), PRP8^{RT} (light pink), and PRP8^{En} domains (salmon) are depicted in their conformations as found in the pre-B, B, B^{act}, and C complexes. The PRP8^{NTDL} (dark grey), which surrounds catalytic center (dark teal) is indicated by a stippled line (unstructured) or a surface representation (structured). In the tri-snRNP (not shown) and pre-B complex, PRP8 is in an open conformation. Release of U1 snRNP and PRP28 results in a slight pivoting of the PRP8$^{RT/En}$, narrowing the gap between PRP8^{NTD} and PRP8^{En} to generate a partially open conformation in the B complex. Following release of the B-specific proteins, PRP8^{En} is clamped onto PRP8^{NTD} to form a closed conformation of PRP8. The catalytic center is nested within the cavity formed by the closed conformation of PRP8 in B^{act}. Following Step 1 of splicing, the remaining proteins wedged between PRP8^{En} and PRP8^{NTD} are destabilized, allowing for the distance between the domains to be even more restricted in the C complex. PDBs 6QX9 (pre-B complex), 6AHD (B complex) complex, 5Z58 (B^{act} complex), and 5YZG (C complex) shown.

The largest conformational change of PRP8 occurs in the transition from the B to the B^{act} complex during activation: the En domain clamps down onto the NTD, while the RT domain is concomitantly shifted upwards relative to the NTD (Fig. 1.11). This closed conformation thus forms a compact "chamber" that harbors the mature catalytic center of the B^{act}. Following Step 1 of splicing (i.e., in the C complex) proteins that reside within the gap between the NTD and En domains have been displaced, allowing the En domain to clamp even more tightly onto the NTD. The fully closed conformation of PRP8 is retained through Step 2 of splicing and is intact in the post-catalytic spliceosome (P complex). PRP8 is thus a dynamic scaffold that facilitates the organization of the entire spliceosome, in particular the active site U2/U6 RNA network.

1.4.2 Cryo-EM structure of the pre-catalytic spliceosome (B complex)

Comparison of the cryo-EM structures of the pre-B and B complexes reveals drastic rearrangements in the molecular architecture of the spliceosome leading up to activation

[196, 201]. In both complexes, the U5 snRNP proteins PRP8, BRR2, SNU114, U5 Sm, and U5-40K as well as the U5 snRNA form the rigid core of the complex. U4/U6 components are positioned upon the U5 snRNP scaffold, held in place by the U5 protein PRP6 as well as the tri-snRNP proteins DIM1, SNU66, and SAD1 [195].

U2 snRNP is loosely associated to the complex by U2/U6 helix II. In the pre-B complex, all 5 snRNPs are still present and the 5′ss is not yet base-paired to U6 and U5 snRNAs to form the U6/5′ss helix and the exon-U5 interaction found in the activated spliceosome. Moreover, BRR2 is kept in its tri-snRNP orientation, separated from its U4/U6 snRNA substrate. Unwinding of the U1/5′ss interaction allows for the release of DExD-box helicase PRP28 and for the integration of the 5′ss and U6 ACAGA sequence into the core of the spliceosome. The release of PRP28 and U1 snRNP likely allows for the large-scale repositioning of BRR2 onto U4 snRNA, where it is primed to unwind the U4/U6 base-pairing that keeps this interaction intact. Repositioning of BRR2 is also accompanied by the translocation of U4 Sm, including the 3′ domain of U4 snRNA, as well as the repositioning of the entire U2 snRNP and U4/U6 di-snRNP. The tri-snRNP proteins RBM42 and PRPF4B are also dissociated at this stage. The remaining building blocks of the complex are thus the U2 snRNP and almost all components from a heavily remodeled U4/U6.U5 tri-snRNP.

These major rearrangements make way for the integration of a set of proteins that were first identified as co-purifiying with the B complex, the B-specific proteins, which consist of PRP38, SNU23, MFAP1, UBL5, FBP21, SMU1, RED, WBP11 and PQBP1 [202]. Intriguingly, only PRP38, SNU23, MFAP1 and UBL5 have known homologs in *S. cerevisiae* (Prp38, Snu23, and Spp381, Hub1), suggesting that the remaining five proteins may have evolved as regulatory factors to handle the more complex genomes of higher eukaryotes [122]. In the B complex, the $PRP8^{En}$ domain is rotated toward $PRP8^{NTD}$, but still displays an open conformation. PRP38, SNU23, MFAP1, and UBL5 – all of which have homologs in yeast – bind to this large cleft separating the $PRP8^{NTD}$ and $PRP8^{En}$ domains, which was previously occupied by the RecA domains of PRP28 in the pre-B complex. This is a mutually exclusive interaction observed between the pre-B and

B complex, one of many in the assembly of a functional spliceosome. UBL5 is situated very close to the 5′ss, possibly stabilizing this interaction leading up to Step 1 of splicing [203]. PRP38 makes contacts to SNU23, MFAP1, and UBL5, likely acting as a platform upon which these factors are organized [204]. SNU23 passes alongside the U6/5′ss helix. FBP21 is positioned at the interface between the helicase BRR2 and its U4/U6 substrate, appearing to prevent premature unwinding of this duplex. The U2 snRNP, although still loosely associated to the body of the spliceosome by U2/U6 helix II, is partially stabilized by contacts to SMU1 and RED, which form a heterodimer that bridges the U2 protein SF3B3 to BRR2. SMU1 and RED were shown to be important for the splicing of short introns (e.g., 200 nts or fewer) [205]. The remaining B-specific proteins, WBP11 and PQBP1 (also known as NPW38BP and NPW38), were also detected in the biochemical preparation of the B complex but were not observed in the cryo-EM structure, hinting that they may act at a distance or bind transiently to the spliceosome during B complex formation [196].

Despite major conformational and compositional remodeling, the catalytic center of the spliceosome is still not formed in the B complex. For this to occur, the U4/U6 snRNA duplex must be unwound by BRR2, leading to the dissociation of the U4 snRNP; the U5 proteins PRP6, DIM1, SNU66, and SAD1; as well as the U6 LSm proteins from U2/U6 helix II [99]. In particular, PRP6, DIM1, and PRP31 occlude the large cavity between $PRP8^{NTD}$ and $PRP8^{HB}$ in which the catalytic center will be formed. Thus, removal of these factors is a clear prerequisite for activation. The U2 snRNP must subsequently and dock onto $PRP8^{RT/En}$, to allow for new base-pairing interactions between U2 and U6 snRNA to form within the nascent catalytic center (see 1.4.4). The B-specific proteins must be dissociated from the complex, allowing for a conformational change in PRP8 from an open to a closed conformation (Fig. 1.11). This drastic loss of approximately 32 proteins and U4 snRNA from the B complex allows for the incorporation of approximately 28 additional proteins, leading to massive rearrangements in the molecular architecture of the spliceosome *en route* to an activated complex containing a fully formed catalytic center (Fig. 1.12).

1.4.3 Exchange of protein and RNA components during activation

The activation phase comprises the largest change in composition and molecular architecture of the spliceosome (Fig. 1.12) [97]. Formation of the catalytic center – including the U6 ISL and U2/U6 helix Ia and Ib – requires the removal of the U4/U6 proteins and the integration of numerous non-snRNP proteins into the spliceosome. Many of these proteins are organized into subcomplexes that join the spliceosome as preformed units, while certain other proteins are reported to join the spliceosome as independent entities (Fig. 1.12). The major subcomplexes that join during activation to form the B^{act} complex are the Retention and Splicing complex (RES), NineTeen Complex (NTC), and the Intron Binding Complex (IBC) (Fig. 1.13). Additional proteins that join the spliceosome during activation are the NineTeen Complex Related (NTR) proteins, SRRM1, SYF3, PPIL2, PRP17, CWC22/CWC27, RNF113A, SRRM2, PRP2, and GPKOW.

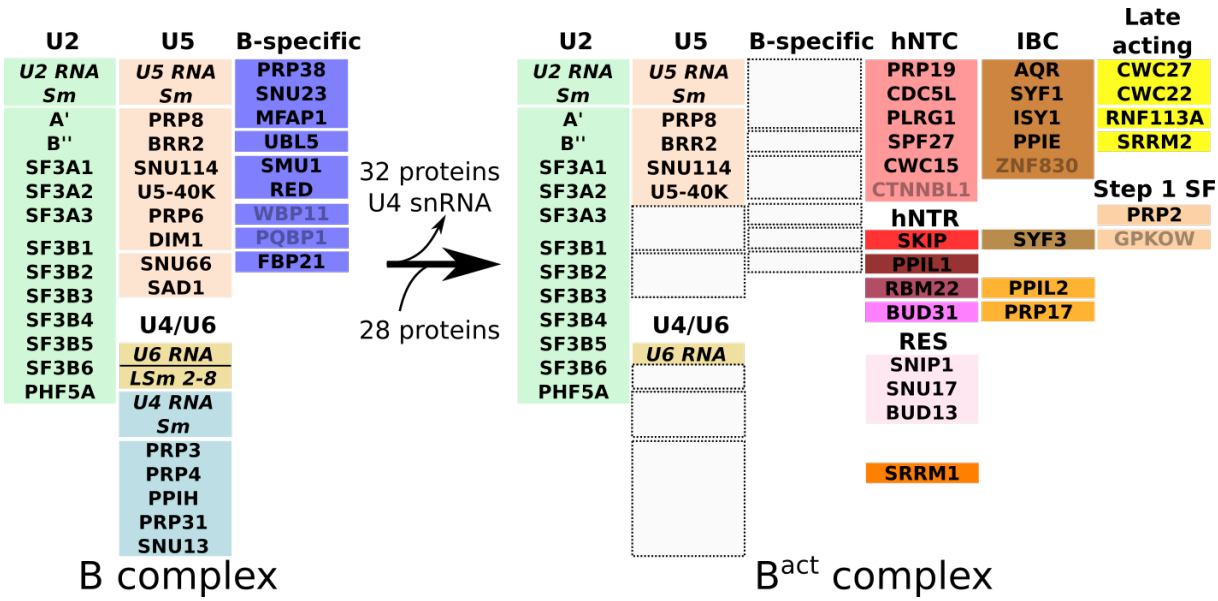

Fig. 1.12: Protein and RNA exchanges during activation. The protein and RNA composition of the B complex (left) and B^{act} complex (right) are shown, with the snRNPs or protein groups labeled above the colored rectangles. Protein components that are organized into a complex are grouped together within a single box. Components that are dissociated from the spliceosome following activation are indicated by a gray box with a stippled border. Lighter colored text within in individual box (i.e., WBP11, PQBP1, CTNNBL1, ZNF830, GPKOW) indicates that the protein has been identified in biochemical preparations of the complex but has not been localized in structures of the spliceosome. Adapted from [131].

1.4.4 Cryo-EM structure of the activated spliceosome (B^{act} complex)

The cryo-EM structure of the human B^{act} complex revealed for the first time the 3D organization of the catalytic center of the human spliceosome prior to splicing [99]. In the activated spliceosome, removal of the B-specific proteins allows PRP8 to adopt a closed conformation, in which the catalytic center is formed (Figs. 1.11 and 1.12). The overall architecture of the B^{act} complex is strikingly distinct from that of the B complex: following unwinding of the U4/U6 snRNA duplex by BRR2, leading to the removal of U4 snRNA and associated proteins (Fig. 1.12 and 1.13), U2 snRNP is shifted towards the body of the U5 snRNP, docking to the $PRP8^{RT/En}$ domain in B^{act} (Fig. 1.12 and 1.13). This major repositioning of U2 snRNP moves the U2/BS helix in closer proximity to the 5′ss, setting the stage for Step 1 of splicing. U2/U6 helix II is drastically repositioned and the U6 LSm proteins, which were previously bound to helix II, are dissociated. Following its unwinding of U4/U6 duplex, BRR2 rotates on its anchor at the $PRP8^{Jab1}$ domain, such that it comes into contact with the SF3B3 protein of U2 snRNP. Thus the bridge between BRR2 and U2 snRNP which was formed by SMU1/RED in the B complex is broken, to be replaced by direct protein-protein interactions between SF3B3 and BRR2 [99, 196].

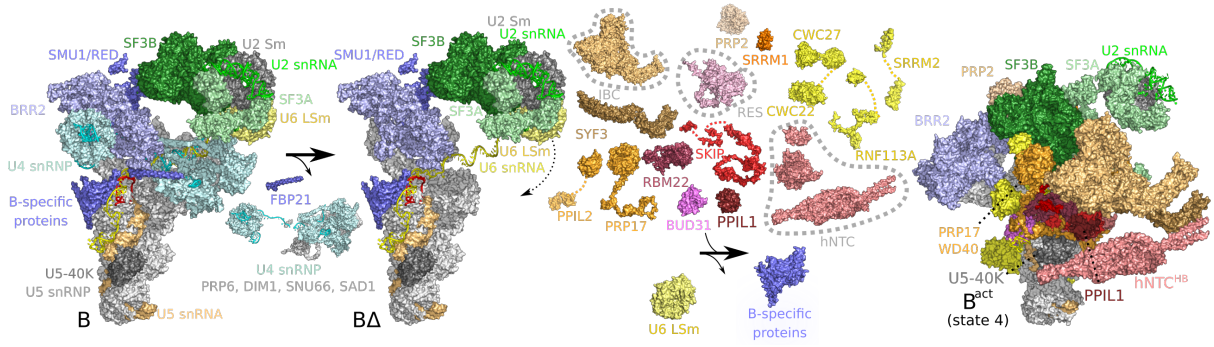

Fig. 1.13: Structural changes involved in B-to-B^{act} transition. Change in the ribonucleoprotein inventory and the molecular architecture of the spliceosome during activation are shown. In the B complex (left; PDB 5O9Z), U4 snRNP (light blue) is still associated and U2 snRNP (green, with SF3b, SF3a, U2 Sm, and U2 snRNA highlighted) is loosely bridged to BRR2 by SMU1/RED. Unwinding of U4/U6 base-pairing allows U4 snRNP, FBP21, and several U4/U6 proteins to be dissociated. The hypothetical intermediate, BΔ is shown (middle structure) to emphasize the translocation of U2 snRNP, U6 LSm, and U6 snRNA downwards towards the body of the spliceosome. Proteins that join the spliceosome at this stage are indicated above the arrow, while proteins that dissociate (B-specific proteins and U6 LSm) are shown below the arrow. Subcomplexes that join the spliceosome (IBC, RES, NTC) are surrounded by a stippled gray line. The B^{act} complex (state 4, unpublished PDB) is shown on the far right. Adapted from [131]

Retention and Splicing Complex

The RES complex is comprised of the proteins SNU17, PML1, and BUD13, and is important for the formation of the activated spliceosome (Fig. 1.14) [116, 206]. The RES complex was shown to bind cooperatively at the pre-catalytic (B complex) stage of assembly and to be stably integrated in the activated spliceosome, prior to becoming more loosely bound to the spliceosome during the conversion to the catalytically activated complex (B*) [207, 208]. In the B^{act} complex, the RES complex binds to the intron within a region spanning from the BS to the 3′ss [207, 209]. SNU17 interacts with the pre-mRNA via its RNA Recognition Motif (RRM), but the mechanism of action of SNU17 and the RES complex as a whole is not well understood [207, 210]. However, spliceosomes lacking RES are disassembled prematurely by the DEAH-box helicase PRP2, which joins the complex during activation and catalyzes the progression from an activated (B^{act}) to a catalytically activated (B*) complex [211]. Therefore, binding of the RES proteins to the intron appears to be a critical checkpoint in formation of the B^{act} complex.

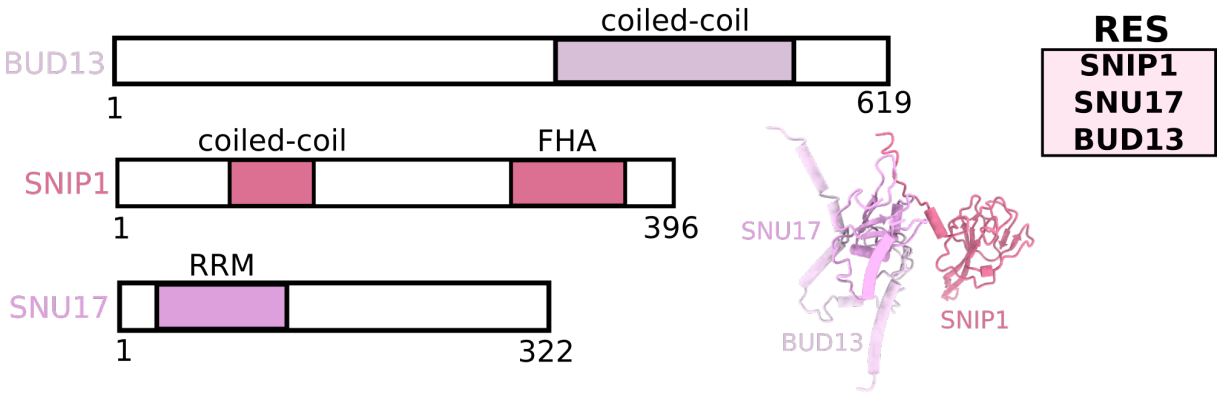

Fig. 1.14: Retention and Splicing (RES) complex. Names of the RES proteins (right), their domain architectures (left) and 3D organization in B^{act} (middle; from PDB 6FF7) are shown. Both BUD13 and SNIP1 contain coiled-coil domains. SNIP1 has a forkhead-associated (FH) domain in its C-terminal. The RNA recognition motif RRM) of SNU17, which associates the RES complex to the pre-mRNA, is indicated.

NineTeen Complex

The NineTeen Complex (NTC) is a preformed subcomplex consisting of PRP19, CDC5L, PLRG1, SPF27, CWC15, and CTNNBL1 (Fig. 1.15) [212, 213]. The NTC associates with the spliceosome during activation and is required for splicing [214]. Association of the

NTC with the spliceosome is implicated in removal of U6 LSm proteins and subsequent stabilization of U6 snRNP within the spliceosome [214]. The structure of the activated complex is consistent with this evidence, as the U6 LSm proteins are not present and the U6 snRNA is tightly integrated into the mature U2/U6 catalytic center. Components of the NTC form extensive contacts throughout the spliceosome in B^{act}, facilitated by the complex domain structures of its constituent proteins [99]. The WD40 domain of PLRG1 ($PLRG1^{WD40}$) docks to $PRP8^{HB}$, near the catalytic center. $PLRG1^{WD40}$ contacts many additional proteins, acting as an organizer of protein-protein interactions. CWC15 (also called Ad002) is wound around $PLRG1^{WD40}$, threading along the $PRP8^{NTDL}$, $PRP8^{HB}$, and $PRP8^{RT}$. The low complexity regions of CWC15 present one of many examples of an intrinsically disordered protein that is structured upon integration into the spliceosome [197, 215].

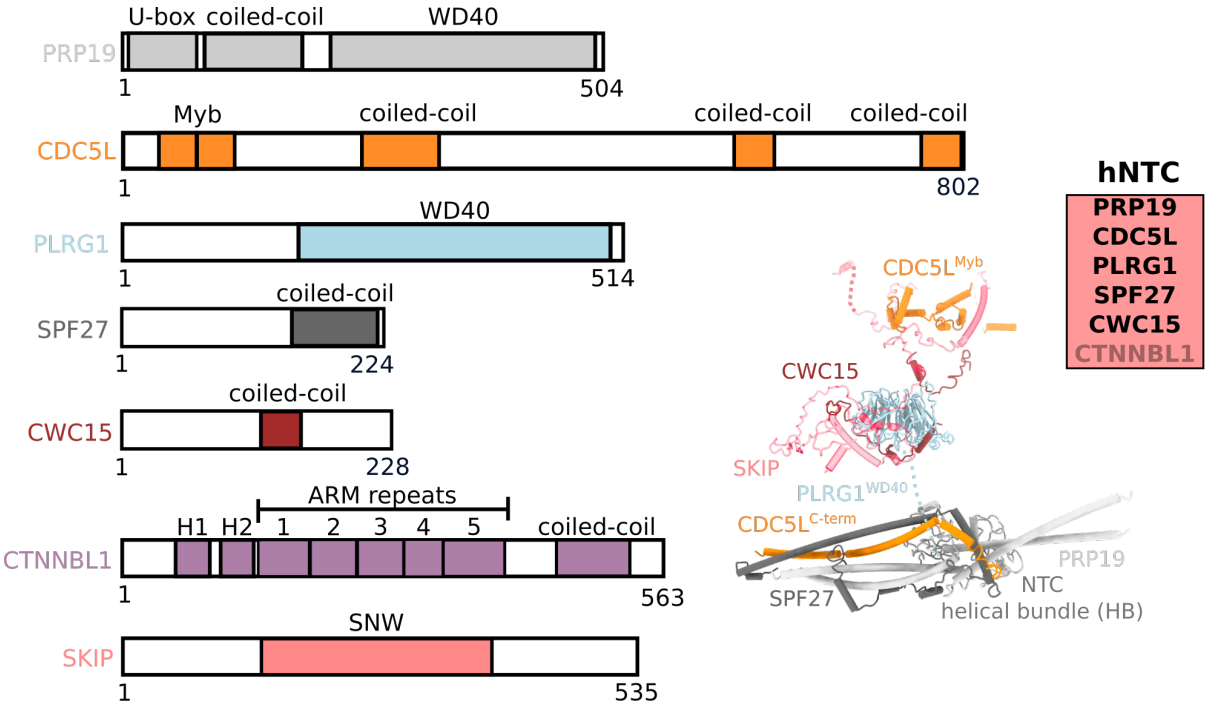

Fig. 1.15: NineTeen Complex (NTC) architecture. Names of proteins making up the NTC (right), their domain architectures (left) and 3D structure (middle; from PDB 6FF7) of NTC proteins identified in B^{act}. Stippled lines indicate flexible linker regions that are not modelled in the structure. The NTR protein, SKIP, associates closely with the core NTC proteins and is therefore also shown. CTNNBL1 is not shown as a structural representation as it not been identified in a structure of the spliceosome. The NTC helical bundle is found only in later stages of B^{act} (e.g., state 4).

CDC5L is a protein that interacts with both the core as well as the periphery of the spliceosome. The N-terminal region of CDC5L contains two domains comprised of short

α-helices, known as Myb domains ($CDC5L^{Myb}$), which have nucleic acid binding activity [216]. $CDC5L^{Myb}$ domains are positioned very close to U2/U6 helix Ib of the active site, and are proposed to facilitate formation of U2/U6 snRNA interactions of the catalytic center [95, 99]. The C-terminal region of CDC5L contains coiled-coil regions that together with PRP19 and SPF27 form the NTC Helical Bundle (NTC^{HB}), which is stably integrated into the spliceosome during late stages of B^{act} formation [99].

CTNNBL1 contains two short helices, H1 and H2, in its N-terminal region, followed by α-helical armadillo repeats (ARM) and a coiled-coil region [217, 218]. CTNNBL1 and CWC15 are the least stably bound components of the NTC and are known to readily dissociate under high ionic strength conditions [215].

NineTeen Complex-Related proteins

Several proteins associate with the NTC and are therefore termed NTC-related (NTR) proteins, which include BUD31, RBM22, SKIP, and PPIL1 [219]. The NTR proteins are thought to not be part of a preformed complex and are thus expected to associate individually to the spliceosome. BUD31 binds near the 5′ stem-loop (SL) of U6 snRNA and is likely involved in stabilization thereof, while RBM22 latches onto the 3′ end of the U6/5′ss helix [220]. RBM22 (Cwc2 in yeast) contains a zinc finger (Znf) domain in its N-terminal region involved in protein-protein and protein-RNA interactions, as well as an RRM domain in its C-terminal region that associates with surrounding proteins and the intron [220, 221].

$RBM22^{RRM}$ forms a bridge between the U5 snRNP core of the spliceosome and a large (ca. 400 kDa) protein subcomplex, the Intron Binding Complex, into which the intron is threaded from the U6/5′ss across $RBM22^{Znf/RRM}$ [99]. SKIP contains an N-terminal region that mediates its interaction with PPIL1 and a C-terminal SNW domain that tethers SKIP to the spliceosome [222–224]. PPIL1 binds to the spliceosome in later stages of activation, prior to the first catalytic step, and interacts with SKIP and RBM22 [99, 220].

Intron Binding Complex

The Intron Binding Complex (IBC), comprised of AQR, SYF1, PPIE, ISY1 and ZNF830 (Fig. 1.16) [225], bridges the 3′ end of the intron at the U6/5′ss helix to the U2/BS helix within U2 snRNP [225]. Connections between RBM22 and AQR stabilize IBC association with the spliceosome [99]. AQR is a large (171 kDa), ATP-dependent RNA helicase that exhibits a multi-domain architecture [225]. ATPase activity of AQR is required for splicing, although the target of AQR's activity is not known [225]. SYF1 contains 14 Half-A-Tetricopeptide (HAT) repeats which form a curved structure of short α-helices [99, 226, 227]. PPIE (also known as CypE) consists of an N-terminal RRM domain as well as a peptidyl-prolyl cis-trans-isomerase (PPIase) domain in its C-terminal region [228]. PPIE is situated at a juncture between the U2 snRNP-associated proteins SF3B2/SF3B4 and SYF1/AQR, forming an additional bridging interaction between the IBC and the body of the spliceosome. PPIE^{RRM} interacts with the intron [120, 199]. As for all PPIases of the spliceosome, the role of $\text{PPIE}^{\text{PPIase}}$ is unclear.

ISY1 contains at least two α-helices in its N-terminal region, which bind to the IBC via interfaces to $\text{PPIE}^{\text{PPIase}}$ [199]. ISY1 was suggested to regulate Step 2 of splicing by modulating activity of the DExD/H-box helicase, PRP16 [229]. ZNF830 is a small protein (106 aa) containing a zinc finger motif and a predicted coiled-coil region, but has not been visualized in reconstructions of the spliceosome and the function of ZNF830 in the spliceosome is unknown [225].

B^{act} proteins

Numerous additional proteins join the spliceosome during activation, including SYF3, PPIL2, and PRP17 but are not known to join as part of preformed subcomplexes. SYF3 (also known as CRNKL1 or CRN) contains 17 HAT repeats and binds to the spliceosome between the IBC and $\text{PLRG}^{\text{WD40}}$ of the NTC, and has been reported to interact with PPIL2 (also known as CYC4) [230]. SYF3 also interacts with U6 snRNA at the catalytic center, potentially contributing to stabilization of the activated complex [99]. PPIL2 has been reported to have E3 ubiquitin ligase activity but its role in the spliceosome is unclear,

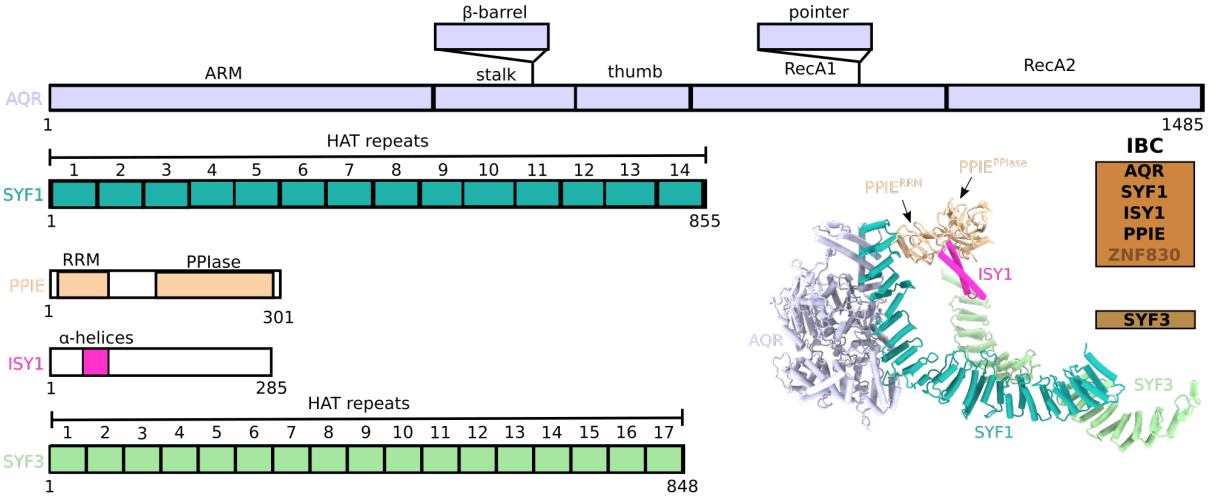

Fig. 1.16: Intron Binding Complex (IBC) and SYF3. Names of IBC proteins and the closely associated SYF3 (right), their domain architectures (left) and 3D representation as identified in B^{act} (PDB 6FF7).

and it has not been identified in any published spliceosome structures as of this writing [231]. PRP17 contains a WD40 domain ($PRP17^{WD40}$) that binds near the catalytic center of the activated spliceosome, and is known to join the spliceosome during late stages of activation [99]. Disordered regions of PRP17 interact tightly with NTR proteins, forming a complex web of interactions.

Several factors associate with the spliceosome during the final stages of activation, including CWC27, CWC22, RNF113A, SRRM1, SRRM2, PRP2, and GPKOW. CWC27 and CWC22 are splicing factors that together form a heterodimer which is reported to stabilize the interaction between the exon junction complex (EJC) and the spliceosome following Step 1 of splicing [232]. In addition, CWC27 contains a PPIase domain that binds between $PRP8^{En}$ and $PRP8^{NTD}$ domains following the dissociation of the B-specific proteins from their binding site in the B complex. RING finger protein 113A (RNF113A), a ubiquitin ligase proposed to polyubiquitinate BRR2 [233], is integrated into the spliceosome during activation and closely associated to the catalytic RNA network [99]. Along with $CWC27^{PPIase}$, RNF113A binds to the narrow gap between the $PRP8^{En}$ and $PRP8^{NTD}$ domains, preventing complete closure of PRP8 (Fig. 1.11).

Serine/arginine repetitive matrix protein 1 (SRRM1, also known as SRm160) binds pre-mRNA and associates with SR proteins involved in splicing [234]. SRRM1 was localized

in the activated spliceosome, but has not been observed in the pre-catalytic spliceosome [98, 99]. It is unclear whether SRRM1 joins the spliceosome during activation, or merely becomes stabilized due to protein-protein interactions as the spliceosome assembles. Serine/arginine repetitive matrix protein 2 (SRRM2, also known as SRm300) forms a heterodimer with SRRM1 (SRm160/300) that is known to be involved in the control of splicing by exonic splicing enhancers (ESEs) by forming contacts to SR proteins and U2 snRNP [235], and was reported to be involved in modulation of alternative splicing [236]. SRMM2 interacts with $PRP8^{SWL}$, and was proposed to contribute to stabilization of $PRP8^{SWL}$, which appears to be initially flexible after repositioning from the B-complex configuration (Fig. 1.10) [99, 237].

PRP2 and GPKOW (Spp2) are also recruited in the formation of B^{act}. PRP2, the DExD/H-box helicase responsible for promoting catalytic activation of the spliceosome (i.e., B* formation), docks to the U2 snRNP component SF3B1 (HEAT domain, repeats 7-8), where it binds the pre-mRNA [238]. Association of PRP2 to the spliceosome is mediated by GPKOW (Spp2), which co-activates PRP2 and is required for its activity [238]. With PRP2 in place, the activated spliceosome can be converted to the B* complex to carry out Step 1 of splicing (Fig. 1.4) [211].

1.4.5 Proteins surrounding the active site RNA in B^{act}

The catalytic center of the spliceosome is surrounded not only by PRP8, but also by many other factors that are repositioned and/or join the spliceosome during the formation of an activated complex [99]. The loop of U6 ISL is docked to $PRP8^{NTD}$ and is contacted by the flexible strand of $PRP8^{NTDL}$, which connects the $PRP8^{NTD}$ to the $PRP8^{HB}$ domain. $SYF3^{N\text{-}term}$ HAT repeats, as well as SKIP and RBM22 (NTR proteins), contact a flipped-out nucleotide of the stem region of the ISL (C60), while CWC15 and PLRG1 (NTC proteins) interact with C68 of the ISL loop region [99].

Helix Ib is fit into a groove of $PRP8^{HB}$, while helix Ia is surrounded by $PRP8^{\alpha\text{-}finger}$, the $CDC5L^{Myb}$ (an NTC protein), as well as SF3A2 and SF3B2 (U2 snRNP proteins).

The 5′ss is bound by RNF113A, following the dissociation of the B-specific proteins. This ensemble of proteins forms a complex network in which protein-RNA interactions stabilize the mature catalytic center within the closed conformation of PRP8. With the exception of SF3A2, SF3B2, and PRP8, all of these proteins become integrated as part of spliceosome activation, implying a role in facilitating formation of the U2/U6 active site RNA.

Surprisingly, both steps of pre-mRNA splicing can be carried out in the absence of proteins by U2/U6 snRNAs alone *in vitro*, although it is unclear whether such a minimal system is as tightly controlled as found *in vivo* in the spliceosome [239–242]. In a protein-free system, Step 1 occurred by hydrolysis as opposed to branching; that is, the BS-A did not act as the nucleophile in the reaction, indicating mechanistic differences in the protein-free, *in vitro* system. Such discrepancies appear to suggest that proteins are likely involved in mediating the proper positioning of the BS-A in the spliceosome [239]. Furthermore, splicing efficiency is increased by NTR, IBC, and RES proteins [225, 243, 244], while the NTC is essential for Step 1 of splicing [245]. The observed regulatory roles for these factors is perhaps unsurprising, as proteins make up approximately 90% of the molecular weight of the spliceosome [246]. All of these events strongly suggest that although protein-free systems in vitro can technically carry out reactions that are similar – but not identical to – pre-mRNA splicing, proteins are required for coordinating the formation of the activated spliceosome and for tightly regulating splicing fidelity and efficiency. While the formation of a catalytically essential, RNA triple helix in the spliceosome is well documented, the potential role of proteins in facilitating folding of U2/U6 active site RNA has not been fully determined. Moreover, the temporal ordering of active site formation – that is, the order in which the ISL, helix Ia, helix Ib, and the resulting triple helix are formed – is unclear.

Structural and biochemical studies have been successful at characterizing snapshots of spliceosomes stalled before and after the catalytic center is formed (i.e., the pre-catalytic B complex and the activated B^{act} complex), but have so far not been able to shed light onto intermediate states during the formation of the catalytic RNA network at the core of the

spliceosome [99, 196, 247, 248]. The molecular architecture of such intermediates would provide insight into the critical steps during activation of the spliceosome, while possibly revealing assembly checkpoints that contribute to accurate folding of the catalytic center. In particular, it would be of interest to isolate complexes trapped at intermediate stages of activation, following the dissociation of U4 snRNP but prior to the stable association of the non-snRNP proteins that become integrated in B^{act} (i.e., RES, NTC/NTR, IBC and later stage assembly proteins). It would also be necessary to analyze the conformation and arrangement of PRP8, and to determine whether its conformational change occurs before or after the binding of the RES, NTC/NTR, IBC and additional factors, and how this may be coordinated by the B-specific proteins. The identification of activation intermediates may also yield new insights into the roles of proteins that co-purify with the spliceosome but have not yet been identified in cryo-EM structures of spliceosome assembly. However, the biochemical preparation of novel structural "snapshots" has been hindered by a lack of available means by which to stall spliceosome assembly at intermediate stages of activation.

1.5 Small molecule chemical inhibitors of pre-mRNA splicing

Small molecule chemical inhibitors provide a method by which to trap spliceosomal complexes at intermediate stages of assembly. A major focus of spliceosome inhibition has been on SF3B1, a core component of the U2 snRNP that accommodates the U2/BS helix and is frequently mutated in numerous cancers [249]. Targeting of SF3B1 has therefore been the subject of many studies due to its relevance in disease. A number of natural products or synthetic derivatives target SF3B1, stalling spliceosome assembly at the A complex stage at low nanomolar or micromolar concentrations [250]. Many of these compounds share a similar structural backbone consisting of a conjugated diene [251]. These compounds are relatively well characterized through structure-activity relationships, in

some cases detailing the mechanism of how these inhibitors work by blocking a functionally relevant conformational change within SF3B1 [252].

Several examples exist of spliceosome inhibitors that block assembly by interfering with post-translational modifications of proteins involved in splicing [250]. It has been shown that inhibition of Ser/Thr protein phosphatase-1 (PP1) and protein phosphatase 2A (PP2A) by the natural products okadaic acid, tautomycin, and microcrystin-LR block splicing reactions [253]. Use of a phosphatase inhibitor cocktail (PhosSTOP, Roche) was shown to arrest spliceosome assembly at a pre-catalytic stage, resulting in a mixture of pre-B and B complexes [203]. Blockage of histone acetyltransferases, histone deacetylases and SR protein kinases all results in arrest of splicing activity [83, 254]. Thus, interruption of the post-translational modifications of proteins involved in splicing has profound effects on the assembly and function of the spliceosome, and presents an additional way to study assembly intermediates. A large-scale screen of small molecules identified a distinct class of quinone derivatives that stall splicing at low micromolar concentrations [255]. The most potent of these compounds (measured by the *in vitro* 50% inhibitory concentration (IC_{50}) in the assay) were BN82685 (7 μM), NSC663284 (10 μM), and NSC95397 (21 μM). BN82685 was shown to inhibit exon ligation (Step 2), thus stalling the spliceosome at approximately the C complex stage [255]. However, the accumulation of spliceosome assembly intermediates was not reported. A later high throughput screen of over 160,000 compounds identified eight small molecules that inhibited *in vitro* pre-mRNA splicing [256]. Fortuitously, one of these eight compounds was NSC95397. Further investigation of NSC95397 (also referred to as compound 297) revealed that at concentrations at or above 75 μM, a mixture of what appeared to be A and B complexes resulted, indicating that the compound stalls splicing at a pre-catalytic stage [256]. Deeper analysis of the resulting affinity purified complexes revealed stoichiometric amounts of U2, U5 and U6 snRNA and a decrease in U1 and U4 snRNAs relative to the B complex [256]. Additionally, mass spectrometry analysis indicated the presence of the NTC/NTR proteins as well as the B-specific proteins, but underrepresentation of later stage assembly factors (i.e., RNF113A and PRP17) [256]. Taken together, these findings indicated that NSC95397

could potentially stall spliceosome assembly at an intermediate stage of activation, following the assembly of the B complex yet prior to the formation of a mature B^{act} complex, providing a putative tool to study elusive states of the human spliceosome during the formation of its catalytic center. In summary, the two key questions of such an investigation are as follows:

- Are there structurally distinct intermediate stages of activation at which formation of the catalytic center can be observed?
- How do proteins facilitate the folding of active site U2/U6 snRNAs in the spliceosome?

1.6 Aim of this study

In eukaryotes, DNA is transcribed into precursor mRNA (pre-mRNA), which contains protein-coding exons separated by non-coding introns. Prior to export of mRNA from the nucleus to the cytoplasm, introns are excised from the pre-mRNA by the spliceosome: a highly dynamic, multimegadalton machine that is assembled de novo for each intron to be spliced. The spliceosome is constructed of small nuclear ribonucleoprotein (snRNP) building blocks – U1, U2, and the U4/U6.U5 tri-snRNP, in which U4 and U6 snRNAs are base-paired and U5 is associated by protein-protein interactions – which bind to their pre-mRNA substrate in a stepwise manner, proceeding through a series of complexes until an catalytically competent complex is formed. Progression through the spliceosome assembly pathway is driven by the action of ATP-dependent RNA helicases that remodel the RNA-protein and RNA-RNA network of the complex. Over 150 proteins are associated with the spliceosome, making it one of the most complicated molecular machines in the nucleus. While over 90% of the molecular mass of the spliceosome is contributed by proteins, the remaining 10% consists of RNA. This relatively minor contribution by RNA is responsible for the catalysis of RNA splicing, making the spliceosome a ribozyme. Unlike many enzymes, the catalytic center of the spliceosome is not a preformed entity, but must be actively sculpted by remodeling base-pairing interactions in a step known as activation. Activation is the single largest change in the molecular architecture and composition of the spliceosome. During activation, the pre-catalytic spliceosome – which contains U2, U4, U5 and U6 snRNPs – is drastically altered by the action of an ATP-dependent helicase, BRR2, which unwinds the base-pairing of the U4/U6 snRNA duplex, allowing for the liberation of the U4 snRNP and associated proteins, as well as for large-scale rearrangements in the molecular architecture of the protein and RNA content of the spliceosome. It is as this step that at least 28 additional factors, some of which are organized into salt-stable subcomplexes, join the spliceosome. At some point during activation, a group of proteins known as the B-specific proteins is destabilized from the complex, allowing for the central scaffold protein of the spliceosome, PRP8, to undergo

a conformational change from an open to a closed conformation. It is within a compact cavity of PRP8 that the U2/U6 RNA network of the active site is formed, resulting in an activated spliceosome.

Structural studies by cryo-EM have elucidated the molecular architecture of complexes prior to and following activation, providing major insights into the conformational and compositional rearrangements that must occur for an activated complex to be formed. However, the temporal ordering in the exchange of proteins is not well understood. Moreover, the potential role of proteins in mediating the folding of U2/U6 RNA of the catalytic center is unclear. Thus, it would be advantageous to capture "snapshots" of the spliceosome at discrete points along its activation pathway, which would likely reveal new insights into the integration and destabilization of proteins that may assist in the formation of an active complex. Despite this clear need for finer sampling of the spliceosome activation step, no such structures have been reported. Several approaches to stall the splicing using chemical inhibitors have been introduced, suggesting possibly viable new methods to trap assembly at biochemical and structural stages of activation. High throughput screening of small molecule libraries successfully identified quinone derivatives as compounds that inhibited splicing, blocking assembly at distinct stages. Further investigation of this phenomenon suggested that one of these compounds, NSC95397, appeared to stall spliceosome activation. We therefore set out to determine the structure of spliceosomes stalled by NSC95397, with the goal of identifying previously uncharacterized structural stages of activation. The aim of this thesis is to unravel the spliceosome activation step using a small molecule chemical inhibitor, NSC95397, and to determine the 3D molecular architectures of the resulting complexes with cryo-EM.

2 Materials and Methods

2.1 Materials

2.1.1 *Software*

Software	Reference
COW	unpublished
Coot	[257, 258]
cryoSPARC	[259]
Gautomatch	unpublished
Gctf	[260]
Mascot	[261]
MolProbity	[262]
MotionCor2	[160]
PHENIX	[258, 263]
pLink2.3.5	[264, 265]
Protein Data Bank	[266]
PyMOL	unpublished
RELION 2.0-1 and 3.0	[267, 268]
SpliProt3D	[269]
SWISS-MODEL	[270]
UCSF Chimera	[271]
UCSF ChimeraX	[272]
UniProt	[273]

2.1.2 *Equipment*

Equipment	Manufacturer
Amicon 50 kDa filters	Merck
C_s corrector	CEOS GmbH
Falcon 3EC Direct Electron Detector	Thermo Fisher Scientific
LTQ Orbitrap XL mass spectrometer	Thermo Fisher Scientific
MBP Trap HP column	GE Healthcare
Q Exactive HF-X mass spectrometer	Thermo Fisher Scientific
Quantifoil 3.5/1 Cu grids	Quantifoil
Superdex Peptide PC3.2/30 column	GE Healthcare
Surespin TLS-630 rotor	Thermo Fisher Scientific
Titan Krios electron microscope	Thermo Fisher Scientific
TST41.14 rotor	Kontron
Typhoon phosphoimager	GE Healthcare
Vitrobot Mark IV	Thermo Fisher Scientific

2.2 Methods

Biochemical experiments (see 2.2.1, 2.2.2, 2.2.3, 2.2.4, 2.2.5, 2.2.7) were carried out by Dr. Majety Leelaram and Dr. Dmitry Agafonov (Department of Cellular Biochemistry, MPI-BPC), and crosslink analysis was performed by Dr. Olex Dybkov (Department of Cellular Biochemistry, MPI-BPC) in collaboration with Prof. Dr. Henning Urlaub (Bioanalytical Mass Spectrometry, MPI-BPC) (see 2.2.6).

2.2.1 *In vitro splicing*

HeLa S3 cells (Helmholtz Zentrum für Infektionsforschung, Braunschweig) were obtained, tested negative for mycoplasma, and used to prepare splicing extracts [274]. MINX-MS2 pre-mRNA was prepared in vitro transcription by T7 RNA polymerase, with α^{32}P-UTP (for radiolabeling of pre-mRNA) and m^{7}G(5′)ppp(5′)G-cap present. To carry out *in vitro* splicing reactions, the resulting ^{32}P-labeled MINX-MS2 pre-mRNA (5.0 nM) was incubated at 30°C for specified time points (Fig. 3.1) in the presence of 40% (v/v) HeLa

nuclear extract, 20 mM HEPES-KOH pH 7.9, 3 mM $MgCl_2$, 65 mM KCl, 2 mM adenosine triphosphate (ATP), and 20 mM creatine phosphate as well as increasing amounts (Fig. 3.1) of the splicing inhibitor splicing inhibitor NSC95397 (Sigma Aldrich), which was solubilized in dimethyl sulfoxide (DMSO). Following the splicing reaction, NSC95397 was quenched with a twofold molar excess of cysteine for 2 min at 30°C and then put on ice. RNA isolation was carried out using phenol-chloroform extraction and a subsequent ethanol precipitation step, and the purified RNAs were examined using a 14% denaturing polyacrylamide gel. Splicing products, including intermediates and unspliced pre-mRNA, were visualized with a Typhoon phosphoimager (GE Healthcare). Formation of spliceosomal complexes was monitored using 1.5% low-melting agarose gel electrophoresis containing 0.4 mg/ml heparin [275].

2.2.2 *MS2 affinity purification of spliceosomes*

To prepare for affinity purification of spliceosomes stalled by the action of NSC95397, a 10-fold molar excess of MS2-MBP fusion protein was incubated with α^{32}P-labeled MINX-MS2 pre-mRNA (5 nM) for 30 min on ice. The pre-mRNA, bound by MS2-MBP, was used as the substrate in a splicing reaction in the presence of 160 µM of NSC95397, incubated for 60 min at 30°C, and subsequently quenched with 320 µM cysteine for 2 min. Centrifugation was used to remove aggregates and the salt concentration was increased to 165 mM by addition of sodium chloride (NaCl). For affinity purification, the resulting splicing reaction was loaded on an MBP Trap HP column (GE Healthcare) preequilibrated with G-150 buffer (20 mM HEPES-KOH pH 7.9, 1.5 mM $MgCl_2$, 150 mM NaCl). G-150 buffer was used to wash the column and 10 mM maltose (in G-150 buffer) was used to elute the spliceosomes. Following isolation of the affinity purified complexes, a further gradient purification step was used to separate complexes by their sedimentation coefficients. Samples were loaded onto a 36-ml linear 10 to 30% (v/v) glycerol gradient (G-150 buffer) and centrifuged for 15 hours at 4°C and 25,000 rotations per minute (rpm) using a Surespin TLS-630 rotor (Thermo Fisher Scientific). A peristaltic pump was used

to fractionate the gradients from the bottom of each tube, and the radioactivity of each fraction was measured with Cherenkov counting. Peak gradient fractions were analyzed for RNA composition using denaturing 4 to 12% NuPAGE gels (Life Technologies), which were subsequently stained with SYBR Gold for detection and visualization.

2.2.3 *2D gel electrophoresis and mass spectrometry*

2D gel electrophoresis, as described in [276], was used to separate all proteins in the purified sample larger than 25 kDa. Following staining of the gel with Coomassie, individual proteins spots were excised, digested with trypsin, and extracted. The extracted peptides were subjected to a liquid chromatography-coupled electrospray ionization mass spectrometer (LTQ Orbitrap XL). Protein identities were determined by comparison of fragment spectra against the UniProt database [273], with Mascot used as a search engine [261].

2.2.4 *Chase of pre-B^{act} complexes with micrococcal nuclease-treated extract*

To digest RNAs in the sample, micrococcal nuclease (MN) was added to HeLa nuclear extract (as described in [197]). B and pre-B^{act} complexes were formed separately on α^{32}P-labeled MINX-MS2 pre-mRNA, affinity purified, and incubated under the following two conditions: 1) in splicing buffer alone (20 mM HEPES-KOH pH 7.9, 65 mM KCl, 3 mM $MgCl_2$, 2 mM ATP, 20 mM creatine phosphate); 2) in the presence of HeLa nuclear extract that had been treated with 20% (v/v) MN. Unlabeled MINX-MS2 pre-mRNA was added in 20-fold excess to block the reassembly of spliceosomes on radiolabeled pre-mRNA. Reactions were kept at 30°C for timepoints from 0 to 60 min. RNA from the resulting samples was isolated, analyzed on a 14% denaturing polyacrylamide gel, and visualized with a Typhoon phosphoimager (GE Healthcare).

2.2.5 *Western blotting*

Proteins from the purified complexes were separated on denaturing 4 to 12% NuPAGE gels and transferred onto nitrocellulose membranes (Amersham Protran 0.2, GE Healthcare). Antibodies against the following human proteins were used to probe the membrane: hSnu114 [277], LSm4 [57], hPrp19 and AQR [213], hCwc22 (Proteintech Europe, Manchester, UK catalog number 26898-1-AP), hRNF113A (Sigma Aldrich, catalog number HPA000160, RRID: AB_1079821), and hPPIL2 (ThermoFisher Scientific, catalog number PA5-29239, RRID: AB_2546715). An ECL detection kit (GE Healthcare) was used to detect bound antibody.

2.2.6 *Cross-linking of pre-B^{act} complexes and cross-link identification*

Following MS2 affinity purification, an additional step was used to further purify spliceosomes, by introducing a 10 to 30% (v/v) glycerol gradient centrifuged at 23,300 rpm for 16 hours 40 min at 4°C in a TST41.14 rotor (Kontron). Peak fractions having approximately 15 pmol of the pre-B^{act} complexes were combined and cross-linked with 150 µM BS3 at 25°C for 30 min. Samples were subsequently quenched using 1.5 mM Tris-HCl pH 8.0 at 25°C for 10 min, placed on ice for 20 min, and pelleted using ultracentrifugation and a S100AT4 rotor (Thermo Fisher Scientific) and analyzed as previously reported [197]. Peptide extraction was carried out in a reverse-phase manner with Sep-Pak Vac tC18 1-cc cartridges (Waters) and peptides were fractionated using gel filtration by the use of a Superdex Peptide PC3.2/30 column (GE Healthcare). A Thermo Fisher Scientific Q Exactive HF-X mass spectrometer was used to analyze fifty-µl fractions in triplicate, corresponding to an elution volume of 1.2 to 1.8 ml. The pLink2.3.5 search engine (`pfind.ict.ac.cn/software/pLink`) was used to identify protein-protein crosslinks, and the results were filtered by false discovery rates of 1, 3, and 5% [264, 265]. The score for crosslinks is shown as the negative value of the common logarithm of the original pLink score: Score = -log10("pLink Score").

2.2.7 *Purification and buffer exchange*

To prepare samples for cryo-EM, peak fractions of samples purified by MS2 affinity selection and glycerol gradient centrifugation were combined. These combined samples were concentrated and buffer exchanged in G-150 buffer (without glycerol) using an Amicon 50-kD cut-off filter as reported in [98]. For further purification, buffer-exchanged samples were loaded on a 36-ml linear 10 to 30% (v/v) glycerol gradient in G150 buffer in the presence of 0 to 0.15% glutaraldehyde (for fixation) [278], centrifuged at 25,000 rpm at 4°C for 15 hours using a Surespin TLS-630 rotor (Thermo Fisher Scientific). The gradient was fractionated and quenched immediately using 100 mM aspartate (pH 7.0). Peak fractions were combined and subjected to one more round of concentration and buffer exchange as before. Control experiments were carried out by following the same protocol as described, only without addition of glutaraldehyde to the gradient, thus allowing the protein and RNA content of the final samples to be analyzed using 2D gel electrophoresis and denaturing polyacrylamide gel electrophoresis.

2.2.8 *Cryo-EM sample preparation and data acquisition*

Approximately 30 µl of purified sample was added to a well of a custom Teflon block kept on ice. A piece of continuous carbon film (prepared in-house) was placed directly on top of the sample and incubated for 1 minute, allowing the purified samples to adsorb to the carbon film. The carbon film was adhered to the evaporated carbon-coated face of a QUANTIFOIL 3.5/1 grid (Quantifoil, Jena, Germany) and 4 µl of distilled-deionized water was added to the same side as the sample to prevent dehydration. Grids were blotted and plunge-frozen in liquid ethane at ambient conditions of 100% humidity and 4°C using a Vitrobot Mark IV (Thermo Fisher Scientific). For data acquisition, a Titan Krios electron microscope (Thermo Fisher Scientific) equipped with a C_s corrector (CEOS, Heidelberg) and Falcon III direct electron detector (Thermo Fisher Scientific) operated in integration mode was used. Images were acquired at a nominal magnification of 59,000x, resulting in a pixel size of 1.16 Å. Each individual micrograph was comprised of 40 individual frames

and a total fluence of 45 e/Å^2 acquired over an exposure time of 1 s, divided into 20 fractions (2.25 e/Å^2 per fraction). 10,000 micrographs were collected in an automated acquisition, carried out using spot-scanning in a 3x3 array for each 3.5 µm hole. Grid preparation and data acquisition were carried out by Dr. Karl Bertram (Department of Structural Dynamics, MPI-BPC).

2.2.9 *Image processing*

Motion correction and dose-weighting of micrographs was carried out with MotionCor2 [160], followed by calculation of CTF parameters with Gctf [260]. Each micrograph was inspected manually using the Micrograph Quality Checker (MQC) of COW (`www.cow-em.de/`) and micrographs were judged by optical contrast, nominal resolution, and astigmatism. Micrographs showing predominantly crystalline ice, ethane contamination, or particle aggregates were exluded from further processing, resulting in 4676 out of 10,000 micrographs retained. Gautomatch (`https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/#gauto`) was used for reference-free identification of 503,208 particle images, which were extracted from micrographs with 2x binning and split into three equally proportioned subgroups prior to five rounds of 2D classification with RELION 2.0 [267]. 2D class averages showing clear features were chosen (72,193 particle images) for the determination of an initial model, which was calculated using ab initio reconstruction within cryoSPARC [259]. The best resolved initial model was low-pass filtered to 60 Å and used for an initial reference in RELION 2.1 in a 3D classification using 15 classes (regularization parameter, "T", set to 4). Two of the resulting classes exhibited distinguishable features and were better resolved (13 to 15 Å) than the other 13 classes (>24 Å). The two well defined 3D classes consisted of of 108,932 (26% of particle images) and 56,177 (14% of particle images) particles. Each class was refined using RELION 2.1 without masks to nominal resolutions of of 6.3 and 6.4 Å (data not shown). Angular orientations were adopted from these refinements and used to re-extract unbinned particle images from dose-weighted micrographs (box size

384 pixels). The extracted particles were refined against up-scaled references without masking using RELION 3.0 [268], resulting in an improvement in resolution to 5.6 and 6.5 Å. At this point, a soft mask encompassing the entirety of each particle was generated from each reconstruction and applied to carry out a masked refinement step, improving the average resolutions to 4.4 Å (not shown) and 4.6 Å. It was suspected at this point that there may be further structural heterogeneity in the reconstructions. Therefore, soft masks were generated to include the most stable core of each complex, and applied to each reconstruction in a 3D classification without re-alignment, instead using refined particle coordinates obtained in the previous refinement step. For this step, particle images from each of the two populations were divided into four classes and the regularization parameter (T) was set to 20. For both populations, this classification step resulted in a single class having a large majority of particle images, containing containing 84,539 (77% of particles classified) and 39,336 (70% of particles classified). These classes also displayed higher resolution features than the less populated subsets and were thus chosen for further structure determination. As previously, each class was refined initially without a mask, this time resulting in average resolutions of 5.1 and 6.0 Å. The reconstructions were then refined with a soft mask encompassing the entire particle, resulting in average resolutions of 4.3 and 4.6 Å. To further improve two large regions of the map, multibody refinement was carried out using the refined coordinates of the previous refinement step [166]. Two "bodies" were defined: body 1, consisting of the core of each complex; and body 2, which encompasses the 5′ domain of the U2 snRNP, particularly the SF3b proteins. For the first group of particles, body 1 was determined to an average resolution of 3.9 Å, whereas the body 2 resolution did not improve. For the second group of particles, body 1 was resolved to 4.2 Å and body 2 was resolved to 4.5 Å. Although CTF refinement was attempted, it did not improve the resolution. To test for the possible effect of the box size on the resolution of the reconstruction, refinements of both complexes were also carried out in a larger box (512 pixels), but this also failed to result in resolution improvements. The Fourier shell correlation (FSC) 0.143 criterion was used for resolution determination

[169]. Sharpening of maps using automatically determined B-factors and local resolution variability were carried out in RELION 3.0.

2.2.10 *Model building and refinement*

The protein and RNA components used for building integrative structural models are detailed in Figure B.2. Based on prior knowledge and our protein-protein crosslinking dataset (Table B.1), model files from the Protein Data Bank (PDB) [266] for individual proteins of the spliceosomal B [98, 203] and B^{act} [99, 237], as well as X-ray crystal structures and NMR structures of various single components, were docked as rigid-bodies into the cryo-EM maps using UCSF Chimera v1.12 [271]. A model of the B-specific protein WBP11/NPW38BP was generated using SWISS-MODEL [270] and placed within a helical density within the pre-$B^{act\text{-}1}$ (see Results) reconstruction, consistent with numerous crosslinks. Models of CTNNBL1, PPIL2, and TCERG-1 (coiled-doil domain and FF4 to 6 domains) were downloaded from the SpliProt3D predicted structure database [269] and docked into densities consistent with intermolecular crosslinks. The remining protein components incorporated into the models were taken from previously published structures (Figure B.2). *De novo* modeling of several regions of MFAP1, PRP8, and SF3B2 were guided by well-defined EM densities continuing from previously identified, well documented portions of each protein as observed in the B or B^{act} complexes. Chains were truncated to polyalanine/polyproline traces and subsequently docked as rigid bodies (for regions with secondary structure) or were docked and manually adjusted (disordered regions) using Coot [257, 258]. U5 snRNA was docked as a rigid-body and its fit was further improved using the jiggle-fit and model morphing scripts in Coot (`https://www2.mrc-lmb.cam.ac.uk/groups/murshudov/content/em_fitting/em_fitting.html`). For modeling of the U2/U6 RNA network, the U6/5′ss helix and U6 5′SL regions (nucleotides 1 to 46) from a B^{act} complex [237] were fit into both maps, followed by manual adjustment of the 5′ss and a single-stranded region of U6 snRNA (nucleotides C25 to A30). U6 RNA nucleotides A47 to U51 were rebuilt into

the unmasked and unsharpened EM map of the pre-B^{act-2} complex. U2/U6 helix Ib was docked as a rigid body into the pre-B^{act-2} map, guided by clear map features consistent with the phosphate backbone of several nucleotides within the helix. U6 ISL was taken from a published hB^{act} cryo-EM structure [99] and docked into both maps, followed by manual rebuilding of nucleotides C60 to C61 and A73 to G75 (G58 to C61 and A73 to A76 for pre-B^{act-1}) in Coot.

U2/U6 helix II [99] was rigid-body fit into the pre-B^{act-2} reconstruction, based on the proximity of this density to U6 ISL and U2/U6 helix Ib. Based on comparison to previous structures [99, 237], U2 nucleotides of the stem-loop II (U46-U65) and several of stem-loop IIb (G68 to C73 and G79 to A85) were placed in well-defined densities. Individual chains and then the composite models of both complexes were processed by real-space refinement using PHENIX [263]. The resulting models were subjected to MolProbity for analysis of model geometry [262], and FSCs between the map and the model were calculated using phenix.mtriage [258]. Statistics from model refinement are shown in Figure B.1. Model building was overseen by Dr. Berthold Kastner (Department of Cellular Biochemistry, MPI-BPC).

3 Results

3.1 Biochemical characterization of spliceosomes stalled during activation

A small molecule chemical inhibitor was used to stall spliceosome assembly at an intermediate stage of activation (Fig. 3.1A). The inhibitor, NSC95397, had been previously identified as a splicing inhibitor in a high-throughput screen [71] and further characterized as a compound that likely stalled spliceosome assembly during activation but prior to the formation of a mature B^{act} complex [256]. At a concentration of 150 µM, splicing was completely blocked, as shown by the RNA products (Fig. 3.1B). Formation of spliceosomal complexes in the presence of NSC95397 was analyzed with agarose gels (Fig. 3.1C). The RNA composition of the purified spliceosomes suggested a complex with decreased amounts of U4 snRNA and stoichiometric amounts of U2, U6 and U5 snRNAs (Fig. 3.1D). Comparison of the RNA profiles of the purified spliceosomes to pre-catalytic spliceosomes (B complex) and activated spliceosomes (B^{act} complex) indicated that the stalled complex contained decreased amounts of U4 snRNA relative to the B complex, while containing stoichiometric amounts of U2, U6 and U5 snRNAs, as in the B^{act} complex (Fig. 3.1D). U1 snRNA was detected in the B complex and stalled complex, but was strongly decreased in B^{act} (Fig. 3.1D). Analysis of the protein composition of the complex of interest was carried out with 2D gel electrophoresis followed by mass spectrometry for all proteins larger than 25 kDa in the sample (Fig. 3.1E). Western blotting for complex-specific markers revealed that the stalled complex contained increased levels PRP19, PPIL2, and AQR relative to the B complex (Fig. 3.1F). The stalled complexes were purified and subjected to chase experiments in nuclear extract in which snRNAs have been digested by micrococcal nuclease treatment. Chase experiments demonstrated

that the pre-B^{act} complexes were still functional and not dead-end intermediates (Fig. 3.1G). Biochemical experiments (see 2.2.1, 2.2.2, 2.2.3, 2.2.4, 2.2.5, 2.2.7 and 3.1) were carried out by Dr. Majety Leelaram and Dr. Dmitry Agafonov (Department of Cellular Biochemistry, MPI-BPC), and crosslink analysis was performed by Dr. Olex Dybkov (Department of Cellular Biochemistry, MPI-BPC) in collaboration with Prof. Dr. Henning Urlaub (Bioanalytical Mass Spectrometry, MPI-BPC) (see 2.2.6, B.1, and B.2).

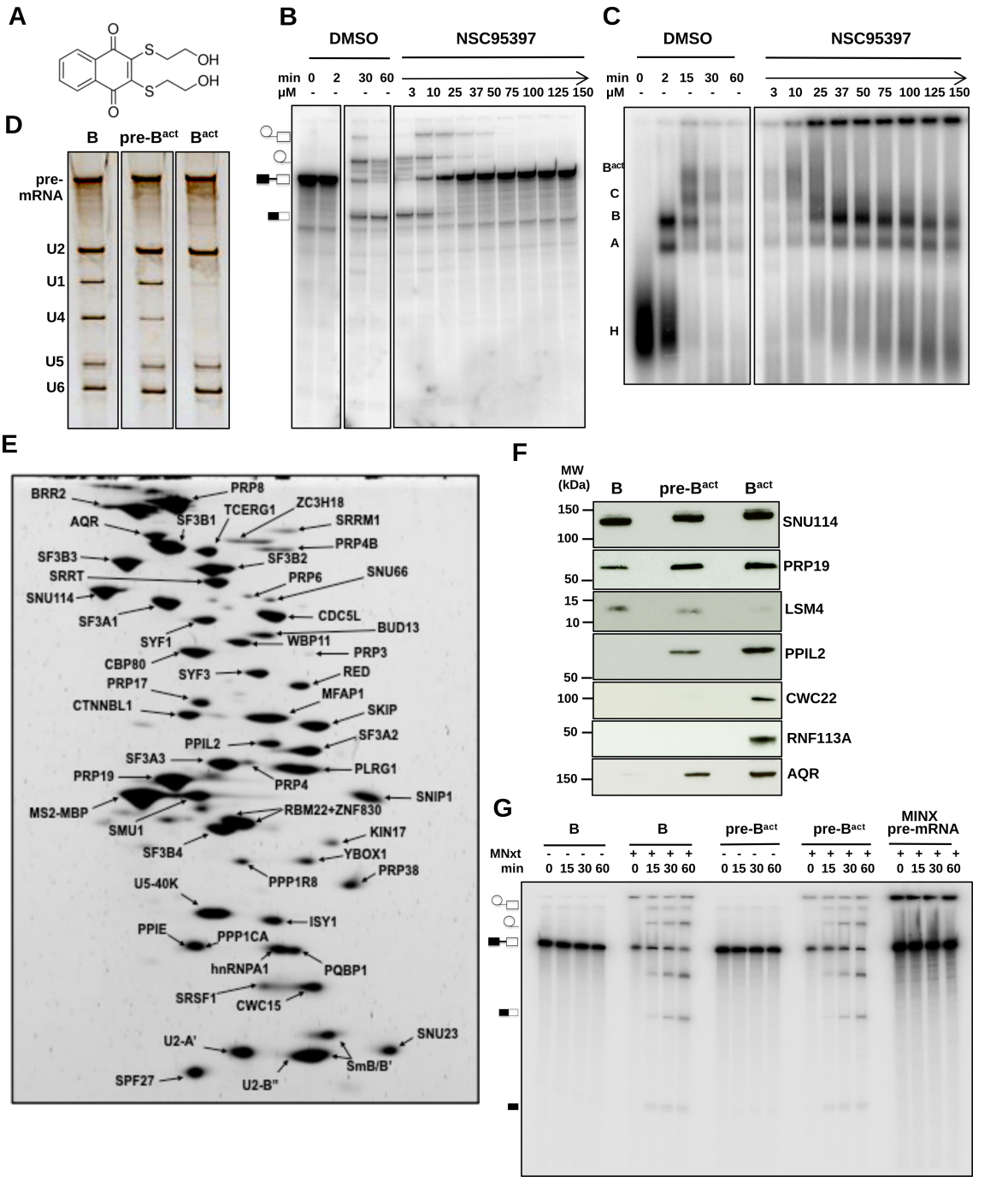

Fig. 3.1: Biochemical characterization of pre-B^{act}. (A) Chemical structure of NSC95397. (B) RNA products assessed by denaturing PAGE after 60 min incubation with 3-150 µM of NSC95397 in HeLa nuclear extract. In the diagrams on the far-left hand side of the figure, the 5′ exon is represented as a dark rectangle, the intron is represented as a thin black line, and the 3′ exon is shown as a white rectangle. From top to bottom are an intron lariat-3′ exon, intron lariat only, unspliced pre-mRNA, and mRNA. (C) Formation of spliceosomal complexes in the presence of 3-150 µM of NSC95397 in HeLa nuclear extract after 60 min monitored by agarose gel electrophoresis. The far-left panel indicates labels of distinct, well characterized spliceosome intermediates for comparison. (D) RNA composition of B, pre-B^{act} and B^{act} complexes following affinity purification and gradient ultracentifigation to purify each complex assembled for 60 min in HeLa nuclear extract containing 150 µM NSC95397. (E) Protein composition of purified pre-B^{act} complexes assessed by 2D gel electrophoresis followed by mass spectrometry to identify the individual protein spots. Proteins larger 25 kDa are separated on the gel. (F) Western blotting of several proteins contained in B, pre-B^{act} and B^{act}. Dissociation of LSm proteins and recruitment of PRP19 (NTC component), PPIL2, CWC22, RNF113A, and AQR (IBC component) occurs during activation [276]. SNU114 (integral component of U5 snRNP) is included as a loading control. (G) Stalled pre-B^{act} complexes can be rescued in micrococcal-nuclease treated HeLa nuclear extract (MNxt). The B complex and unspliced MINX pre-mRNA are included for comparison. Affinity purified pre-B^{act} complexes splice MINX pre-mRNA in MNxt. From [131]. Reprinted with permission from AAAS.

3.2 Cryo-EM structures of two novel activation intermediates

The purified pre-B^{act} sample was analyzed by cryo-EM, and computational sorting of 503,208 particle images revealed two distinct spliceosomal structures at intermediate stages of activation (Fig. 3.1A-C). Neither reconstruction suffered from strongly preferred orientation, indicating sufficient angular sampling (Fig. 3.1D,E). The average resolution of the overall complexes was 4.3 and 4.6 Å (Fig. 3.1F,G) and the map-model Fourier Shell Correlation (FSC) for the core and SF3B-focused maps was 4.2 and 4.6 Å (Fig. 3.1H). Using focused classifications, the core structures were resolved to 3.9 and 4.2 Å (Fig. 3.1A,F,G). The local resolution varies widely in each reconstruction, ranging from approximately 3.5 Å at the cores to >20 Å at the peripheries (Fig. 3.1I,J). Grid preparation and data acquisition were carried out by Dr. Karl Bertram (Department of Structural Dynamics, MPI-BPC) (see 2.2.8 and 3.2B).

A model of each complex was built using an integrative approach. Published cryo-EM and crystallographic structures were docked into moderately resolved (ca. 4-10 Å) map regions and XL-MS data was used to guide placement of less well resolved regions at flexible interfaces or the periphery of the complexes (Table B.2). Each spliceosomal intermediate has a relatively similar core structure, comprised of an approximately triangular body (Fig. 3.3). The resulting models allow for inspection of previously uncharacterized intermediates of spliceosome activation. The resulting reconstructions are denoted pre-B$^{act\text{-}1}$ and pre-B$^{act\text{-}2}$, as they represent pre-activated complexes which lack U4 snRNP but do not yet have a mature catalytic center (see sections 3.3, 3.4, 3.5, 3.6, 3.7). Model building was overseen by Dr. Berthold Kastner (Department of Cellular Biochemistry, MPI-BPC) (see 2.2.10.

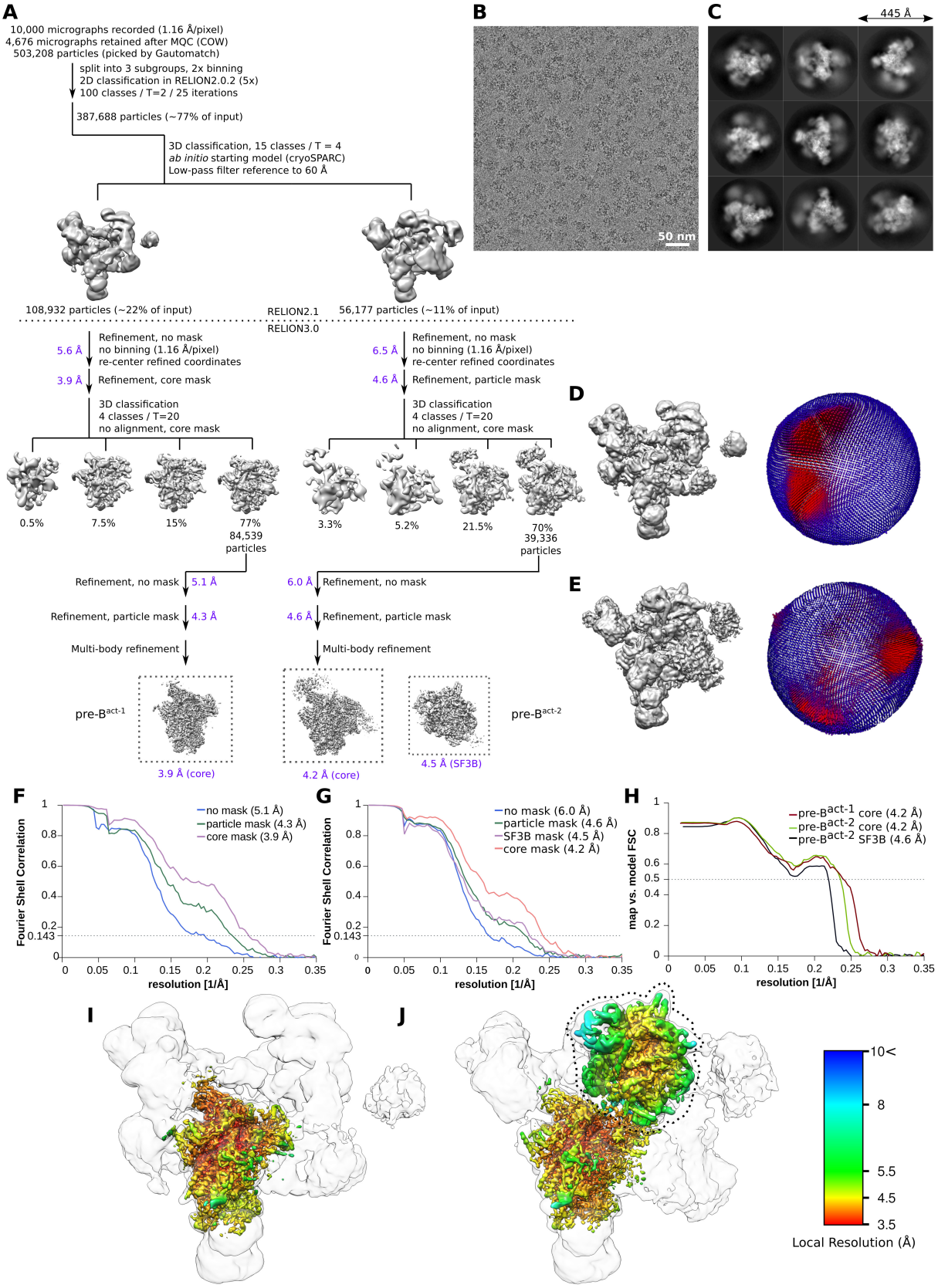

Fig. 3.2: Image processing and reconstruction of cryo-EM data. (A) Computational sorting scheme of cryo-EM data. For a detailed explanation of data acquisition and image processing, please refer to sections 2.2.8 and 2.2.9 of "Methods." (B) Cryo-EM micrograph and (C) representative 2D class averages. (D) Overall cryo-EM reconstruction (left) and angular distribution (right) of pre-B^{act-1} and (E) overall cryo-EM reconstruction (left) and angular distribution (right) of pre-B^{act-2}, where red shading indicates relative number of particle images assigned to a particular orientation. Resolution estimates of pre-B^{act-1} (F), pre-B^{act-2} (G) at the 0.143 Fourier Shell Correlation (FSC) threshold. Map-to-model FSC (0.5 threshold) of pre-B^{act-1} and pre-B^{act-2}. Local resolution estimates of pre-B^{act-1} (I) and pre-B^{act-2} (J) masked reconstructions (colored densities) and overall reconstructions (gray overlays). From [131]. Reprinted with permission from AAAS.

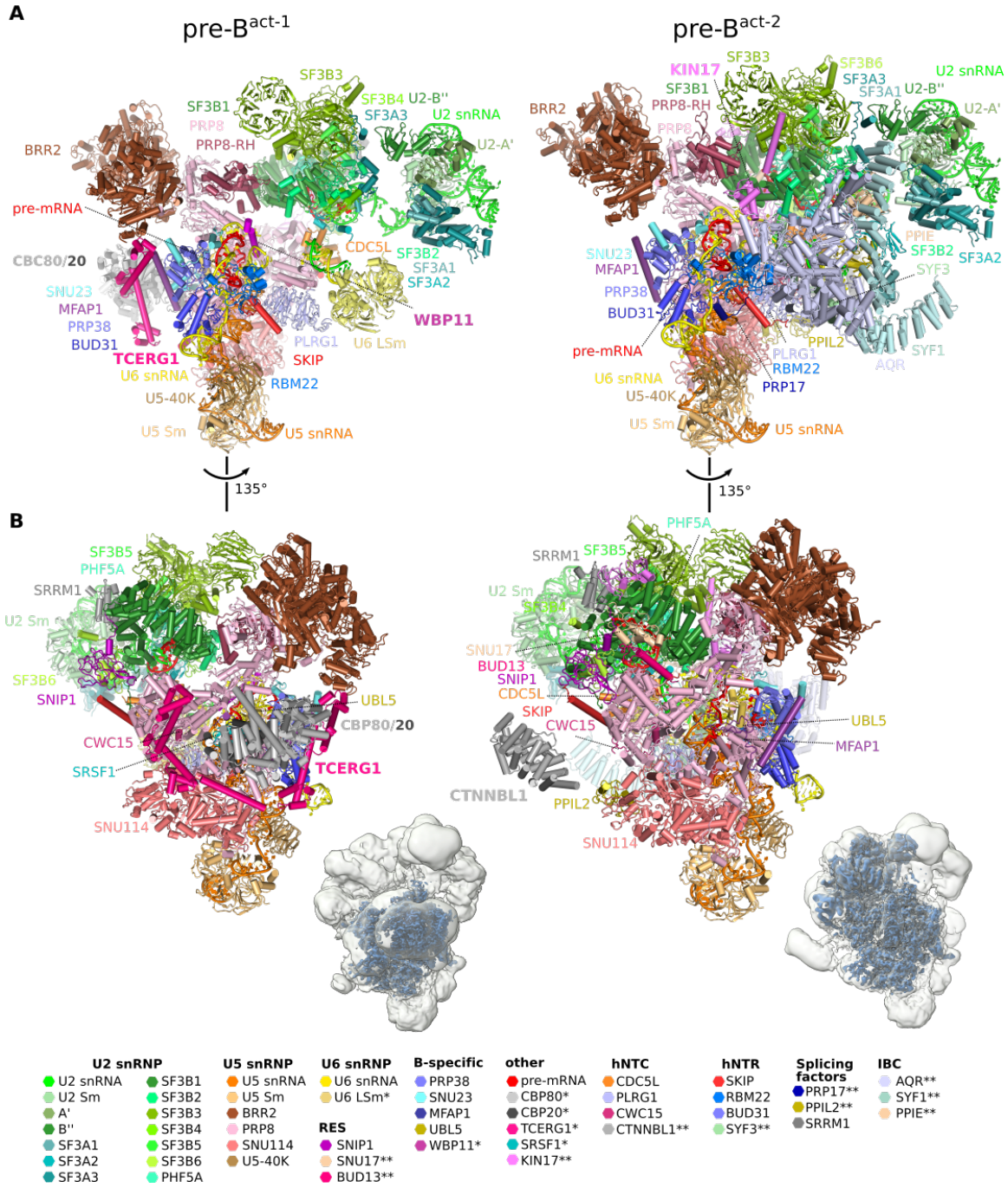

Fig. 3.3: Structural overview of pre-B^{act-1} and pre-B^{act-2} Model composition of pre-B^{act-1} (A,B; left panel) and pre-B^{act-2} (A,B; right panel). All modeled components are labeled within the 3D models and indicated in the table at the bottom of the figure. Components belonging to snRNPs or specific subcomplexes or groups (RES, B-specific, hNTC, hNTR, splicing factors, IBC) are indicated in bold text. Cryo-EM maps are shown as insets (B; left and right), with the higher resolved masked reconstructions shown in dark blue and the overall reconstructions shown in light gray overlay. From [131]. Reprinted with permission from AAAS.

3.3 PRP8 maintains an open conformation in pre-B^{act}

Analysis of the pre-B^{act} structures indicates that the central scaffold protein of the spliceosome, PRP8, is in an open conformation (Fig. 3.4). At some point of spliceosome activation, PRP8 undergoes a large-scale conformational change from open to closed, resulting in a compact arrangement found in the activated spliceosome (B^{act}) (Fig. 1.11). During this conformational change, a large portion of PRP8's C-terminal region (PRP8$^{RT/En}$) clamps down onto its N-terminal domain (PRP8^{NTD}). Conversion from the open to the closed conformation of PRP8 requires loss of U4 snRNP and dissociation of the B-specific proteins (see section 1.4.1). The closed conformation forms a chamber that accommodates the active site, and is therefore a critical event during spliceosome activation.

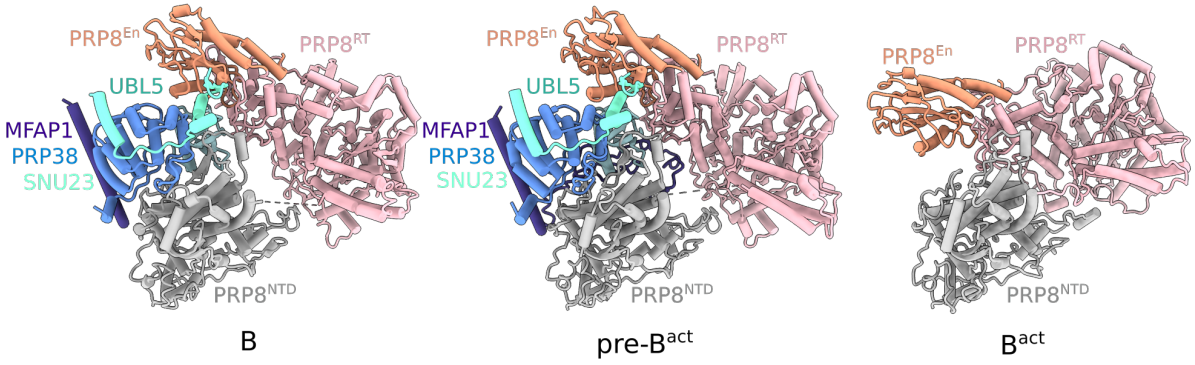

Fig. 3.4: PRP8 is held in an open conformation by B-specific proteins in pre-B^{act}. The B-specific proteins MFAP1, PRP38, SNU23, and UBL5 are still bound between PRP8^{En}and PRP8^{NTD}in both pre-B$^{act\text{-}1}$ and pre-B$^{act\text{-}2}$ (middle panel), as in the B complex (left; PDBs 5O9Z and 6AHD – from 6AHD, only UBL5, as UBL5 is not modeled in PDB 5O9Z). This prevents open-to-closed conformational change of PRP8. All complexes are aligned to the PRP8^{NTD}of B^{act} (PDB 6FF7).

The B-specific proteins MFAP1, PRP38, SNU23, and UBL5 must be dissociated from their binding site between the PRP8^{En} and PRP8^{NTD} domains in order for PRP8 to be converted from an open to a closed conformation during activation (Fig. 3.4). In pre-B^{act}, the B-specific proteins UBL5, PRP38, MFAP1, and SNU23 are still present, holding PRP8 in an open conformation (Fig. 3.4). UBL5 is situated between PRP8^{En} and PRP8^{NTD} domains, preventing closure of PRP8. SNU23 and PRP38 remain bound alongside the U6/5′ss helix. MFAP1 contains an α-helix that is associated with PRP38 as well as a C-terminal region of 80 amino acids that forms multiple interactions with UBL5, loop I of U5 snRNA, PRP8^{SWL}, and the 5′ exon. Thus, the B-specific proteins MFAP1, PRP38,

SNU23, and UBL5 remain bound even after the dissociation of U4 snRNP, holding PRP8 in the same open conformation as in the pre-catalytic spliceosome (B complex).

PRP8^{SWL} points upwards (i.e., away from U5 Sm) in the B complex and downwards in B^{act}, indicating dynamic repositioning of this domain following activation [122]. In pre-B^{act}, PRP8^{SWL} remains in its B complex position and forms contacts with the C-terminal region of MFAP1, which likely clamps PRP8^{SWL} in place (Fig. 3.5). PRP8^{SWL} repositioning appears to coincide with the closed conformation of PRP8, which can only occur following the displacement of B-specific proteins.

The distal C-terminal domains of PRP8^{RH} and PRP8^{Jab1} domains were localized in both pre-B^{act-1} and pre-B^{act-2}, and in largely different positions as found in the B or B^{act} complex, further underscoring the drastic repositioning of these domains at every stage of spliceosome assembly (Fig. 3.5). PRP8^{RH} is rotated between pre-B^{act-1} and pre-B^{act-2}, occupying a gap between BRR2 and the U2 snRNP proteins SF3B1/SF3B3 (Fig. 3.5), replacing the bridge of SMU1/RED that previously bridged BRR2 and U2 snRNP in the B complex. The PRP8^{Jab1} domain is anchored to BRR2, maintaining the association of BRR2 to the spliceosome following its unwinding of the U4/U6 snRNA duplex.

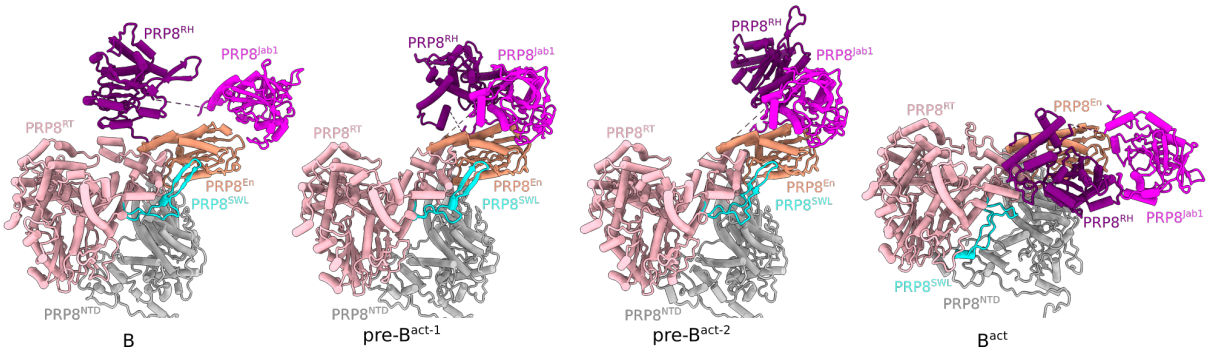

Fig. 3.5: PRP8^{RH} and PRP8^{Jab1} occupying structurally unique positions in both states of pre-B^{act}. In both pre-B^{act-1} and pre-B^{act-2}, PRP8^{RH} and PRP8^{Jab1} occupy positions and orientations previously not observed in the spliceosome. PRP8^{SWL} is also shown for comparison among the B (PDB 5O9Z), pre-B^{act-1}, pre-B^{act-2}, and B^{act} (PDB 6FF7) structures. In the B complex and pre-B^{act} complexes, PRP8^{SWL} points upwards. In B^{act}, PRP8^{SWL} is repositioned and points downwards. All structures are aligned to PRP8^{En}(aa 1579-1754) of B^{act}.

3.4 Novel repositioning of BRR2 and U2 snRNP

BRR2, the ATP-dependent RNA helicase responsible for unwinding the U4/U6 RNA duplex to allow for release of U4 snRNP during activation, has undergone a complex rotational movement following release of U4 snRNP, yet has not assumed its final position corresponding as in B^{act} (Fig. 3.6). BRR2 has not been documented in this orientation relative to the rest of the spliceosomal body, and exhibits a similar orientation in both pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$. The catalytically active region of BRR2 ($BRR2^{N\text{-}term}$) is facing away from $PRP8^{SWL}$, while $BRR2^{C\text{-}term}$ points toward $PRP8^{SWL}$. In B^{act}, $BRR2^{N\text{-}term}$ contacts the U2 snRNP 5′ domain[1]; in pre-B^{act}, BRR2 is not directly contacting the U2 snRNP 5′ domain, but is loosely bridged to U2 snRNP by $PRP8^{RH}$.

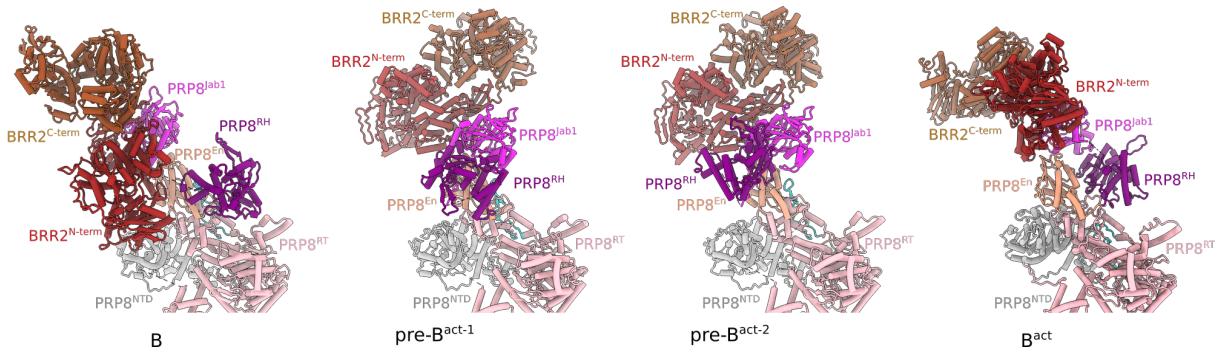

Fig. 3.6: BRR2 is repositioned to a novel orientation in pre-B^{act}. In both pre-B^{act} structures, BRR2 has been rotated upon its contact with the PRP8, $PRP8^{Jab1}$. BRR2 occupies a similar position in both pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$, distinct from its position in the B complex (PDB 5O9Z) and B^{act} (PDB 6FF7). All structures are aligned to $PRP8^{En}$(aa 1579-1754) of B^{act}.

U2 snRNP, essential for the construction of the spliceosome's catalytic center, is initially separated from the core of the spliceosome by U4 snRNP in the B complex (Fig. 3.7, upper left panel). Following unwinding of the U4/U6 RNA duplex by BRR2 to result in dissociation of U4 snRNP, U2 snRNP has begun to dock with the U5 snRNP at the core of the spliceosome, but has not yet been tightly integrated into the complex as in B^{act} (Fig. 3.7, right panels). U2 snRNP is thus flexible in both pre-B^{act} structures, but especially so in pre-$B^{act\text{-}1}$. Many proteins of the U2 snRNP 5′ domain in pre-$B^{act\text{-}2}$ can be clearly mapped to the density based on clearly resolved secondary structure elements,

[1] "U2 snRNP 5′ domain" refers to the SF3b and PHF5A proteins and 5′ half of the U2 snRNA, including portions contributing to helix I, helix II, the U2/BS helix; "U2 snRNP 3′ domain" encompasses the Sm, A', B", SF3a proteins and the 3′ half of the U2 snRNA; see Figs. 1.3 and 3.3.

including those of the larger components SF3B1 and SF3B3. The U2 snRNP 3′ domain is likely not yet stably docked and therefore not as well resolved as in the B^{act} complex. $PRP8^{RH}$ separates SF3B1/SF3B3 and BRR2 in both pre-B^{act} structures (Fig. 3.7, right panel). The U2/BS helix is also distinguishable in both pre-B^{act} reconstructions, although is better resolved in pre-$B^{act\text{-}2}$.

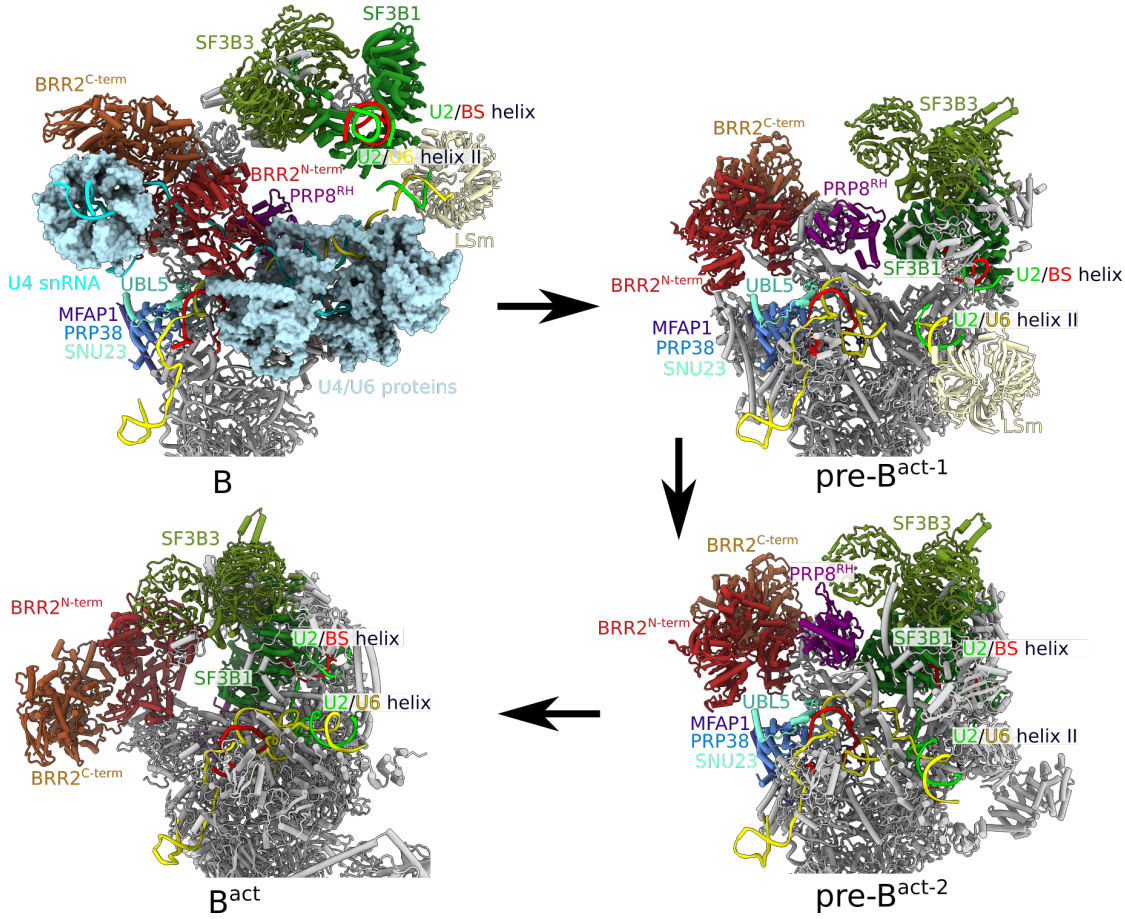

Fig. 3.7: Repositioning of BRR2 and docking of U2 snRNP during activation. The 3′ domain of U2 snRNP and several proteins are omitted from the figure for clarity. Release of U4 snRNA allows for rearrangement of BRR2 and components of the U2 snRNP. Relevant protein and RNA components from the B complex (PDB 5O9Z), pre-$B^{act\text{-}1}$, pre-$B^{act\text{-}2}$, and B^{act} (PDB 6FF7) are highlighted (clockwise from upper left to bottom right quadrant). The U4/U6 proteins, bound to U4 snRNA, occlude a large cavity and prevent docking of U2 snRNP onto PRP8 and the rest of the spliceosome. Following U4/U6 snRNA unwinding, BRR2 becomes repositioned and U2 snRNP as well as U2/U6 helix II and the LSm proteins are moved toward the body of the spliceosome (grey). U2 snRNP becomes progressively more tightly associated with the spliceosome in pre-$B^{act\text{-}2}$ and finally in B^{act}, in which the B-specific proteins are dissociated and PRP8 is in a closed conformation, and BRR2 directly contacts SF3B3 of U2 snRNP.

3.5 Recruitment of proteins at distinct stages of activation in pre-B^{act}

3.5.1 RES and SRRM1

The Retention and Splicing complex (RES) is required for splicing and consists of three proteins: SNIP1, SNU17, and BUD13 (Fig. 1.14) [116, 206]. RES is not present in the B complex, but joins the spliceosome at a yet undetermined stage of activation. The FH domain of SNIP1 in pre-B^{act-1} is localized at an interface between U2 and U5 snRNPs, where it interacts with PRP8^{RT}, SKIP, and SF3B6 (Fig. 3.8, left panel). Densities for BUD13 and SNU17 were not detected in pre-B^{act-1}, but become apparent in pre-B^{act-2}. BUD13 and SNU17 are positioned adjacent to one another, with SNU17 forming contacts to SF3B1 and BUD13 binding within a gap between SNU17 and PRP8^{RT} (Fig. 3.8, middle panel). The position of RES complex proteins agrees closely with their binding sites in the B^{act} (Fig. 3.8, right panel) [99].

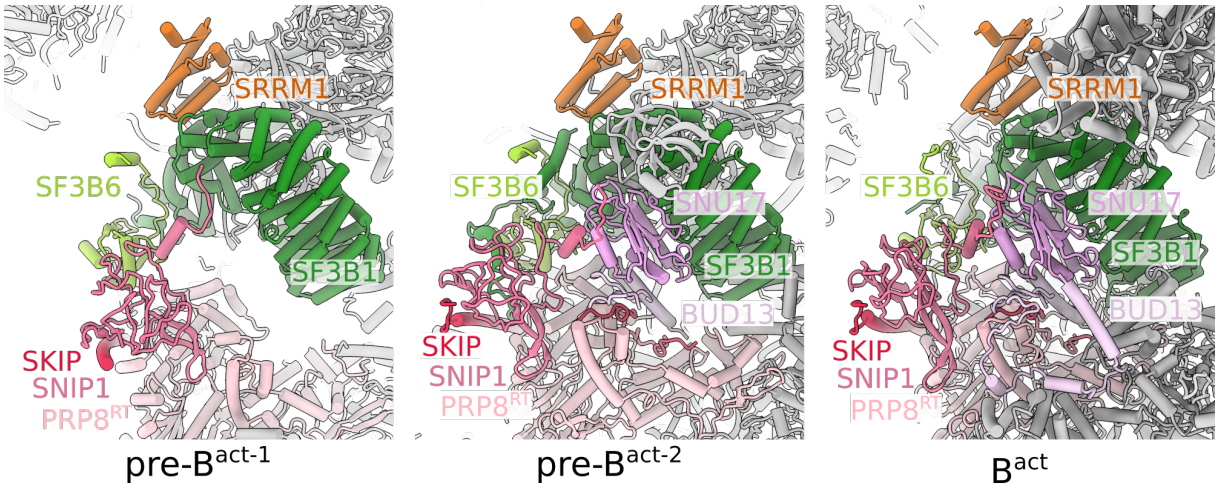

Fig. 3.8: RES integration into pre-B^{act-1} and pre-B^{act-2}. RES proteins are gradually integrated into pre-B^{act-2}, leading to B^{act}. In pre-B^{act-1}, only SNIP1 is observed. SNU17 and several fragments of BUD13 are stabilized in pre-B^{act-2}. In B^{act}, additional regions of BUD13 become stabilized. SKIP (NTR protein) and SRRM1 are shown for orientation. All structures are aligned on SF3B1 of B^{act} (PDB 6FF7).

Nearby the RES complex proteins, SRRM1 is bound. SRRM1 is a protein involved in regulation of splicing (see section 1.4.3) that is associated with B^{act}. SRRM1 interacts

with the U2 snRNP by binding to the SF3B1$^{\mathrm{HEAT}}$ domain adjacent to SF3B6, consistent with the positioning of this protein relative to neighboring proteins in B$^{\mathrm{act}}$ (Fig. 3.8).

3.5.2 NTC

The NineTeen Complex (NTC) is a preformed unit consisting of PRP19, CDC5L, PLRG1, SPF27, CWC15, and CTNNBL1 that joins the spliceosome during activation (Fig. 1.15). The NTC is involved in the removal of the U6 LSm proteins and the stabilization of newly formed U6 snRNA interactions within the spliceosome during activation [214]. Of these proteins, parts of CDC5L, PLRG1, CWC15, and CTNNBL1 can be localized in both or at least one of the pre-B^{act} reconstructions (CTNNBL1 is only localized in pre-B^{act-2}) (Fig. 3.9). CDC5L^{Myb} domains bridge SF3B1 of U2 snRNP to PRP8^{HB}, likely contributing to docking of U2 snRNP to the U5 snRNP. PLRG1^{WD40} is stably docked near the core of the spliceosome, lodged beneath PRP8^{HB}, providing an interface for other proteins such as SKIP (see section 3.5.3) to bind. A short α-helix of CWC15 could be docked into a density near PRP8^{HB}, based on crosslinks as well as the position of this protein in the published B^{act} cryo-EM structure [99]. The ARM repeats of CTNNBL1, a protein previously unmapped in any spliceosomal structure, could be docked based on crosslinks into a large curved-cylindrical density at the outer periphery of pre-B^{act-2}. CTNNBL1 is linked to the body of the complex by a short α-helix of CDC5L as well as the U-box domain of PPIL2. All components of the NTC are docked to positions that are highly similar to their previously observed positions in B^{act}.

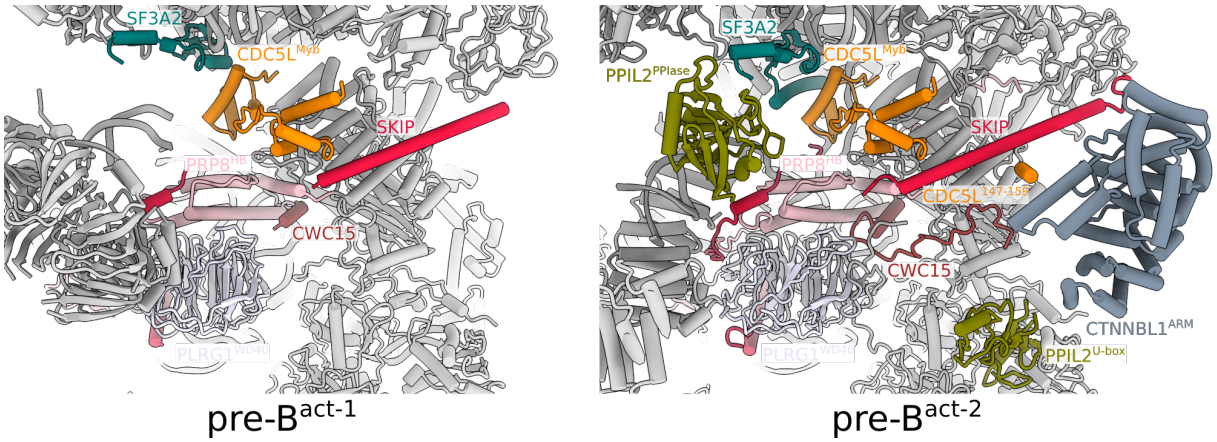

Fig. 3.9: NTC integration into pre-B^{act-1} and pre-B^{act-2}. NTC proteins CDC5L, PLRG1, and CWC15 binding to pre-B^{act} are shown. CDC5L^{Myb} binds above PRP8^{HB}, between SF3A2 (U2 snRNP protein) and SKIP (NTR protein). In pre-B^{act-2} (right panel), CTTNBL1 is attached to the periphery of the complex and additional domains of CWC15 are structured. PPIL2 is shown for orientation. Structures are aligned on PRP8^{NTD}.

3.5.3 NTR proteins

NineTeen Complex Related (NTR) proteins SKIP, RBM22, BUD31, and PPIL1 all bind during spliceosome activation. Of these four proteins, only SKIP, RBM22, and BUD31 were detectable (Fig. 3.10). PPIL1 joins in late stages activation (mature B^{act} and is stabilized by surrounding protein-protein interactions. Several portions of SKIP could be docked throughout the core of each complex (Fig. 3.10); its N-terminal domain interacts with the NTC proteins PLRG1 and CDC5L and the NTR proteins RBM22 and BUD31, as well as $PRP8^{HB}$, while its long helical region spans the outer edge of $PRP8^{RT}$, ending at the RES protein SNIP1. In pre-B^{act-2}, additional regions of SKIP are stably integrated; in particular, a C-terminal strand that is sandwiched between $PRP8^{RT}$ and the RES protein BUD13. SKIP thus forms an extensive "wiring" to connect far-reaching regions of the spliceosome. RBM22 and BUD31 bind near the 5′ SL of U6 snRNA, nestled into an interface on $PRP8^{NTD}$ (Fig. 3.10, left and middle panel). Moreover, RBM22 and BUD31 are bound immediately adjacent to the B-specific proteins MFAP1, PRP38, and SNU23, which are engaged in maintaining the open conformation of PRP8. RBM22 contains a Zinc-finger (Znf) domain and an RNA recognition motif (RRM). $RBM22^{Znf}$ is stably bound in both pre-B^{act-1} and pre-B^{act-2}, yet $RBM22^{RRM}$ is clearly resolved only in pre-B^{act-2}. RBM22 binds the intron distal to the U6/5′ss helix, guiding the intron into the five-membered Intron Binding Complex (IBC).

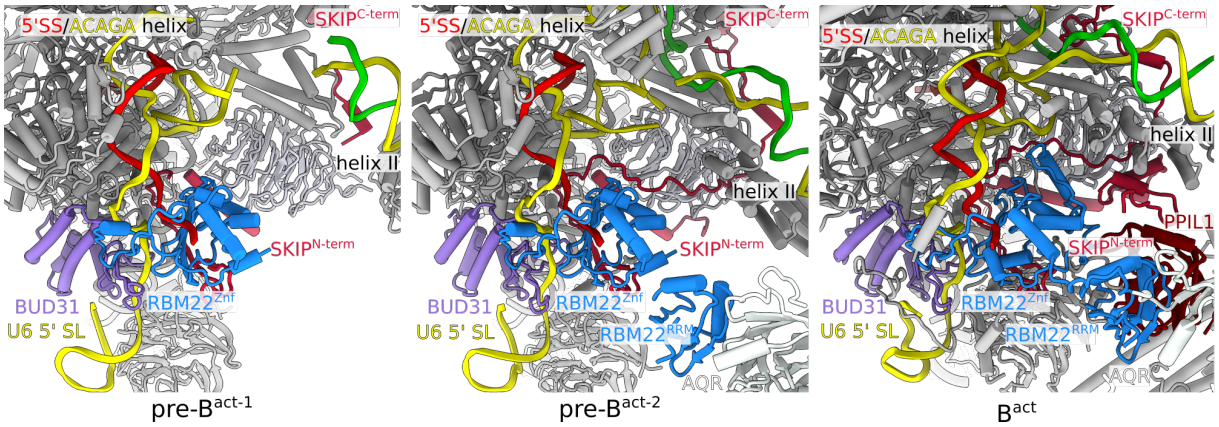

Fig. 3.10: NineTeen complex Related (NTR) proteins. NTR proteins gradually integrated from pre-B^{act-1} to B^{act}. In all structures, BUD31 and $RBM22^{Znf}$ bind to the U6/5′ss helix (i.e., 5′SS/ACAGA helix). $RBM22^{RRM}$ and several portions of SKIP are first observed in pre-B^{act-2} and retained in B^{act}. PPIL1 joins at a later stage of B^{act}. All structures are aligned on $PRP8^{NTD}$of B^{act} (PDB 6FF7).

3.5.4 IBC and SYF3

In the B complex, the 3′ end of U6 and the 5′ end of U2 snRNA are base-paired to form U2/U6 helix II, which remains intact also in the activated spliceosome (B^{act}; see 1.3.5). Helix II is bound by the U6 LSm proteins in the B complex [98]. At some point during spliceosome activation, U6 LSm proteins are replaced by SYF3 and the intron-binding complex (IBC) (see section 3.5.4, [225]). Helix II appears to be flexible in pre-B$^{act\text{-}1}$, as evidenced by a cylindrical-shaped density distal to the U6 ISL and the U2/U6 helix Ib, where helix II is located in B^{act}. Crosslinking data allowed for the placement of the U6 LSm into a globular density near the expected position of helix II (Fig. 3.11), revealing that U6 LSm remains associated with helix II even after loss of U4 snRNP, integration of the NTC, and the large-scale remodeling of the spliceosome's molecular architecture that occurs during spliceosome activation. U6 LSm is replaced by SYF3 and the IBC in pre-B$^{act\text{-}2}$, as its binding site on helix II would otherwise clash with this protein ensemble (Fig. 3.11). The HAT repeats of SYF3 provide a platform on which helix II appears to be stabilized in pre-B$^{act\text{-}2}$ (Fig. 3.19).

The IBC, consisting of five proteins (AQR, SYF1, ISY1, PPIE, and ZNF830; Fig. 1.16) [225], is absent from pre-B$^{act\text{-}1}$, but is stably integrated in pre-B$^{act\text{-}2}$ (Fig. 3.11). IBC components interact with SYF3, upon which U2/U6 helix II is accommodated. In pre-B$^{act\text{-}1}$, helix II is bound by LSm proteins, abrogating binding of SYF3 (Fig. 3.11). Stabilization of the IBC onto the spliceosome is apparently mediated not only by RBM22, but also by SYF3 and SF3B2 and SF3B4. The N-terminal HAT repeats of SYF3 bridge the IBC to the spliceosome by contacting SYF1 (Fig. 3.11). PPIE also bridges the IBC to the spliceosome by providing a stable binding interface to channel the intron from RBM22 to SF3B2/SF3B4 and finally to the SF3B1^{HEAT} domain, where the U2/BS helix is sequestered.

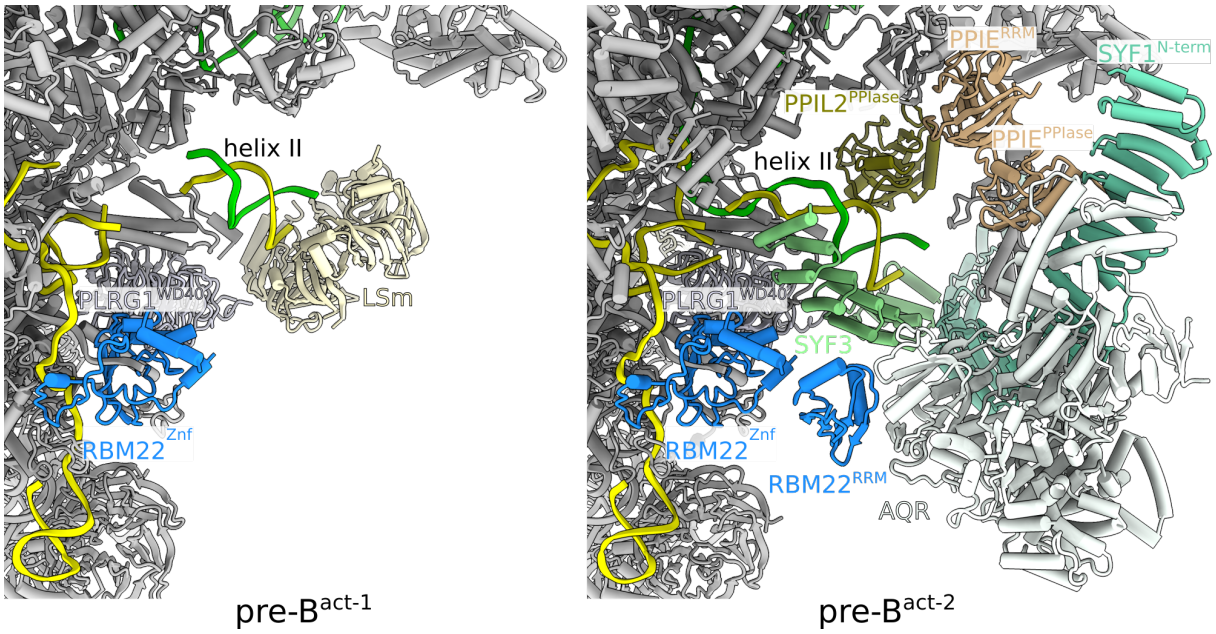

Fig. 3.11: Mutually exclusive interactions of LSm proteins and the IBC/SYF3. In pre-B$^{act\text{-}1}$, LSm proteins are still bound to helix II and interact with PLRG1. LSm proteins are replaced by SYF3 and the IBC (AQR, PPIE, SYF1) in pre-B$^{act\text{-}2}$. PPIL2^{PPIase} also appears first in pre-B$^{act\text{-}2}$. Structures are aligned on PRP8^{NTD}.

3.5.5 PPIL2 and PRP17

Several proteins involved in generation of the B^{act} join during spliceosome activation, including PPIL2 and PRP17. The function of PPIL2 is not well understood. PPIL2 contains a PPIase and a U-box domain [231]. The U-box of PPIL2 binds to SNU114 and crosslinks to the nearby CTNNBL1^{ARM} domain (Fig. 3.9). The PPIase domain of PPIL2 docks near U2/U6 helix II and forms numerous crosslinks to SYF3 (Fig. 3.11 and Table B.1). Densities corresponding to the PPIL2 U-box and PPIase domains are evident in pre-B$^{act\text{-}2}$, yet absent in pre-B$^{act\text{-}1}$, suggesting that PPIL2 may be recruited only at intermediate stages of spliceosome activation. PRP17 is a protein that becomes structured during the formation of the B^{act}. Several short α-helices of PRP17 can be localized in early states of B^{act} while its WD40 domain is stably integrated near the catalytic center at the latter stages of B^{act} formation [99]. In pre-B$^{act\text{-}2}$, a short fragment of PRP17 can be docked alongside SKIP, similar as in B^{act} (Fig. 3.12). Absence of density for PRP17 in pre-B$^{act\text{-}1}$ may indicate either a gradual structuring of PRP17 during the transition from pre-B$^{act\text{-}1}$ to pre-B$^{act\text{-}2}$ or a recruitment of this protein in the same transition.

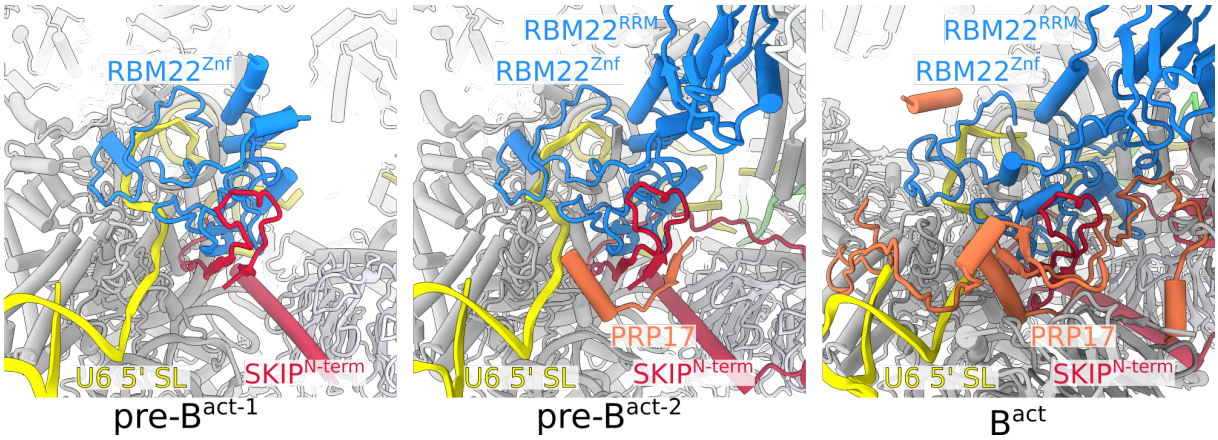

Fig. 3.12: Integration of PRP17 during activation. PRP17 is observed in pre-B^{act-2} but not in pre-B^{act-1}, interacting with the NTR proteins SKIP and RBM22. Additional regions of PRP17 are stabilized in B^{act} (right panel). All structures are aligned to PRP8^{NTD}of B^{act} (PDB 6FF7).

3.6 Identification of previously unobserved factors

3.6.1 Stabilization of pre-B^{act-1} by transiently interacting factors

TCERG-1 is a transcription factor that binds to the C-terminal domain (CTD) of Pol II and has also been identified by multiple studies as a component of the spliceosome (see section 4.2), although no structural evidence detailing the presence of TCERG-1 in the spliceosome has been reported. TCERG-1 contains three WW domains, six FF domains (FF1-FF6) and three coiled-coil domains [279]. TCERG-1 is abundant in the pre-B^{act} sample, as shown by 2D gel electrophoresis followed by tandem mass spectrometry as well as by XL-MS (Fig. 3.1 and Table B.2). In pre-B^{act-1}, two long α-helical densities stretch along the side of the spliceosome, extending from PRP8^{RT} to the RecA2 domain of BRR2^{N-term} (Fig. 3.13). These long helical densities connect compact bundles of short α helices comprising FF domain 1-6. Based on crosslinks, all 6 FF domains of TCERG-1 can be docked onto pre-B^{act-1}. TCERG-1 forms crosslinks to CBP80 as well as to SRSF1, MFAP1, and multiple other proteins (Table B.1). Density for TCERG-1 is completely absent from pre-B^{act-2}, suggesting that it binds only transiently to the pre-B^{act-1}, potentially playing a stabilizing role. TCERG-1 is to our knowledge the first transcription factor that has been observed in the cryo-EM structure of a spliceosome.

The cap-binding complex (CBC) is a bipartite structure composed of two proteins, CBP20 and CBP80, that associates with the 7-methyl guanosine (m^7G) cap at the 5′ end of pre-mRNAs during early spliceosome assembly [280]. The CBC mediates spliceosome assembly in a co-transcriptional manner [281] and facilitates binding of U1 and U5 snRNPs to the pre-mRNA [282]. Although the cap-binding complex is expected to remain bound to the m^7G cap structure and is present in all spliceosome purifications, it has not been localized in any human spliceosome structure to date. In pre-$B^{act\text{-}1}$, but not pre-$B^{act\text{-}2}$, a large density is evident near the exon-binding channel, from which the 5′ end of the pre-mRNA emerges (Fig. 3.13). The crystal structure of the cap-binding complex can be docked into this density based on crosslinks as well as on proximity to the 5′ end of the pre-mRNA. After docking of the cap-binding complex, a low-resolution density between the CBP20 and the 5′ end of the 5′ exon remained. Inspection of crosslinks permitted docking of an RNA binding protein, SRSF1 (Table B.1, Fig. 3.13). SRSF1 is a serine- and arginine-rich (SR) protein containing two RRMs and is known to be implicated in 5′ss recognition by the U1 snRNP (see section 4.2). This is the first report of an SR protein being localized in a spliceosome structure.

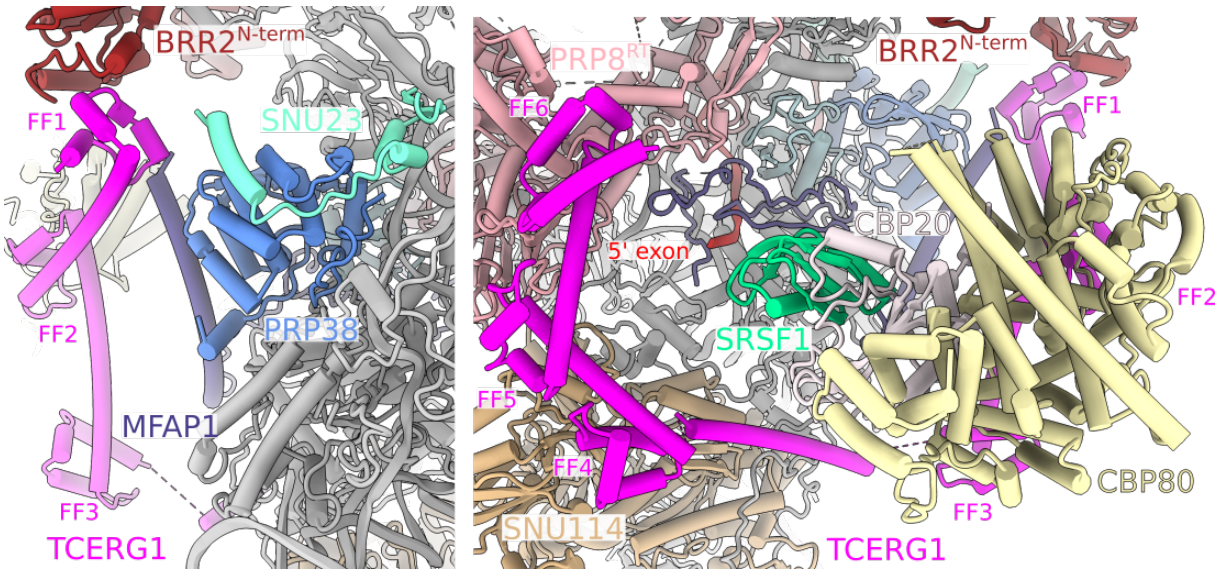

Fig. 3.13: Localization of TCERG-1, CBP80/20, $SRSF1^{RRM2}$ in pre-$B^{act\text{-}1}$. The pre-$B^{act\text{-}1}$ structure is shown in two orientations. TCERG-1 wraps around pre-$B^{act\text{-}1}$ and contains 6 FF domains (FF1-6). FF1 interacts with $BRR2^{N\text{-}term}$ (left panel). In addition to TCERG-1, several other proteins bind to the periphery of pre-$B^{act\text{-}1}$, including the cap-binding complex (CBP80/20), $SRSF1^{RRM2}$.

WBP11 (also known as SIPP1/NPW38BP) is a component of the spliceosome [283]. WBP11 binds to PQBP1 *in vitro* and protein phosphatase 1 (PP1) *in vivo* [283, 284]. The positioning and mechanism by which WBP11 exerts its activity in a biological context is not well understood. Structural evidence of WBP11 has not been reported, although secondary structure predictions of WBP11 exist. In pre-B$^{act\text{-}1}$, an α-helical density intersects the core of the spliceosome, near the nascent catalytic center, situated between the U6/5′ss helix and the U6 ISL. Multiple crosslinks to PRP8, PLRG1, and SNU23 allow for the placement of WBP11 (aa 24-59), a predicted α-helix, at this position (Fig. 3.14) (Table B.2). WBP11 passes along the U6 ISL immediately above U74, a nucleotide that is part of the catalytic center and must be flipped out of the stem-loop in order to participate in triple helix formation. Although crosslinks to the binding partner of WBP11, PQBP1, are also present in the XL-MS dataset (Table B.1), no corresponding density was observed in pre-B$^{act\text{-}1}$. In pre-B$^{act\text{-}2}$, density for WBP11 between the U6/5′ss helix is absent, indicating that WBP11 is displaced from this position during catalytic center formation (see 4.3.2 for detailed discussion).

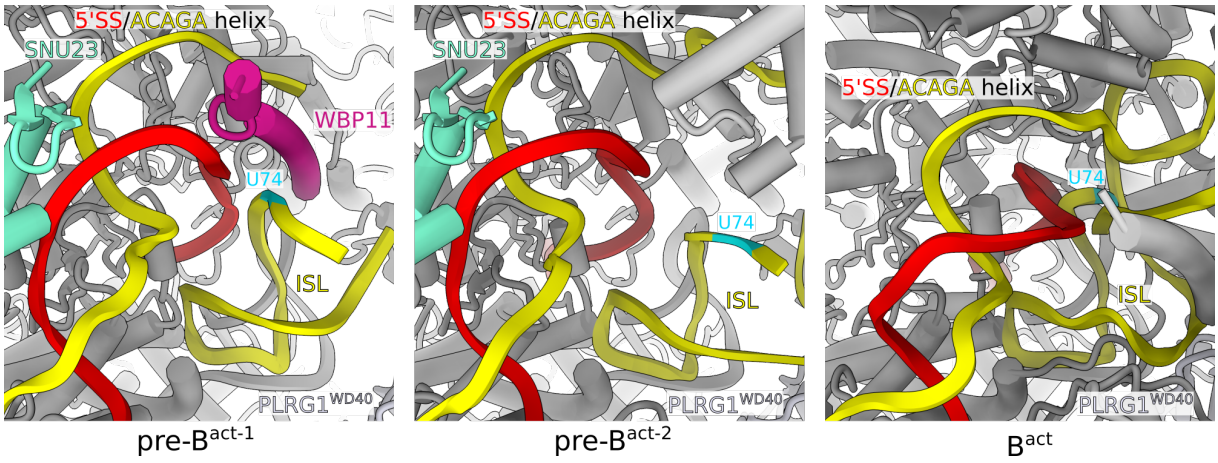

Fig. 3.14: Localization of WBP11 in pre-B$^{act\text{-}1}$. An α-helix of the B-specific protein WBP11 intersects the core of pre-B$^{act\text{-}1}$ but is displaced in pre-B$^{act\text{-}2}$ and B^{act} (PDB 6FF7). WBP11 passes between the U6/5′ss helix (i.e., 5′SS/ACAGA helix) and the U6 ISL, in particular immediately above the phosphate backbone of U6 nt U74.

3.6.2 Coordination of a network of protein-protein interactions in pre-B^{act-2} by the DNA/RNA-binding protein KIN17

KIN17 is an RNA/DNA-binding protein implicated in UV-induced DNA damage repair, although an exact function for this protein has yet to be documented [285, 286]. The solution structure of a KIN17 fragment demonstrated that it forms a winged-helix (WH) motif [287]. KIN17 has not been observed in spliceosomal structures, yet it appears in mass spectrometry data of spliceosomal complexes [75]. Mass spectrometry analysis of the pre-B^{act} sample indicated the presence of KIN17 (Fig. 3.1). In pre-B^{act-2} only, the WH motif can be docked into a globular density between the U6 ISL and $PRP8^{RH}$, based on crosslinks to PRP8 and SNU23 (Fig. 3.15) (Table B.2). An α-helix of KIN17 is connected by a short linker to the WH domain, and is located in a density running between the WD40 domains of SF3B3-BPB and SF3B3-BPC. The SH3 domain of KIN17 was docked into a low-resolution density at the periphery of the complex, guided by crosslinks proceeding C-terminally from the SF3B3-bound α-helix of KIN17 as well as a crosslink between the SH3 domain and SF3B1 (Fig. 3.15, right panel). Binding of $KIN17^{SH3}$ is in a mutually exclusive position compared to the DEAH box helicase PRP2 binding site in B^{act}, suggesting that KIN17 transiently binds and must be displaced to allow for PRP2 to be integrated to the complex *en route* to catalytic activation. Positioning of the KIN17 WH and α-helix may imply a possible role of KIN17 in organizing protein-protein and/or protein-RNA interactions surrounding the active site, as it binds close to the catalytic core of the complex.

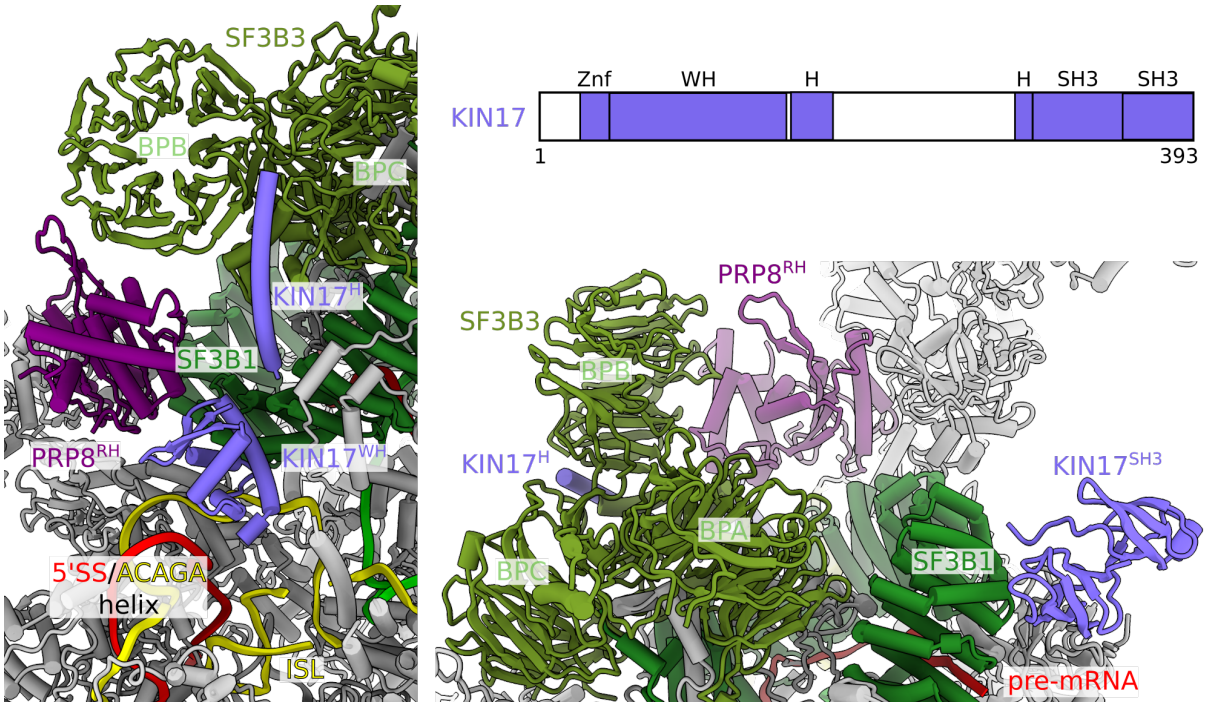

Fig. 3.15: KIN17 binds transiently in pre-B$^{act\text{-}2}$. KIN17 interacts with SF3B3 and PRP8^{RH}, and SF3B1 in pre-B$^{act\text{-}2}$. The domain structure of KIN17 is depicted for orientation (upper right panel). Znf - zinc finger, H - helix. 5′SS/ACAGA helix label indicates U6/5′ss helix.

3.7 Early steps in the stepwise folding of the U2/U6 active site RNA

The spliceosome's catalytic center – comprised of the U6 ISL and U2/U6 helix Ia and Ib as well as interactions with the U6/5′ss helix – begins to form during spliceosome activation. Pre-B^{act} structures reveal new insights into the order in which these secondary structures form *en route* to the fully formed catalytic center of B^{act} (Fig. 3.16). Pre-B$^{act\text{-}1}$ and pre-B$^{act\text{-}2}$ contain similar features, yet also exhibit stark differences in the maturity of the formation of their RNA networks (Fig. 3.18).

3.7.1 Formation of U6 ISL in pre-B$^{act\text{-}1}$

Following liberation of U4 snRNA from the spliceosome, a single-stranded region (nts A56 to A78) of U6 snRNA refolds into the catalytically essential internal stem-loop (ISL).

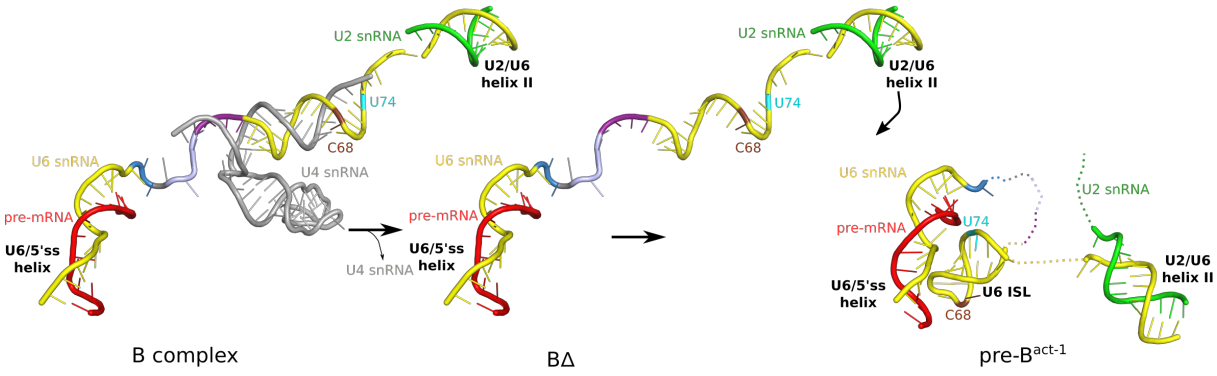

Fig. 3.16: Reorganization of U2/U6 snRNA upon unwinding of U4 snRNA by BRR2. 3D representation of changes in the RNA network that occur during activation. B complex (left; from PDB 5O9Z) contains U4 snRNA, which is dissociated by BRR2 unwindase activity to allow U6 snRNA to engage in new base-pairing interactions. A hypothetical intermediate, BΔ, is shown to illustrate how the absence of U4 snRNA liberates U6 the nucleotide sequence that are rearranged. In pre-B^{act-1}, U6 ISL is formed and U2/U6 helix II has been repositioned. Unstable parts of the catalytic center in pre-B^{act-1} are indicated with stippled lines corresponding to functionally relevant sequences.

The loop of U6 ISL is formed in both pre-B^{act} structures and is nested within the core of the spliceosome. The loop region of the ISL (nts G65 to A69) in pre-B^{act-2} exhibits a similar arrangement to that of B^{act}. The loop region (nts G65 to A69) of the ISL is docked to its final position on $PRP8^{NTD}$, as in B^{act}. $PRP8^{NTDL}$, a flexible strand of basic residues that meanders around the loop of the ISL in the activated spliceosome, is partially structured in pre-B^{act-1} and pre-B^{act-2} (Fig. 1.11). The residues of $PRP8^{NTDL}$ (aa 659 to 676) connecting $PRP8^{NTD}$ to $PRP8^{HB}$ appear to be flexible in both pre-B^{act} reconstructions and was therefore not observed. Several other regions of proteins that encompass the U6 ISL in the fully formed catalytic network of B^{act}, such as SF3A2 and SF3B2 are also not observed. The mid-stem region (nts G75 to A73 and C60 through C61) of the ISL appears to be distorted in both pre-B^{act} complexes relative to B^{act}, with pre-B^{act-1} and pre-B^{act-2} each exhibiting unique conformations in this region 3.17. Nucleotides (A73 and U74) of the mid-stem region important for the construction of the triple helix as well as for metal ion coordination appear to be stacked into the ISL, instead of flipped out as in B^{act}, although the resolution of the mid-stem region does not permit detailed comparison of the configuration of these nucleotides. The lower-stem of the ISL (nts A76 to A78 and U57 to G59), distal to the loop region, is not well resolved and therefore could not be modeled.

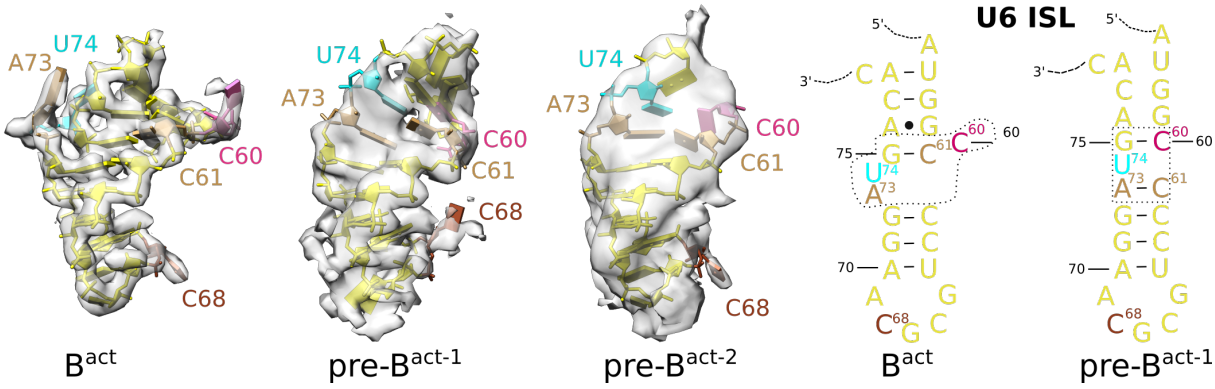

Fig. 3.17: Unique conformations of U6 ISL in pre-B^{act} and proposed alternative base-pairing scheme. Density fits of the U6 ISL in B^{act} (EMD 4255 and PDB 6FF4), pre-B^{act-1}, and pre-B^{act-2} suggest rearrangement of U6 ISL base-pairing, leading to unique conformations of the ISL. The base-pairing scheme of B^{act} and a proposed base-pairing scheme of pre-B^{act-1} (right panel) are indicated. Alternative base-pairing of the mid-stem region (nts C60, C61, A73, and U74) may lead to a more compact (B^{act}) or elongated (pre-B^{act-1}) ISL conformation. Adapted from [131]. Reprinted with permission from AAAS.

3.7.2 Stable formation of U2/U6 helix Ib in pre-B^{act-2}

In the fully formed catalytic center of B^{act}, the U6 ISL is tightly packed against two short RNA duplexes, U2/U6 helix Ia and Ib (Fig. 3.18, right panel). U6 (nt A56) connects the ISL to helix Ib. The U6 nucleotides (A53, G54, and C55) of helix Ib form the AGC catalytic triad and are essential for splicing catalysis, forming part of the RNA triple helix of the active site (see Fig. 3.18). In addition, the U6/5′ss helix is situated close to helix Ib and the ISL, allowing G46 and A47 of the U6 ACAGA(GA) sequence to form Hoogsteen interactions making up the third strand of the triple helix.

Pre-B^{act-1} does not exhibit a stably formed helix Ib, whereas pre-B^{act-2} shows clear density for helix Ib (Fig. 3.19A,C,E). In pre-B^{act-2}, helix Ib is docked to its final position on PRP8^{HB} and CDC5L^{Myb} as in B^{act} (Fig. 3.19G,F). Basic residues of PRP8 (K774, R781, K1020) surround the flipped-out U2 nts A23 and A24 separating helix Ia and helix Ib nucleotides, clamping helix Ib into place in pre-B^{act-2} (Fig. 3.19G). The second RNA duplex, U2/U6 helix Ia, could not be detected in either pre-B^{act} reconstruction. At low thresholds in an unmasked, unsharpened map of pre-B^{act-2}, a strand-like density extends from the U6 nts of helix Ib to the U6/5′ss helix (Fig. 3.19D). The U6 nucleotides that will later contribute to U2/U6 helix Ia are thus initially in an extended conformation, connecting the U6/5′ss helix and helix Ib prior to base-pairing with U2 snRNA (nts G25,

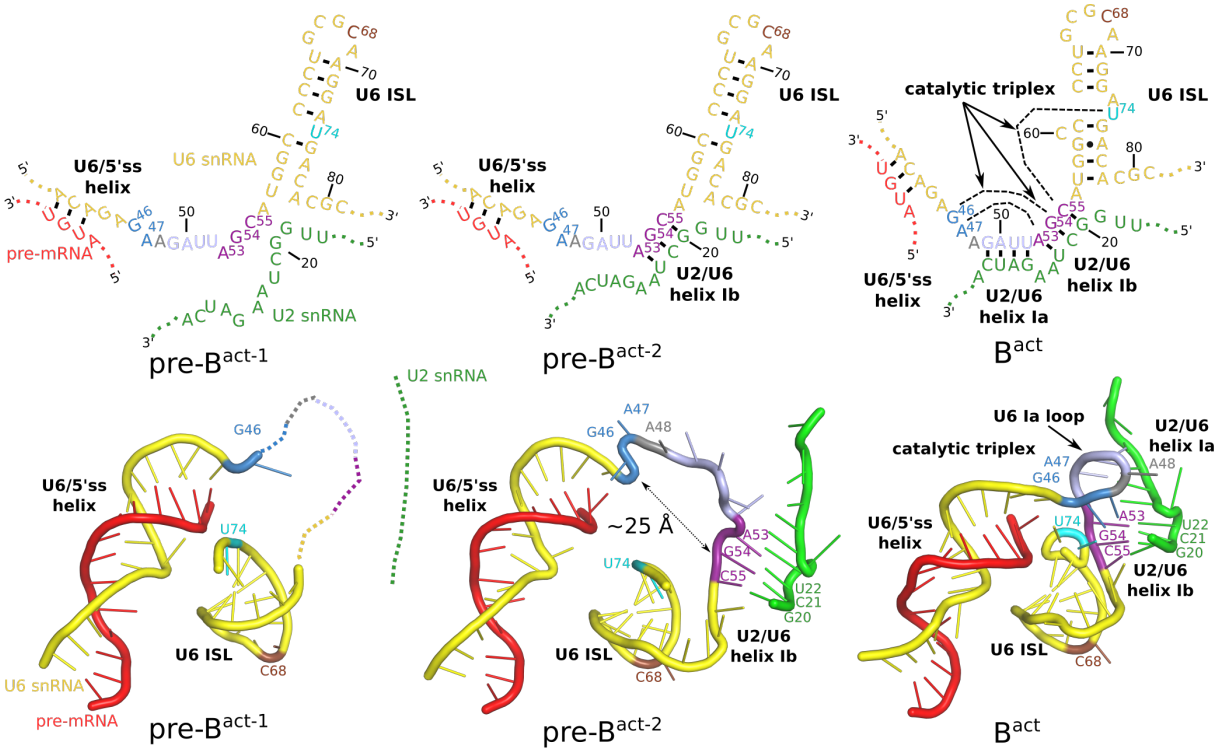

Fig. 3.18: Establishment of catalytic center through interaction of U2/U6 snRNAs. The stepwise formation of the catalytic center is shown as a 2D diagram (upper row) and 3D models (bottom row) from left-to-right for pre-B^{act-1}, pre-B^{act-2}, and B^{act} (from PDB 6FF4). Unstable regions of RNA are indicated with colored stippled lines. The U6 nts C68 and U74 are labeled for orientation. Nucleotides that form the triple helix are indicated in the 2D diagram of B^{act} (upper-right panel). Adapted from [131]. Reprinted with permission from AAAS.

A26, U27, C28) to form helix Ia (Fig. 3.18, right panel). This extended conformation must later be changed to a more compact arrangement to allow for the triple helix to form (Fig. 3.18).

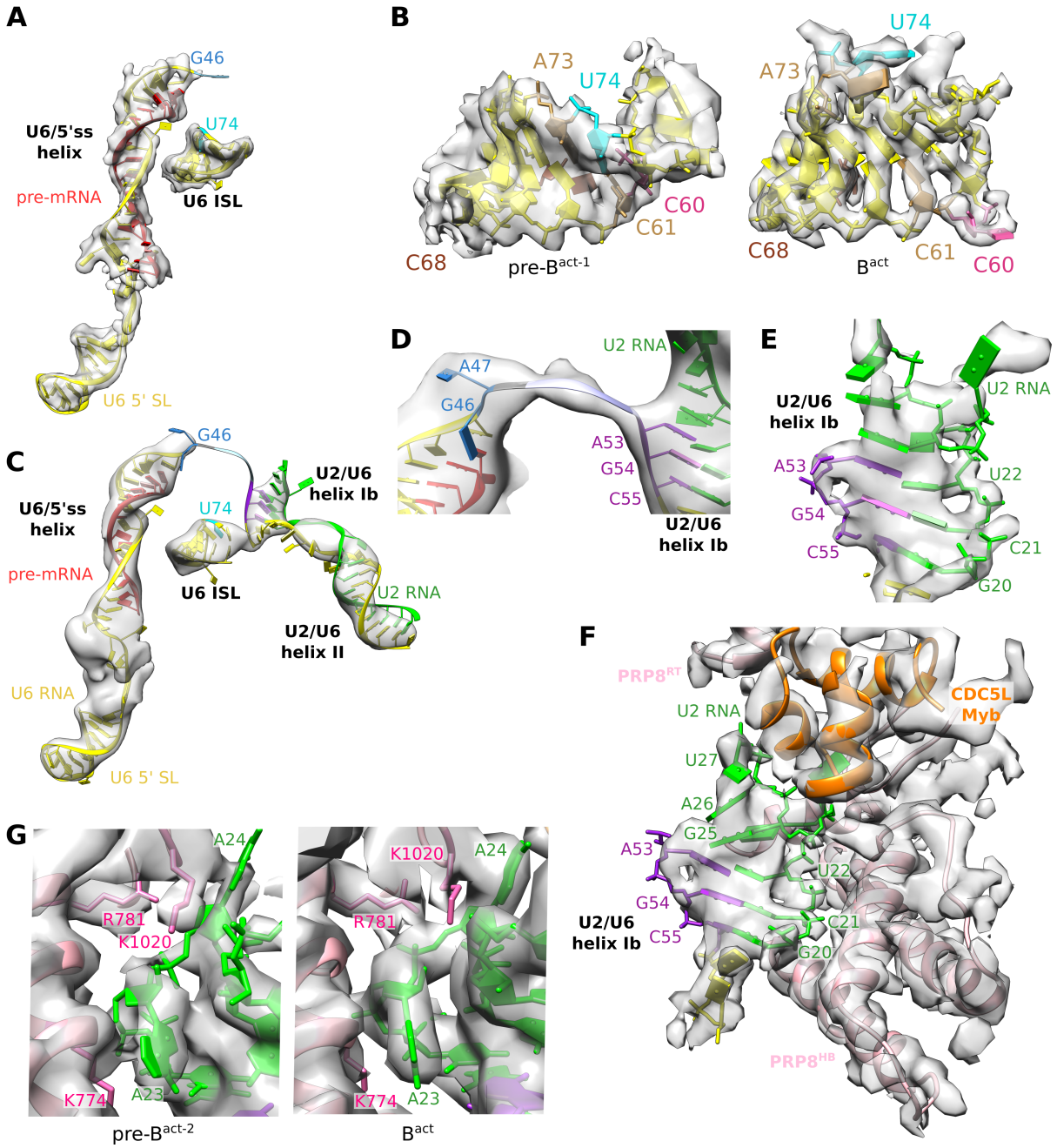

Fig. 3.19: EM densities of RNA structures involved in formation of the active site. Cryo-EM densities of RNA structures of pre-B^{act-1} and pre-B^{act-2}. (A) U6/5′ss, U6 5′ SL, and U6 ISL of pre-B^{act-1}. (B) Comparison of U6 ISL model and density from pre-B^{act-1} (left) and B^{act} (right; EMD 4255 and PDB 6FF4). (C) U6/5′ss, U6 5′ SL, and U6 ISL, U2/U6 helix Ib, helix II of pre-B^{act-2}. Note that the linker region (U6 loop Ia) between U6/5′ss and helix Ib is not visible at this threshold but is shown in (D) at a lower map threshold. (E) Fit of U2/U6 helix Ib in pre-B^{act-2}. (F) Docking of U2/U6 helix Ib to $PRP8^{HB}$ and organization of NTC protein $CDC5L^{Myb}$ in pre-B^{act-2}. (G) Comparison of basic residues of PRP8 (R774, R781, K1020) that surround the flipped-out U2 nts A23 and A24, which separate U2 helix Ib and Ia, in pre-B^{act-2} and B^{act} (EMD 4255 and PDB 6FF4). From [131]. Reprinted with permission from AAAS.

4 Discussion

The pre-B^{act} structures reported here provide new insights into the complex assembly pathway of the spliceosome during activation. By stalling the spliceosome during the formation of its catalytic center, we were able to observe novel aspects of several strategies the spliceosome uses to form the U2/U6 RNA network at intermediate stages of activation, ensuring accurate folding of the active site. These strategies (discussed in detail in the following sections) include large-scale dynamics in the organization of protein-protein interactions; potential links to co-transcriptional splicing; mutually exclusive interactions; and gradual stabilization of proteins surrounding the active site. All of these modalities appear to contribute to the protein-guided folding of U2/U6 RNA, beginning with the U6 ISL, followed by the U2/U6 helix Ib, and finally the U2/U6 helix Ia and resulting catalytic triplex. In addition to highlighting details of the human spliceosome's activation step, our structures bring up questions for further studies into the necessity and specific roles of various proteins or protein domains in productive formation of activated spliceosomes.

4.1 Functional implications of large-scale dynamics of BRR2 and PRP8

BRR2 is the only DExD/H-box ATP-dependent helicase that remains bound to the spliceosome following its unwindase action and occupies distinct positions in the tri-snRNP, B complex, and B^{act}, as well as in later stage complexes) [99, 122, 195, 196]. BRR2 is kept tethered to the spliceosome by its connection to $PRP8^{Jab1}$. Retention of BRR2 following U4 snRNP dissociation suggests BRR2 may additionally carry out other roles in the spliceosome besides ATP-dependent helicase activity. Prior to U4/U6 snRNA

unwinding, BRR2 is kept in an inhibited state by the C-terminal region of PRP8^{Jab1} [191]. In the pre-B^{act}, BRR2 is repositioned following U4/U6 snRNP unwinding and is kept in a very different position than in the B or B^{act} complex, with its C-terminal region being rotated relative to its position in B^{act} (Fig. 3.6). BRR2 must be tightly regulated prior to activation to prevent premature unwinding of the U4/U6 snRNA duplex. This regulation is carried out by multiple factors and domains, including but not limited to PRP8^{RH}, PRP8^{Jab1}, and FBP21. The trigger for BRR2 unwindase activity is unclear, but appears to be closely tied to changes in the factors that regulate BRR2. While the pre-B^{act} structures do not offer explicit insight into BRR2 regulation, they do indicate that factors governing BRR2 regulation are repositioned following release of U4 snRNA by BRR2. BRR2 repositioning appears to be closely tied to the dynamics of PRP8^{Jab1}, which is mobile in all spliceosome intermediates from the tri-snRNP to B^{act} (Fig. 3.5).

PRP8^{Jab1} contains an intrinsically disordered region that is inserted into BRR2 to prevent binding of RNA [191]. This unstructured region of PRP8^{Jab1} must be removed from BRR2 to allow for it to be loaded onto the U4/U6 snRNA duplex. BRR2 is proposed to become inactivated following U4 release, again by the C-terminus of PRP8^{Jab1} [191]. In pre-B^{act-1}, TCERG-1 binds the N-terminal region of BRR2 (Fig. 3.13). This is a previously uncharacterized protein-protein interaction that could stabilize or regulate the activity of BRR2, or at least aid in its initial repositioning. Following dissociation of TCERG-1, BRR2 remains in a similar position in pre-B^{act-2}, anchored to its binding site on PRP8^{Jab1}. This positioning is in stark contrast to the position of BRR2 in the B^{act}, in which BRR2 is again rotated, forming connections to SF3B3 (Fig. 3.6). This interaction appears to be important for stabilizing the U2 snRNP following its docking to PRP8 (Fig. 3.7). These repositioning events appear to be coupled to PRP8^{RH} and PRP8^{Jab1} repositioning (Fig. 3.5). Nonetheless, it is unclear whether BRR2 repositioning is a consequence or the cause of PRP8^{RH} and PRP8^{Jab1} dynamics, or what forces may drive these large scale rearrangements.

The pre-B^{act} complexes are the first snapshots of what happens to PRP8^{RH} immediately following U4 snRNA dissociation. In addition to its dynamic repositioning throughout

the assembly and function of the spliceosome, PRP8^{RH} exhibits several regulatory functions. PRP8^{RH} regulates BRR2 activity by preventing BRR2 binding to U4 snRNA [193]. Moreover, PRP8^{RH} contains a motif that undergoes a conformational toggle between a β-finger and a loop, and this toggle is implicated in control of splicing fidelity as well as Mg^{2+} binding that stimulates splicing (Fig. 1.11) [194, 288]. Furthermore, PRP8^{RH} was suggested to be involved in organization of the spliceosome's catalytic center [188]. Consistent with a role in catalytic center formation, PRP8^{RH} is positioned near the U6 ISL in both pre-B^{act} structures. In pre-B^{act-1}, the β-finger is positioned above WBP11 and the U6 ISL and points outward from the core of the spliceosome. In pre-B^{act-2}, the β-finger is rotated farther away from the catalytic center and KIN17. In addition, the entire PRP8^{RH} appears to be more flexible in pre-B^{act-1}, implying that it may be destabilized prior to being drastically repositioned in B^{act}. The proximity of PRP8^{RH} and its rotation in the pre-B^{act-1} to pre-B^{act-2} transition suggests that it may be involved in guiding the formation of U6 ISL by modulating protein-RNA interactions surrounding the nascent catalytic center.

Finally, PRP8^{RH} positioning and dynamics appear to facilitate docking of U2 snRNP to PRP8 (Fig. 3.7). In the B complex, the functionally important SMU1/RED proteins likely stabilize the U2 snRNP on the spliceosome by bridging it to BRR2, as U2 snRNP would otherwise be connected only via very flexible connections mediated by U2 snRNA [98]. This bridge appears to be replaced by PRP8^{RH} in both pre-B^{act-1} and pre-B^{act-2}, with PRP8^{RH} being situated between BRR2 and U2 snRNP (see section 4.3.1). Moreover, it is reasonable to suggest that the rotation of PRP8^{RH} from pre-B^{act-1} to pre-B^{act-2} (section 3.3) may be involved in the more stable docking of U2 snRNP, particularly SF3B1^{HEAT}, to the latter.

Repositioning of the Switch loop (SWL) of PRP8 (PRP8^{SWL}) (aa 1329-1367) also occurs during activation (Fig. 1.11 and 3.5). Prior to activation, PRP8^{SWL} maintains contact with the PRP8^{En}, while in the activated and catalytic spliceosomes it undergoes approximately a 180°rotation to face the PRP8^{SWL} [200]. This change is proposed to transition the PRP8^{SWL} from an "inactive" to an "active" conformation [200]. In the inactive confor-

mation of the B complex, PRP8^{SWL} is contacted by a flexible strand of SNU66 (approximately aa 142-160) [98]. This contact is disrupted in pre-B^{act}, but PRP8^{SWL} remains in place. Thus, while SNU66 may contribute to locking PRP8^{SWL} in the inactive form, other factors must contribute to preventing PRP8^{SWL} repositioning. PRP8^{SWL} may remain in the inactive form due to the presence of the B-specific proteins (in particular UBL5 and MFAP1), which are bound immediately beneath the PRP8^{SWL}. Displacement of UBL5 and MFAP1 appears to coincide with the transition of PRP8^{SWL} to the active conformation, in which it PRP8^{SWL} closely interacts with SRRM2 that binds to the activated complex (summarized in Fig. 4.3). This finding is consistent with a previous suggestion based on structural evidence, which stated that either SNU66 or the B-specific proteins could be responsible for regulating the repositioning of PRP8^{SWL} [200]. PRP8^{SWL} should be more thoroughly investigated to understand whether its conformational dynamics play important roles in spliceosome activation and/or catalysis.

4.2 TCERG-1 and potential links to co-transcriptional RNA splicing

TCERG-1 (also known as CA150) is reported to bind the Pol II CTD and to regulate transcription and to interact with the spliceosome, although the exact nature by which TCERG-1 exerts its activity in the spliceosome is unknown [289–291]. Although TCERG-1 was identified as a component of the spliceosome, it has not been mapped to any 3D structure of the spliceosome until now. Binding of TCERG-1 to pre-B$^{act\text{-}1}$ suggests that it may be recruited in the early stages of transitioning from the B to pre-B$^{act\text{-}1}$ complex, following dissociation of U4 snRNP (Fig. 3.13). The extensive contacts between TCERG-1 to BRR2, PRP8, CPB80/20, and the B-specific proteins suggest that it may be involved in stabilizing or guiding the intermediate state of the pre-B$^{act\text{-}1}$ spliceosome. Contacts to BRR2^{NC} indicate that TCERG-1 may also help stabilize the repositioning of BRR2 after it unwinds the U4/U6 duplex and thereby undergoes a complex rotational movement

(Figs. 3.13 and 3.7). Binding of TCERG-1 to pre-B^{act-1} appears to represent a transient interaction with the spliceosome, as TCERG-1 is destabilized in pre-B^{act-2}.

In addition to TCERG-1, WBP11 in pre-B^{act-1} provides another example of a potential link between the splicing and transcription machineries. Both TCERG-1 and WBP11 bind the Pol II CTD, although it is not known whether these interactions occur simultaneously. WBP11 also interacts with PQBP1 (also called NPW38) and protein phosphatase 1 (PP1) [283, 284]. PQBP1 has been reported to interact not only with the spliceosome, but also with the transcription machinery. WBP11 and PQBP1 form a complex in which PQBP1 binds the spliceosomal protein DIM1 (also called U5-15 kDa; associated with the B complex) and Pol II CTD [292–294]. It is unsurprising that PQBP1 has not been observed in the B complex, as the C-terminal region of PQBP1 that interacts with DIM1 is intrinsically disordered [295]. PQBP1 binds to the C-terminal of WBP11, while only the short N-terminal α-helix of WBP11 was identified in pre-B^{act-1}. Although PQBP1 is detectable within the mass spectrometry data of the pre-B^{act} sample, it could not be localized within the cryo-EM reconstruction of either pre-B^{act} complex. WBP11 was shown to bind and weakly inhibit PP1, and the inhibition increased when WBP11 was phosphorylated [283]. The interplay between WBP11, PQBP1, PP1, and the Pol II CTD should be more thoroughly investigated to understand the role of possible crosstalk between transcription and pre-mRNA splicing, and how this may be intimately tied to association of WBP11 to the spliceosome during activation. It is noteworthy that both TCERG-1 and WBP11 dissociate from the spliceosome, or at least to become destabilized, in pre-B^{act-2}. These transient interactions may suggest that early steps of spliceosome activation could be closely tied with transcription, and that these two processes could be mutually affected at this stage of spliceosome assembly. Nonetheless, more experiments will be required to understand how TCERG-1 may be recruited to the spliceosome and whether it may mediate crosstalk between transcription and splicing.

Another link in pre-B^{act-1} between the transcription factor TCERG-1 and the spliceosome is SRSF1 (Fig. 3.13). SRSF1 (also called ASF/SF2) is a serine-arginine (SR) protein that binds RNA and influences spliceosome assembly by aiding U1 snRNP recognition

of the 5′ss [296]. Domains FF4 and FF5 of TCERG-1 were shown to interact with the splicing factor SRSF1, thereby assisting in its localization it to nuclear speckles where spliceosome components are thought to be concentrated [297]. The RRMs (RRM1 and RRM2) of SRSF1 were demonstrated to interact with U1-70K^{RRM}, while the RS domain in the C-terminal region of SRSF1 was shown to bind to Pol II CTD [298, 299]. As Pol II CTD is known to bind a plethora of splicing factors, it is unknown whether the binding of TCERG-1 and SRSF1 to Pol II CTD is simultaneous and/or coordinated. Nonetheless, it is noteworthy that both TCERG-1 and SRSF1^{RRM2} are both localized in the pre-B$^{act\text{-}1}$ complex. While TCERG-1 clearly becomes destabilized in pre-B$^{act\text{-}2}$, it is unclear whether SRSF1^{RRM2} also dissociates from the pre-mRNA, or whether it is simply not visualized due to the apparent flexibility of this region in pre-B$^{act\text{-}2}$. Binding of TCERG-1 to pre-B$^{act\text{-}1}$ but not pre-B$^{act\text{-}2}$ would appear to be consistent with a model in which TCERG-1 binds transiently to the transcription machinery or the spliceosome, as previously suggested [300]. Further experiments should be conducted to understand the interplay between TCERG-1 and spliceosome activation, as well as the nature of possible feedback mechanisms between transcription and splicing and how these may be mediated by TCERG-1.

4.3 Mutually exclusive interactions during spliceosome activation

The pre-B^{act} complexes reveal previously undocumented transitions in the activation of the spliceosome, an event consisting of the dissociation of approximately 32 proteins and U4 snRNA and the association of at least 28 proteins. Many of the proteins bound to the B complex occupy binding sites that are later occupied by different factors in B^{act}, thus being mutually exclusive interactions.

4.3.1 Dissociation of B-specific proteins is a prerequisite for PRP8 conformational change and integration of B^{act} proteins

Removal of B-specific proteins is necessary for the large-scale conformational change of PRP8 from an open to a closed conformation (Fig. 3.4). The B-specific proteins would furthermore clash with several B^{act}-specific factors: CWC27 and RNF113A. It is unclear what governs the removal of the B-specific proteins to allow for conformational change of PRP8 as well as for the association of the B^{act}-specific factors, including CWC22/CWC27 (which forms a heterodimer), RNF113A, and SRRM2.

Several B-specific proteins (SMU1, RED, and FBP21) are not visualized in pre-B^{act} structures, yet were localized in the B complex. SMU1 and RED together form a dimer that bridges BRR2 to SF3B3, potentially stabilizing the flexibly associated U2 snRNP [196]. SMU1 and RED were shown to be required for splicing of short introns (approximately 100 nts or fewer) [205]. SMU1 contains a WD40 domain while RED contains two short α-helices and extensive disordered regions that crosslink to many components of the spliceosome [196, 301]. Knockdown of SMU1 and/or RED prevents activation of the spliceosome, stalling assembly at the B complex stage [205]. Progressive shortening of the sequence between the 5′ss and the BS of the intron resulted in splicing being blocked when SMU1/RED were absent, suggesting the latter could be involved in mediating structural rearrangements required for positioning of the intron [205]. Both SMU1 and RED are detected in the mass-spectrometry data of the purified pre-B^{act}, suggesting they remain present in the complex but are loosely associated and therefore undetectable in the EM densities (Table B.1). In the human B complex, bridging of BRR2 and SF3B3 of U2 snRNP may assist in the docking of U2 onto the spliceosome during activation. In the pre-B^{act} complex, this bridge appears to be either destabilized or not present, despite the presence of SMU1/RED in the sample. This apparently loose association between SMU1/RED and the spliceosome could suggest that SMU1/RED act transiently at the B complex stage during activation, helping to guide U2 snRNP into place, and are subsequently destabilized. Furthermore, BRR2 and SF3B3 are in a different orientation

relative to one another in the pre-B^{act} structures than in the B complex, leading to these components being separated by a larger gap (see Fig. 3.7). The gap between BRR2 and SF3B3 is occupied by PRP8^{RH}, which appears to be dynamic even between pre-B$^{act\text{-}1}$ and pre-B$^{act\text{-}2}$ (Fig. 3.5). Consistent with its proposed role in activation, it is thus reasonable to hypothesize that SMU1/RED thus acts transiently and subsequently becomes destabilized from its original binding site between BRR2 and SF3B3 after guiding U2 snRNP into place, and that its bridging interactions may be replaced by the repositioned PRP8^{RH} in pre-B^{act}.

While nearly all B-specific proteins are abundant in the pre-B^{act} sample, FBP21 was not detectable (Tables B.1 and B.2). This likely is due to the close association of FBP21 with U4/U6 snRNP-related proteins, including PRP6, DIM1, CypH, which are dissociated during activation [276]. FBP21 binds to a site that is sandwiched between BRR2 and PRP6/DIM1 in the B complex [98]. Following unwinding of U4/U6 snRNA duplex, BRR2 becomes repositioned and U4 snRNP is dissociated from the spliceosome to allow activation to occur. During this transition, the interaction between FBP21 and BRR2 appears to be disrupted, possibly suggesting that FBP21 remains associated with U4 snRNP components and is thereby displaced. Importantly, only the N-terminal Znf domain of FBP21 has been localized in the spliceosome, while FBP21 has shown to have additional interactions with numerous spliceosomal proteins, including BRR2, which it was reported to regulate by binding of an intrinsically disordered region to BRR2$^{C\text{-}term}$ [126]. FBP21 was shown to interact with WBP11 (also known as NPW38BP/SIPP1) with its two tandem WW domains [302]. Both proteins are classified as B-specific proteins due to their characteristic association with the B complex. Although FBP21 is not abundant in the pre-B^{act} samples, it may nevertheless be involved in positioning of WBP11 into the nascent catalytic center of the spliceosome during the transition from the B to the pre-B^{act} complex.

B-specific proteins that block the conformational change of PRP8 consist of PRP38, MFAP1, SNU23, and UBL5. These proteins are deeply lodged within the cleft between PRP8^{En} and PRP8^{NTD}. PRP38 is of particular importance in the organization of the

B-specific proteins, as it forms direct contacts to MFAP1, SNU23, and UBL5 – consistent with its proposed role as a platform that helps coordinate multiple proteins during spliceosome activation [204]. Previous studies in yeast indicated that PRP38 is required for splicing, particularly at the activation step. Intriguingly, depletion of PRP38 did not inhibit pre-spliceosome formation, but abrogated spliceosome activation by preventing the release of U4 snRNA [303]. This still did not clarify whether PRP38 remained bound to the spliceosome after U4 snRNA dissociation. In contrast, the pre-B^{act} complexes provide biochemical and structural evidence that PRP38 (and associated MFAP1, SNU23, UBL5) indeed remains bound after U4 snRNA is released. Given the requirement of PRP38 in splicing, it appears that PRP38 and the remaining associated B-specific proteins could facilitate U4 snRNA release by holding PRP8 in an open conformation, which may favorably position the U4/U6 snRNA duplex for unwinding by BRR2.

4.3.2 Displacement of WBP11 from pre-B^{act-1} allows for rearrangements of U6 ISL *en route* to a mature catalytic center

WBP11, bound to the nascent catalytic center only in pre-B^{act-1}, must be dislodged from the core of the spliceosome to allow for maturation of the active site (Fig. 3.14). The position of the WBP11 would clash with the folding of U6 snRNA, particularly the formation of helix Ia. Strikingly, WBP11 passes along the backbone of the U6 ISL and is situated above a functionally relevant nucleotide, U74, that becomes bulged out of the ISL in the B^{act} (Fig. 3.14). U74 becomes stacked with G46 and interacts with C55 to contribute to formation of the triple helix (Fig. 3.18). U74 appears to be stacked into the ISL in pre-B^{act-1}. This arrangement may suggest that WBP11 could initially aid in the formation or placement of the ISL, and becomes displaced following the docking of the ISL to PRP8^{NTD}. How WBP11 may become displaced from the spliceosomal core is yet unclear. However, the presence of WBP11 in pre-B^{act-1} is in line with a report documenting WBP11 as a part of the spliceosome [283]. Interestingly, a C-terminal region of WBP11 (aa 180-372) was documented to be the portion of this protein that stalls splicing catalysis, while the α-helix of WBP11 mapped in the pre-B^{act-1} is located in the N-terminal region (aa 2-59) [283]. Thus the N-terminal α-helix appears to play a role in spliceosome activation (or regulation), while the C-terminal region may be involved in regulation of splicing catalysis. In summary, the position of WBP11 in pre-B^{act-1} implies that WBP11 exerts its regulatory activity at least in part during formation of the active site, perhaps by functioning as a checkpoint to ensure proper folding of the catalytic RNA network.

4.3.3 Exchange of U6 LSm for IBC, SYF3, and PPIL2

In pre-B^{act-1}, U6 LSm proteins are still bound to U2/U6 helix II (see section 3.5.4). In pre-B^{act-2}, U6 LSm is replaced by SYF3, which cradles U2/U6 helix II and coordinates

the protein-protein interactions stabilizing the IBC (Fig. 3.11). Of SYF3, only the N-terminal HAT repeats are stably bound to pre-$B^{act\text{-}2}$, while the C-terminal HAT repeats appear to be flexible and only integrated in B^{act}. It is unclear whether $SYF3^{C\text{-}term}$ is indispensable for splicing activation or other steps during assembly. $SYF3^{N\text{-}term}$ provides not only a binding site for U2/U6 helix II, but also a platform for one of its interaction partners, PPIL2. SYF3 was shown to be important for spliceosome function *in vitro*, as immunodepletion resulted in inhibition of splicing [304]. In addition to providing an interface for the binding of U2/U6 helix II, PPIL2 and the IBC, SYF3 also appears to be important for stabilization of the catalytic center. In B^{act}, $SYF3^{N\text{-}term}$ becomes pushed toward the ISL and comes into contact with the flipped out C60 base of the ISL as well as with a strand of SKIP (Fig. 4.2). These contacts are not yet established in pre-$B^{act\text{-}2}$ but appear to be primed to occur following the transition of PRP8 from an open to a closed conformation.

The functional significance of the interaction between $SYF3^{N\text{-}term}$ and $PPIL2^{PPIase}$ (Fig. 3.11) is less clear, albeit consistent with a previous report revealing binding between these domains [230]. The proline isomerase activity of PPIL2 has been reported to be nonfunctional, suggesting that $PPIL2^{PPIase}$ may play more of a scaffolding role than a catalytic one in the spliceosome [305]. In contrast, the U-box domain ($PPIL2^{U\text{-}box}$) exhibits E3 ubiquitin ligase activity [306], although its target is unknown. $PPIL2^{U\text{-}box}$ binds to a cleft situated on SNU114, directly opposite to the Armadillo repeats of CTNNBL1 in pre-$B^{act\text{-}2}$ (Fig. 3.9). PPIL2 appears to join the spliceosome during activation and become destabilized following Step 1 of splicing, although the significance of this transient binding remains unclear [276]. An *in vitro* approach showed a concentration dependent inhibition of spliceosome catalysis by PPIL2, suggesting that its dissociation may be necessary to allow splicing to occur, yet it is unknown whether this occurs similarly *in vivo* [307].

Binding of $SYF3^{N\text{-}term}$ to PPIL2 was found not to influence the ubiquitin ligase activity of PPIL2 [230]. Both domains of PPIL2 were reported to bind to ZNF830, which is detected in the sample by mass spectrometry but whose function in splicing is unclear and which is the only component of the IBC that has not been localized in a spliceosomal

reconstruction (Fig. 3.1) [306]. PPIL2 thus appears to coordinate multiple protein-protein interactions in the spliceosome prior to activation, joining first at the pre-$B^{act\text{-}2}$ and remaining bound until the complex is fully activated. Furthermore, the $PPIL2^{U\text{-}box}$ could maintain structural and functional interactions required for activation. Probing of the ubiquitin ligase activity of $PPIL2^{U\text{-}box}$ in the context of splicing will be necessary to better understand the role of PPIL2 in spliceosome assembly and activation.

4.3.4 Destabilization of KIN17 liberates PRP2 binding site

KIN17 is a known component of the spliceosome but was localized for the first time in the reconstruction of pre-$B^{act\text{-}2}$ reported here (Fig. 3.15) [75]. Previous *in vitro* studies have determined the structure of KIN17 domains and have provided some indications of its RNA binding activity and potential protein-protein interactions, but did not offer evidence as to the location or putative functions of KIN17 in the spliceosome (see section 3.6.2).

Absence of density for KIN17 in pre-$B^{act\text{-}1}$ suggests that KIN17 is recruited during transition from pre-$B^{act\text{-}1}$ to pre-$B^{act\text{-}2}$. The position of $KIN17^{WH}$ is incompatible with the position of the U6/5′ss helix in B^{act}, and must therefore be displaced following pre-$B^{act\text{-}2}$ formation. Dissociation of KIN17 from the spliceosome thus appears to be a prerequisite for final folding of the catalytic center, yet due to its proximity thereof, may assist in guiding maturation of the active site.

$KIN17^{WH}$ does not have known RNA binding activity, and was suggested to be involved in protein-protein interactions, although its function in RNA processing is currently unknown [287]. The position of $KIN17^{WH}$ in pre-$B^{act\text{-}2}$ suggests that it may involved in positioning proteins that surround and possibly assist formation of the catalytic center. The nearby helix of KIN17, which intersects two WD40 domains of SF3B3 of the U2 snRNP, appears to coordinate protein-protein interactions that could be involved in docking of U2 snRNP onto the spliceosome. There is currently no reported function of this helix of KIN17. The C-terminal SH3 domain of KIN17 ($KIN17^{SH3}$) is loosely docked

to the HEAT repeats 7-8 of SF3B1 of the U2 snRNP. KIN17^{SH3} binding appears to occur transiently, as it would need to be displaced to liberate the binding site for PRP2 in the B^{act}. KIN17^{SH3} binds RNA *in vitro*, but this activity is undocumented in the context of the spliceosome [308]. Positioning of KIN17^{SH3} in pre-B^{act-2} would allow for access to the intron sequence, which is bound by the adjacent RES complex.

KIN17 has been characterized as a component of DNA replication machinery and is known to be upregulated following UV-induced DNA damage [285, 286]. Despite identification of KIN17 in such processes, a mechanistic understanding of its action is still lacking. Nonetheless, identification of KIN17 in the pre-B^{act-2} seems to indicate that it may be shared among multiple molecular machineries in the nucleus.

4.4 Stabilization of proteins within the spliceosome during activation

4.4.1 RES proteins: BUD13, SNU17, SNIP1

Constituents of the RES complex, required for spliceosome activation, also appear to be gradually stabilized at the periphery on an interface between the loosely docked U2 snRNP and PRP8^{RT} (Fig. 3.8). Of the three proteins in the RES complex, only the FH domain of SNIP1 is observed in pre-B$^{act\text{-}1}$. The binding site of SNIP1^{FH} at the interface between U2 and U5 snRNP components – in particular SF3B1, SF3B6, and PRP8^{RT} – could likely assist in the stabilization of U2 snRNP onto the spliceosome. SNIP1 also interacts with the NTR protein, SKIP, which appears to anchor it to the complex. The two other components of the RES complex, BUD13 and SNIP1, are likely attached but not localized due to flexibility of this unstable interface. BUD13 and SNU17 are more stably integrated into the complex in pre-B$^{act\text{-}2}$, with BUD13 sandwiched between a flexible strand of SKIP and SNU17, and SNU17 accommodated in a gap between SF3B1^{HEAT} and BUD13 (Fig. 3.8). SNU17^{RRM} interacts with the intron where it emerges from SF3B1^{HEAT}, consistent with the known RNA binding activity of SNU17 downstream of the branch site [207]. The gradual stabilization of RES binding observed in the transition from pre-B$^{act\text{-}1}$ to pre-B$^{act\text{-}2}$ is consistent with its reported cooperative binding properties, becoming more tightly associated as the spliceosome becomes progressively activated [207, 208]. However, RES is not yet as stabilized in pre-B$^{act\text{-}2}$ as in the B^{act}, indicated by the absence of the C-terminus of BUD13 as observed in B^{act} [99]. In the latter, a short α-helix (aa 530-538) interacts with the B^{act} protein CWC22 [99]. In addition, two short C-terminal α-helices (aa 474-495) point toward the helicase PRP2, which is docked to SF3B1^{HEAT} and is required for catalytic activation [99, 309]. Deletion of SNU17 greatly impairs RES complex formation, which leads to lower efficiency of spliceosome activation as well as premature disassembly of spliceosomes by PRP2 [310]. The pre-B^{act} structures demonstrate that RES becomes

gradually stabilized on the spliceosome during activation, and that this event appears to precede stable PRP2 binding to its position in B^{act}.

4.4.2 NTC binding and stabilization of U6 snRNA

The NTC complex (see section 3.5.2), consisting of PRP19, CDC5L, PLRG1, SPF27, CWC15, and CTNNBL1, joins the spliceosome following dissociation of the U4 snRNP. Previous studies indicated that association of NTC is associated with the loss of U6 LSm while conferring added stability to U6 snRNA within the spliceosome [214].

Intriguingly, several NTC components are already stably integrated even as U6 LSm remains bound in pre-$B^{act\text{-}1}$. This observation is consistent with a previous proposal that U6 LSm dissociation is likely dependent on NTC binding [214]. Instead of a concerted or mutually exclusive effect, the NTC proteins already begin to bind before U6 LSm has left the complex (see section 3.5.2). Thus, destabilization of U6 LSm must not be a prerequisite for addition of the NTC [214] (Fig. 3.11). The stable binding of additional NTC domains in pre-$B^{act\text{-}2}$, as well as the loss of U6 LSm in this structure, appear to be involved in gradual stabilization of U6 snRNA (Fig. 3.9), as the RNA network of pre-$B^{act\text{-}2}$ is more mature than in pre-$B^{act\text{-}1}$ (Fig. 3.18). This stabilization of U6 snRNA is likely in part contributed by the $CDC5L^{Myb}$, a part of the NTC (Fig. 3.19). $CDC5L^{Myb}$ domains are bound near $PRP8^{HB}$, at an interface between U2 and U5 snRNPs (Fig. 3.9). In pre-$B^{act\text{-}2}$, $CDC5L^{Myb}$ and $PRP8^{HB}$ domains accommodate U2/U6 helix Ib, coordinating the phosphate backbone of U2 snRNA (nts G20-U22) with multiple basic residues (Fig. 3.19). This interaction appears to be stably established at the pre-$B^{act\text{-}2}$ and is also present in B^{act}. The absence of clear density for U2/U6 helix Ib in pre-B^{act} may suggest that base-pairing between the constituent nucleotides (U2 nts G20-U22 and U6 nts A53-C55) has not yet been established, or that the positioning of U2/U6 helix Ib may already be formed but not yet docked to its binding site on the $CDC5L^{Myb}$ and $PRP8^{HB}$ domains. U2/U6 helix Ib is separated from the U6/5′ss helix by approximately 2.5 nm in pre-$B^{act\text{-}2}$, but is brought into proximity with these nucleotides in B^{act}. U6 helix Ia loop, visualized as

a low-resolution density in pre-B^{act-2}, appears to connect U6/5′ss to helix Ib (Fig. 3.18). Separation of U6/5′ss and helix Ib thus appears to result in an extended conformation of the U6 helix Ia nucleotides, preventing base-pairing thereof with U2 nts G25-C28 to form U2/U6 helix Ia (Figs. 3.19 and 3.18). In addition, $CDC5L^{Myb}$ domains appear to provide an additional link between the flexible U2 snRNP (via contacts with SF3A2) and the rest of the spliceosome, possibly facilitating the stable docking of the U2 snRNP (Fig. 3.9).

$PLRG1^{WD40}$ is stably bound beneath $PRP8^{HB}$ in both pre-B^{act-1} and pre-B^{act-2} (Fig. 3.9) and is connected to the NTC via the C-terminal region of CDC5L, which is not visible in the pre-B^{act} complexes but is a connection that is required for splicing to occur [311]. $PLRG1^{WD40}$ forms multiple contacts with intrinsically disordered proteins, including SKIP and CWC15 (Fig. 3.9 and 3.10). The explicit function of PLRG1 is unclear, aside from its apparent role as a binding platform for otherwise very flexible splicing factors to bind, consistent with the proposed role of WD40 domains as organizers of protein-protein interactions [312].

Only a short α-helix of CWC15 (aa 218-226) can be localized in pre-B^{act-1}, whereas a larger segment connected to this helix is bound to pre-B^{act-2} (Fig. 3.9). As with other members of the NTC, CWC15 appears to be gradually stabilized on the spliceosome in the later stages of activation. [1]

PRP19 and SPF27 are not visualized in reconstructions of pre-B^{act}. This is unsurprising, as even in the B^{act}, densities for the PRP19 and SPF27 are not stably bound until later stages of assembly, when the NTC helical bundle (NTC^{HB}) becomes visible (see Fig. 4.3). Together with $CDC5L^{C\text{-}term}$, PRP19 and SPF27 form the NTC^{HB} that is accommodated by the translocation of U5-40K and PPIL1 in B^{act} (state 4) [99]. Additional contacts to the IBC, which is present in pre-B^{act-2}, likely facilitate positioning of NTC^{HB}. Thus PRP19 and SPF27, while detectable in the sample by mass spectrometry, appear to be the last members of the NTC to stably bind the spliceosome due to rearrangements in

[1] "stages of activation" refers to states 1-8 of the B^{act} reported in [99], detailing structurally distinct states of B^{act} assembly.

protein-protein contacts that occur only later in the assembly pathway (see Fig. 3.1 and Table B.1).

CTNNBL1 is an essential component of the NTC, but has not been observed in any spliceosomal structure besides pre-$B^{act\text{-}2}$. CTNNBL1 coordinates multiple protein-protein interactions between CWC15 and CDC5L and thereby contributes to stability of the NTC, but is loosely bound to these components and therefore readily dissociates at higher salt concentrations (e.g., >150 mM) [215, 313]. In the pre-$B^{act\text{-}2}$, a short α-helix containing segment (aa 147-159) of CDC5L mediates the connection of the $CTNNBL^{ARM}$ domain to the body of the spliceosome (Fig. 3.9). The integration of CTNNBL1 and its role in spliceosome activation is an exciting topic for new research and should be analyzed by additional experiments in more detail.

NTC binding contribution to U6 snRNA stabilization appears to be more a consequence of overall complex stabilization rather than as a result of direct interactions with U6 snRNA. However, it was previously shown that the NTC proteins are required for the removal of U6 LSm from the spliceosome [304]. The pre-B^{act} complexes highlight a possible role for the NTC component $PLRG1^{WD40}$ in mediating this transaction by providing a interface on which U6 LSm is initially bound and subsequently replaced with SYF3 and the IBC, offering a putative structural explanation consistent with previous biochemical studies. This interaction leads to the stabilization of U2/U6 helix II and thus could potentially contribute to the overall stability of the U6 snRNA within the spliceosome.

4.4.3 NTR proteins (SKIP, RBM22) and PRP17

In addition to the gradual stabilization of NTC proteins such as those comprising the NTC^{HB} (see section 4.4.2), several spliceosomal proteins involved in spliceosome activation appear to join the spliceosome and become progressively stabilized along binding interfaces throughout the complex. SKIP, PRP17, and CWC15 – all of which are present in B^{act} – are partially stabilized on the pre-B^{act}. Both SKIP and RBM22 are NTC-related (NTR) proteins.

SKIP is an intrinsically disordered protein (IDP) that was shown to join the spliceosome during activation [223]. Furthermore, SKIP is thought to be implicated in the recruitment of PPIL1, a PPIase that is integrated into the spliceosome in later stages of activation (Fig. 4.3) [222, 314]. SKIP appears to gradually undergo a disorder-order transition as it binds throughout the spliceosome, consistent with an *in vitro* study showing that an N-terminal fragment of SKIP becomes ordered upon binding to PPIL1 [223]. This transition is even more pronounced in the spliceosome, with many regions of SKIP stably binding to multiple interfaces throughout the pre-B^{act} complexes.

The SNW domain of SKIP contains α-helical regions as well as flexible strands (Fig. 1.142). Multiple α-helices of SKIP are bound to $PLRG1^{WD40}$ and along the boundary between U2 and U5 snRNPs (Figs. 3.9 and 3.10). These α-helices are connected by flexible strands, as in the B^{act} [99] (Fig. 3.10). Flexible strands of SKIP appear to be further stabilized in the transition from pre-$B^{act\text{-}1}$ to pre-$B^{act\text{-}2}$.

$RBM22^{RRM}$ integration appears to be facilitated by binding of the IBC (Fig. 3.11). Density for $RBM22^{RRM}$ is not present in pre-$B^{act\text{-}1}$, suggesting that this domain remains flexible until it is stabilized by protein-protein contacts with AQR of the IBC in pre-$B^{act\text{-}2}$. $RBM22^{Znf}$ is evident in both pre-B^{act} structures. The N-terminus of SKIP is already bound near the PPIL1 binding site, situated on $RBM22^{Znf}$ and $PRP8^{NTD}$. An α-helix of SKIP (aa 137-161) binds to $PLRG1^{WD40}$, opposite $RBM22^{Znf}$. Another longer α-helix of SKIP (aa 282-325) wraps around the exterior portion of $PRP8^{RT}$ near the junction between U2 snRNP components and PRP8, and near $SNIP1^{FH}$ (Figs. 3.8 and 3.10). In pre-$B^{act\text{-}2}$, additional strands as well as a short α-helix are resolved, linking the $RBM22^{Znf}$-bound N-terminal region to binding sites on $PLRG1^{WD40}$, sandwiched tightly between SYF3 and U2/U6 helix II and protruding into a gap between the U2 nucleotides of U2/U6 helix Ib, $CDC5L^{Myb}$, and $PRP8^{HB}$. Finally, a flexible strand in the C-terminal region (aa 405-438) winds through a gap separating $PRP8^{RT}$ and the RES protein BUD13. The apparent disorder-order transition of SKIP thus appears to be detectable even in the transition between pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$, with multiple regions becoming ordered across distinct domain interfaces. In addition, ordering of SKIP coincides with stabilization of

other functionally important interactions in pre-$B^{act\text{-}2}$, including U2/U6 helix Ib and helix II, binding of SYF3, RES complex, and docking of U2 snRNP. In addition, one α-helix of SKIP (aa 282-325) appears to serve as a binding platform for a short fragment of CDC5L that extends towards $CTNNBL1^{ARM}$, perhaps mediating the interaction between CDC5L and the loosely associated CTNNBL1. It is unclear whether SKIP ordering is a prerequisite for these assembly events to occur, or whether SKIP binding is part of a concerted process in which SKIP and its interacting partners are mutually stabilized. The gradual ordering of SKIP across functionally important components of the spliceosome suggest that its disordered nature allows it to establish a dynamic network that may assist in spliceosome assembly and activation. PRP17 is a splicing factor that associates with the spliceosome during activation and has been reported to facilitate excision of introns containing 200 nts or more [315–317]. Furthermore, PRP17 was proposed to play a role in modulation of splicing efficiency, as it is not required for steps 1 and 2 of splicing but appears to exert a kinetic effect at both reactions [317]. In pre-$B^{act\text{-}1}$, density for PRP17 is not observed, while in pre-$B^{act\text{-}2}$ a small fragment of PRP17 (aa 114-131) is wedged between BUD31, $RBM22^{Znf}$, SKIP, and a single-stranded region of U6 snRNA (Fig. 3.12). This binding site is consistent with that of PRP17 in B^{act}, although in the latter, PRP17 becomes more stably integrated and forms multiple connections to PPIL1, U5-40K, and RNF113A (Fig. 4.3) [99]. PPIL1 can catalyze proline isomerization in PRP17, although this reaction was shown not to be essential for the function of either component, suggesting that PPIL1 might play more of a supporting than an enzymatic role in this binary interaction [318]. The ordered domains of $PRP17^{WD40}$ and several short α-helices are linked together by intrinsically disordered strands; similar to SKIP, PRP17 has several regions that undergo a gradual disorder-order transition upon integration to the spliceosome [220]. The presence of a short fragment of PRP17 may suggest that the cooperativity of this transition may be strongly influenced by the presence of PPIL1, which is absent in both pre-B^{act} structures. However, it is worth noting that both PRP17 and SKIP, two proteins that converge upon PPIL1, already begin to become ordered in the spliceosome prior to stable integration of PPIL1 (Fig. 3.10).

4.4.4 PRP8^{NTDL}, SF3B2, SF3A2, and CWC15 closely interact with U2/U6 RNA at catalytic center

A number of proteins directly surrounding the nascent catalytic center contain domains that appear to be flexible and therefore unobserved in pre-B^{act}, yet highly ordered in B^{act} (Fig. 4.1). This suggests disorder-order transitions in regions of proteins which are already docked near the nascent catalytic center in pre-B^{act}. PRP8^{NTDL}, a strand connecting the PRP8^{NTD} and PRP8^{HB} domains that wraps around U6 ISL in B^{act}, is not visualized in either pre-B^{act} reconstruction (Figs. 1.11 and 4.1). This suggests that the PRP8^{NTDL} is initially flexible and becomes stably positioned in later stages of activation. Furthermore, in B^{act}, PRP8^{NTDL} is sandwiched within a channel formed by the U6 ISL loop and PLRG1^{WD40}. In the open conformation of PRP8, as observed in pre-B^{act}, this channel is wider and may not allow for the same protein-protein interactions between PLRG1^{WD40} and PRP8^{NTDL}. Thus, conformational change may help to stabilize PRP8^{NTDL} by tightly compacting the channel through which it is threaded, possibly leading to stabilizing interactions with both the loop of U6 ISL as well as PLRG1^{WD40}.

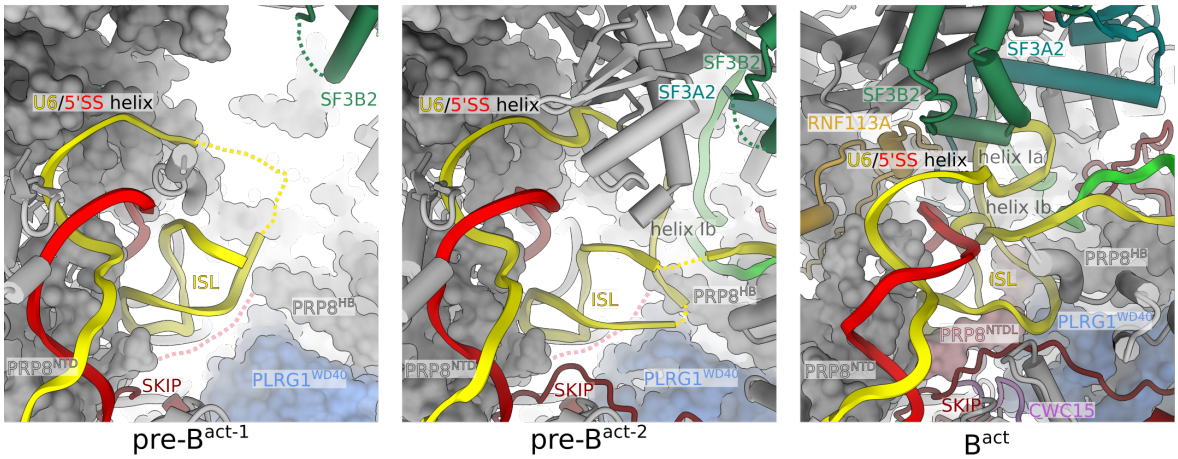

Fig. 4.1: Structuring of proteins surrounding catalytic center during activation. Proteins that surround the catalytic center in B^{act}(right panel; PDB 6FF7) are not yet in place in pre-B^{act-1} and pre-B^{act-2} (left, middle panels). Protein or RNA regions that are not yet stabilized in pre-B^{act-1} and pre-B^{act-2} are indicated with stippled lines. In B^{act}, multiple regions of SKIP, PRP8^{NTDL}, CWC15, and the U2 snRNP proteins SF3B2 and SF3A2 are stabilized. These same proteins or protein regions (i.e., PRP8^{NTDL}), while present in both pre-B^{act-1} and pre-B^{act-2}, are not yet stabilized. In contrast, RNF113A binds only to B^{act}and is not detectable in the biochemical preparation of pre-B^{act-1} (see Fig. 3.1D).

SF3B2 exhibits a short segment (aa 546-562) containing an α-helix (aa 547-559) that is positioned immediately above the catalytic center in B^{act} (Fig. 4.1), but is not visible in pre-B^{act}. This short segment appears to become stabilized in the compact arrangement of the catalytic center and is positioned immediately alongside the functionally important U2/U6 helix Ia. Absence of density for this α-helix of SF3B2 may point to involvement of this domain in protein-RNA and/or protein-protein interactions involved in formation of the catalytic center.

Stabilization of SF3B2 (aa 546-562) and SF3A2 (aa 1-18) near the catalytic center may also be partially mediated by RNF113A, which is sandwiched between $PRP8^{En}$ and $PRP8^{NTD}$, where it binds immediately above the U6/5′ss and adjacent to SF3B2 (aa 546-562) and SF3A2 (aa 1-18) (Fig. 4.1). RNF113A is absent from pre-B^{act}, as its binding site is occluded by B-specific proteins. Binding of RNF113A may thus contribute not only to stabilization of SF3B2 and SF3A2 around the catalytic center, but also to the formation of the catalytic center itself.

SF3A2 also contains an N-terminal strand (aa 1-18) that is stably incorporated into the core of the spliceosome in later stages of activation, during the pre-$B^{act\text{-}2}$ to B^{act} transition, as this region is visible in B^{act} but not resolved in pre-$B^{act\text{-}1}$ (Fig. 4.1, right panel). This strand winds along the backside of U2/U6 helix Ia and Ib, passes near the U6/5′ss, and ends near the flipped out U40 nucleotide of U5 loop I. The multiple contacts to the catalytic center suggest that SF3A2 could be involved in positioning and/or stabilization of the active site. Flexibility of the SF3A2 N-terminal region may be at least in part explained by the incomplete docking of the 5′ domain of U2 snRNP in pre-$B^{act\text{-}2}$, as SF3A2 is associated with U2 snRNP.

In B^{act}, CWC15, a member of the NTC, has an N-terminal strand (aa 18-50) that passes between a gap between $PLRG1^{WD40}$ and U5 snRNA and contacts the loop of U6 ISL while forming multiple interactions with the nearby SKIP and RBM22 (Fig. 4.1, right panel). As only farther C-terminal regions of CWC15 could be localized in pre-B^{act}, it appears that CWC15 undergoes a gradual disorder-order transition as it associates with

the spliceosome. CWC15 may thereby provide additional stabilization for the docking of the loop U6 ISL onto PRP8^{NTD}, while potentially contributing to stability of other protein-protein interactions within NTC and NTR components.

4.5 Protein-guided folding of the U2/U6 RNA-based catalytic center

4.5.1 Protein-RNA interactions may contribute to rearrangements of the U6 ISL stem region during activation

The U6 ISL stem region (nts U57-U64 and A70-A78) contains three nucleotides that are bulged out of the ISL in the B^{act}: C60, which is contacted by SKIP and SYF3 (Fig. 4.2); A73, the function of which is unclear; and U74, which coordinates two divalent metal ions and interacts with C55 of U6 snRNA and G20 of U2 snRNA to become part of the triple helix (Fig. 1.9). These three flipped out nucleotides cause distortions in the 3D structure of the ISL and likely contribute to its compact arrangement (Fig. 3.18). Despite a similar U6 ISL loop region in the pre-B^{act} and B^{act}, the U6 ISL stem exhibits marked conformational differences in the pre-B$^{act\text{-}1}$, pre-B$^{act\text{-}2}$, and B^{act} complexes (see Fig. 3.17). These differences suggest flexibility of the U6 ISL stem region, which may be required for repositioning of nts C60, A73, and U74.

Rearrangements in U6 ISL are consistent with numerous studies showing alternative conformations and base-pairing within the U6 ISL stem region, in particular the nucleotides C67, A79, and U80 in yeast (C61, A73, and U74 in human) [319–321]. The mid-stem nucleotides (C67, A79, and U80) undergo conformational transitions in a pH-dependent manner *in vitro*; U80 is flipped out of the ISL when A79 is protonated (pH 5.7), but U80 is stacked into the ISL when A79 is deprotonated (pH 7.0) [319]. Protonation of A79 and the flipping out of U80 from the ISL results in disruption of base-pairing between C67-U80 and the establishment of a wobble interaction between C67-A79 [321]. U80 (U74

in human) is a functionally important nucleotide that binds two divalent metal ions and forms part of the catalytically essential triple helix [128]. Changes in conformation and base-pairing thus distort the mid-stem region of U6 ISL, while leaving the conformation and base-pairing of the loop region intact [320]. These studies provided insight into the dynamics of functionally important, conserved regions in the mid-stem of U6 ISL, but were all performed *in vitro* using isolated fragments of U6 snRNA; thus, many intermolecular interactions with U6 snRNA, as observed within the spliceosome, could not be probed in such experiments.

In the B^{act}, which contains a fully formed catalytic center, both A79 and U80 (A73 and U74 in human) are flipped out of the ISL [99, 248]. C67 (C61 in human) is stacked into the ISL, where it base-pairs with G81 (G75 in human), while C66 (C60 in human) is flipped out of the ISL [99] [248]. C60 (human) is flipped out of the ISL in B^{act}, but exhibits different orientations based on the surrounding protein environment. Between state 2 and state 8 of B^{act}, repositioning of C60 coincides with the structuring of a flexible strand of SKIP, leading to an interaction between the SKIP residue P219 and C60 in state 8 [99]. The structuring of SKIP is also likely mediated by SYF3 and RBM22, to which SKIP forms numerous contacts. In pre-$B^{act\text{-}2}$, the ISL-interacting region of SKIP is structured, but is more distant from C60 than in B^{act} due to the open conformation of PRP8, preventing interaction between P219 of SKIP and C60 of U6 snRNA (Fig. 4.2). While the resolution of the ISL in pre-B^{act} does not permit detailed analysis whether C60 is stacked into the ISL or is flipped out, it is clear that interaction C60 and SKIP cannot occur in pre-$B^{act\text{-}2}$ due to separation of SKIP and the ISL (Fig. 4.2).

Flipping out of C66 (C60 in human) and A79 (A73 in human) from the ISL was not observed in NMR structures of *in vitro* assembled U6 ISL, thus implying that these alternative conformations likely occur as a result of protein-RNA or additional RNA-RNA interactions formed in the spliceosome [319–321].

The mid-stem region of U6 ISL from B^{act} did not fit well into the density of the pre-B^{act} reconstructions, although the loop region (nts C62-G74) is similar in all three complexes.

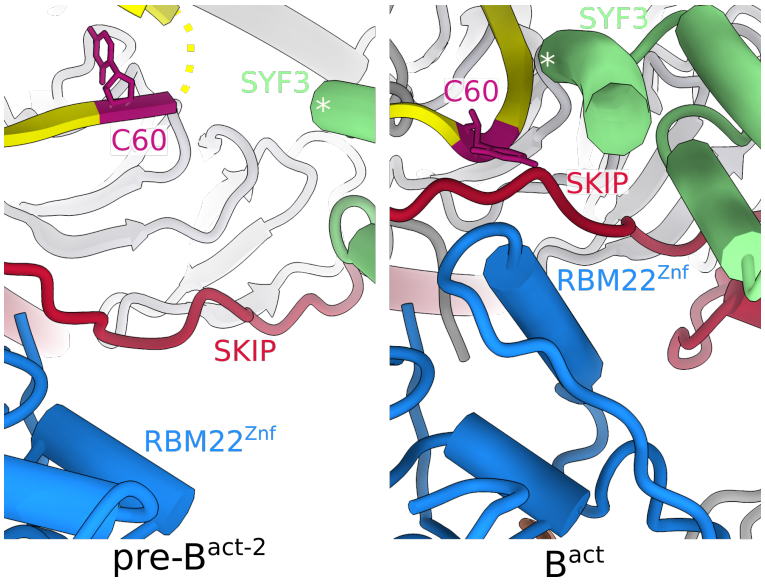

Fig. 4.2: U6-C60 interactions with SKIP, RBM22, SYF3. SKIP, RBM22, and SYF3 are present in both pre-B^{act-2} and B^{act} (PDB 6FF7), yet are farther away from U6 nt C60 in pre-B^{act-2}. Conformational change of PRP8 is likely responsible for rearranging this protein-RNA interaction in B^{act}, but the interaction appears to be primed in pre-B^{act-2}. $SYF3^{N238}$ is marked with an asterisk for orientation.

In particular, the distortions to the helical structure that would be introduced by flipping out of C60, A73, and U74 from the ISL do not appear likely in pre-B^{act}, as these conformational changes would lead to a more compact form of the ISL. The resolution of the ISL becomes progressively worse in regions farther from the loop, which is docked to $PRP8^{NTDL}$. This lower resolved region corresponds to the mid-stem region, in which the alternative base-pairing and conformations have been known to occur *in vitro*, as well as the lower stem region, which links the ISL to the U2/U6 helix Ib and U2/U6 helix II. While the exact base-pairing patterns in the U6 ISL of pre-B^{act-1} and pre-B^{act-2} cannot be precisely mapped, the more elongated form of the density suggests that A73, U74, and C60 map be stacked within the ISL, likely resulting in a base-pairing arrangement that is highly similar to that observed *in vitro* (see Fig. 3.17) [319].

Conformational dynamics of the U6 ISL between pre-B^{act-1} and pre-B^{act-2} suggest that the environment surrounding the ISL may be important for modulating its structure. In both pre-B^{act} complexes, many proteins that surround the catalytic center in B^{act} are not yet in place. WBP11 (see Fig. 3.14) passes immediately along the phosphate backbone of A73 and U74 in pre-B^{act-1}, preventing these nucleotides from being flipped out of the

ISL. Moreover, this positioning of WBP11 prevents potential interactions between A73 and the 5′ss by separating the U6/5′ss helix and the ISL. However, it is conceivable that even in the absence of WBP11, A73 and U74 may not be flipped out of the helix due to the expanded arrangement of the catalytic center, which results from the open conformation of PRP8 that separates $PRP8^{HB}$ from $PRP8^{NTD}$. In pre-B^{act} structures, the U6/5′ss helix is positioned farther away from the ISL and U2/U6 helix Ib (pre-$B^{act\text{-}2}$ only), which would likely prevent the interaction of A73 with G44, A45, and 5′ss even in the absence of WBP11, as in pre-$B^{act\text{-}2}$. This "expanded" configuration of U6/5′ss, the ISL, and U2/U6 helix Ib also prevents G46 and A47, located one and two nucleotides downstream of the ACAGA(GA) sequence of U6 snRNA, from interacting with helix Ib nucleotides to form the catalytic triple helix. In fact, the distance between G46 and the G54 is approximately 2.5 nm, totally preventing Hoogsteen interactions between these nucleotides as found in the triple helix. Moreover, U74 must be flipped out of the ISL to form stacking interactions with G46 as part of the triple helix, and this interaction cannot occur when U74 and G46 are physically separated. Thus, it is reasonable to speculate that A73 and U74 are flipped out of the ISL when the RNA network of the spliceosome is brought into the compact orientation of the B^{act}, but remain either flexible or stacked into the ISL while PRP8 is in an open conformation as in pre-B^{act}.

4.5.2 Conformational change of PRP8 allows for the formation of U2/U6 helix Ia and subsequently the triple helix

The overall arrangement of the nascent catalytic network could be in part explained by the positioning of the RNA network onto PRP8, the central scaffold of the spliceosome (see section 1.4.1). The open conformation of PRP8 in pre-B^{act} separates $PRP8^{HB}$ from $PRP8^{NTD}$, onto which the U6/5′ss helix and U2/U6 helix Ib are docked (Figs. 1.11 and 4.1). Formation of U2/U6 helix Ia as well as the catalytically essential triple-helix are likely facilitated by the large-scale conformational change of PRP8, in which removal of the B-specfic proteins allows for conversion from the open to closed conformation (Fig.

3.4). The closed conformation of PRP8 respositions $PRP8^{HB}$ along with $PRP8^{RT}$ and $PRP8^{En}$, such that these domains are brought into closer proximity with the $PRP8^{NTD}$. $PRP8^{HB}$ and the associated $CDC5L^{Myb}$ drive helix Ib closer to the the U6/5′ss helix, bringing them into contact to allow for the rearrangement of the U6 Ia loop to base-pair with U2 nts to form U2/U6 helix Ia (Fig. 3.18). Stabilization of U2/U6 helix Ia likely allows for the final steps in the folding of the catalytic center, namely the formation of the triple helix: U6 ISL stem is rearranged and bases A73 and U74 are flipped out of the ISL to participate in stacking interactions and triple helix formation, respectively (Fig. 3.17). Finally, the $PRP8^{NTDL}$ becomes structured, potentially contributing to stabilization of the ISL prior to catalysis (Figs. 1.11 and 4.1). Thus a large-scale conformational change of PRP8 generates a compact cavity at the core of the spliceosome, allowing for base-pairing between U74 of U6 ISL, helix Ib, and G46 and A47 one and two nts downstream of the U6/5′ss helix to form the mature catalytic center of the B^{act} (Fig. 3.18).

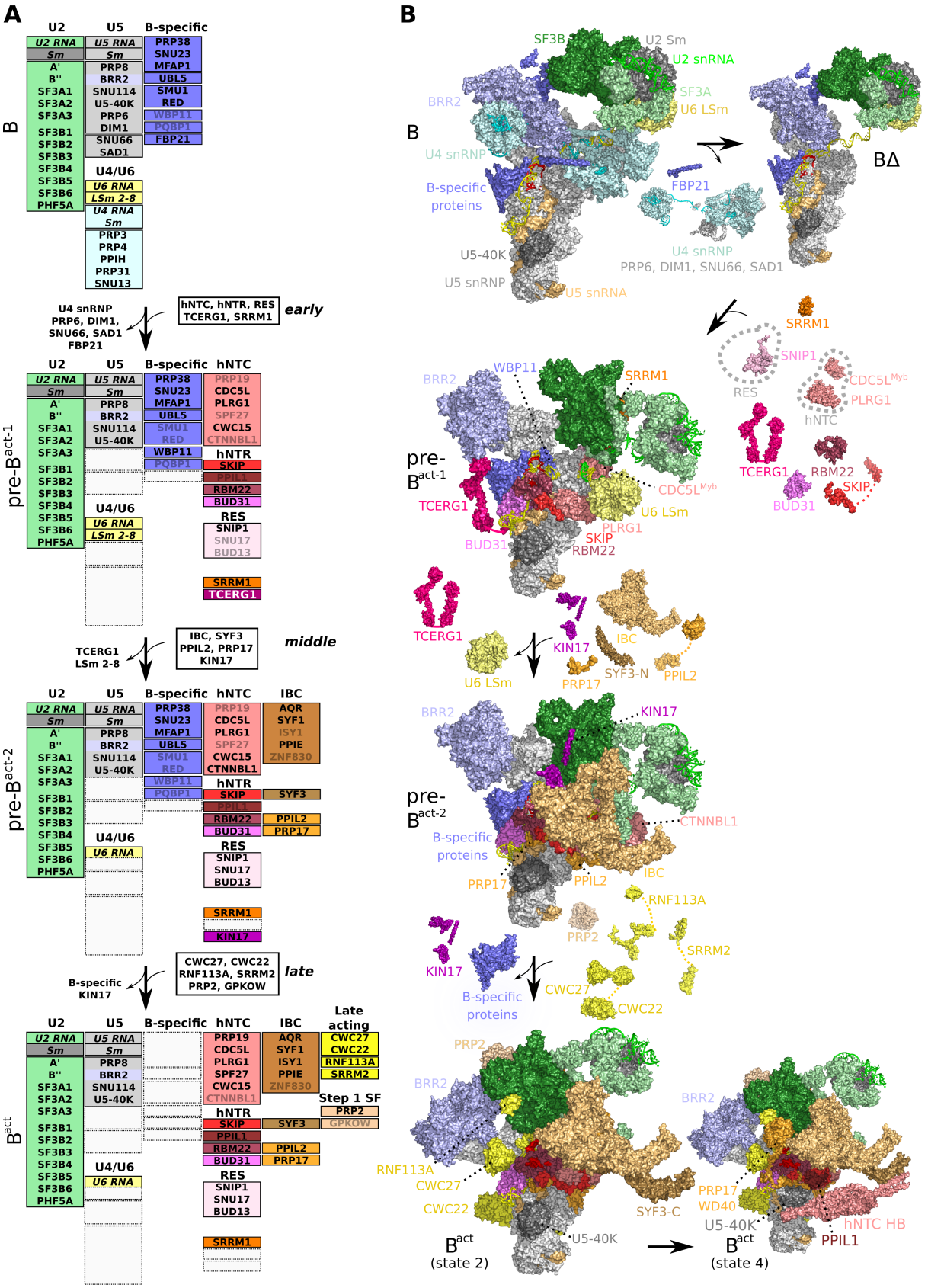

Fig. 4.3: Assembly pathway of pre-B^{act-1} and pre-B^{act-2} during activation. (A) Protein and RNA components of the B (PDB 5O9Z), pre-B^{act-1}, pre-B^{act-2}, B^{act} (unpublished PDB models) complexes are shown (top-to-bottom). Exchange of protein and RNA components is indicated by the vertical arrow, with factors that associate to the complex outlined in a box and factors that dissociate from the complex shown to the left of each arrow. (B) 3D representation of assembly pathway from B to B^{act}. In the top panel, a hypothetical intermediate BΔ is shown for clarity. In the bottom panel, states 2 and 4 of B^{act} are shown to illustrate the maturation of the B^{act}, in which several proteins or additional domains are stably integrated (PRP17^{WD40}, PPIL1, NTC^{HB}). From [131]. Reprinted with permission from AAAS.

4.6 Further considerations on spliceosome activation

Spliceosome activation constitutes one of the most drastic changes in protein and RNA inventory during the assembly of the spliceosome, with approximately 32 proteins and U4 snRNA leaving and at least 28 proteins joining the complex (Fig. 1.12). Multiple remodeling events are required to form the active site, with perhaps the most critical being the removal of U4 snRNP to allow for refolding of U6 into the catalytic network (Figs. 1.6 and 1.7). The pre-B^{act} structures offer many novel insights into this massive change in spliceosome inventory and configuration, as summarized in Figure 4.3. While DExD/H-box ATP-dependent helicases drive many transitions from one complex to another in the splicing cycle (see section 1.3.5), the driving force between discrete transitions in the ribonucleoprotein content and architecture of the spliceosome is less well understood.

4.6.1 Energy sources governing pre-B^{act} transitions

Transitions in the spliceosome assembly pathway are driven in large part by DExD/H-box ATP-dependent helicases. BRR2's ATPase activity catalyzes unwinding of the U4/U6 snRNA duplex, liberating from U4 snRNP from the spliceosome and allowing for large-scale rearrangements and compositional changes in the spliceosome to occur that ultimately result in the formation of B^{act} (Fig. 1.6). ATP hydrolysis provides a "driving force" for proceeding from one complex to the next in the splicing cycle, yet it is unclear what governs transitions between pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$, especially regarding the dissociation of certain factors and the recruitment of others.

An even greater degree of structural and compositional heterogeneity was documented in the B^{act}, of which at least eight unique states were present [99]. Transitions between states were consistent with previous biochemical studies detailing the assembly of the spliceosome, suggesting that each structural state represented a "snapshot" *en route* to a mature B^{act} complex [99]. Nonetheless, transitions from one state to the next could not be explained by the energy provided from ATP hydrolysis as there is no known

ATPase (e.g., an ATP-dependent helicase) that acts between these complexes. The high degree of heterogeneity in the sample was attributed to stochastic sampling of multiple conformations, mediated by thermal energy present in the sample prior to crosslinking and vitrification [99]. In a study of the 70S ribosome, similarly low energy barriers (< 3 k_BT) were observed in the energy landscape of the conformational changes of the complex, explaining that it can sample a variety of kinetic states independent of NTP hydrolysis [322].

By this logic, one would expect that the pre-B^{act} may also be able to sample a B^{act} conformation even in the absence of NTP hydrolysis, as GTP and ATP are not present in the buffer upon chemical crosslinking of the sample. However, this would require at the very least the dissociation of the B-specific proteins to allow for the closed conformation of PRP8 and the recruitment of B^{act} specific factors. While this may indeed occur in solution, the absence of B^{act} class averages in the cryo-EM dataset as well as the low concentration of factors associated with B^{act} formation suggests that this would at least be a minor population. Thus, one consequence of spliceosome inhibition by NSC95397 may be a kinetic trapping of the complexes, such that transitions between various states are slowed down, biasing the energy landscapes toward pre-B^{act}-like states.

4.6.2 Directionality of spliceosome activation phase

Steps 1 and 2 (intron lariat formation and exon ligation) of pre-mRNA splicing are reversible, as shown biochemically [117]. It was shown that mRNA can be converted with high fidelity into pre-mRNA by changing buffer conditions of the purified sample, namely the salt concentration/type and the pH [117]. If splicing is reversible, what prevents "backwards" assembly of the spliceosome? Dissociation of U4 snRNP has been reported to be essentially irreversible, indicating that following disruption of U4/U6 basepairing by BRR2 in the B complex, all subsequent complexes will contain only U2, U5, and U6 snRNPs (Fig. 1.6) [323]. Re-integration of U4 snRNP into the spliceosome after activation was shown to be very unlikely [323]. For U4 snRNP to be fully reintegrated to

the spliceosome, U6 intramolecular base-pairing and U2/U6 intermolecular base-pairing would have to be disrupted, likely by a helicase, in order to allow for U4 snRNA to base-pair with U6 snRNA as in the B complex. Although the composition and molecular architectures of the pre-B^{act-1} and pre-B^{act-2} strongly suggest that these intermediates are formed consecutively, it cannot be entirely ruled out that pre-B^{act-1} may be a breakdown product of pre-B^{act-2}. Backwards assembly from pre-B^{act-2} to pre-B^{act-1} would likely require the replacement of multiple mutually exclusive interactions and the destabilization of the multiple domains (i.e., such as those of NTC proteins) that become more ordered from pre-B^{act-1} to pre-B^{act-2}. NTC proteins were nonetheless shown to bind reversibly to the spliceosome, thus backwards assembly from pre-B^{act-2} to pre-B^{act-1} cannot be entirely ruled out [324]. The assembly pathway of the pre-B^{act} could be more closely analyzed by performing a time-series in the incubation of the splicing inhibitor, NSC95397, and comparing the relative abundances of different protein markers as a function of incubation time. Another way could be to perform a similar biochemical analysis but as a function of inhibitor concentration (Fig. 3.1).

4.6.3 Alternative spliceosome assembly pathways?

Early spliceosome assembly has been reported to proceed not only by the sequential addition of U1 followed by U2 to the pre-mRNA, but also U2 followed by U1 binding [325]. Another report indicated that the NTC complex, which is generally thought to join the spliceosome during activation (i.e., after the integration of the tri-snRNP), was recruited to the pre-mRNA via Pol II CTD prior to the integration of the tri-snRNP [326]. Both pieces of evidence suggest that in addition to the canonical splicing cycle, there may be alternative spliceosome assembly pathways at discrete steps of assembly. This could bring into question whether pre-B^{act} structures assemble in a canonical or an off-pathway manner. Steps preceding the formation of the pre-B^{act-1} could be initiated with either U1 or U2 binding the pre-mRNA first, although further experiments would be required to determine whether this alternative assembly event occurs during pre-B^{act}

formation. Conversely, it appears highly unlikely that the NTC is integrated into the pre-B^{act} complex via an alternative pathway, but rather that it joins the spliceosome after tri-snRNP addition. Comparison of the pre-B^{act} structures strongly suggests that pre-$B^{act\text{-}1}$ is a precursor to pre-$B^{act\text{-}2}$, as evidenced by the gradual structuring of several proteins of the NTC (Fig. 3.9). Moreover, the temporal ordering of component addition to each structure is in line with the well established canonical pathway of spliceosome assembly, including PRP17, the RES complex, the IBC complex, NTR proteins, and other splicing factors such as SRRM1 (Figs. 3.12, 3.8, and 3.10) [116, 214, 220, 221, 223, 225, 276]. Thus while early events in prespliceosome formation could potentially proceed via an alternative pathway, it appears unlikely that downstream assembly events proceed noncanonically during pre-B^{act} assembly. Nonetheless, it cannot be entirely ruled out that pre-$B^{act\text{-}1}$ is a breakdown product of pre-$B^{act\text{-}2}$, or that there is an equilibrium between the two complexes.

4.6.4 Possible mechanisms by which the small molecule NSC95397 may stall spliceosome assembly

NSC95397 was identified as an inhibitor of spliceosome assembly in a small molecule screen [71]. NSC95397 was documented previously to be a CDC25 dual-specificity phosphatase inhibitor that can bind covalently to cysteine residues [327, 328]. CDC25 is involved in control of cell cycle checkpoints, but does not have a known target in the spliceosome and has not been shown to associate with the spliceosome [329]. In a previous study, several known CDC25 inhibitors were assayed in addition to NSC95397 [71]. Small molecules containing a similar organic structure of a 1,4-napthoquinone backbone were also shown to inhibit splicing, although not as potently as NSC95397. Another CDC25 inhibitor (PM-20) was structurally distinct, not bearing a 1,4-napthoquinone backbone, but did not inhibit splicing. Moreover, the IC_{50} of splicing inhibition by NSC95397 is much higher (50 fold) than that for CDC25 inhibition. Taken together, these results suggested that stalling of spliceosome assembly by NSC95397 is likely due to a mechanism other than CDC25 inhibition [71].

Additional structure activity relationship studies have analyzed quinone-based molecules in more detail. It was shown that these derivatives alter protein function in two ways: 1) covalent binding to cysteine residues; 2) production of reactive oxygen species (ROS) that interfere with protein function by oxidizing reactive side chains, including the thiol moiety of cysteines [330, 331]. In one case, NSC95397 bound covalently to cysteine residues and thereby contributed to disruption of protein-protein interactions [328]. NSC95397 was found to generate H_2O_2, a type of ROS, in the presence of dithiothreitol (DTT) [332], illustrating that it is capable of redox cycling. An *in vitro* study of quinone derivatives structurally similar to NSC95397 revealed that these compounds inhibited spliceosome assembly in a dose-dependent manner, but that this block could be partially removed by addition of DTT [333], which presumably prevents further oxidation of cysteine residues by ROS. An earlier study showed that the quinone-containing compound, β-lapachone, generated ROS that stalled snRNP assembly by promoting the formation of disulfide crosslinks between survival of motor neuron complexes (SMN), which are involved in the Sm core biosynthesis [334]. Likewise, it was shown that NSC95397 reduces snRNP assembly *in vitro*, and this effect was proposed to result from generation of ROS [71]. It was therefore suggested that the main mechanism of splicing inhibition by quinone derivatives is likely due to ROS-induced oxidation of cysteine residues of one or more proteins involved in spliceosome assembly, rather than covalent modification of cysteine residues [250, 333]. Nonetheless, it is conceivable that both of these effects may be at play. Incubation of splicing reactions with quinone derivatives resulted in accumulation of spliceosomes stalled at the B complex, suggesting that the target of these small molecules is likely a factor involved in activation [333]. This is in line with our results with the pre-B^{act}, which is in a precatalytic state and migrates similar to a B complex on a native agarose gel. However, only upon closer biochemical dissection and structural analysis does it become evident that the complex has progressed past the B complex *en route* to activation.

It has previously been shown that phosphatase inhibition can stall spliceosome assembly [199]. A cocktail of phosphatase inhibitors was used to stall assembly at the pre-B/B

complex level, suggesting that dephosphorylation of at least one unidentified target is required for spliceosome activation [199]. Stalling of spliceosome assembly prior to activation, induced by small molecule quinone derivatives, may point to a disruption of reversible protein phosphorylation, perhaps by the inhibition of one or more phosphatases. It is conceivable that NSC95397 may stall spliceosome assembly by preventing dephosphorylation of *trans*-acting proteins involved in spliceosome regulation, but not on spliceosome proteins directly. Alternatively, NSC95397 might exert its effect by preventing dephosphorylation of one or more yet uncharacterized target(s) bound to the spliceosome, thus blocking splicing catalysis by inhibiting dephosphorylation, or by a combination of both activities. The B-specific proteins prevent conformational change of $PRP8^{RT/EN}$ domains and therefore stall the spliceosome prior to activation. Although highly speculative, NSC95397 could potentially be involved in regulation of this protein exchange, possibly by influencing the post-translational modifications of the B-specific proteins directly or other proteins which govern their dissociation from the spliceosome. An examination of the post-translational modifications of the B-specific proteins before and after incubation with the NSC95397 may be warranted. Alternatively, a more general approach could be taken to profile post-translational modifications of the global sample proteome with and without incubation with NSC95397. Potential regulators of such transitions could be purified and supplemented to the inhibited splicing reaction, to determine whether they are sufficient to relieve the assembly block induced by NSC95397. Furthermore, it would be highly interesting to test whether splicing inhibition by NSC95397 can be rescued by DTT, which would likely indicate that the mechanism by which it inhibits spliceosome assembly is at least partially through the generation of ROS. However, it may be that NSC95397 exerts its activity by nonspecific mechanisms and not by selective inhibition of a single target. Identification of the target(s) of NSC95397 may uncover novel interactions that regulate spliceosome activation.

4.7 Classification of dynamic complexes in single particle cryo-EM

Using image classification techniques, it was possible to identify two distinct macromolecular complexes within a single biochemical purification detailed in this study (Figs. 3.2 and 3.3) [131]. Classification techniques used in single particle cryo-EM are able to sort for compositional and conformational heterogeneity in a dataset, as well as for fine differences in portions of a single refined complex, as recently shown for an novel intermediate state of the human C complex (C_i) [335]. However, classification of flexible regions remains a challenging task, the success and specific outcome of which can vary largely based on many user-defined parameters. Several variables that are likely to vary from one user to the next include the use and design of masks; selection, modification, and filtering of the initial reference structure; rotational parameter specification during multi-body refinement; efficiency of 2D/3D classification; choice of regularization parameter (T); number of classes; number of iterations; and angular search options. The choice of software used in image classification and map reconstruction also likely has an impact on the exact makeup of particles included in the final reconstruction, as well as their angular assignments determined during refinement steps.

It would be advantageous to have algorithms which remove user-induced bias of which parts of a macromolecule are "important," instead determining every possible region of a complex to the highest possible resolution in an unsupervised manner. Such a method would be particularly powerful if it could refine multiple regions of a flexible macromolecular complex without the need for highly variable, user-defined masks as used in multi-body refinements or focused classifications.

Shortly following publication of the manuscript corresponding to this work, a new algorithm for "nonuniform refinement" of cryo-EM reconstructions was published [336]. In refinements of macromolecular complexes, poor alignments in flexible regions often worsen the alignment of the entire reconstruction. While masked refinements partially ameliorate

this problem, they must be manually defined and they introduce user-specific variability as well as artifacts at masks interfaces. Nonuniform refinement implements a novel strategy in which different parts of a reconstruction are filtered based on their alignment accuracy, instead of filtering the entire reconstruction in the same manner. This allows stable portions to become progressively well aligned without overfitting noisy regions in the periphery of the complex - all without the introduction of user-implemented masks. In test cases of publicly available data, nonuniform refinement was shown to improve the local resolution not only of the centrally located, stable portions of the reconstructions, but also those within peripheral regions.

A limitation of nonuniform refinement is that it is a refinement step only, and does not perform classification of the data into multiple subsets. Thus, nonuniform refinement assumes that the data has been sufficiently classified to remove compositional and conformational heterogeneity, even though this is probably not the case for most datasets. Deep learning approaches have emerged in the last year as novel ways to classify cryo-EM datasets having high structural heterogeneity, and will likely continue to bring the field closer to a complete picture of the entire array of structural states sampled by the spliceosome and other dynamic macromolecular assemblies [337, 338]. In another method to characterize structural heterogeneity, 3D principal component analysis (PCA) approaches have been used to calculate energy landscapes of large macromolecules, uncovering heterogeneous assembly intermediates in single biochemical preparations and mapping them to a Boltzmann distribution [99, 167, 339, 340]. In principle, 3D PCA approaches could also be preceded with unsupervised *ab initio* structure determination, possibly helping to detect classes that would otherwise not be properly classified if using only a single reference. More accurate 3D classifications will lead to reconstructions that are closer to representing the "ground truth" of a given conformation or composition of a complex, by improving the homogeneity of the final particle subset to be refined instead of refining small parts of a heterogeneous mixture of similar structures.

Although cryo-EM studies of the spliceosome and other large biological complexes have made great strides in determining the molecular architectures of very dynamic samples,

it is clear that there is substantial room for improvement if atomic or "near atomic" (e.g., better than 2 Å) resolutions are to be reached. Such resolutions will bring cryo-EM investigations of challenging biochemical samples into the realm of chemistry, at which point chemical modifications and catalytic mechanisms can be interrogated with accuracy and precision. Better microscope hardware leading to significant boost in micrograph quality represent a clear path forward in collecting better raw data [183, 184]; likewise, more robust classification methods should be developed to deal with the compositional and conformational heterogeneity of the spliceosome and other dynamic complexes at high resolution.

5 Conclusion and Outlook

The work presented herein sought to address two major questions:

- Are there structurally distinct intermediate stages of activation at which formation of the catalytic center can be observed?
- How do proteins facilitate the folding of active site U2/U6 snRNAs in the spliceosome?

By employing a small molecule inhibitor of pre-mRNA splicing, it was possible to stall spliceosome activation at a previously uncharacterized stage between a precatalytic B complex and the activated B^{act} complex. Analysis of the single biochemical preparation uncovered two distinct structural states of the spliceosome – pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$ – each lacking a mature catalytic center while bearing the U6 ISL. Importantly, our structures have implications for step-wise formation of the catalytic center's active site: the ISL forms first, followed by helix Ib, and ends with formation of helix Ia and subsequently the triple helix. The pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$ provide the first structural insights into the ordering by which tertiary interactions of the spliceosome's catalytic center is assembled.

Our structures also provided new insights into conformational change, mutually exclusive protein-protein interactions, and gradual stabilization of numerous proteins which appear to facilitate the folding of active site U2/U6 snRNAs in the core of the spliceosome. Both pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$ surprisingly showed that the central scaffold protein of the spliceosome, PRP8, remains in an open conformation even as the catalytic center begins to form. Comparison of our structures to the B^{act} further suggests that the closed conformation is the final step in the formation of the catalytic center, acting as a "mold" in which $PRP8^{HB}$ pushes U2/U6 helix Ib towards the U6/5′ss and the ISL, removing the

spatial restraint between these regions and thus allowing helix Ia and the triple helix to form. Although mutually-exclusive protein-protein interactions are a well-documented feature of spliceosome assembly, the pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$ provide the first structural evidence for several interactions, such as the concomitant exchange of LSm proteins for SYF3/IBC proteins at U2/U6 helix II; WBP11 docking and helix Ia formation; KIN17 and PRP2 binding to SF3B1. Furthermore, the pre-$B^{act\text{-}1}$ and pre-$B^{act\text{-}2}$ reveal clues as to the gradual stabilization of multiple protein subcomplexes or groups on the spliceosome, such as the NTC, NTR, RES, and the IBC. In addition, our structures indicate that several protein regions are poised for stabilization in B^{act} ($PRP8^{NTDL}$, SF3B2, SF3A2, and CWC15), initially remaining flexible as a possible means of either permitting or guiding initial refolding U2/U6 snRNAs into the catalytic center. By revealing discrete structural snapshots of spliceosome activation, we were able to uncover previously uncharacterised protein-RNA interactions that appear to be part of a concerted process that ensures proper folding of the catalytic center.

While the pre-B^{act} structures deliver novel snapshots of human spliceosome assembly, they also bring up new questions that could become topics for future investigation. First, what is/are the target(s) of the small molecule chemical inhibitor NSC95397? The answer to this question may uncover hitherto unclarified mechanisms of splicing regulation. Second, what governs the highly ordered exchange of proteins that have mutually exclusive interactions during activation? Of particular importance in this category are the B-specific proteins, whose dissociation is an obligate step to allow for PRP8 closure. Third, are the protein regions undergoing disorder-order transitions surrounding the catalytic centers necessary for splicing? Truncation or mutation of residues interacting with the U2/U6 snRNA network would be one means of examining this question in more detail. Fourth, what is responsible for the apparent structural plasticity of the U6 ISL? Addressing this question will likely help to clarify studies related to the conformational dynamics of the ISL, but in the context of the spliceosome instead of an *in vitro*, RNA-only system.

It will be essential to examine the formation of the spliceosome's catalytic center in more depth and at higher resolutions. Pushing the core of the pre-B^{act} structures and other

complexes to sub-2 Å resolutions will likely unlock unanticipated structural modalities that the spliceosome uses to fine-tune the construction of its catalytic center. In addition, much higher resolutions would enable the modeling of the divalent metal cations coordinated by the phosphate backbone of nucleotides within the catalytic center, as well as structural water molecules.

A Abbreviations

aa	amino acid
ATP	adenosine triphosphate
BS	branch site
BS-A	branch site adenosine
BSL	branch site stem-loop
cryo-EM	cryo electron microscopy
Cs	spherical aberration coefficient
CTD	C-terminal domain
CTF	contrast transfer function
DExD/H	Aspartate-Glutamate-x-Aspartate/Histidine
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EM	electron microscopy
FSC	Fourier Shell Correlation
GTP	guanosine triphosphate
HEAT	**H**untingtin, **e**longation factor 3, PR/**A** subunit of protein phosphatase 2A and the **T**OR lipid kinase
IBC	Intron Binding Complex
IC_{50}	50% inhibitory concentration
ISL	internal stem-loop
kDa	kilodalton
MDa	megadalton
mRNA	messenger RNA
MW	molecular weight
m_3G	2,2,7-trimethylguanosine
m^7G	7-methyl guanosine
NMR	nuclear magnetic resonance

NTC	NineTeen Complex
nt	nucleotide
NTD	N-terminal domain
NTR	NineTeen Complex-Related
PAGE	polyacrylamide gel electrophoresis
PCA	principal component analysis
PDB	Protein Data Bank
Pol	polymerase
pre-mRNA	precursor mRNA
PSF	point spread function
RES	Retention and Splicing
RNA	ribonucleic acid
RNP	ribonucleoprotein
ROS	reactive oxygen species
RRM	RNA recognition motif
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SNR	signal-to-noise ratio
SR	serine-arginine rich
ss	splice site
XL-MS	cross-linking mass spectrometry
2D	two-dimensional
3D	three-dimensional

B Supplementary information

Table B.1: Intermolecular crosslinks. Crosslink analyses by pLink2.3.5 at FDR 1, 3 and 5% are indicated. Number of CSMs (crosslinked peptide spectrum matches) and highest score are shown for each peptide. "Residue 1" and "Residue 2" denote crosslinked residues from Protein 1 and Protein 2. Crosslink analysis was performed by Dr. Olex Dybkov (Department of Cellular Biochemistry, MPI-BPC) in collaboration with Prof. Dr. Henning Urlaub (Bioanalytical Mass Spectrometry, MPI-BPC) (see 2.2.6).

Protein 1		Protein 2				CSMs			$Score_{max}$		
Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
AQR	O60306	PRP19	Q9UMS4	759	200		4	9		0.246	0.246
		SYF1	Q9HCS7	954	2	10	10	11	0.991	0.991	0.991
BRR2	O75643	CBP80	Q09161	1169	65		2	2		0.609	0.609
		CDC5L	Q99459	60	170	9	10	10	1.030	1.030	1.030
				73	170		2	2		0.353	0.353
				83	170		3	3		0.334	0.334
		MFAP1	P55081	1142	249	3	7	8	0.516	0.516	0.516
					250	27	36	37	1.485	1.485	1.485
					256	2	3	4	0.795	0.795	0.795
				1145	249	2	3	4	0.963	0.963	0.963
					250	14	24	28	1.556	1.556	1.556
					256			3			0.454
				1146	242	40	41	41	2.453	2.453	2.453
					249	13	26	30	0.813	0.813	0.813
					250	22	28	30	2.382	2.382	2.382
					256	18	18	18	2.300	2.300	2.300
				1169	250		2	3		0.335	0.335
					256		2	3		0.399	0.399
		PQBP1	O60828	60	87	2	3	3	1.266	1.266	1.266
				73	2		3	3		0.386	0.386
					18	2	2	3	1.087	1.087	1.087
				85	2	7	7	7	2.323	2.323	2.323
					18	2	2	2	2.851	2.851	2.851
				103	18	2	2	2	0.950	0.950	0.950
		PRP8	Q6P2Q9	14	1306		2	3		0.548	0.548
				73	1994	2	10	11	0.522	0.522	0.522
				1039	1636		5	6		0.535	0.535
					1659	15	19	20	1.289	1.289	1.289
					2049	3	3	3	1.131	1.131	1.131
					2070	6	10	10	1.031	1.031	1.031
				1049	1866	13	13	13	3.092	3.092	3.092
					1984	2	2	2	1.289	1.289	1.289
					2070	2	3	3	1.655	1.655	1.655
				1141	1636		2	2		1.184	1.184
				1146	1636	9	9	9	1.635	1.635	1.635
					1649	7	8	8	2.375	2.375	2.375
		RED	Q13123	1874	520	6	6	6	1.016	1.016	1.016
		SMU1	Q2TAY7	971	308	14	17	18	1.051	1.051	1.051
				975	308	3	3	3	2.201	2.201	2.201
				1544	308		2	2		0.279	0.279
				1552	308	31	50	53	1.971	1.971	1.971
				1711	308			2			0.450
				1715	308	11	15	16	1.197	1.197	1.197
		SNU23	Q96NC0	85	70	2	2	2	2.348	2.348	2.348
				479	8	1	4	6	1.354	1.354	1.354
				557	8	12	21	24	1.121	1.121	1.121
				696	160	4	6	7	0.472	0.472	0.472
				729	147			2			0.568
				733	138	5	7	7	0.643	0.643	0.643
					147	4	5	5	1.856	1.856	1.856
				804	8	4	4	4	0.759	0.759	0.759
					15	6	8	8	1.189	1.189	1.189
					18			2			0.344
		TCERG1	O14776	60	570	3	5	6	0.683	0.683	0.683
				440	634		3	3		1.936	1.936
		WBP11	Q9Y2W2	14	168	5	5	5	0.851	0.851	0.851
				60	168	2	4	4	0.754	0.754	0.754
					169	4	5	6	1.501	1.501	1.501
				85	48			2			0.189
					59	2	2	2	2.734	2.734	2.734

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
					168	2	2	2	14.531	14.531	14.531
BUD13	Q9BRD0	PRP8	Q6P2Q9	65	1831	14	14	14	2.929	2.929	2.929
					1840	20	21	21	2.421	2.421	2.421
				68	1831	40	43	45	3.605	3.605	3.605
					1840	24	24	24	3.027	3.027	3.027
					2108	3	3	3	2.055	2.055	2.055
				605	1505		2	2		1.180	1.180
		SKIP	Q13573	605	441	22	23	23	3.246	3.246	3.246
BUD31	P41223	TCERG1	O14776	68	711		4	5		0.521	0.521
CBP20	P52298	CBP80	Q09161	7	17	8	11	14	0.796	0.796	0.796
					37	3	6	6	0.608	0.608	0.608
					67	6	9	14	1.536	1.536	1.536
					327	7	13	14	0.719	0.719	0.719
				34	20	4	4	5	0.919	0.919	0.919
				38	657		3	3		0.751	0.751
				67	607	20	31	33	0.932	0.932	0.932
					654	3	5	7	0.532	0.532	0.532
					657		2	3		0.567	0.567
				68	607	124	127	127	3.640	3.640	3.640
				78	511	6	6	6	3.174	3.174	3.174
		SRSF1	Q07955	7	165	21	22	22	2.728	2.728	2.728
CBP80	Q09161	BRR2	O75643	65	1169		2	2		0.609	0.609
		CBP20	P52298	17	7	8	11	14	0.796	0.796	0.796
				20	34	4	4	5	0.919	0.919	0.919
				37	7	3	6	6	0.608	0.608	0.608
				67	7	6	9	14	1.536	1.536	1.536
				327	7	7	13	14	0.719	0.719	0.719
				511	78	6	6	6	3.174	3.174	3.174
				607	67	20	31	33	0.932	0.932	0.932
					68	124	127	127	3.640	3.640	3.640
				654	67	3	5	7	0.532	0.532	0.532
				657	38		3	3		0.751	0.751
					67		2	3		0.567	0.567
		SKIP	Q13573	607	503	4	4	4	2.187	2.187	2.187
					509	2	3	3	1.072	1.072	1.072
		SRSF1	Q07955	17	165	4	4	4	1.131	1.131	1.131
				20	165	3	3	3	1.855	1.855	1.855
					193	4	5	5	1.189	1.189	1.189
		TCERG1	O14776	187	756	1	1	1	0.510	0.510	0.510
		ZC3H18	Q86VM9	221	319	19	21	22	1.585	1.585	1.585
CCDC12	Q8WUD4	MFAP1	P55081	123	249	3	7	10	0.619	0.619	0.619
					250	11	15	19	1.172	1.172	1.172
					256	5	11	11	0.818	0.818	0.818
		SYF1	Q9HCS7	23	532	3	6	6	0.634	0.634	0.634
				42	532	5	5	5	1.178	1.178	1.178
				94	593	2	2	2	3.169	3.169	3.169
		SYF3	Q9BZJ0	117	607		3	3		0.700	0.700
				126	549		3	4		0.399	0.399
CDC5L	Q99459	BRR2	O75643	170	60	9	10	10	1.030	1.030	1.030
					73		2	2		0.353	0.353
					83		3	3		0.334	0.334
		CTNNBL1	Q8WYA6	28	56	8	9	9	1.192	1.192	1.192
				218	458	3	3	3	2.145	2.145	2.145
		CWC15	Q9P013	135	152		2	2		0.286	0.286
		PLRG1	O43660	432	113		2	2		0.336	0.336
				522	80	8	9	10	1.157	1.157	1.157
				598	31	5	6	7	2.178	2.178	2.178
					41	3	3	3	1.493	1.493	1.493
					68	2	3	3	1.340	1.340	1.340
				601	31	15	15	15	10.025	10.025	10.025
					41	2	2	2	0.749	0.749	0.749
					62	3	3	3	1.546	1.546	1.546
					68	15	16	16	3.260	3.260	3.260
					80	3	3	3	16.912	16.912	16.912
				708	80		2	3		0.383	0.383
				733	135		2	2		1.173	1.173
		PPIL2	Q13356	466	507	3	4	4	0.394	0.394	0.394
					508		4	4		0.310	0.310
		PRP17	O60508	200	333	6	12	12	0.566	0.566	0.566
		PRP19	Q9UMS4	380	244	2	2	2	0.753	0.753	0.753
				432	179		2	2		0.693	0.693
					244		2	2		0.634	0.634
				487	192	7	7	8	2.404	2.404	2.404
				522	192	6	6	6	1.479	1.479	1.479
					244	4	5	5	0.912	0.912	0.912
				570	122	9	23	27	0.675	0.675	0.675
				631	122	3	3	3	1.567	1.567	1.567
				685	179	2	4	4	0.741	0.741	0.741
					192			3			0.257
					244		9	9		0.381	0.381
					266	2	4	4	0.412	0.412	0.412
				718	76	72	76	77	6.685	6.685	6.685
				771	244		3	3		0.508	0.508

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
				782	179		2	2		0.306	0.306
					192	5	6	6	1.379	1.379	1.379
					244	28	33	34	1.343	1.343	1.343
					261	3	3	3	1.265	1.265	1.265
					266	10	11	14	0.925	0.925	0.925
				795	244	7	8	8	1.168	1.168	1.168
		PRP8	Q6P2Q9	7	892	37	51	52	1.770	1.770	1.770
				28	727	29	36	39	1.463	1.463	1.463
		SF3A2	Q15428	28	10	5	6	6	1.289	1.289	1.289
					42	9	13	13	1.355	1.355	1.355
		SF3B1	O75533	76	454	2	2	2	0.887	0.887	0.887
				106	454	5	5	5	3.348	3.348	3.348
		SF3B6	Q9Y3B4	124	29	26	26	26	9.871	9.871	9.871
		SKIP	Q13573	28	255	6	9	9	1.319	1.319	1.319
					258	16	25	25	1.281	1.281	1.281
				47	258		3	6		0.327	0.327
				124	311	8	16	19	0.641	0.641	0.641
		SPF27	O75934	380	168	2	2	3	1.331	1.331	1.331
				685	97	6	17	22	0.689	0.689	0.689
				686	97	15	17	17	2.361	2.361	2.361
				771	177	74	126	151	1.449	1.449	1.449
				782	191	8	11	12	1.148	1.148	1.148
		SYF1	Q9HCS7	294	654	19	19	20	3.245	3.245	3.245
				522	532	8	8	8	1.502	1.502	1.502
		SYF3	Q9BZJ0	466	568			2			0.277
		U5-40K	Q96DI7	795	270	2	3	5	0.515	0.515	0.515
CTNNBL1	Q8WYA6	CDC5L	Q99459	56	28	8	9	9	1.192	1.192	1.192
				458	218	3	3	3	2.145	2.145	2.145
		CWC15	Q9P013	297	91	3	4	8	0.539	0.539	0.539
				372	91	2	3	3	0.823	0.823	0.823
				527	18			2			0.250
		PPIL2	Q13356	56	90	3	3	4	0.545	0.545	0.545
		PRP19	Q9UMS4	534	192	4	4	5	2.205	2.205	2.205
		PRP8	Q6P2Q9	27	796	16	18	18	3.179	3.179	3.179
				31	796	13	13	13	2.264	2.264	2.264
		SF3B1	O75533	56	6	2	2	2	1.561	1.561	1.561
				83	6	6	7	7	2.140	2.140	2.140
				84	3			2			0.302
					6	12	13	13	1.691	1.691	1.691
				91	3		2	2		0.376	0.376
					6	6	14	20	2.710	2.710	2.710
				95	6		3	3		0.823	0.823
CWC15	Q9P013	CDC5L	Q99459	152	135		2	2		0.286	0.286
		CTNNBL1	Q8WYA6	18	527			2			0.250
				91	297	3	4	8	0.539	0.539	0.539
					372	2	3	3	0.823	0.823	0.823
		PLRG1	O43660	18	363	15	16	16	1.804	1.804	1.804
				28	320	3	3	3	1.764	1.764	1.764
		PRP8	Q6P2Q9	91	1144		2	2		0.317	0.317
				199	1158	2	3	4	0.543	0.543	0.543
				205	987		3	4		0.379	0.379
		RBM22	Q9NW64	18	76	5	7	7	1.223	1.223	1.223
				28	40	3	4	4	0.842	0.842	0.842
		SKIP	Q13573	18	153	16	16	16	2.622	2.622	2.622
					217	1	5	5	0.535	0.535	0.535
				28	158		2	2		0.932	0.932
					204	2	3	4	0.621	0.621	0.621
					217	3	4	5	0.573	0.573	0.573
				183	311	2	2	3	0.816	0.816	0.816
		ZNF830	Q96NB3	147	26		2	2		0.405	0.405
HNRNPA1	P09651	PHF5A	Q7RTV0	15	104		2	2		0.288	0.288
ISY1	Q9ULR0	SF3A1	Q15459	105	30	2	4	4	0.502	0.502	0.502
					37	5	6	6	0.664	0.664	0.664
		SF3B2	Q13435	101	148		2	2		0.248	0.248
		SYF1	Q9HCS7	190	2	4	6	6	2.763	2.763	2.763
				260	2			2			0.660
		U2-A'	P09661	266	193	17	17	17	2.627	2.627	2.627
					221	6	6	6	2.095	2.095	2.095
KIN17	O60870	MFAP1	P55081	231	67	3	3	3	0.719	0.719	0.719
		PRP8	Q6P2Q9	87	2034		2	3		0.280	0.280
		SF3B1	O75533	271	729		2	2		0.434	0.434
					733	3	6	6	7.933	7.933	7.933
		SF3B3	Q15393	231	26	13	15	15	1.957	1.957	1.957
					296	2	2	2	1.531	1.531	1.531
		SNU23	Q96NC0	120	70	3	3	3	2.324	2.324	2.324
LSm2	Q9Y333	PLRG1	O43660	88	510	4	5	5	0.915	0.915	0.915
		SKIP	Q13573	8	236	2	3	4	0.430	0.430	0.430
LSm4	Q9Y4Z0	LSm8	O95777	1	28	14	14	14	2.730	2.730	2.730
				80	28	7	7	7	3.106	3.106	3.106
LSm6	P62312	PLRG1	O43660	2	7	1	1	1	0.472	0.472	0.472
		RED	Q13123	5	30	6	7	8	1.153	1.153	1.153
				77	30			2			0.456
		SF3A1	Q15459	5	131	4	4	4	1.310	1.310	1.310

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
LSm8	O95777	LSm4	Q9Y4Z0	28	1	14	14	14	2.730	2.730	2.730
					80	7	7	7	3.106	3.106	3.106
MBPMS2	P0AEX9-r	PRP8	Q6P2Q9	47	452			2			0.209
				438	2293	5	17	19	0.472	0.472	0.472
		SF3B1	O75533	274	656	2	3	3	1.690	1.690	1.690
		SF3B2	Q13435	47	448	2	3	3	0.676	0.676	0.676
		SmE	P62304	278	1			2			0.661
		SRRT	Q9BXP5	103	286	7	8	10	1.052	1.052	1.052
				180	286	25	31	40	1.136	1.136	1.136
				363	286		4	6		0.357	0.357
		U2-A'	P09661	103	192			3			0.191
MFAP1	P55081	BRR2	O75643	242	1146	40	41	41	2.453	2.453	2.453
				249	1142	3	7	8	0.516	0.516	0.516
					1145	2	3	4	0.963	0.963	0.963
					1146	13	26	30	0.813	0.813	0.813
				250	1142	27	36	37	1.485	1.485	1.485
					1145	14	24	28	1.556	1.556	1.556
					1146	22	28	30	2.382	2.382	2.382
					1169		2	3		0.335	0.335
				256	1142	2	3	4	0.795	0.795	0.795
					1145			3			0.454
					1146	18	18	18	2.300	2.300	2.300
					1169		2	3		0.399	0.399
		CCDC12	Q8WUD4	249	123	3	7	10	0.619	0.619	0.619
				250	123	11	15	19	1.172	1.172	1.172
				256	123	5	11	11	0.818	0.818	0.818
		KIN17	O60870	67	231	3	3	3	0.719	0.719	0.719
		PRP8	Q6P2Q9	238	2293	9	14	16	0.848	0.848	0.848
				242	1649	25	30	31	1.453	1.453	1.453
					2293	46	53	54	2.052	2.052	2.052
				249	1649	41	43	46	2.123	2.123	2.123
				250	1649	17	17	17	2.174	2.174	2.174
				256	1636	2	2	2	0.687	0.687	0.687
					1649			2			0.207
		SF3B1	O75533	9	807	15	15	15	1.516	1.516	1.516
		SF3B3	Q15393	67	26		2	2		0.248	0.248
		SMU1	Q2TAY7	67	308	17	21	21	1.104	1.104	1.104
		SRSF1	Q07955	329	179	13	18	19	1.295	1.295	1.295
		SYF1	Q9HCS7	285	539	4	4	4	0.809	0.809	0.809
		TCERG1	O14776	415	878	4	9	10	1.176	1.176	1.176
PHF5A	Q7RTV0	HNRNPA1	P09651	104	15		2	2		0.288	0.288
		SF3B1	O75533	108	946		2	2		0.289	0.289
		SF3B3	Q15393	13	137	7	9	9	1.464	1.464	1.464
		SF3B5	Q9BWJ5	13	17	6	7	7	1.145	1.145	1.145
PLRG1	O43660	CDC5L	Q99459	31	598	5	6	7	2.178	2.178	2.178
					601	15	15	15	10.025	10.025	10.025
				41	598	3	3	3	1.493	1.493	1.493
					601	2	2	2	0.749	0.749	0.749
				62	601	3	3	3	1.546	1.546	1.546
				68	598	2	3	3	1.340	1.340	1.340
					601	15	16	16	3.260	3.260	3.260
				80	522	8	9	10	1.157	1.157	1.157
					601	3	3	3	16.912	16.912	16.912
					708		2	3		0.383	0.383
				113	432		2	2		0.336	0.336
				135	733		2	2		1.173	1.173
		CWC15	Q9P013	320	28	3	3	3	1.764	1.764	1.764
				363	18	15	16	16	1.804	1.804	1.804
		LSm2	Q9Y333	510	88	4	5	5	0.915	0.915	0.915
		LSm6	P62312	7	2	1	1	1	0.472	0.472	0.472
		PPIL1	Q9Y3C6	62	80	2	3	3	0.573	0.573	0.573
				68	80	8	8	9	2.608	2.608	2.608
		PRP19	Q9UMS4	1	122	14	16	17	1.899	1.899	1.899
					244	4	5	5	1.546	1.546	1.546
				7	244	2	3	3	1.343	1.343	1.343
		PRP8	Q6P2Q9	268	670	3	9	11	0.584	0.584	0.584
					674			2			0.442
				320	666	4	8	8	0.978	0.978	0.978
					670	24	27	28	1.358	1.358	1.358
		RBM22	Q9NW64	363	40	3	3	3	0.754	0.754	0.754
		RED	Q13123	320	553	5	5	6	1.345	1.345	1.345
		SKIP	Q13573	113	81	2	2	2	1.745	1.745	1.745
					97		2	2		0.584	0.584
				135	97	2	2	3	0.416	0.416	0.416
				173	81		3	3		0.710	0.710
					97	2	4	4	0.779	0.779	0.779
					110		3	4		0.397	0.397
				180	97	5	5	5	1.618	1.618	1.618
					110	3	3	3	0.500	0.500	0.500
					115		3	3		0.423	0.423
					122	10	11	12	0.832	0.832	0.832
				320	158	35	38	40	2.214	2.214	2.214

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
					170		2	2		0.323	0.323
				363	153	3	4	5	1.083	1.083	1.083
					158	14	16	19	1.286	1.286	1.286
				396	122	4	6	6	0.731	0.731	0.731
				510	110	2	6	7	0.694	0.694	0.694
					236	2	2	6	0.421	0.421	0.421
					240	5	6	6	0.974	0.974	0.974
		SNU114	Q15029	237	602	2	3	4	0.524	0.524	0.524
		SPF27	O75934	62	136	12	12	13	1.939	1.939	1.939
				68	136	11	12	12	3.413	3.413	3.413
				80	168	2	2	2	2.075	2.075	2.075
				113	168	8	11	12	1.103	1.103	1.103
				135	136	2	2	2	1.855	1.855	1.855
					168	8	9	9	1.343	1.343	1.343
		WBP11	Q9Y2W2	268	10		4	5		0.649	0.649
PPIE	Q9UNP9	SF3A1	Q15459	104	210			2			0.446
PPIL1	Q9Y3C6	PLRG1	O43660	80	62	2	3	3	0.573	0.573	0.573
					68	8	8	9	2.608	2.608	2.608
		SKIP	Q13573	58	81	27	27	27	2.753	2.753	2.753
					97	2	4	4	1.204	1.204	1.204
				80	193	5	5	5	1.332	1.332	1.332
				91	48		3	3		0.308	0.308
				158	97	4	7	7	1.290	1.290	1.290
					193			2			0.370
		SPF27	O75934	80	136	8	8	8	2.484	2.484	2.484
					168	9	10	10	2.210	2.210	2.210
				158	168	2	2	3	0.477	0.477	0.477
		U5-40K	Q96DI7	80	1	26	33	35	1.630	1.630	1.630
					6		6	7		0.430	0.430
					8	8	8	8	3.983	3.983	3.983
					18	13	13	13	1.974	1.974	1.974
					270		2	2		0.245	0.245
					275	17	18	18	2.524	2.524	2.524
					322	4	4	4	2.753	2.753	2.753
					349	3	3	3	1.773	1.773	1.773
PPIL2	Q13356	CDC5L	Q99459	507	466	3	4	4	0.394	0.394	0.394
				508	466		4	4		0.310	0.310
		CTNNBL1	Q8WYA6	90	56	3	3	4	0.545	0.545	0.545
		PRP8	Q6P2Q9	226	727	9	9	9	1.925	1.925	1.925
				230	727	6	10	10	0.649	0.649	0.649
				231	727	9	10	10	1.226	1.226	1.226
				245	727	5	5	5	8.264	8.264	8.264
		SF3A2	Q15428	418	91	2	2	2	0.699	0.699	0.699
		SF3B1	O75533	90	6	3	3	3	1.154	1.154	1.154
		SKIP	Q13573	277	240		2	2		0.292	0.292
		SYF3	Q9BZJ0	460	445			2			0.354
				462	445	8	8	12	0.858	0.858	0.858
					485	2	3	3	0.483	0.483	0.483
				490	569		2	3		0.531	0.531
					602	1	2	2	0.510	0.510	0.510
		ZNF830	Q96NB3	313	228	8	8	8	1.629	1.629	1.629
					234	29	30	32	2.622	2.622	2.622
				450	228	5	7	9	1.095	1.095	1.095
					234	5	5	5	2.125	2.125	2.125
				454	228	2	3	3	0.679	0.679	0.679
				460	234		3	3		0.544	0.544
				462	228	4	4	4	1.222	1.222	1.222
					234			1			0.637
PPP1CA	P62136	PPP1R8	Q12972	26	78	2	3	4	0.545	0.545	0.545
					129	3	3	4	1.094	1.094	1.094
					234	3	4	4	0.882	0.882	0.882
				305	234	6	7	7	0.796	0.796	0.796
PPP1R8	Q12972	PPP1CA	P62136	78	26	2	3	4	0.545	0.545	0.545
				129	26	3	3	4	1.094	1.094	1.094
				234	26	3	4	4	0.882	0.882	0.882
					305	6	7	7	0.796	0.796	0.796
		SF3A1	Q15459	234	115	3	3	3	0.758	0.758	0.758
		SNU23	Q96NC0	334	155	6	9	9	1.503	1.503	1.503
					167	3	3	3	1.039	1.039	1.039
		WBP11	Q9Y2W2	234	610		1	1		0.291	0.291
PQBP1	O60828	BRR2	O75643	2	73		3	3		0.386	0.386
					85	7	7	7	2.323	2.323	2.323
				18	73	2	2	3	1.087	1.087	1.087
					85	2	2	2	2.851	2.851	2.851
					103	2	2	2	0.950	0.950	0.950
				87	60	2	3	3	1.266	1.266	1.266
		PRP8	Q6P2Q9	2	1831	36	45	49	2.057	2.057	2.057
		WBP11	Q9Y2W2	2	168	2	2	2	1.542	1.542	1.542
				18	168	4	4	4	9.348	9.348	9.348
					169	2	2	2	1.262	1.262	1.262
				87	168		3	3		0.836	0.836
					169	8	9	9	1.887	1.887	1.887
PRP17	O60508	CDC5L	Q99459	333	200	6	12	12	0.566	0.566	0.566

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		PRP8	Q6P2Q9	151	29	2	2	2	0.550	0.550	0.550
				551	2034	5	5	5	1.881	1.881	1.881
PRP19	Q9UMS4	AQR	O60306	200	759		4	9		0.246	0.246
		CDC5L	Q99459	76	718	72	76	77	6.685	6.685	6.685
				122	570	9	23	27	0.675	0.675	0.675
					631	3	3	3	1.567	1.567	1.567
				179	432		2	2		0.693	0.693
					685	2	4	4	0.741	0.741	0.741
					782		2	2		0.306	0.306
				192	487	7	7	8	2.404	2.404	2.404
					522	6	6	6	1.479	1.479	1.479
					685			3			0.257
					782	5	6	6	1.379	1.379	1.379
				244	380	2	2	2	0.753	0.753	0.753
					432		2	2		0.634	0.634
					522	4	5	5	0.912	0.912	0.912
					685		9	9		0.381	0.381
					771		3	3		0.508	0.508
					782	28	33	34	1.343	1.343	1.343
					795	7	8	8	1.168	1.168	1.168
				261	782	3	3	3	1.265	1.265	1.265
				266	685	2	4	4	0.412	0.412	0.412
					782	10	11	14	0.925	0.925	0.925
		CTNNBL1	Q8WYA6	192	534	4	4	5	2.205	2.205	2.205
		PLRG1	O43660	122	1	14	16	17	1.899	1.899	1.899
				244	1	4	5	5	1.546	1.546	1.546
					7	2	3	3	1.343	1.343	1.343
		SF3A1	Q15459	266	2	2	2	2	1.516	1.516	1.516
		SmD2	P62316	192	51	3	9	9	1.241	1.241	1.241
		SPF27	O75934	76	85	3	6	6	2.109	2.109	2.109
				122	168	8	8	8	1.841	1.841	1.841
				179	168	18	20	20	2.075	2.075	2.075
				192	168	14	15	15	3.667	3.667	3.667
					218	3	3	3	0.664	0.664	0.664
				244	47	10	10	10	1.851	1.851	1.851
					168	41	47	49	2.180	2.180	2.180
					177	4	4	8	1.370	1.370	1.370
					191	2	2	2	0.972	0.972	0.972
					218	8	14	15	0.750	0.750	0.750
				265	168		2	2		0.335	0.335
				266	47	3	3	3	1.336	1.336	1.336
					168	16	16	16	3.048	3.048	3.048
					218			2			0.290
		SYF1	Q9HCS7	266	708	3	4	4	0.992	0.992	0.992
		U5-40K	Q96DI7	122	226		2	3		0.555	0.555
					270		6	9		0.451	0.451
					275	17	17	18	1.655	1.655	1.655
				179	1	2	2	2	0.608	0.608	0.608
				192	270		3	4		0.457	0.457
					275	21	22	22	3.772	3.772	3.772
				244	1		2	2		0.588	0.588
					275	17	17	17	1.749	1.749	1.749
				266	275		2	2		0.596	0.596
				425	275		3	4		0.286	0.286
		ZC3H18	Q86VM9	179	918	9	10	10	2.517	2.517	2.517
					921	7	7	7	1.175	1.175	1.175
PRP38	Q8NAV1	RBM22	Q9NW64	7	170	3	3	3	0.822	0.822	0.822
		SNU23	Q96NC0	7	39	2	3	3	1.154	1.154	1.154
					45	19	41	47	1.083	1.083	1.083
					102	26	33	37	2.767	2.767	2.767
				50	132		5	6		0.407	0.407
				92	123		2	2		0.843	0.843
PRP4B	Q13523	ZC3H18	Q86VM9	99	933	2	2	2	0.956	0.956	0.956
					948	26	34	42	1.058	1.058	1.058
					952	32	42	45	1.825	1.825	1.825
				117	936	34	35	35	2.034	2.034	2.034
PRP8	Q6P2Q9	BRR2	O75643	1306	14		2	3		0.548	0.548
				1636	1039		5	6		0.535	0.535
					1141		2	2		1.184	1.184
					1146	9	9	9	1.635	1.635	1.635
				1649	1146	7	8	8	2.375	2.375	2.375
				1659	1039	15	19	20	1.289	1.289	1.289
				1866	1049	13	13	13	3.092	3.092	3.092
				1984	1049	2	2	2	1.289	1.289	1.289
				1994	73	2	10	11	0.522	0.522	0.522
				2049	1039	3	3	3	1.131	1.131	1.131
				2070	1039	6	10	10	1.031	1.031	1.031
					1049	2	3	3	1.655	1.655	1.655
		BUD13	Q9BRD0	1505	605		2	2		1.180	1.180
				1831	65	14	14	14	2.929	2.929	2.929
					68	40	43	45	3.605	3.605	3.605
				1840	65	20	21	21	2.421	2.421	2.421
					68	24	24	24	3.027	3.027	3.027

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
				2108	68	3	3	3	2.055	2.055	2.055
		CDC5L	Q99459	727	28	29	36	39	1.463	1.463	1.463
				892	7	37	51	52	1.770	1.770	1.770
		CTNNBL1	Q8WYA6	796	27	16	18	18	3.179	3.179	3.179
					31	13	13	13	2.264	2.264	2.264
		CWC15	Q9P013	987	205		3	4		0.379	0.379
				1144	91		2	2		0.317	0.317
				1158	199	2	3	4	0.543	0.543	0.543
		KIN17	O60870	2034	87		2	3		0.280	0.280
		MBPMS2	P0AEX9-r	452	47			2			0.209
				2293	438	5	17	19	0.472	0.472	0.472
		MFAP1	P55081	1636	256	2	2	2	0.687	0.687	0.687
				1649	242	25	30	31	1.453	1.453	1.453
					249	41	43	46	2.123	2.123	2.123
					250	17	17	17	2.174	2.174	2.174
					256			2			0.207
				2293	238	9	14	16	0.848	0.848	0.848
					242	46	53	54	2.052	2.052	2.052
		PLRG1	O43660	666	320	4	8	8	0.978	0.978	0.978
				670	268	3	9	11	0.584	0.584	0.584
					320	24	27	28	1.358	1.358	1.358
				674	268			2			0.442
		PPIL2	Q13356	727	226	9	9	9	1.925	1.925	1.925
					230	6	10	10	0.649	0.649	0.649
					231	9	10	10	1.226	1.226	1.226
					245	5	5	5	8.264	8.264	8.264
		PQBP1	O60828	1831	2	36	45	49	2.057	2.057	2.057
		PRP17	O60508	29	151	2	2	2	0.550	0.550	0.550
				2034	551	5	5	5	1.881	1.881	1.881
		RED	Q13123	666	541		1	1		0.300	0.300
				670	534	11	15	15	1.554	1.554	1.554
					541	1	8	12	0.532	0.532	0.532
					544		3	5		0.935	0.935
					553		4	6		0.394	0.394
				674	534		2	4		0.402	0.402
					553			2			0.203
				746	520	9	10	10	2.787	2.787	2.787
					553	4	4	4	1.725	1.725	1.725
				774	520	11	13	13	1.085	1.085	1.085
				1020	520	38	41	42	2.572	2.572	2.572
				1958	191			2			0.321
					194			3			0.527
					198	3	3	3	0.919	0.919	0.919
		SF3A2	Q15428	769	10	13	17	19	1.256	1.256	1.256
				853	10		2	2		1.348	1.348
				892	10	2	3	3	1.144	1.144	1.144
		SF3B1	O75533	837	80	3	3	3	3.101	3.101	3.101
					81	3	3	3	3.159	3.159	3.159
				892	468	6	6	6	1.909	1.909	1.909
				1222	80	3	3	3	1.957	1.957	1.957
				1505	943		3	3		0.251	0.251
				1838	1008		2	2		0.543	0.543
				1958	963	14	15	15	1.837	1.837	1.837
					1008	2	2	2	3.511	3.511	3.511
				1978	1014	6	12	12	1.542	1.542	1.542
				1984	1014		3	3		0.682	0.682
				2034	1008	5	9	9	1.559	1.559	1.559
		SF3B2	Q13435	892	894		3	4		0.340	0.340
				1020	891		2	2		0.319	0.319
				1801	790	8	8	8	1.529	1.529	1.529
					815	21	21	21	3.426	3.426	3.426
				1958	815	2	3	3	17.629	17.629	17.629
				1993	857	4	8	9	0.577	0.577	0.577
					870			2			0.573
					877		2	2		0.336	0.336
		SF3B3	Q15393	1958	1191		3	4		1.151	1.151
		SKIP	Q13573	666	158	19	39	50	0.968	0.968	0.968
					217	7	7	8	0.918	0.918	0.918
				670	158	2	5	5	0.561	0.561	0.561
				702	236	22	26	26	2.128	2.128	2.128
					246		2	2		0.704	0.704
				705	246	2	8	10	0.481	0.481	0.481
					258			2			0.292
				721	255		1	2		0.269	0.269
				727	258	2	2	2	2.103	2.103	2.103
					266	2	4	10	0.701	0.701	0.701
				774	258	3	3	3	1.190	1.190	1.190
				796	255			2			0.352
					258	21	21	22	3.650	3.650	3.650
					266	6	28	35	0.592	0.592	0.592
				847	416	5	6	6	1.083	1.083	1.083
				892	266		6	7		0.687	0.687

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
				1020	255		2	3		0.230	0.230
					258	12	16	17	1.700	1.700	1.700
				1344	441	38	39	39	4.538	4.538	4.538
					452	13	14	15	3.912	3.912	3.912
					456	6	6	6	2.045	2.045	2.045
				1505	441	17	18	18	1.648	1.648	1.648
		SMU1	Q2TAY7	1801	107	14	14	15	1.600	1.600	1.600
				1838	107		2	3		0.285	0.285
				2108	107	27	32	33	2.317	2.317	2.317
		SNU114	Q15029	218	405	23	27	28	3.670	3.670	3.670
					409		2	4		0.623	0.623
				366	341	2	2	2	1.012	1.012	1.012
					358		2	4		0.252	0.252
					359	62	66	67	2.466	2.466	2.466
				796	64	4	4	4	15.455	15.455	15.455
		SNU23	Q96NC0	1838	70	9	17	17	0.683	0.683	0.683
				2031	70	3	3	3	3.525	3.525	3.525
				2034	70	49	50	50	3.906	3.906	3.906
		TCERG1	O14776	1210	1024	6	16	23	0.768	0.768	0.768
		U5-40K	Q96DI7	29	286	22	22	22	2.107	2.107	2.107
				36	286	2	3	5	2.014	2.014	2.014
				43	286	29	44	50	2.810	2.810	2.810
				50	131		2	2		1.633	1.633
					349	11	11	11	7.715	7.715	7.715
		WBP11	Q9Y2W2	674	10		3	3		0.307	0.307
					13		3	3		0.598	0.598
				746	10	5	8	9	0.983	0.983	0.983
				774	10		2	2		0.351	0.351
					13	2	4	4	1.084	1.084	1.084
				2031	51		5	11		0.570	0.570
				2034	48		4	5		0.343	0.343
					51		8	9		0.360	0.360
		ZC3H18	Q86VM9	366	499	2	2	2	0.913	0.913	0.913
					500	6	9	11	0.807	0.807	0.807
					510		2	2		0.310	0.310
RBM22	Q9NW64	CWC15	Q9P013	40	28	3	4	4	0.842	0.842	0.842
				76	18	5	7	7	1.223	1.223	1.223
		PLRG1	O43660	40	363	3	3	3	0.754	0.754	0.754
		PRP38	Q8NAV1	170	7	3	3	3	0.822	0.822	0.822
		SKIP	Q13573	40	153	2	2	2	1.302	1.302	1.302
					158	4	7	8	0.666	0.666	0.666
					170			3			0.194
				78	217		1	2		0.386	0.386
				104	193	2	2	2	1.020	1.020	1.020
				109	193		3	4		0.883	0.883
				114	193	13	18	19	1.559	1.559	1.559
				139	193	2	7	9	0.799	0.799	0.799
				149	193	2	3	4	0.815	0.815	0.815
				286	81	7	8	8	3.029	3.029	3.029
		SNU23	Q96NC0	76	8		2	2		0.534	0.534
				149	15	2	2	2	2.532	2.532	2.532
				158	15	2	5	5	0.627	0.627	0.627
RED	Q13123	BRR2	O75643	520	1874	6	6	6	1.016	1.016	1.016
		LSm6	P62312	30	5	6	7	8	1.153	1.153	1.153
					77			2			0.456
		PLRG1	O43660	553	320	5	5	6	1.345	1.345	1.345
		PRP8	Q6P2Q9	191	1958			2			0.321
				194	1958			3			0.527
				198	1958	3	3	3	0.919	0.919	0.919
				520	746	9	10	10	2.787	2.787	2.787
					774	11	13	13	1.085	1.085	1.085
					1020	38	41	42	2.572	2.572	2.572
				534	670	11	15	15	1.554	1.554	1.554
					674		2	4		0.402	0.402
				541	666		1	1		0.300	0.300
					670	1	8	12	0.532	0.532	0.532
				544	670		3	5		0.935	0.935
				553	670		4	6		0.394	0.394
					674			2			0.203
					746	4	4	4	1.725	1.725	1.725
		SF3B1	O75533	112	943		3	3		0.281	0.281
				137	816		2	4		0.285	0.285
		SF3B2	Q13435	78	547	6	8	8	1.041	1.041	1.041
					870	6	8	8	1.860	1.860	1.860
				83	547	6	7	7	1.881	1.881	1.881
		SF3B3	Q15393	209	1191	4	5	5	0.422	0.422	0.422
		SKIP	Q13573	520	258	2	2	2	0.904	0.904	0.904
				544	217		2	2		0.317	0.317
		SMU1	Q2TAY7	191	107	3	4	4	0.858	0.858	0.858
				194	107	21	24	24	1.402	1.402	1.402
				197	107	3	3	3	1.895	1.895	1.895
				198	107	10	12	12	2.382	2.382	2.382
				209	337		2	2		0.330	0.330

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
				331	107	3	3	3	0.787	0.787	0.787
		WBP11	Q9Y2W2	534	10	4	5	5	1.069	1.069	1.069
SF3A1	Q15459	ISY1	Q9ULR0	30	105	2	4	4	0.502	0.502	0.502
				37	105	5	6	6	0.664	0.664	0.664
		LSm6	P62312	131	5	4	4	4	1.310	1.310	1.310
		PPIE	Q9UNP9	210	104			2			0.446
		PPP1R8	Q12972	115	234	3	3	3	0.758	0.758	0.758
		PRP19	Q9UMS4	2	266	2	2	2	1.516	1.516	1.516
		SF3A2	Q15428	131	204			2			0.570
				258	101	7	10	10	0.661	0.661	0.661
				259	101	2	4	5	0.486	0.486	0.486
				264	101		2	2		0.517	0.517
					190	5	8	8	0.939	0.939	0.939
		SF3A3	Q12874	131	29		5	7		0.363	0.363
					201	4	4	6	2.130	2.130	2.130
					212	2	3	3	0.657	0.657	0.657
					219	6	15	20	0.790	0.790	0.790
				188	89	47	61	67	1.810	1.810	1.810
					92	119	186	204	2.107	2.107	2.107
				251	29	11	18	19	2.407	2.407	2.407
				258	29	21	27	30	1.869	1.869	1.869
				259	29	7	9	10	0.631	0.631	0.631
		SF3B2	Q13435	210	275	3	5	7	0.660	0.660	0.660
				217	275		3	5		0.325	0.325
		SRRT	Q9BXP5	399	286		2	4		0.280	0.280
		SYF1	Q9HCS7	2	708	5	5	5	1.492	1.492	1.492
				55	549	11	11	11	1.613	1.613	1.613
		U2-A'	P09661	188	193	18	18	18	3.097	3.097	3.097
					221	10	10	10	2.230	2.230	2.230
				223	205	2	8	11	0.421	0.421	0.421
		WBP11	Q9Y2W2	102	610	15	29	31	1.682	1.682	1.682
					614	4	7	7	0.502	0.802	0.802
				105	610		3	3		0.478	0.478
					614	9	11	11	1.115	1.115	1.115
				115	599	2	2	2	1.264	1.264	1.264
					610	23	29	30	1.198	1.198	1.198
					614	20	20	20	2.278	2.278	2.278
					626	2	2	2	2.172	2.172	2.172
				131	610	4	4	4	0.781	0.781	0.781
					614	4	4	4	1.208	1.208	1.208
					626	2	2	2	1.057	1.057	1.057
SF3A2	Q15428	CDC5L	Q99459	10	28	5	6	6	1.289	1.289	1.289
				42	28	9	13	13	1.355	1.355	1.355
		PPIL2	Q13356	91	418	2	2	2	0.699	0.699	0.699
		PRP8	Q6P2Q9	10	769	13	17	19	1.256	1.256	1.256
					853		2	2		1.348	1.348
					892	2	3	3	1.144	1.144	1.144
		SF3A1	Q15459	101	258	7	10	10	0.661	0.661	0.661
					259	2	4	5	0.486	0.486	0.486
					264		2	2		0.517	0.517
				190	264	5	8	8	0.939	0.939	0.939
				204	131			2			0.570
		SF3A3	Q12874	76	390	4	6	6	0.918	0.918	0.918
				91	390	24	26	26	1.591	1.591	1.591
		SF3B1	O75533	10	454	8	8	8	2.086	2.086	2.086
					496		3	4		0.804	0.804
					499	2	4	6	0.473	0.473	0.473
		SF3B2	Q13435	10	543	6	8	9	0.923	0.923	0.923
					547	5	5	5	0.999	0.999	0.999
					556	6	8	11	0.492	0.492	0.492
					560			3			0.319
					563		2	2		0.784	0.784
					604	2	2	3	0.641	0.641	0.641
				216	268	12	13	16	1.458	1.458	1.458
					275	6	6	6	1.296	1.296	1.296
		SF3B4	Q15427	190	183		2	2		0.308	0.308
		SF3B6	Q9Y3B4	10	29	2	2	2	1.603	1.603	1.603
					105	3	5	6	0.786	0.786	0.786
					106	6	6	6	0.875	0.875	0.875
					116	3	5	7	0.672	0.672	0.672
		SKIP	Q13573	10	258	7	7	7	0.931	0.931	0.931
		WBP11	Q9Y2W2	118	556	27	36	38	2.024	2.024	2.024
					557	32	53	59	0.733	0.733	0.733
					565	31	79	103	0.719	0.719	0.719
SF3A3	Q12874	SF3A1	Q15459	29	131		5	7		0.363	0.363
					251	11	18	19	2.407	2.407	2.407
					258	21	27	30	1.869	1.869	1.869
					259	7	9	10	0.631	0.631	0.631
				89	188	47	61	67	1.810	1.810	1.810
				92	188	119	186	204	2.107	2.107	2.107
				201	131	4	4	6	2.130	2.130	2.130
				212	131	2	3	3	0.657	0.657	0.657
				219	131	6	15	20	0.790	0.790	0.790

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		SF3A2	Q15428	390	76	4	6	6	0.918	0.918	0.918
					91	24	26	26	1.591	1.591	1.591
		SF3B2	Q13435	69	280	37	49	57	3.963	3.963	3.963
				461	486	4	6	9	0.477	0.477	0.477
		SF3B3	Q15393	463	1074	35	81	100	0.824	0.824	0.824
				466	984		3	3		0.488	0.488
					1074	32	59	66	0.840	0.840	0.840
				496	974	3	8	10	0.678	0.913	0.913
		U2-A'	P09661	64	179	4	4	4	1.268	1.268	1.268
					193	2	3	3	1.758	1.758	1.758
				69	56	2	2	2	1.204	1.204	1.204
					179	31	32	32	3.188	3.188	3.188
				92	193	4	5	5	0.961	0.961	0.961
				97	30	95	101	103	25.397	25.397	25.397
		U2-B"	P08579	97	57	3	5	5	0.721	0.721	0.721
				463	103			5			0.209
SF3B1	O75533	CDC5L	Q99459	454	76	2	2	2	0.887	0.887	0.887
					106	5	5	5	3.348	3.348	3.348
		CTNNBL1	Q8WYA6	3	84			2			0.302
					91		2	2		0.376	0.376
				6	56	2	2	2	1.561	1.561	1.561
					83	6	7	7	2.140	2.140	2.140
					84	12	13	13	1.691	1.691	1.691
					91	6	14	20	2.710	2.710	2.710
					95		3	3		0.823	0.823
		KIN17	O60870	729	271		2	2		0.434	0.434
				733	271	3	6	6	7.933	7.933	7.933
		MBPMS2	P0AEX9-r	656	274	2	3	3	1.690	1.690	1.690
		MFAP1	P55081	807	9	15	15	15	1.516	1.516	1.516
		PHF5A	Q7RTV0	946	108		2	2		0.289	0.289
		PPIL2	Q13356	6	90	3	3	3	1.154	1.154	1.154
		PRP8	Q6P2Q9	80	837	3	3	3	3.101	3.101	3.101
					1222	3	3	3	1.957	1.957	1.957
				81	837	3	3	3	3.159	3.159	3.159
				468	892	6	6	6	1.909	1.909	1.909
				943	1505		3	3		0.251	0.251
				963	1958	14	15	15	1.837	1.837	1.837
				1008	1838		2	2		0.543	0.543
					1958	2	2	2	3.511	3.511	3.511
					2034	5	9	9	1.559	1.559	1.559
				1014	1978	6	12	12	1.542	1.542	1.542
					1984		3	3		0.682	0.682
		RED	Q13123	816	137		2	4		0.285	0.285
				943	112		3	3		0.281	0.281
		SF3A2	Q15428	454	10	8	8	8	2.086	2.086	2.086
				496	10		3	4		0.804	0.804
				499	10	2	4	6	0.473	0.473	0.473
		SF3B2	Q13435	1025	556			2			0.375
					870		8	10		0.454	0.454
				1086	857	12	12	12	4.584	4.584	4.584
				1292	486	4	15	18	0.739	0.739	0.739
		SF3B3	Q15393	513	137	8	9	9	1.406	1.406	1.406
		SF3B5	Q9BWJ5	513	17	24	29	30	1.205	1.205	1.205
		SF3B6	Q9Y3B4	80	7		2	2		1.369	1.369
				81	7	9	12	13	1.222	1.222	1.222
				175	116			3			0.204
				195	116		3	4		0.315	0.315
				413	29	4	4	4	1.077	1.077	1.077
				430	7	10	11	14	1.245	1.245	1.245
					41	6	6	6	2.306	2.306	2.306
				454	29	26	28	31	3.790	3.790	3.790
				866	7	13	15	15	1.892	1.892	1.892
				943	41	17	31	32	0.994	0.994	0.994
				946	41		6	6		0.488	0.488
		SKIP	Q13573	81	317	3	3	3	1.297	1.297	1.297
					319	2	3	3	1.183	1.183	1.183
					323	2	2	2	1.539	1.539	1.539
				943	441	1	1	1	0.414	0.414	0.414
		SNIP1	Q8TAD8	816	216		1	1		0.572	0.572
		TCERG1	O14776	1	794			1			0.184
		U2-A'	P09661	175	193		2	2		0.263	0.263
SF3B2	Q13435	ISY1	Q9ULR0	148	101		2	2		0.248	0.248
		MBPMS2	P0AEX9-r	448	47	2	3	3	0.676	0.676	0.676
		PRP8	Q6P2Q9	790	1801	8	8	8	1.529	1.529	1.529
				815	1801	21	21	21	3.426	3.426	3.426
					1958	2	3	3	17.629	17.629	17.629
				857	1993	4	8	9	0.577	0.577	0.577
				870	1993			2			0.573
				877	1993		2	2		0.336	0.336
				891	1020		2	2		0.319	0.319
				894	892		3	4		0.340	0.340

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		RED	Q13123	547	78	6	8	8	1.041	1.041	1.041
					83	6	7	7	1.881	1.881	1.881
				870	78	6	8	8	1.860	1.860	1.860
		SF3A1	Q15459	275	210	3	5	7	0.660	0.660	0.660
					217		3	5		0.325	0.325
		SF3A2	Q15428	268	216	12	13	16	1.458	1.458	1.458
				275	216	6	6	6	1.296	1.296	1.296
				543	10	6	8	9	0.923	0.923	0.923
				547	10	5	5	5	0.999	0.999	0.999
				556	10	6	8	11	0.492	0.492	0.492
				560	10			3			0.319
				563	10		2	2		0.784	0.784
				604	10	2	2	3	0.641	0.641	0.641
		SF3A3	Q12874	280	69	37	49	57	3.963	3.963	3.963
				486	461	4	6	9	0.477	0.477	0.477
		SF3B1	O75533	486	1292	4	15	18	0.739	0.739	0.739
				556	1025			2			0.375
				857	1086	12	12	12	4.584	4.584	4.584
				870	1025		8	10		0.454	0.454
		SF3B3	Q15393	604	1074	29	46	47	0.996	0.996	0.996
				605	1074	42	44	44	2.110	2.110	2.110
		SF3B5	Q9BWJ5	486	82	3	6	6	0.711	0.711	0.711
		SmB	P14678	320	57	16	17	20	1.609	1.609	1.609
		SmD2	P62316	387	6		3	3		0.273	0.273
					8	2	2	2	2.907	2.907	2.907
		SYF3	Q9BZJ0	556	229	7	11	11	1.150	1.150	1.150
		U2-A'	P09661	280	179	3	5	7	0.628	0.628	0.628
					191			2			0.203
					193	16	17	18	2.415	2.415	2.415
				320	172	2	3	3	0.433	0.433	0.433
		U2-B"	P08579	352	111	4	7	7	0.931	0.931	0.931
		WBP11	Q9Y2W2	556	98	5	5	6	0.484	0.484	0.484
					105	3	7	8	0.643	0.643	0.643
SF3B3	Q15393	KIN17	O60870	26	231	13	15	15	1.957	1.957	1.957
				296	231	2	2	2	1.531	1.531	1.531
		MFAP1	P55081	26	67		2	2		0.248	0.248
		PHF5A	Q7RTV0	137	13	7	9	9	1.464	1.464	1.464
		PRP8	Q6P2Q9	1191	1958		3	4		1.151	1.151
		RED	Q13123	1191	209	4	5	5	0.422	0.422	0.422
		SF3A3	Q12874	974	496	3	8	10	0.678	0.913	0.913
				984	466		3	3		0.488	0.488
				1074	463	35	81	100	0.824	0.824	0.824
					466	32	59	66	0.840	0.840	0.840
		SF3B1	O75533	137	513	8	9	9	1.406	1.406	1.406
		SF3B2	Q13435	1074	604	29	46	47	0.996	0.996	0.996
					605	42	44	44	2.110	2.110	2.110
		SF3B5	Q9BWJ5	137	17	43	62	67	4.104	4.104	4.104
SF3B4	Q15427	SF3A2	Q15428	183	190		2	2		0.308	0.308
SF3B5	Q9BWJ5	PHF5A	Q7RTV0	17	13	6	7	7	1.145	1.145	1.145
		SF3B1	O75533	17	513	24	29	30	1.205	1.205	1.205
		SF3B2	Q13435	82	486	3	6	6	0.711	0.711	0.711
		SF3B3	Q15393	17	137	43	62	67	4.104	4.104	4.104
SF3B6	Q9Y3B4	CDC5L	Q99459	29	124	26	26	26	9.871	9.871	9.871
		SF3A2	Q15428	29	10	2	2	2	1.603	1.603	1.603
				105	10	3	5	6	0.786	0.786	0.786
				106	10	6	6	6	0.875	0.875	0.875
				116	10	3	5	7	0.672	0.672	0.672
		SF3B1	O75533	7	80		2	2		1.369	1.369
					81	9	12	13	1.222	1.222	1.222
					430	10	11	14	1.245	1.245	1.245
					866	13	15	15	1.892	1.892	1.892
				29	413	4	4	4	1.077	1.077	1.077
					454	26	28	31	3.790	3.790	3.790
				41	430	6	6	6	2.306	2.306	2.306
					943	17	31	32	0.994	0.994	0.994
					946		6	6		0.488	0.488
				116	175			3			0.204
					195		3	4		0.315	0.315
SKIP	Q13573	BUD13	Q9BRD0	441	605	22	23	23	3.246	3.246	3.246
		CBP80	Q09161	503	607	4	4	4	2.187	2.187	2.187
				509	607	2	3	3	1.072	1.072	1.072
		CDC5L	Q99459	255	28	6	9	9	1.319	1.319	1.319
				258	28	16	25	25	1.281	1.281	1.281
					47		3	6		0.327	0.327
				311	124	8	16	19	0.641	0.641	0.641
		CWC15	Q9P013	153	18	16	16	16	2.622	2.622	2.622
				158	28		2	2		0.932	0.932
				204	28	2	3	4	0.621	0.621	0.621
				217	18	1	5	5	0.535	0.535	0.535
					28	3	4	5	0.573	0.573	0.573
				311	183	2	2	3	0.816	0.816	0.816
		LSm2	Q9Y333	236	8	2	3	4	0.430	0.430	0.430
		PLRG1	O43660	81	113	2	2	2	1.745	1.745	1.745

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
					173		3	3		0.710	0.710
				97	113		2	2		0.584	0.584
					135	2	2	3	0.416	0.416	0.416
					173	2	4	4	0.779	0.779	0.779
					180	5	5	5	1.618	1.618	1.618
				110	173		3	4		0.397	0.397
					180	3	3	3	0.500	0.500	0.500
					510	2	6	7	0.694	0.694	0.694
				115	180		3	3		0.423	0.423
				122	180	10	11	12	0.832	0.832	0.832
					396	4	6	6	0.731	0.731	0.731
				153	363	3	4	5	1.083	1.083	1.083
				158	320	35	38	40	2.214	2.214	2.214
					363	14	16	19	1.286	1.286	1.286
				170	320		2	2		0.323	0.323
				236	510	2	2	6	0.421	0.421	0.421
				240	510	5	6	6	0.974	0.974	0.974
		PPIL1	Q9Y3C6	48	91		3	3		0.308	0.308
				81	58	27	27	27	2.753	2.753	2.753
				97	58	2	4	4	1.204	1.204	1.204
					158	4	7	7	1.290	1.290	1.290
				193	80	5	5	5	1.332	1.332	1.332
					158			2			0.370
		PPIL2	Q13356	240	277		2	2		0.292	0.292
		PRP8	Q6P2Q9	158	666	19	39	50	0.968	0.968	0.968
					670	2	5	5	0.561	0.561	0.561
				217	666	7	7	8	0.918	0.918	0.918
				236	702	22	26	26	2.128	2.128	2.128
				246	702		2	2		0.704	0.704
					705	2	8	10	0.481	0.481	0.481
				255	721		1	2		0.269	0.269
					796			2			0.352
					1020		2	3		0.230	0.230
				258	705			2			0.292
					727	2	2	2	2.103	2.103	2.103
					774	3	3	3	1.190	1.190	1.190
					796	21	21	22	3.650	3.650	3.650
					1020	12	16	17	1.700	1.700	1.700
				266	727	2	4	10	0.701	0.701	0.701
					796	6	28	35	0.592	0.592	0.592
					892		6	7		0.687	0.687
				416	847	5	6	6	1.083	1.083	1.083
				441	1344	38	39	39	4.538	4.538	4.538
					1505	17	18	18	1.648	1.648	1.648
				452	1344	13	14	15	3.912	3.912	3.912
				456	1344	6	6	6	2.045	2.045	2.045
		RBM22	Q9NW64	81	286	7	8	8	3.029	3.029	3.029
				153	40	2	2	2	1.302	1.302	1.302
				158	40	4	7	8	0.666	0.666	0.666
				170	40			3			0.194
				193	104	2	2	2	1.020	1.020	1.020
					109		3	4		0.883	0.883
					114	13	18	19	1.559	1.559	1.559
					139	2	7	9	0.799	0.799	0.799
					149	2	3	4	0.815	0.815	0.815
				217	78		1	2		0.386	0.386
		RED	Q13123	217	544		2	2		0.317	0.317
				258	520	2	2	2	0.904	0.904	0.904
		SF3A2	Q15428	258	10	7	7	7	0.931	0.931	0.931
		SF3B1	O75533	317	81	3	3	3	1.297	1.297	1.297
				319	81	2	3	3	1.183	1.183	1.183
				323	81	2	2	2	1.539	1.539	1.539
				441	943	1	1	1	0.414	0.414	0.414
		SNIP1	Q8TAD8	323	355	5	10	12	0.663	0.663	0.663
				339	342	32	43	49	1.789	1.789	1.789
					353	2	2	2	1.637	1.637	1.637
					355	21	24	24	1.528	1.528	1.528
				344	342	8	18	23	0.547	0.547	0.547
				379	342	5	11	12	0.735	0.735	0.735
		SYF3	Q9BZJ0	23	388	2	2	2	1.679	1.679	1.679
		U5-40K	Q96DI7	48	275	4	4	4	2.240	2.240	2.240
					322	5	6	6	1.276	1.276	1.276
				81	18	5	6	6	1.837	1.837	1.837
					275	6	6	6	2.195	2.195	2.195
				97	18	2	3	3	0.691	0.691	0.691
					322	2	2	2	1.031	1.031	1.031
				110	18	1	2	2	0.560	0.560	0.560
				153	349	6	7	10	1.331	1.331	1.331
				170	349		2	3		0.850	0.850
				193	1	7	9	9	2.990	2.990	2.990
SmB	P14678	SF3B2	Q13435	57	320	16	17	20	1.609	1.609	1.609
		SmD1	P62314	5	41		6	10		0.421	0.421
				8	41	14	16	16	1.298	1.298	1.298

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		SmD3	P62318	88	84	2	4	4	0.729	0.729	0.729
					87			2			0.225
SmD1	P62314	SmB	P14678	41	5		6	10		0.421	0.421
					8	14	16	16	1.298	1.298	1.298
		U5-40K	Q96DI7	86	226	2	2	3	0.520	0.520	0.520
SmD2	P62316	PRP19	Q9UMS4	51	192	3	9	9	1.241	1.241	1.241
		SF3B2	Q13435	6	387		3	3		0.273	0.273
				8	387	2	2	2	2.907	2.907	2.907
		SmF	P62306	8	8	20	21	23	1.581	1.581	1.581
				18	8	7	7	7	1.539	1.539	1.539
		SNU114	Q15029	6	352	1	5	7	0.378	0.378	0.378
SmD3	P62318	SmB	P14678	84	88	2	4	4	0.729	0.729	0.729
				87	88			2			0.225
		U2-B"	P08579	84	93	3	3	3	1.443	1.443	1.443
					101			2			0.290
SmE	P62304	MBPMS2	P0AEX9-r	1	278			2			0.661
		SmG	P62308	12	10		2	6		0.351	0.351
					15	2	4	4	0.679	0.679	0.679
SmF	P62306	SmD2	P62316	8	8	20	21	23	1.581	1.581	1.581
					18	7	7	7	1.539	1.539	1.539
SmG	P62308	SmE	P62304	10	12		2	6		0.351	0.351
				15	12	2	4	4	0.679	0.679	0.679
SMU1	Q2TAY7	BRR2	O75643	308	971	14	17	18	1.051	1.051	1.051
					975	3	3	3	2.201	2.201	2.201
					1544		2	2		0.279	0.279
					1552	31	50	53	1.971	1.971	1.971
					1711			2			0.450
					1715	11	15	16	1.197	1.197	1.197
		MFAP1	P55081	308	67	17	21	21	1.104	1.104	1.104
		PRP8	Q6P2Q9	107	1801	14	14	15	1.600	1.600	1.600
					1838		2	3		0.285	0.285
					2108	27	32	33	2.317	2.317	2.317
		RED	Q13123	107	191	3	4	4	0.858	0.858	0.858
					194	21	24	24	1.402	1.402	1.402
					197	3	3	3	1.895	1.895	1.895
					198	10	12	12	2.382	2.382	2.382
					331	3	3	3	0.787	0.787	0.787
				337	209		2	2		0.330	0.330
SNIP1	Q8TAD8	SF3B1	O75533	216	816		1	1		0.572	0.572
		SKIP	Q13573	342	339	32	43	49	1.789	1.789	1.789
					344	8	18	23	0.547	0.547	0.547
					379	5	11	12	0.735	0.735	0.735
				353	339	2	2	2	1.637	1.637	1.637
				355	323	5	10	12	0.663	0.663	0.663
					339	21	24	24	1.528	1.528	1.528
SNU114	Q15029	PLRG1	O43660	602	237	2	3	4	0.524	0.524	0.524
		PRP8	Q6P2Q9	64	796	4	4	4	15.455	15.455	15.455
				341	366	2	2	2	1.012	1.012	1.012
				358	366		2	4		0.252	0.252
				359	366	62	66	67	2.466	2.466	2.466
				405	218	23	27	28	3.670	3.670	3.670
				409	218		2	4		0.623	0.623
		SmD2	P62316	352	6	1	5	7	0.378	0.378	0.378
		TCERG1	O14776	914	992	5	6	6	1.894	1.894	1.894
				963	992			2			0.782
		ZC3H18	Q86VM9	244	510	4	6	8	0.795	0.795	0.795
				359	500		2	2		0.547	0.547
				694	510		2	3		0.636	0.636
SNU23	Q96NC0	BRR2	O75643	8	479	1	4	6	1.354	1.354	1.354
					557	12	21	24	1.121	1.121	1.121
					804	4	4	4	0.759	0.759	0.759
				15	804	6	8	8	1.189	1.189	1.189
				18	804			2			0.344
				70	85	2	2	2	2.348	2.348	2.348
				138	733	5	7	7	0.643	0.643	0.643
				147	729			2			0.568
					733	4	5	5	1.856	1.856	1.856
				160	696	4	6	7	0.472	0.472	0.472
		KIN17	O60870	70	120	3	3	3	2.324	2.324	2.324
		PPP1R8	Q12972	155	334	6	9	9	1.503	1.503	1.503
				167	334	3	3	3	1.039	1.039	1.039
		PRP38	Q8NAV1	39	7	2	3	3	1.154	1.154	1.154
				45	7	19	41	47	1.083	1.083	1.083
				102	7	26	33	37	2.767	2.767	2.767
				123	92		2	2		0.843	0.843
				132	50		5	6		0.407	0.407
		PRP8	Q6P2Q9	70	1838	9	17	17	0.683	0.683	0.683
					2031	3	3	3	3.525	3.525	3.525
					2034	49	50	50	3.906	3.906	3.906
		RBM22	Q9NW64	8	76		2	2		0.534	0.534
				15	149	2	2	2	2.532	2.532	2.532
					158	2	5	5	0.627	0.627	0.627

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		U5-40K	Q96DI7	136	8		1	2		0.235	0.235
		WBP11	Q9Y2W2	8	169		2	2		0.377	0.377
				45	76	4	7	8	1.092	1.092	1.092
				70	48		4	6		0.325	0.325
					59	2	2	2	3.695	3.695	3.695
				102	76	18	23	23	2.737	2.737	2.737
SPF27	O75934	CDC5L	Q99459	97	685	6	17	22	0.689	0.689	0.689
					686	15	17	17	2.361	2.361	2.361
				168	380	2	2	3	1.331	1.331	1.331
				177	771	74	126	151	1.449	1.449	1.449
				191	782	8	11	12	1.148	1.148	1.148
		PLRG1	O43660	136	62	12	12	13	1.939	1.939	1.939
					68	11	12	12	3.413	3.413	3.413
					135	2	2	2	1.855	1.855	1.855
				168	80	2	2	2	2.075	2.075	2.075
					113	8	11	12	1.103	1.103	1.103
					135	8	9	9	1.343	1.343	1.343
		PPIL1	Q9Y3C6	136	80	8	8	8	2.484	2.484	2.484
				168	80	9	10	10	2.210	2.210	2.210
					158	2	2	3	0.477	0.477	0.477
		PRP19	Q9UMS4	47	244	10	10	10	1.851	1.851	1.851
					266	3	3	3	1.336	1.336	1.336
				85	76	3	6	6	2.109	2.109	2.109
				168	122	8	8	8	1.841	1.841	1.841
					179	18	20	20	2.075	2.075	2.075
					192	14	15	15	3.667	3.667	3.667
					244	41	47	49	2.180	2.180	2.180
					265		2	2		0.335	0.335
					266	16	16	16	3.048	3.048	3.048
				177	244	4	4	8	1.370	1.370	1.370
				191	244	2	2	2	0.972	0.972	0.972
				218	192	3	3	3	0.664	0.664	0.664
					244	8	14	15	0.750	0.750	0.750
					266			2			0.290
		SRRM1	Q8IYB3	97	217		2	2		0.321	0.321
		U5-40K	Q96DI7	168	8	5	5	5	2.865	2.865	2.865
					18	3	4	4	0.756	0.756	0.756
					226	6	8	8	1.298	1.298	1.298
					270	6	9	10	0.617	0.617	0.617
					275	43	46	47	3.253	3.253	3.253
					322	3	3	3	1.409	1.409	1.409
				177	275	2	5	6	1.211	1.211	1.211
SRRM1	Q8IYB3	SPF27	O75934	217	97		2	2		0.321	0.321
SRRT	Q9BXP5	MBPMS2	P0AEX9-r	286	103	7	8	10	1.052	1.052	1.052
					180	25	31	40	1.136	1.136	1.136
					363		4	6		0.357	0.357
		SF3A1	Q15459	286	399		2	4		0.280	0.280
		ZC3H18	Q86VM9	723	211		3	3		0.488	0.488
				730	188		3	4		0.324	0.324
				744	188	2	3	3	1.091	1.091	1.091
SRSF1	Q07955	CBP20	P52298	165	7	21	22	22	2.728	2.728	2.728
		CBP80	Q09161	165	17	4	4	4	1.131	1.131	1.131
					20	3	3	3	1.855	1.855	1.855
				193	20	4	5	5	1.189	1.189	1.189
		MFAP1	P55081	179	329	13	18	19	1.295	1.295	1.295
		TCERG1	O14776	179	1016		4	6		0.488	0.488
					1024	2	3	3	0.691	0.691	0.691
		UBL5	Q9BZL1	179	1			2			1.037
SYF1	Q9HCS7	AQR	O60306	2	954	10	10	11	0.991	0.991	0.991
		CCDC12	Q8WUD4	532	23	3	6	6	0.634	0.634	0.634
					42	5	5	5	1.178	1.178	1.178
				593	94	2	2	2	3.169	3.169	3.169
		CDC5L	Q99459	532	522	8	8	8	1.502	1.502	1.502
				654	294	19	19	20	3.245	3.245	3.245
		ISY1	Q9ULR0	2	190	4	6	6	2.763	2.763	2.763
					260			2			0.660
		MFAP1	P55081	539	285	4	4	4	0.809	0.809	0.809
		PRP19	Q9UMS4	708	266	3	4	4	0.992	0.992	0.992
		SF3A1	Q15459	549	55	11	11	11	1.613	1.613	1.613
				708	2	5	5	5	1.492	1.492	1.492
		SYF3	Q9BZJ0	482	388	25	30	34	1.075	1.075	1.075
					427	2	3	4	0.607	0.607	0.607
					475	2	3	4	0.691	0.691	0.691
		U2-A'	P09661	2	205	5	6	6	1.644	1.644	1.644
				50	193		5	6		0.368	0.368
		ZNF830	Q96NB3	482	279	5	9	9	0.851	0.851	0.851
				654	216	7	7	7	2.416	2.416	2.416
SYF3	Q9BZJ0	CCDC12	Q8WUD4	549	126		3	4		0.399	0.399
				607	117		3	3		0.700	0.700
		CDC5L	Q99459	568	466			2			0.277
		PPIL2	Q13356	445	460			2			0.354
					462	8	8	12	0.858	0.858	0.858

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
				485	462	2	3	3	0.483	0.483	0.483
				569	490		2	3		0.531	0.531
				602	490	1	2	2	0.510	0.510	0.510
		SF3B2	Q13435	229	556	7	11	11	1.150	1.150	1.150
		SKIP	Q13573	388	23	2	2	2	1.679	1.679	1.679
		SYF1	Q9HCS7	388	482	25	30	34	1.075	1.075	1.075
				427	482	2	3	4	0.607	0.607	0.607
				475	482	2	3	4	0.691	0.691	0.691
		ZNF830	Q96NB3	445	228	3	4	4	0.587	0.587	0.587
					234	14	15	15	2.437	2.437	2.437
				460	234	9	12	15	1.573	1.573	1.573
				475	234	1	4	6	0.670	0.670	0.670
TCERG1	O14776	BRR2	O75643	570	60	3	5	6	0.683	0.683	0.683
				634	440		3	3		1.936	1.936
		BUD31	P41223	711	68		4	5		0.521	0.521
		CBP80	Q09161	756	187	1	1	1	0.510	0.510	0.510
		MFAP1	P55081	878	415	4	9	10	1.176	1.176	1.176
		PRP8	Q6P2Q9	1024	1210	6	16	23	0.768	0.768	0.768
		SF3B1	O75533	794	1			1			0.184
		SNU114	Q15029	992	914	5	6	6	1.894	1.894	1.894
					963			2			0.782
		SRSF1	Q07955	1016	179		4	6		0.488	0.488
				1024	179	2	3	3	0.691	0.691	0.691
		U2-B"	P08579	495	109		5	6		0.271	0.271
				634	108	1	1	1	0.490	0.490	0.490
					109		2	2		0.816	0.816
		WBP11	Q9Y2W2	570	169	2	2	2	2.062	2.062	2.062
		ZC3H18	Q86VM9	895	510		2	2		0.391	0.391
				1016	622	3	4	4	0.802	0.802	0.802
U2-A'	P09661	ISY1	Q9ULR0	193	266	17	17	17	2.627	2.627	2.627
				221	266	6	6	6	2.095	2.095	2.095
		MBPMS2	P0AEX9-r	192	103			3			0.191
		SF3A1	Q15459	193	188	18	18	18	3.097	3.097	3.097
				205	223	2	8	11	0.421	0.421	0.421
				221	188	10	10	10	2.230	2.230	2.230
		SF3A3	Q12874	30	97	95	101	103	25.397	25.397	25.397
				56	69	2	2	2	1.204	1.204	1.204
				179	64	4	4	4	1.268	1.268	1.268
					69	31	32	32	3.188	3.188	3.188
				193	64	2	3	3	1.758	1.758	1.758
					92	4	5	5	0.961	0.961	0.961
		SF3B1	O75533	193	175		2	2		0.263	0.263
		SF3B2	Q13435	172	320	2	3	3	0.433	0.433	0.433
				179	280	3	5	7	0.628	0.628	0.628
				191	280			2			0.203
				193	280	16	17	18	2.415	2.415	2.415
		SYF1	Q9HCS7	193	50		5	6		0.368	0.368
				205	2	5	6	6	1.644	1.644	1.644
		U2-B"	P08579	30	111	2	2	2	1.466	1.466	1.466
U2-B"	P08579	SF3A3	Q12874	57	97	3	5	5	0.721	0.721	0.721
				103	463			5			0.209
		SF3B2	Q13435	111	352	4	7	7	0.931	0.931	0.931
		SmD3	P62318	93	84	3	3	3	1.443	1.443	1.443
				101	84			2			0.290
		TCERG1	O14776	108	634	1	1	1	0.490	0.490	0.490
				109	495		5	6		0.271	0.271
					634		2	2		0.816	0.816
		U2-A'	P09661	111	30	2	2	2	1.466	1.466	1.466
U5-40K	Q96DI7	CDC5L	Q99459	270	795	2	3	5	0.515	0.515	0.515
		PPIL1	Q9Y3C6	1	80	26	33	35	1.630	1.630	1.630
				6	80		6	7		0.430	0.430
				8	80	8	8	8	3.983	3.983	3.983
				18	80	13	13	13	1.974	1.974	1.974
				270	80		2	2		0.245	0.245
				275	80	17	18	18	2.524	2.524	2.524
				322	80	4	4	4	2.753	2.753	2.753
				349	80	3	3	3	1.773	1.773	1.773
		PRP19	Q9UMS4	1	179	2	2	2	0.608	0.608	0.608
					244		2	2		0.588	0.588
				226	122		2	3		0.555	0.555
				270	122		6	9		0.451	0.451
					192		3	4		0.457	0.457
				275	122	17	17	18	1.655	1.655	1.655
					192	21	22	22	3.772	3.772	3.772
					244	17	17	17	1.749	1.749	1.749
					266		2	2		0.596	0.596
					425		3	4		0.286	0.286
		PRP8	Q6P2Q9	131	50		2	2		1.633	1.633
				286	29	22	22	22	2.107	2.107	2.107
					36	2	3	5	2.014	2.014	2.014
					43	29	44	50	2.810	2.810	2.810
				349	50	11	11	11	7.715	7.715	7.715

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		SKIP	Q13573	1	193	7	9	9	2.990	2.990	2.990
				18	81	5	6	6	1.837	1.837	1.837
					97	2	3	3	0.691	0.691	0.691
					110	1	2	2	0.560	0.560	0.560
				275	48	4	4	4	2.240	2.240	2.240
					81	6	6	6	2.195	2.195	2.195
				322	48	5	6	6	1.276	1.276	1.276
					97	2	2	2	1.031	1.031	1.031
				349	153	6	7	10	1.331	1.331	1.331
					170		2	3		0.850	0.850
		SmD1	P62314	226	86	2	2	3	0.520	0.520	0.520
		SNU23	Q96NC0	8	136		1	2		0.235	0.235
		SPF27	O75934	8	168	5	5	5	2.865	2.865	2.865
				18	168	3	4	4	0.756	0.756	0.756
				226	168	6	8	8	1.298	1.298	1.298
				270	168	6	9	10	0.617	0.617	0.617
				275	168	43	46	47	3.253	3.253	3.253
					177	2	5	6	1.211	1.211	1.211
				322	168	3	3	3	1.409	1.409	1.409
UBL5	Q9BZL1	SRSF1	Q07955	1	179			2			1.037
WBP11	Q9Y2W2	BRR2	O75643	48	85			2			0.189
				59	85	2	2	2	2.734	2.734	2.734
				168	14	5	5	5	0.851	0.851	0.851
					60	2	4	4	0.754	0.754	0.754
					85	2	2	2	14.531	14.531	14.531
				169	60	4	5	6	1.501	1.501	1.501
		PLRG1	O43660	10	268		4	5		0.649	0.649
		PPP1R8	Q12972	610	234		1	1		0.291	0.291
		PQBP1	O60828	168	2	2	2	2	1.542	1.542	1.542
					18	4	4	4	9.348	9.348	9.348
					87		3	3		0.836	0.836
				169	18	2	2	2	1.262	1.262	1.262
					87	8	9	9	1.887	1.887	1.887
		PRP8	Q6P2Q9	10	674		3	3		0.307	0.307
					746	5	8	9	0.983	0.983	0.983
					774		2	2		0.351	0.351
				13	674		3	3		0.598	0.598
					774	2	4	4	1.084	1.084	1.084
				48	2034		4	5		0.343	0.343
				51	2031		5	11		0.570	0.570
					2034		8	9		0.360	0.360
		RED	Q13123	10	534	4	5	5	1.069	1.069	1.069
		SF3A1	Q15459	599	115	2	2	2	1.264	1.264	1.264
				610	102	15	29	31	1.682	1.682	1.682
					105		3	3		0.478	0.478
					115	23	29	30	1.198	1.198	1.198
					131	4	4	4	0.781	0.781	0.781
				614	102	4	7	7	0.502	0.802	0.802
					105	9	11	11	1.115	1.115	1.115
					115	20	20	20	2.278	2.278	2.278
					131	4	4	4	1.208	1.208	1.208
				626	115	2	2	2	2.172	2.172	2.172
					131	2	2	2	1.057	1.057	1.057
		SF3A2	Q15428	556	118	27	36	38	2.024	2.024	2.024
				557	118	32	53	59	0.733	0.733	0.733
				565	118	31	79	103	0.719	0.719	0.719
		SF3B2	Q13435	98	556	5	5	6	0.484	0.484	0.484
				105	556	3	7	8	0.643	0.643	0.643
		SNU23	Q96NC0	48	70		4	6		0.325	0.325
				59	70	2	2	2	3.695	3.695	3.695
				76	45	4	7	8	1.092	1.092	1.092
					102	18	23	23	2.737	2.737	2.737
				169	8		2	2		0.377	0.377
		TCERG1	O14776	169	570	2	2	2	2.062	2.062	2.062
ZC3H18	Q86VM9	CBP80	Q09161	319	221	19	21	22	1.585	1.585	1.585
		PRP19	Q9UMS4	918	179	9	10	10	2.517	2.517	2.517
				921	179	7	7	7	1.175	1.175	1.175
		PRP4B	Q13523	933	99	2	2	2	0.956	0.956	0.956
				936	117	34	35	35	2.034	2.034	2.034
				948	99	26	34	42	1.058	1.058	1.058
				952	99	32	42	45	1.825	1.825	1.825
		PRP8	Q6P2Q9	499	366	2	2	2	0.913	0.913	0.913
				500	366	6	9	11	0.807	0.807	0.807
				510	366		2	2		0.310	0.310
		SNU114	Q15029	500	359		2	2		0.547	0.547
				510	244	4	6	8	0.795	0.795	0.795
					694		2	3		0.636	0.636
		SRRT	Q9BXP5	188	730		3	4		0.324	0.324
					744	2	3	3	1.091	1.091	1.091
				211	723		3	3		0.488	0.488
		TCERG1	O14776	510	895		2	2		0.391	0.391
				622	1016	3	4	4	0.802	0.802	0.802
ZNF830	Q96NB3	CWC15	Q9P013	26	147		2	2		0.405	0.405

Continued on next page

Table B.1 – *Continued from previous page*

Name	UniProt ID	Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		PPIL2	Q13356	228	313	8	8	8	1.629	1.629	1.629
					450	5	7	9	1.095	1.095	1.095
					454	2	3	3	0.679	0.679	0.679
					462	4	4	4	1.222	1.222	1.222
				234	313	29	30	32	2.622	2.622	2.622
					450	5	5	5	2.125	2.125	2.125
					460		3	3		0.544	0.544
					462			1			0.637
		SYF1	Q9HCS7	216	654	7	7	7	2.416	2.416	2.416
				279	482	5	9	9	0.851	0.851	0.851
		SYF3	Q9BZJ0	228	445	3	4	4	0.587	0.587	0.587
				234	445	14	15	15	2.437	2.437	2.437
					460	9	12	15	1.573	1.573	1.573
					475	1	4	6	0.670	0.670	0.670

Table B.2: Intramolecular crosslinks. Crosslink analyses by pLink2.3.5 at FDR 1, 3 and 5% are indicated. Number of CSMs (crosslinked peptide spectrum matches) and highest score are shown for each peptide. "Residue 1" and "Residue 2" denote crosslinked residues within a single protein. A single crosslinking experiment was performed and analyzed with mass spectrometry using three technical replicates. Crosslink analysis was performed by Dr. Olex Dybkov (Department of Cellular Biochemistry, MPI-BPC) in collaboration with Prof. Dr. Henning Urlaub (Bioanalytical Mass Spectrometry, MPI-BPC) (see 2.2.6).

Protein				CSMs			$Score_{max}$		
Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
AQR	O60306	8	29	5	5	5	1.817	1.817	1.817
			36	5	5	5	2.105	2.105	2.105
			52		2	4		0.795	0.795
			56	19	20	21	2.489	2.489	2.489
		29	8	5	5	5	1.817	1.817	1.817
		36	8	5	5	5	2.105	2.105	2.105
			38	13	13	13	3.099	3.099	3.099
			115	4	5	5	0.643	0.643	0.643
		37	115	3	4	4	0.960	0.960	0.960
		38	36	13	13	13	3.099	3.099	3.099
			114	28	28	28	3.227	3.227	3.227
			115	4	6	6	1.250	1.250	1.250
		52	8		2	4		0.795	0.795
		56	8	19	20	21	2.489	2.489	2.489
			63	15	15	15	1.854	1.854	1.854
		63	56	15	15	15	1.854	1.854	1.854
		114	38	28	28	28	3.227	3.227	3.227
		115	36	4	5	5	0.643	0.643	0.643
			37	3	4	4	0.960	0.960	0.960
			38	4	6	6	1.250	1.250	1.250
		188	201	6	6	6	1.864	1.864	1.864
		189	195	3	3	3	1.993	1.993	1.993
		192	195	50	55	55	3.037	3.037	3.037
			202	2	2	2	0.733	0.733	0.733
			762	13	13	13	1.549	1.549	1.549
		195	189	3	3	3	1.993	1.993	1.993
			192	50	55	55	3.037	3.037	3.037
			202	45	47	47	5.718	5.718	5.718
		201	188	6	6	6	1.864	1.864	1.864
		202	192	2	2	2	0.733	0.733	0.733
			195	45	47	47	5.718	5.718	5.718
		206	493	2	3	3	1.509	1.509	1.509
		228	246	35	36	36	1.711	1.711	1.711
		246	228	35	36	36	1.711	1.711	1.711
		493	206	2	3	3	1.509	1.509	1.509
		558	1045	2	3	3	1.166	1.166	1.166
			1051	56	60	60	4.214	4.214	4.214
		662	1045		2	2		0.462	0.462
		759	762	19	20	20	1.238	1.238	1.238
		760	762	2	2	2	0.909	0.909	0.909
		762	192	13	13	13	1.549	1.549	1.549
			759	19	20	20	1.238	1.238	1.238
			760	2	2	2	0.909	0.909	0.909
		796	865	19	23	23	2.213	2.213	2.213
			890	5	8	8	1.276	1.276	1.276
			1003	2	2	2	0.795	0.795	0.795
		865	796	19	23	23	2.213	2.213	2.213
		890	796	5	8	8	1.276	1.276	1.276
		1003	796	2	2	2	0.795	0.795	0.795
		1025	1045	8	8	8	2.229	2.229	2.229
			1051	6	6	6	2.263	2.263	2.263

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		1045	558	2	3	3	1.166	1.166	1.166
			662		2	2		0.462	0.462
			1025	8	8	8	2.229	2.229	2.229
			1051	75	77	77	2.850	2.850	2.850
		1051	558	56	60	60	4.214	4.214	4.214
			1025	6	6	6	2.263	2.263	2.263
			1045	75	77	77	2.850	2.850	2.850
		1234	1256	8	11	11	1.164	1.164	1.164
		1256	1234	8	11	11	1.164	1.164	1.164
BRR2	O75643	1	1711		2	2		0.225	0.225
		14	46	38	39	39	3.985	3.985	3.985
			55	30	34	34	1.935	1.935	1.935
			60	28	28	28	2.908	2.908	2.908
			73	9	10	10	1.822	1.822	1.822
			83	5	5	5	2.027	2.027	2.027
			85	5	5	5	2.679	2.679	2.679
			151	3	3	3	0.635	0.635	0.635
			349	2	2	2	2.321	2.321	2.321
			1874	3	3	3	1.291	1.291	1.291
		46	14	38	39	39	3.985	3.985	3.985
			55	93	101	101	1.853	1.853	1.853
			60	129	137	137	3.400	3.400	3.400
			73	3	3	3	2.184	2.184	2.184
			1603	3	3	3	1.432	1.432	1.432
		55	14	30	34	34	1.935	1.935	1.935
			46	93	101	101	1.853	1.853	1.853
			60	122	130	130	1.998	1.998	1.998
			70	7	9	10	0.802	0.802	0.802
			83	8	8	8	2.137	2.137	2.137
			103	9	9	9	1.088	1.088	1.088
			107	2	2	2	0.893	0.893	0.893
		60	14	28	28	28	2.908	2.908	2.908
			46	129	137	137	3.400	3.400	3.400
			55	122	130	130	1.998	1.998	1.998
			70	20	23	23	1.292	1.292	1.292
			73	24	25	25	2.330	2.330	2.330
			83	13	13	14	1.757	1.757	1.757
			103	4	4	4	2.294	2.294	2.294
			107	8	13	13	1.385	1.385	1.385
			155		2	2		1.082	1.082
		70	55	7	9	10	0.802	0.802	0.802
			60	20	23	23	1.292	1.292	1.292
			83	2	2	2	0.617	0.617	0.617
			155		2	2		0.620	0.620
		73	14	9	10	10	1.822	1.822	1.822
			46	3	3	3	2.184	2.184	2.184
			60	24	25	25	2.330	2.330	2.330
			83	13	13	13	1.771	1.771	1.771
			85	3	3	3	3.539	3.539	3.539
			103	2	2	2	0.534	0.534	0.534
			155	6	6	6	1.497	1.497	1.497
		83	14	5	5	5	2.027	2.027	2.027
			55	8	8	8	2.137	2.137	2.137
			60	13	13	14	1.757	1.757	1.757
			70	2	2	2	0.617	0.617	0.617
			73	13	13	13	1.771	1.771	1.771
			103	8	8	8	2.288	2.288	2.288
			107	12	14	15	2.163	2.163	2.163
			155		2	2		0.357	0.357
		85	14	5	5	5	2.679	2.679	2.679
			73	3	3	3	3.539	3.539	3.539
			155	3	3	3	4.038	4.038	4.038
		103	55	9	9	9	1.088	1.088	1.088
			60	4	4	4	2.294	2.294	2.294
			73	2	2	2	0.534	0.534	0.534
			83	8	8	8	2.288	2.288	2.288
			107	3	3	3	1.323	1.323	1.323
		107	55	2	2	2	0.893	0.893	0.893
			60	8	13	13	1.385	1.385	1.385
			83	12	14	15	2.163	2.163	2.163
			103	3	3	3	1.323	1.323	1.323
			178	3	3	3	1.456	1.456	1.456
		147	155	22	43	43	1.076	1.076	1.076
		151	14	3	3	3	0.635	0.635	0.635
			155	95	130	131	1.684	1.684	1.684
			177	10	10	10	1.088	1.088	1.088
			1049	26	27	27	1.171	1.171	1.171
		155	60		2	2		1.082	1.082
			70		2	2		0.620	0.620
			73	6	6	6	1.497	1.497	1.497
			83		2	2		0.357	0.357
			85	3	3	3	4.038	4.038	4.038
			147	22	43	43	1.076	1.076	1.076

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			151	95	130	131	1.684	1.684	1.684
			1172		3	3		0.233	0.233
		177	151	10	10	10	1.088	1.088	1.088
		178	107	3	3	3	1.456	1.456	1.456
		254	256	31	32	32	3.300	3.300	3.300
			487	2	2	2	0.556	0.556	0.556
		255	349	23	23	23	3.695	3.695	3.695
			358	6	6	6	3.151	3.151	3.151
			368	44	44	44	1.869	1.869	1.869
			944	12	12	12	2.552	2.552	2.552
		256	254	31	32	32	3.300	3.300	3.300
			319	5	7	7	2.131	2.131	2.131
			349	2	2	2	2.677	2.677	2.677
			358	11	12	12	3.565	3.565	3.565
			368	34	37	38	2.197	2.197	2.197
			487	2	3	3	1.893	1.893	1.893
			729			2			0.180
			944	11	11	11	1.776	1.776	1.776
			1134		2	2		0.304	0.304
			1556	2	2	2	2.590	2.590	2.590
		285	966	2	2	2	2.279	2.279	2.279
			1874	8	8	8	3.147	3.147	3.147
		294	349	30	30	30	4.022	4.022	4.022
		319	256	5	7	7	2.131	2.131	2.131
		349	14	2	2	2	2.321	2.321	2.321
			255	23	23	23	3.695	3.695	3.695
			256	2	2	2	2.677	2.677	2.677
			294	30	30	30	4.022	4.022	4.022
			695	2	2	2	1.214	1.214	1.214
			944	3	4	4	2.364	2.364	2.364
			1544	2	3	3	0.379	0.379	0.379
		358	255	6	6	6	3.151	3.151	3.151
			256	11	12	12	3.565	3.565	3.565
			944	3	3	3	3.470	3.470	3.470
		368	255	44	44	44	1.869	1.869	1.869
			256	34	37	38	2.197	2.197	2.197
			426	8	8	8	1.389	1.389	1.389
			944	33	33	33	2.410	2.410	2.410
			975	3	3	3	1.561	1.561	1.561
			1556	5	5	5	1.443	1.443	1.443
		426	368	8	8	8	1.389	1.389	1.389
			451	3	3	3	2.538	2.538	2.538
			944	51	52	52	2.941	2.941	2.941
		440	695	3	6	7	0.545	0.545	0.545
			696	78	88	90	4.284	4.284	4.284
			699	3	9	9	0.725	0.725	0.725
			745	13	13	13	2.935	2.935	2.935
		451	426	3	3	3	2.538	2.538	2.538
			487	121	135	136	2.465	2.465	2.465
			745	7	8	8	2.703	2.703	2.703
			864	57	57	57	13.644	13.644	13.644
		453	479	3	4	4	1.400	1.400	1.400
			487	363	378	378	7.933	7.933	7.933
			745		2	2		1.507	1.507
		466	479	174	177	178	13.693	13.693	13.693
			487	82	94	94	6.448	6.448	6.448
		479	453	3	4	4	1.400	1.400	1.400
			466	174	177	178	13.693	13.693	13.693
			487	4	5	5	1.064	1.064	1.064
			557	32	33	33	4.500	4.500	4.500
		487	254	2	2	2	0.556	0.556	0.556
			256	2	3	3	1.893	1.893	1.893
			451	121	135	136	2.465	2.465	2.465
			453	363	378	378	7.933	7.933	7.933
			466	82	94	94	6.448	6.448	6.448
			479	4	5	5	1.064	1.064	1.064
			671	16	19	19	1.901	1.901	1.901
		557	479	32	33	33	4.500	4.500	4.500
		577	599	26	30	31	1.623	1.623	1.623
			1567	69	82	83	1.594	1.594	1.594
		592	599	74	93	95	2.202	2.202	2.202
		599	577	26	30	31	1.623	1.623	1.623
			592	74	93	95	2.202	2.202	2.202
		671	487	16	19	19	1.901	1.901	1.901
		695	349	2	2	2	1.214	1.214	1.214
			440	3	6	7	0.545	0.545	0.545
		696	440	78	88	90	4.284	4.284	4.284
		699	440	3	9	9	0.725	0.725	0.725
		718	804	13	13	13	2.953	2.953	2.953
			864	19	19	19	12.473	12.473	12.473
		729	256			2			0.180
		745	440	13	13	13	2.935	2.935	2.935
			451	7	8	8	2.703	2.703	2.703

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			453		2	2		1.507	1.507
		770	1134	33	33	33	3.247	3.247	3.247
		804	718	13	13	13	2.953	2.953	2.953
		838	1025	2	2	2	1.110	1.110	1.110
		864	451	57	57	57	13.644	13.644	13.644
			718	19	19	19	12.473	12.473	12.473
		914	1556	4	4	4	2.544	2.544	2.544
		944	255	12	12	12	2.552	2.552	2.552
			256	11	11	11	1.776	1.776	1.776
			349	3	4	4	2.364	2.364	2.364
			358	3	3	3	3.470	3.470	3.470
			368	33	33	33	2.410	2.410	2.410
			426	51	52	52	2.941	2.941	2.941
		966	285	2	2	2	2.279	2.279	2.279
			974	13	27	28	1.173	1.173	1.173
			1552	2	4	4	0.678	0.678	0.678
		971	975	57	60	61	2.966	2.966	2.966
			1544		2	2		0.321	0.321
			1552	110	122	125	1.868	1.868	1.868
			1711	273	281	281	2.977	2.977	2.977
			1715	2	2	2	0.961	0.961	0.961
		974	966	13	27	28	1.173	1.173	1.173
			1552	2	2	2	1.675	1.675	1.675
			1711	3	3	3	1.570	1.570	1.570
		975	368	3	3	3	1.561	1.561	1.561
			971	57	60	61	2.966	2.966	2.966
			1552	40	40	40	3.898	3.898	3.898
			1556	13	13	13	3.605	3.605	3.605
		1006	1049	69	71	71	13.732	13.732	13.732
		1025	838	2	2	2	1.110	1.110	1.110
		1049	151	26	27	27	1.171	1.171	1.171
			1006	69	71	71	13.732	13.732	13.732
		1120	1134	165	172	173	5.734	5.734	5.734
			1176	3	4	4	0.815	0.815	0.815
		1134	256		2	2		0.304	0.304
			770	33	33	33	3.247	3.247	3.247
			1120	165	172	173	5.734	5.734	5.734
			1146	2	2	2	2.993	2.993	2.993
			1169	43	47	47	2.346	2.346	2.346
			1172	29	31	31	2.106	2.106	2.106
			1176	406	409	410	3.816	3.816	3.816
			1213	3	3	3	2.295	2.295	2.295
		1141	1146	6	6	6	1.130	1.130	1.130
		1142	1169	10	11	11	2.400	2.400	2.400
		1146	1134	2	2	2	2.993	2.993	2.993
			1141	6	6	6	1.130	1.130	1.130
			1176	2	2	2	1.425	1.425	1.425
		1169	1134	43	47	47	2.346	2.346	2.346
			1142	10	11	11	2.400	2.400	2.400
			1176	66	70	70	2.027	2.027	2.027
		1172	155		3	3		0.233	0.233
			1134	29	31	31	2.106	2.106	2.106
		1176	1120	3	4	4	0.815	0.815	0.815
			1134	406	409	410	3.816	3.816	3.816
			1146	2	2	2	1.425	1.425	1.425
			1169	66	70	70	2.027	2.027	2.027
		1213	1134	3	3	3	2.295	2.295	2.295
		1242	1567	39	39	39	2.705	2.705	2.705
			1883	2	2	2	2.238	2.238	2.238
		1244	1567		2	2		0.469	0.469
			1883	10	10	10	2.049	2.049	2.049
		1294	1498	11	12	12	10.047	10.047	10.047
			1716	43	44	44	14.963	14.963	14.963
		1404	1421	187	188	189	7.255	7.255	7.255
			2059	133	133	133	5.372	5.372	5.372
		1417	2059	72	72	72	3.930	3.930	3.930
			2080	2	2	2	1.018	1.018	1.018
		1421	1404	187	188	189	7.255	7.255	7.255
			1443	23	23	23	4.403	4.403	4.403
			2059	498	502	503	41.652	41.652	41.652
		1440	2080	22	26	26	1.379	1.379	1.379
		1443	1421	23	23	23	4.403	4.403	4.403
			1743	4	4	4	9.326	9.326	9.326
			2080	7	9	9	2.728	2.728	2.728
		1498	1294	11	12	12	10.047	10.047	10.047
		1544	349	2	3	3	0.379	0.379	0.379
			971		2	2		0.321	0.321
			1556	2	2	2	2.290	2.290	2.290
		1552	966	2	4	4	0.678	0.678	0.678
			971	110	122	125	1.868	1.868	1.868
			974	2	2	2	1.675	1.675	1.675
			975	40	40	40	3.898	3.898	3.898
			1711	23	24	24	1.896	1.896	1.896

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		1556	256	2	2	2	2.590	2.590	2.590
			368	5	5	5	1.443	1.443	1.443
			914	4	4	4	2.544	2.544	2.544
			975	13	13	13	3.605	3.605	3.605
			1544	2	2	2	2.290	2.290	2.290
			1595	3	3	3	2.428	2.428	2.428
		1567	577	69	82	83	1.594	1.594	1.594
			1242	39	39	39	2.705	2.705	2.705
			1244		2	2		0.469	0.469
		1595	1556	3	3	3	2.428	2.428	2.428
			1610	60	62	62	3.650	3.650	3.650
		1603	46	3	3	3	1.432	1.432	1.432
		1610	1595	60	62	62	3.650	3.650	3.650
		1710	2059		2	2		0.469	0.469
		1711	1		2	2		0.225	0.225
			971	273	281	281	2.977	2.977	2.977
			974	3	3	3	1.570	1.570	1.570
			1552	23	24	24	1.896	1.896	1.896
			1716		3	3		0.335	0.335
		1715	971	2	2	2	0.961	0.961	0.961
		1716	1294	43	44	44	14.963	14.963	14.963
			1711		3	3		0.335	0.335
		1743	1443	4	4	4	9.326	9.326	9.326
		1841	1878	14	17	17	1.477	1.477	1.477
		1874	14	3	3	3	1.291	1.291	1.291
			285	8	8	8	3.147	3.147	3.147
			1883	2	2	2	0.921	0.921	0.921
			1961	17	17	17	1.724	1.724	1.724
		1878	1841	14	17	17	1.477	1.477	1.477
			1977	18	53	54	0.654	0.654	0.654
		1883	1242	2	2	2	2.238	2.238	2.238
			1244	10	10	10	2.049	2.049	2.049
			1874	2	2	2	0.921	0.921	0.921
		1961	1874	17	17	17	1.724	1.724	1.724
		1977	1878	18	53	54	0.654	0.654	0.654
		2059	1404	133	133	133	5.372	5.372	5.372
			1417	72	72	72	3.930	3.930	3.930
			1421	498	502	503	41.652	41.652	41.652
			1710		2	2		0.469	0.469
		2080	1417	2	2	2	1.018	1.018	1.018
			1440	22	26	26	1.379	1.379	1.379
			1443	7	9	9	2.728	2.728	2.728
			2087	4	5	5	0.654	0.654	0.654
			2089	44	58	59	1.526	1.526	1.526
			2091	46	48	48	4.314	4.314	4.314
		2087	2080	4	5	5	0.654	0.654	0.654
		2089	2080	44	58	59	1.526	1.526	1.526
		2091	2080	46	48	48	4.314	4.314	4.314
BUD13	Q9BRD0	286	314		3	3		0.374	0.374
		314	286		3	3		0.374	0.374
		330	339	2	3	3	1.151	1.151	1.151
		333	340	7	7	7	2.024	2.024	2.024
		339	330	2	3	3	1.151	1.151	1.151
		340	333	7	7	7	2.024	2.024	2.024
		417	427	3	3	3	1.077	1.077	1.077
		427	417	3	3	3	1.077	1.077	1.077
		466	474	7	7	7	2.766	2.766	2.766
		474	466	7	7	7	2.766	2.766	2.766
			499	2	2	2	1.045	1.045	1.045
		499	474	2	2	2	1.045	1.045	1.045
			520	3	3	3	1.143	1.143	1.143
		520	499	3	3	3	1.143	1.143	1.143
		550	553	2	2	2	0.761	0.761	0.761
			555	5	5	5	1.408	1.408	1.408
		553	550	2	2	2	0.761	0.761	0.761
		555	550	5	5	5	1.408	1.408	1.408
		596	604	5	6	6	1.109	1.109	1.109
		604	596	5	6	6	1.109	1.109	1.109
BUD31	P41223	9	86	3	3	3	1.799	1.799	1.799
		28	40	12	13	13	1.252	1.252	1.252
		40	28	12	13	13	1.252	1.252	1.252
		66	68	9	10	10	2.412	2.412	2.412
		68	66	9	10	10	2.412	2.412	2.412
		86	9	3	3	3	1.799	1.799	1.799
CBP20	P52298	34	38	12	12	12	3.135	3.135	3.135
			120	56	59	59	2.927	2.927	2.927
		38	34	12	12	12	3.135	3.135	3.135
			120	8	8	8	3.107	3.107	3.107
		75	78	20	21	21	2.711	2.711	2.711
			120	23	24	24	3.666	3.666	3.666
		78	75	20	21	21	2.711	2.711	2.711
		120	34	56	59	59	2.927	2.927	2.927
			38	8	8	8	3.107	3.107	3.107

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			75	23	24	24	3.666	3.666	3.666
CBP80	Q09161	17	20	121	129	130	5.127	5.127	5.127
			37	7	7	7	2.035	2.035	2.035
			67	45	47	47	2.258	2.258	2.258
			327			1			0.128
		20	17	121	129	130	5.127	5.127	5.127
			41	3	3	3	2.556	2.556	2.556
			67	97	102	102	2.802	2.802	2.802
			327	6	7	7	0.921	0.921	0.921
			671	2	2	3	0.914	0.914	0.914
		37	17	7	7	7	2.035	2.035	2.035
			82	8	8	8	1.997	1.997	1.997
		41	20	3	3	3	2.556	2.556	2.556
			330	24	24	24	4.070	4.070	4.070
			342	27	28	28	2.134	2.134	2.134
		67	17	45	47	47	2.258	2.258	2.258
			20	97	102	102	2.802	2.802	2.802
		82	37	8	8	8	1.997	1.997	1.997
		188	707		2	2		0.324	0.324
		204	238	2	2	2	0.530	0.530	0.530
			240	3	3	3	0.717	0.717	0.717
			241	2	2	2	0.688	0.688	0.688
		238	204	2	2	2	0.530	0.530	0.530
		240	204	3	3	3	0.717	0.717	0.717
		241	204	2	2	2	0.688	0.688	0.688
		327	17			1			0.128
			20	6	7	7	0.921	0.921	0.921
		330	41	24	24	24	4.070	4.070	4.070
		342	41	27	28	28	2.134	2.134	2.134
		511	557	6	6	6	3.386	3.386	3.386
			568	10	10	10	3.908	3.908	3.908
		557	511	6	6	6	3.386	3.386	3.386
		568	511	10	10	10	3.908	3.908	3.908
			607	67	68	68	4.514	4.514	4.514
		574	698	10	10	10	4.147	4.147	4.147
		607	568	67	68	68	4.514	4.514	4.514
		647	654	10	14	14	1.612	1.612	1.612
			657	12	15	16	1.599	1.599	1.599
			698	39	40	40	2.398	2.398	2.398
		650	657	258	262	262	3.459	3.459	3.459
			698	15	16	16	1.961	1.961	1.961
		654	647	10	14	14	1.612	1.612	1.612
			665	7	7	7	1.876	1.876	1.876
			698	146	150	150	4.646	4.646	4.646
		657	647	12	15	16	1.599	1.599	1.599
			650	258	262	262	3.459	3.459	3.459
			698	5	5	5	2.212	2.212	2.212
		663	671	4	6	6	1.021	1.021	1.021
			684	25	29	29	2.405	2.405	2.405
		665	654	7	7	7	1.876	1.876	1.876
			671	17	27	28	1.142	1.142	1.142
			684	128	141	142	3.605	3.605	3.605
			698	19	20	20	3.958	3.958	3.958
		671	20	2	2	3	0.914	0.914	0.914
			663	4	6	6	1.021	1.021	1.021
			665	17	27	28	1.142	1.142	1.142
			684	34	41	43	1.329	1.329	1.329
		684	663	25	29	29	2.405	2.405	2.405
			665	128	141	142	3.605	3.605	3.605
			671	34	41	43	1.329	1.329	1.329
			698	8	8	8	3.164	3.164	3.164
		698	574	10	10	10	4.147	4.147	4.147
			647	39	40	40	2.398	2.398	2.398
			650	15	16	16	1.961	1.961	1.961
			654	146	150	150	4.646	4.646	4.646
			657	5	5	5	2.212	2.212	2.212
			665	19	20	20	3.958	3.958	3.958
			684	8	8	8	3.164	3.164	3.164
		707	188		2	2		0.324	0.324
CCDC12	Q8WUD4	23	28		3	4		0.598	0.598
		28	23		3	4		0.598	0.598
			32	6	6	6	1.270	1.270	1.270
			34	6	7	7	1.421	1.421	1.421
		32	28	6	6	6	1.270	1.270	1.270
		34	28	6	7	7	1.421	1.421	1.421
			42	3	4	4	0.736	0.736	0.736
		42	34	3	4	4	0.736	0.736	0.736
			126			2			0.143
		53	78	4	4	4	1.778	1.778	1.778
		70	78	38	40	40	2.166	2.166	2.166
		71	78	23	25	25	1.287	1.287	1.287
		78	53	4	4	4	1.778	1.778	1.778
			70	38	40	40	2.166	2.166	2.166

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			71	23	25	25	1.287	1.287	1.287
			87	8	8	8	13.725	13.725	13.725
			126	2	2	2	1.054	1.054	1.054
		87	78	8	8	8	13.725	13.725	13.725
		122	126	13	13	13	1.947	1.947	1.947
		126	42			2			0.143
			78	2	2	2	1.054	1.054	1.054
			122	13	13	13	1.947	1.947	1.947
CDC5L	Q99459	7	47	2	4	4	0.810	0.810	0.810
			124	2	2	2	0.553	0.553	0.553
		20	28	85	88	88	3.752	3.752	3.752
			60	6	6	6	3.668	3.668	3.668
			135	2	2	2	2.037	2.037	2.037
		28	20	85	88	88	3.752	3.752	3.752
			47	13	17	17	2.737	2.737	2.737
			60	22	22	22	2.367	2.367	2.367
			218	2	2	2	1.670	1.670	1.670
		47	7	2	4	4	0.810	0.810	0.810
			28	13	17	17	2.737	2.737	2.737
		59	70	16	16	16	1.544	1.544	1.544
		60	20	6	6	6	3.668	3.668	3.668
			28	22	22	22	2.367	2.367	2.367
		70	59	16	16	16	1.544	1.544	1.544
		76	106	44	45	45	3.678	3.678	3.678
			124	13	13	13	1.963	1.963	1.963
		106	76	44	45	45	3.678	3.678	3.678
			122	7	7	7	6.565	6.565	6.565
			124	12	12	12	2.927	2.927	2.927
		122	106	7	7	7	6.565	6.565	6.565
		124	7	2	2	2	0.553	0.553	0.553
			76	13	13	13	1.963	1.963	1.963
			106	12	12	12	2.927	2.927	2.927
		135	20	2	2	2	2.037	2.037	2.037
			187	2	2	2	0.700	0.700	0.700
		170	174	5	5	5	1.296	1.296	1.296
		174	170	5	5	5	1.296	1.296	1.296
			187	4	4	4	0.625	0.625	0.625
		187	135	2	2	2	0.700	0.700	0.700
			174	4	4	4	0.625	0.625	0.625
			200	15	18	18	3.849	3.849	3.849
			218	2	3	3	0.494	0.494	0.494
		200	187	15	18	18	3.849	3.849	3.849
			218	9	10	11	2.073	2.073	2.073
			219	12	14	14	1.696	1.696	1.696
			255	6	6	6	2.236	2.236	2.236
		218	28	2	2	2	1.670	1.670	1.670
			187	2	3	3	0.494	0.494	0.494
			200	9	10	11	2.073	2.073	2.073
			255	8	9	9	1.063	1.063	1.063
			294	7	7	7	5.084	5.084	5.084
		219	200	12	14	14	1.696	1.696	1.696
			268	2	2	2	1.368	1.368	1.368
			294	3	3	3	4.939	4.939	4.939
		255	200	6	6	6	2.236	2.236	2.236
			218	8	9	9	1.063	1.063	1.063
			264	3	3	3	1.902	1.902	1.902
			268	12	14	14	1.432	1.432	1.432
			271	14	16	16	1.971	1.971	1.971
			294	3	3	3	2.125	2.125	2.125
		264	255	3	3	3	1.902	1.902	1.902
			270	29	34	34	1.733	1.733	1.733
			271	40	43	44	1.821	1.821	1.821
		268	219	2	2	2	1.368	1.368	1.368
			255	12	14	14	1.432	1.432	1.432
			271	57	60	61	1.747	1.747	1.747
		270	264	29	34	34	1.733	1.733	1.733
			294	27	27	27	4.696	4.696	4.696
		271	255	14	16	16	1.971	1.971	1.971
			264	40	43	44	1.821	1.821	1.821
			268	57	60	61	1.747	1.747	1.747
			294		2	2		3.618	3.618
			312	2	2	2	1.872	1.872	1.872
		290	294	121	129	129	5.572	5.572	5.572
			312	12	12	12	1.436	1.436	1.436
		291	294	13	15	15	2.245	2.245	2.245
		294	218	7	7	7	5.084	5.084	5.084
			219	3	3	3	4.939	4.939	4.939
			255	3	3	3	2.125	2.125	2.125
			270	27	27	27	4.696	4.696	4.696
			271		2	2		3.618	3.618
			290	121	129	129	5.572	5.572	5.572
			291	13	15	15	2.245	2.245	2.245
			380	3	3	3	3.560	3.560	3.560

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		312	271	2	2	2	1.872	1.872	1.872
			290	12	12	12	1.436	1.436	1.436
		380	294	3	3	3	3.560	3.560	3.560
			432	13	13	13	1.839	1.839	1.839
			447	5	5	5	2.869	2.869	2.869
			466	4	4	4	1.422	1.422	1.422
			487	3	3	3	3.194	3.194	3.194
		432	380	13	13	13	1.839	1.839	1.839
			447	53	54	54	2.707	2.707	2.707
			466	16	16	16	2.358	2.358	2.358
			487	3	3	3	2.080	2.080	2.080
			522	12	16	16	1.483	1.483	1.483
		447	380	5	5	5	2.869	2.869	2.869
			432	53	54	54	2.707	2.707	2.707
			487	6	6	6	2.644	2.644	2.644
		466	380	4	4	4	1.422	1.422	1.422
			432	16	16	16	2.358	2.358	2.358
			487	12	12	12	3.477	3.477	3.477
		487	380	3	3	3	3.194	3.194	3.194
			432	3	3	3	2.080	2.080	2.080
			447	6	6	6	2.644	2.644	2.644
			466	12	12	12	3.477	3.477	3.477
			522		2	2		1.389	1.389
			708	2	2	2	1.433	1.433	1.433
			718	2	2	2	2.489	2.489	2.489
		500	522	2	2	2	1.850	1.850	1.850
		522	432	12	16	16	1.483	1.483	1.483
			487		2	2		1.389	1.389
			500	2	2	2	1.850	1.850	1.850
			532	11	11	11	3.516	3.516	3.516
			535	11	12	12	3.953	3.953	3.953
			539	4	4	4	2.045	2.045	2.045
			598	11	11	11	1.830	1.830	1.830
			601	13	14	14	2.749	2.749	2.749
			626	13	13	13	2.223	2.223	2.223
		532	522	11	11	11	3.516	3.516	3.516
			539	12	12	12	1.615	1.615	1.615
			601	2	2	2	1.411	1.411	1.411
			623	8	27	28	1.007	1.007	1.007
			626	69	72	72	2.116	2.116	2.116
			630	7	7	7	1.361	1.361	1.361
		535	522	11	12	12	3.953	3.953	3.953
			539	55	59	59	2.813	2.813	2.813
			543	2	21	23	0.438	0.438	0.438
			598	3	8	8	0.480	0.480	0.480
			601	11	13	13	1.287	1.287	1.287
			623	4	26	27	0.497	0.497	0.497
			626	98	123	124	2.001	2.001	2.001
			630	20	24	24	1.583	1.583	1.583
			631	6	6	6	1.923	1.923	1.923
		539	522	4	4	4	2.045	2.045	2.045
			532	12	12	12	1.615	1.615	1.615
			535	55	59	59	2.813	2.813	2.813
			623	44	116	127	1.190	1.190	1.190
			626	262	265	266	3.657	3.657	3.657
			630	49	49	49	3.272	3.272	3.272
		543	535	2	21	23	0.438	0.438	0.438
			626	21	28	28	1.222	1.222	1.222
			630		3	3		0.324	0.324
		570	577	14	15	15	12.094	12.094	12.094
		577	570	14	15	15	12.094	12.094	12.094
			599		2	2		0.415	0.415
		598	522	11	11	11	1.830	1.830	1.830
			535	3	8	8	0.480	0.480	0.480
			694	5	7	7	0.577	0.577	0.577
			708	17	19	21	1.698	1.698	1.698
			712	23	24	24	1.415	1.415	1.415
		599	577		2	2		0.415	0.415
		601	522	13	14	14	2.749	2.749	2.749
			532	2	2	2	1.411	1.411	1.411
			535	11	13	13	1.287	1.287	1.287
			694	13	14	14	1.862	1.862	1.862
			708	45	48	48	2.895	2.895	2.895
			712	27	30	31	1.537	1.537	1.537
		623	532	8	27	28	1.007	1.007	1.007
			535	4	26	27	0.497	0.497	0.497
			539	44	116	127	1.190	1.190	1.190
		626	522	13	13	13	2.223	2.223	2.223
			532	69	72	72	2.116	2.116	2.116
			535	98	123	124	2.001	2.001	2.001
			539	262	265	266	3.657	3.657	3.657
			543	21	28	28	1.222	1.222	1.222
			631	19	19	19	2.288	2.288	2.288

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		630	532	7	7	7	1.361	1.361	1.361
			535	20	24	24	1.583	1.583	1.583
			539	49	49	49	3.272	3.272	3.272
			543		3	3		0.324	0.324
		631	535	6	6	6	1.923	1.923	1.923
			626	19	19	19	2.288	2.288	2.288
		685	694	7	8	8	1.451	1.451	1.451
		694	598	5	7	7	0.577	0.577	0.577
			601	13	14	14	1.862	1.862	1.862
			685	7	8	8	1.451	1.451	1.451
		708	487	2	2	2	1.433	1.433	1.433
			598	17	19	21	1.698	1.698	1.698
			601	45	48	48	2.895	2.895	2.895
			712	25	28	28	1.352	1.352	1.352
			718	11	11	11	2.718	2.718	2.718
		712	598	23	24	24	1.415	1.415	1.415
			601	27	30	31	1.537	1.537	1.537
			708	25	28	28	1.352	1.352	1.352
			716	2	3	3	0.613	0.613	0.613
		716	712	2	3	3	0.613	0.613	0.613
		718	487	2	2	2	2.489	2.489	2.489
			708	11	11	11	2.718	2.718	2.718
		771	782	39	39	39	2.329	2.329	2.329
		782	771	39	39	39	2.329	2.329	2.329
CTNNBL1	Q8WYA6	16	27	21	25	25	1.161	1.161	1.161
			31	11	15	15	4.313	4.313	4.313
		27	16	21	25	25	1.161	1.161	1.161
			31	84	98	102	2.661	2.661	2.661
		31	16	11	15	15	4.313	4.313	4.313
			27	84	98	102	2.661	2.661	2.661
			56	25	27	27	2.432	2.432	2.432
			57	15	15	15	2.433	2.433	2.433
		56	31	25	27	27	2.432	2.432	2.432
			91	2	4	4	0.753	0.753	0.753
		57	31	15	15	15	2.433	2.433	2.433
		83	91	29	39	39	2.041	2.041	2.041
		84	95	13	13	13	2.521	2.521	2.521
		91	56	2	4	4	0.753	0.753	0.753
			83	29	39	39	2.041	2.041	2.041
			95	59	60	60	2.074	2.074	2.074
			102	102	109	109	3.520	3.520	3.520
		95	84	13	13	13	2.521	2.521	2.521
			91	59	60	60	2.074	2.074	2.074
			102	136	144	144	4.435	4.435	4.435
		102	91	102	109	109	3.520	3.520	3.520
			95	136	144	144	4.435	4.435	4.435
		247	250		4	4		0.207	0.207
			252	8	9	9	2.481	2.481	2.481
		250	247		4	4		0.207	0.207
		252	247	8	9	9	2.481	2.481	2.481
		297	347	3	4	4	1.168	1.168	1.168
		332	372	36	36	36	4.008	4.008	4.008
		347	297	3	4	4	1.168	1.168	1.168
		372	332	36	36	36	4.008	4.008	4.008
		401	458	33	33	33	4.563	4.563	4.563
		457	463	4	5	5	0.763	0.763	0.763
			527	13	14	14	1.627	1.627	1.627
		458	401	33	33	33	4.563	4.563	4.563
		463	457	4	5	5	0.763	0.763	0.763
		527	457	13	14	14	1.627	1.627	1.627
			534	177	186	186	2.471	2.471	2.471
		534	527	177	186	186	2.471	2.471	2.471
			554	45	63	63	1.201	1.201	1.201
		554	534	45	63	63	1.201	1.201	1.201
CWC15	Q9P013	18	40	42	44	44	2.574	2.574	2.574
			42	26	31	31	1.542	1.542	1.542
			91	6	6	6	2.171	2.171	2.171
		28	40	63	72	72	2.169	2.169	2.169
			42	54	59	59	1.354	1.354	1.354
			91	15	18	18	1.742	1.742	1.742
		40	18	42	44	44	2.574	2.574	2.574
			28	63	72	72	2.169	2.169	2.169
			91		1	1		0.772	0.772
		42	18	26	31	31	1.542	1.542	1.542
			28	54	59	59	1.354	1.354	1.354
		91	18	6	6	6	2.171	2.171	2.171
			28	15	18	18	1.742	1.742	1.742
			40		1	1		0.772	0.772
		183	195	6	7	7	3.140	3.140	3.140
		195	183	6	7	7	3.140	3.140	3.140
		205	221	6	6	6	1.671	1.671	1.671
		221	205	6	6	6	1.671	1.671	1.671
			226	3	5	5	0.791	0.791	0.791

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		226	221	3	5	5	0.791	0.791	0.791
HNRNPA1	P09651	3	15	54	56	56	6.364	6.364	6.364
			52	8	8	8	1.575	1.575	1.575
			106	28	29	30	4.136	4.136	4.136
			113	17	18	18	2.782	2.782	2.782
			166	4	4	4	2.850	2.850	2.850
			350	38	38	38	3.562	3.562	3.562
		8	15	54	54	54	4.418	4.418	4.418
			106	12	12	12	1.920	1.920	1.920
			113	7	7	7	1.253	1.253	1.253
			166	4	4	4	0.952	0.952	0.952
			350	25	25	25	1.632	1.632	1.632
		15	3	54	56	56	6.364	6.364	6.364
			8	54	54	54	4.418	4.418	4.418
			106	43	43	43	7.045	7.045	7.045
			166	9	9	9	3.569	3.569	3.569
			350	23	23	23	3.795	3.795	3.795
		52	3	8	8	8	1.575	1.575	1.575
			106	18	19	19	1.460	1.460	1.460
			113	5	5	5	0.654	0.654	0.654
			350	16	19	19	0.917	0.917	0.917
		78	106	3	4	4	2.228	2.228	2.228
			113	2	4	4	0.865	0.865	0.865
			144		2	2		0.260	0.260
			166	3	3	3	0.962	0.962	0.962
			350		2	2		0.366	0.366
		105	350	6	7	7	1.842	1.842	1.842
		106	3	28	29	30	4.136	4.136	4.136
			8	12	12	12	1.920	1.920	1.920
			15	43	43	43	7.045	7.045	7.045
			52	18	19	19	1.460	1.460	1.460
			78	3	4	4	2.228	2.228	2.228
			350	66	70	70	4.046	4.046	4.046
		113	3	17	18	18	2.782	2.782	2.782
			8	7	7	7	1.253	1.253	1.253
			52	5	5	5	0.654	0.654	0.654
			78	2	4	4	0.865	0.865	0.865
			144	5	6	7	1.240	1.240	1.240
			350	30	33	33	2.402	2.402	2.402
		144	78		2	2		0.260	0.260
			113	5	6	7	1.240	1.240	1.240
		166	3	4	4	4	2.850	2.850	2.850
			8	4	4	4	0.952	0.952	0.952
			15	9	9	9	3.569	3.569	3.569
			78	3	3	3	0.962	0.962	0.962
			350	16	17	17	2.505	2.505	2.505
		350	3	38	38	38	3.562	3.562	3.562
			8	25	25	25	1.632	1.632	1.632
			15	23	23	23	3.795	3.795	3.795
			52	16	19	19	0.917	0.917	0.917
			78		2	2		0.366	0.366
			105	6	7	7	1.842	1.842	1.842
			106	66	70	70	4.046	4.046	4.046
			113	30	33	33	2.402	2.402	2.402
			166	16	17	17	2.505	2.505	2.505
ISY1	Q9ULR0	24	41	20	23	23	2.925	2.925	2.925
			44	6	8	8	0.910	0.910	0.910
		26	41	7	8	8	1.039	1.039	1.039
			44	13	16	16	0.690	0.690	0.690
		41	24	20	23	23	2.925	2.925	2.925
			26	7	8	8	1.039	1.039	1.039
		44	24	6	8	8	0.910	0.910	0.910
			26	13	16	16	0.690	0.690	0.690
			55	6	7	7	1.307	1.307	1.307
		55	44	6	7	7	1.307	1.307	1.307
		84	92	24	24	24	2.919	2.919	2.919
			101	12	15	15	2.055	2.055	2.055
			105	70	70	70	4.132	4.132	4.132
			112		2	2		0.397	0.397
			121	125	127	127	4.039	4.039	4.039
		92	84	24	24	24	2.919	2.919	2.919
			105	20	20	20	3.077	3.077	3.077
		101	84	12	15	15	2.055	2.055	2.055
			121	9	9	9	1.669	1.669	1.669
		105	84	70	70	70	4.132	4.132	4.132
			92	20	20	20	3.077	3.077	3.077
			121	60	61	61	3.068	3.068	3.068
		112	84		2	2		0.397	0.397
		121	84	125	127	127	4.039	4.039	4.039
			101	9	9	9	1.669	1.669	1.669
			105	60	61	61	3.068	3.068	3.068
		179	190	5	8	8	0.857	0.857	0.857
		190	179	5	8	8	0.857	0.857	0.857

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
KIN17	O60870	42	324			1			$9.921 \cdot 10^{-2}$
		231	242	3	3	3	2.518	2.518	2.518
		240	242	13	13	13	2.771	2.771	2.771
		242	231	3	3	3	2.518	2.518	2.518
			240	13	13	13	2.771	2.771	2.771
		252	258	2	2	2	1.328	1.328	1.328
		258	252	2	2	2	1.328	1.328	1.328
		288	302	14	14	14	3.385	3.385	3.385
		292	300	8	10	10	1.991	1.991	1.991
		293	324	2	3	3	0.618	0.618	0.618
		300	292	8	10	10	1.991	1.991	1.991
			302	5	5	5	2.772	2.772	2.772
		302	288	14	14	14	3.385	3.385	3.385
			300	5	5	5	2.772	2.772	2.772
			341	1	1	1	0.911	0.911	0.911
			391	8	8	8	2.085	2.085	2.085
		317	391	4	4	4	2.258	2.258	2.258
		324	42			1			$9.921 \cdot 10^{-2}$
			293	2	3	3	0.618	0.618	0.618
		341	302	1	1	1	0.911	0.911	0.911
			391	7	7	7	3.600	3.600	3.600
		391	302	8	8	8	2.085	2.085	2.085
			317	4	4	4	2.258	2.258	2.258
			341	7	7	7	3.600	3.600	3.600
LSm2	Q9Y333	1	88	26	27	27	5.800	5.800	5.800
			94	9	9	9	1.459	1.459	1.459
		8	88	21	21	21	2.235	2.235	2.235
			94	3	3	3	1.605	1.605	1.605
		88	1	26	27	27	5.800	5.800	5.800
			8	21	21	21	2.235	2.235	2.235
		94	1	9	9	9	1.459	1.459	1.459
			8	3	3	3	1.605	1.605	1.605
LSm4	Q9Y4Z0	1	80	48	49	50	4.814	4.814	4.814
		80	1	48	49	50	4.814	4.814	4.814
LSm6	P62312	5	77	4	4	4	1.609	1.609	1.609
		13	77	5	5	5	2.114	2.114	2.114
		77	5	4	4	4	1.609	1.609	1.609
			13	5	5	5	2.114	2.114	2.114
LSm7	Q9UK45	4	9	12	17	17	1.772	1.772	1.772
		6	8	3	3	3	1.583	1.583	1.583
			9	29	31	31	2.383	2.383	2.383
		7	9	4	4	4	1.299	1.299	1.299
		8	6	3	3	3	1.583	1.583	1.583
		9	4	12	17	17	1.772	1.772	1.772
			6	29	31	31	2.383	2.383	2.383
			7	4	4	4	1.299	1.299	1.299
MBPMS2	P0AEX9-r	2	314		2	2		0.191	0.191
		7	26	6	7	7	3.399	3.399	3.399
			35	4	4	4	1.156	1.156	1.156
			47	5	5	5	2.548	2.548	2.548
		16	30		3	3		0.224	0.224
			35		2	2		0.507	0.507
			296	15	15	15	3.181	3.181	3.181
			298	13	13	13	2.415	2.415	2.415
		26	7	6	7	7	3.399	3.399	3.399
			30	301	308	308	5.173	5.173	5.173
			35	124	140	140	4.497	4.497	4.497
			43	14	14	14	1.925	1.925	1.925
			47	5	5	5	3.144	3.144	3.144
			296	22	22	22	4.393	4.393	4.393
		27	47	2	2	2	3.225	3.225	3.225
		30	16		3	3		0.224	0.224
			26	301	308	308	5.173	5.173	5.173
			43	11	15	15	2.011	2.011	2.011
			47	24	26	26	3.153	3.153	3.153
			120	7	9	10	2.031	2.031	2.031
			176		4	5		0.823	0.823
			296	3	3	3	1.506	1.506	1.506
		35	7	4	4	4	1.156	1.156	1.156
			16		2	2		0.507	0.507
			26	124	140	140	4.497	4.497	4.497
			176	7	13	14	0.594	0.594	0.594
		43	26	14	14	14	1.925	1.925	1.925
			30	11	15	15	2.011	2.011	2.011
			47	6	6	6	0.796	0.796	0.796
			176		6	6		0.454	0.454
			298	2	2	2	2.520	2.520	2.520
			424		2	2		0.866	0.866
			438	4	7	7	2.894	2.894	2.894
			447	7	13	13	1.127	1.127	1.127
		47	7	5	5	5	2.548	2.548	2.548
			26	5	5	5	3.144	3.144	3.144

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			27	2	2	2	3.225	3.225	3.225
			30	24	26	26	3.153	3.153	3.153
			43	6	6	6	0.796	0.796	0.796
			84	8	8	8	7.914	7.914	7.914
			89	2	2	2	3.285	3.285	3.285
			103	8	8	8	5.653	5.653	5.653
			120	3	3	3	4.505	4.505	4.505
			176	3	6	6	1.808	1.808	1.808
			438	22	28	28	1.972	1.972	1.972
			442	7	9	9	3.511	3.511	3.511
			447	54	55	56	1.970	1.970	1.970
		84	47	8	8	8	7.914	7.914	7.914
			84		4	6		1.858	1.858
			103	28	31	31	5.437	5.437	5.437
			128	24	29	29	1.406	1.406	1.406
			138	4	4	4	1.373	1.373	1.373
			176	2	4	4	0.487	0.487	0.487
			252	3	3	3	5.156	5.156	5.156
			296	2	2	2	1.759	1.759	1.759
			306	450	452	452	9.343	9.343	9.343
			314	7	8	8	1.263	1.263	1.263
			327	7	7	7	4.788	4.788	4.788
		89	47	2	2	2	3.285	3.285	3.285
			306	55	55	55	2.985	2.985	2.985
		103	47	8	8	8	5.653	5.653	5.653
			84	28	31	31	5.437	5.437	5.437
			128	6	6	6	4.145	4.145	4.145
			138	9	11	11	2.134	2.134	2.134
			176	580	591	591	5.504	5.504	5.504
			252	3	3	3	4.583	4.583	4.583
			257	2	2	2	4.072	4.072	4.072
			274	127	128	128	27.568	27.568	27.568
			447	36	37	37	2.076	2.076	2.076
		120	30	7	9	10	2.031	2.031	2.031
			47	3	3	3	4.505	4.505	4.505
			138	5	5	5	1.785	1.785	1.785
			141	19	19	19	2.031	2.031	2.031
			176	12	13	13	1.532	1.532	1.532
			180	3	4	4	3.159	3.159	3.159
			203		2	2		2.081	2.081
			274	2	2	2	2.684	2.684	2.684
			314	110	111	111	3.150	3.150	3.150
			424	5	6	6	1.966	1.966	1.966
			438	8	11	11	0.718	0.718	0.718
			442		2	2		1.371	1.371
			447	19	21	22	1.787	1.787	1.787
		128	84	24	29	29	1.406	1.406	1.406
			103	6	6	6	4.145	4.145	4.145
			145	3	3	3	3.766	3.766	3.766
			327	163	163	163	27.444	27.444	27.444
		138	84	4	4	4	1.373	1.373	1.373
			103	9	11	11	2.134	2.134	2.134
			120	5	5	5	1.785	1.785	1.785
			145	10	10	10	3.420	3.420	3.420
			176	6	6	6	1.803	1.803	1.803
			180	12	12	12	1.921	1.921	1.921
			190	10	10	10	1.853	1.853	1.853
			201	39	39	39	2.404	2.404	2.404
			203	347	351	351	16.203	16.203	16.203
			220	32	37	38	4.234	4.234	4.234
			252	3	3	3	2.776	2.776	2.776
		141	120	19	19	19	2.031	2.031	2.031
			145	9	10	10	2.669	2.669	2.669
			203	30	31	31	10.593	10.593	10.593
		145	128	3	3	3	3.766	3.766	3.766
			138	10	10	10	3.420	3.420	3.420
			141	9	10	10	2.669	2.669	2.669
			180		2	2		0.623	0.623
		171	180	12	12	12	2.564	2.564	2.564
		176	30		4	5		0.823	0.823
			35	7	13	14	0.594	0.594	0.594
			43		6	6		0.454	0.454
			47	3	6	6	1.808	1.808	1.808
			84	2	4	4	0.487	0.487	0.487
			103	580	591	591	5.504	5.504	5.504
			120	12	13	13	1.532	1.532	1.532
			138	6	6	6	1.803	1.803	1.803
			176		2	2		1.471	1.471
			180	7	10	10	1.233	1.233	1.233
			190	15	15	15	1.524	1.524	1.524
			201	11	11	11	2.074	2.074	2.074
			274		3	3		0.368	0.368
			363	1	1	1	0.434	0.434	0.434

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			424	8	8	9	1.255	1.255	1.255
			438	6	8	8	0.667	0.667	0.667
			447	19	24	24	1.734	1.734	1.734
		180	120	3	4	4	3.159	3.159	3.159
			138	12	12	12	1.921	1.921	1.921
			145		2	2		0.623	0.623
			171	12	12	12	2.564	2.564	2.564
			176	7	10	10	1.233	1.233	1.233
			363	71	71	71	4.976	4.976	4.976
			438	5	5	5	1.115	1.115	1.115
			447	16	17	17	1.716	1.716	1.716
		190	138	10	10	10	1.853	1.853	1.853
			176	15	15	15	1.524	1.524	1.524
			252	53	55	55	39.666	39.666	39.666
			363	52	52	52	4.388	4.388	4.388
			438	1	2	2	0.770	0.770	0.770
			447	10	10	10	1.888	1.888	1.888
		201	138	39	39	39	2.404	2.404	2.404
			176	11	11	11	2.074	2.074	2.074
			447	6	6	6	1.779	1.779	1.779
		203	120		2	2		2.081	2.081
			138	347	351	351	16.203	16.203	16.203
			141	30	31	31	10.593	10.593	10.593
			240	3	3	3	4.177	4.177	4.177
			447	4	4	4	0.885	0.885	0.885
		220	138	32	37	38	4.234	4.234	4.234
			296	2	2	2	5.049	5.049	5.049
		240	203	3	3	3	4.177	4.177	4.177
			296	7	7	7	20.695	20.695	20.695
		252	84	3	3	3	5.156	5.156	5.156
			103	3	3	3	4.583	4.583	4.583
			138	3	3	3	2.776	2.776	2.776
			190	53	55	55	39.666	39.666	39.666
			306	2	2	2	1.746	1.746	1.746
		257	103	2	2	2	4.072	4.072	4.072
		274	103	127	128	128	27.568	27.568	27.568
			120	2	2	2	2.684	2.684	2.684
			176		3	3		0.368	0.368
			296	7	8	8	2.452	2.452	2.452
			298	5	5	5	2.014	2.014	2.014
			306	2	3	3	0.586	0.586	0.586
			424	5	5	5	1.579	1.579	1.579
			438	10	10	10	1.768	1.768	1.768
			447	16	16	16	2.040	2.040	2.040
		296	16	15	15	15	3.181	3.181	3.181
			26	22	22	22	4.393	4.393	4.393
			30	3	3	3	1.506	1.506	1.506
			84	2	2	2	1.759	1.759	1.759
			220	2	2	2	5.049	5.049	5.049
			240	7	7	7	20.695	20.695	20.695
			274	7	8	8	2.452	2.452	2.452
			314	66	69	69	1.839	1.839	1.839
		298	16	13	13	13	2.415	2.415	2.415
			43	2	2	2	2.520	2.520	2.520
			274	5	5	5	2.014	2.014	2.014
			306	3	4	4	1.095	1.095	1.095
		306	84	450	452	452	9.343	9.343	9.343
			89	55	55	55	2.985	2.985	2.985
			252	2	2	2	1.746	1.746	1.746
			274	2	3	3	0.586	0.586	0.586
			298	3	4	4	1.095	1.095	1.095
			314		6	6		0.648	0.648
		314	2		2	2		0.191	0.191
			84	7	8	8	1.263	1.263	1.263
			120	110	111	111	3.150	3.150	3.150
			296	66	69	69	1.839	1.839	1.839
			306		6	6		0.648	0.648
		327	84	7	7	7	4.788	4.788	4.788
			128	163	163	163	27.444	27.444	27.444
		363	176	1	1	1	0.434	0.434	0.434
			180	71	71	71	4.976	4.976	4.976
			190	52	52	52	4.388	4.388	4.388
			424	3	3	3	1.624	1.624	1.624
			438	4	5	5	1.086	1.086	1.086
			447	16	16	16	1.296	1.296	1.296
		424	43		2	2		0.866	0.866
			120	5	6	6	1.966	1.966	1.966
			176	8	8	9	1.255	1.255	1.255
			274	5	5	5	1.579	1.579	1.579
			363	3	3	3	1.624	1.624	1.624
			438	20	25	25	1.912	1.912	1.912
			442	75	76	76	3.037	3.037	3.037
			447	29	29	29	2.512	2.512	2.512

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		438	43	4	7	7	2.894	2.894	2.894
			47	22	28	28	1.972	1.972	1.972
			120	8	11	11	0.718	0.718	0.718
			176	6	8	8	0.667	0.667	0.667
			180	5	5	5	1.115	1.115	1.115
			190	1	2	2	0.770	0.770	0.770
			274	10	10	10	1.768	1.768	1.768
			363	4	5	5	1.086	1.086	1.086
			424	20	25	25	1.912	1.912	1.912
			438	4	6	6	0.702	0.702	0.702
			442	169	174	174	3.223	3.223	3.223
			447	18	18	18	2.214	2.214	2.214
		442	47	7	9	9	3.511	3.511	3.511
			120		2	2		1.371	1.371
			424	75	76	76	3.037	3.037	3.037
			438	169	174	174	3.223	3.223	3.223
		447	43	7	13	13	1.127	1.127	1.127
			47	54	55	56	1.970	1.970	1.970
			103	36	37	37	2.076	2.076	2.076
			120	19	21	22	1.787	1.787	1.787
			176	19	24	24	1.734	1.734	1.734
			180	16	17	17	1.716	1.716	1.716
			190	10	10	10	1.888	1.888	1.888
			201	6	6	6	1.779	1.779	1.779
			203	4	4	4	0.885	0.885	0.885
			274	16	16	16	2.040	2.040	2.040
			363	16	16	16	1.296	1.296	1.296
			424	29	29	29	2.512	2.512	2.512
			438	18	18	18	2.214	2.214	2.214
MFAP1	P55081	230	238	41	42	42	3.734	3.734	3.734
		238	230	41	42	42	3.734	3.734	3.734
			249	6	6	6	1.866	1.866	1.866
		242	250	16	27	27	2.369	2.369	2.369
		249	238	6	6	6	1.866	1.866	1.866
		250	242	16	27	27	2.369	2.369	2.369
		329	334	6	12	14	0.644	0.644	0.644
			357	17	17	17	1.095	1.095	1.095
		334	329	6	12	14	0.644	0.644	0.644
		357	329	17	17	17	1.095	1.095	1.095
		381	390	2	2	2	1.366	1.366	1.366
		390	381	2	2	2	1.366	1.366	1.366
		415	420	18	18	18	2.219	2.219	2.219
		420	415	18	18	18	2.219	2.219	2.219
PHF5A	Q7RTV0	3	95	2	2	2	1.358	1.358	1.358
		95	3	2	2	2	1.358	1.358	1.358
PLRG1	O43660	41	62	7	8	8	1.937	1.937	1.937
			66	2	2	2	1.472	1.472	1.472
			68	2	2	2	2.962	2.962	2.962
		62	41	7	8	8	1.937	1.937	1.937
			68	20	20	20	3.188	3.188	3.188
			80	7	7	7	7.317	7.317	7.317
		66	41	2	2	2	1.472	1.472	1.472
			113	4	5	5	1.137	1.137	1.137
		68	41	2	2	2	2.962	2.962	2.962
			62	20	20	20	3.188	3.188	3.188
			113	4	4	4	3.244	3.244	3.244
		80	62	7	7	7	7.317	7.317	7.317
			135	2	2	2	3.558	3.558	3.558
		113	66	4	5	5	1.137	1.137	1.137
			68	4	4	4	3.244	3.244	3.244
			135	19	21	21	10.687	10.687	10.687
			180	5	5	5	2.155	2.155	2.155
		135	80	2	2	2	3.558	3.558	3.558
			113	19	21	21	10.687	10.687	10.687
			173	10	10	10	1.224	1.224	1.224
			180	9	10	10	1.819	1.819	1.819
			510	4	4	4	1.541	1.541	1.541
		173	135	10	10	10	1.224	1.224	1.224
			181	28	32	32	4.588	4.588	4.588
			510	7	9	10	8.955	8.955	8.955
		180	113	5	5	5	2.155	2.155	2.155
			135	9	10	10	1.819	1.819	1.819
			505	15	19	19	1.478	1.478	1.478
			510	48	54	54	2.127	2.127	2.127
		181	173	28	32	32	4.588	4.588	4.588
			505	2	4	4	2.392	2.392	2.392
			510	8	9	9	2.148	2.148	2.148
		187	505	3	3	3	1.526	1.526	1.526
		279	320	9	9	9	2.252	2.252	2.252
		320	279	9	9	9	2.252	2.252	2.252
		372	396	24	27	27	1.625	1.625	1.625
		396	372	24	27	27	1.625	1.625	1.625
			510	4	4	4	1.484	1.484	1.484

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		505	180	15	19	19	1.478	1.478	1.478
			181	2	4	4	2.392	2.392	2.392
			187	3	3	3	1.526	1.526	1.526
			512		3	4		0.287	0.287
		510	135	4	4	4	1.541	1.541	1.541
			173	7	9	10	8.955	8.955	8.955
			180	48	54	54	2.127	2.127	2.127
			181	8	9	9	2.148	2.148	2.148
			396	4	4	4	1.484	1.484	1.484
		512	505		3	4		0.287	0.287
PPIE	Q9UNP9	5	83	33	38	38	3.798	3.798	3.798
			88	3	10	14	0.710	0.710	0.710
			104	21	24	24	3.344	3.344	3.344
			108	7	17	19	0.975	0.975	0.975
			123		3	3		0.248	0.248
		20	104	2	3	3	1.214	1.214	1.214
		83	5	33	38	38	3.798	3.798	3.798
			88	13	17	17	0.636	0.636	0.636
			104	16	16	16	2.657	2.657	2.657
			108	6	11	11	0.717	0.717	0.717
		88	5	3	10	14	0.710	0.710	0.710
			83	13	17	17	0.636	0.636	0.636
			104	2	2	2	0.781	0.781	0.781
		103	108	7	9	9	1.222	1.222	1.222
		104	5	21	24	24	3.344	3.344	3.344
			20	2	3	3	1.214	1.214	1.214
			83	16	16	16	2.657	2.657	2.657
			88	2	2	2	0.781	0.781	0.781
			114	3	3	3	0.461	0.461	0.461
			134	5	5	5	0.610	0.610	0.610
		108	5	7	17	19	0.975	0.975	0.975
			83	6	11	11	0.717	0.717	0.717
			103	7	9	9	1.222	1.222	1.222
		114	104	3	3	3	0.461	0.461	0.461
		123	5		3	3		0.248	0.248
		134	104	5	5	5	0.610	0.610	0.610
		147	185	48	49	49	3.254	3.254	3.254
		185	147	48	49	49	3.254	3.254	3.254
PPIL1	Q9Y3C6	34	80	5	5	5	1.337	1.337	1.337
		52	158	231	298	303	2.371	2.371	2.371
		80	34	5	5	5	1.337	1.337	1.337
		158	52	231	298	303	2.371	2.371	2.371
PPIL2	Q13356	8	26		9	16		0.301	0.301
		26	8		9	16		0.301	0.301
		230	232	12	12	12	1.678	1.678	1.678
			235		3	3		0.815	0.815
		231	454	2	2	2	1.841	1.841	1.841
			462	2	2	2	0.687	0.687	0.687
		232	230	12	12	12	1.678	1.678	1.678
			450	33	33	33	4.496	4.496	4.496
			454	42	42	42	2.857	2.857	2.857
			462		2	2		0.457	0.457
		235	230		3	3		0.815	0.815
			277	3	3	3	0.969	0.969	0.969
			278	2	3	3	0.556	0.556	0.556
			450	7	7	7	1.901	1.901	1.901
			454	24	27	27	2.274	2.274	2.274
			460	3	7	7	0.556	0.556	0.556
			462	4	4	4	2.318	2.318	2.318
		277	235	3	3	3	0.969	0.969	0.969
		278	235	2	3	3	0.556	0.556	0.556
		302	348	13	14	14	5.409	5.409	5.409
		348	302	13	14	14	5.409	5.409	5.409
		450	232	33	33	33	4.496	4.496	4.496
			235	7	7	7	1.901	1.901	1.901
			460	21	21	22	2.332	2.332	2.332
			462	35	37	37	2.227	2.227	2.227
		454	231	2	2	2	1.841	1.841	1.841
			232	42	42	42	2.857	2.857	2.857
			235	24	27	27	2.274	2.274	2.274
			462	46	50	50	3.097	3.097	3.097
		460	235	3	7	7	0.556	0.556	0.556
			450	21	21	22	2.332	2.332	2.332
		462	231	2	2	2	0.687	0.687	0.687
			232		2	2		0.457	0.457
			235	4	4	4	2.318	2.318	2.318
			450	35	37	37	2.227	2.227	2.227
			454	46	50	50	3.097	3.097	3.097
PPP1CA	P62136	26	305	52	53	53	3.266	3.266	3.266
			327	23	24	26	1.661	1.661	1.661
		301	305	2	2	2	0.502	0.502	0.502
		305	26	52	53	53	3.266	3.266	3.266
			301	2	2	2	0.502	0.502	0.502

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			327	14	14	14	1.559	1.559	1.559
		327	26	23	24	26	1.661	1.661	1.661
			305	14	14	14	1.559	1.559	1.559
PPP1CB	P62140	25	301	4	5	5	1.588	1.588	1.588
			303	15	17	17	1.959	1.959	1.959
			326		2	2		0.233	0.233
		301	25	4	5	5	1.588	1.588	1.588
		303	25	15	17	17	1.959	1.959	1.959
		326	25		2	2		0.233	0.233
PPP1R8	Q12972	22	47	16	17	17	1.952	1.952	1.952
			78	28	29	29	2.520	2.520	2.520
		40	129	53	53	53	2.877	2.877	2.877
		47	22	16	17	17	1.952	1.952	1.952
			78	258	259	259	4.203	4.203	4.203
			81	2	2	2	1.053	1.053	1.053
		78	22	28	29	29	2.520	2.520	2.520
			47	258	259	259	4.203	4.203	4.203
			138	15	15	15	1.778	1.778	1.778
			175	3	3	3	1.305	1.305	1.305
		81	47	2	2	2	1.053	1.053	1.053
			129	2	2	2	0.958	0.958	0.958
			138	5	5	5	0.947	0.947	0.947
		105	175	2	2	2	2.112	2.112	2.112
		129	40	53	53	53	2.877	2.877	2.877
			81	2	2	2	0.958	0.958	0.958
		138	78	15	15	15	1.778	1.778	1.778
			81	5	5	5	0.947	0.947	0.947
			151	4	4	4	2.676	2.676	2.676
		151	138	4	4	4	2.676	2.676	2.676
		175	78	3	3	3	1.305	1.305	1.305
			105	2	2	2	2.112	2.112	2.112
			343	6	8	8	1.142	1.142	1.142
		234	236	19	23	23	2.849	2.849	2.849
		236	234	19	23	23	2.849	2.849	2.849
		334	344		2	2		0.537	0.537
		337	344		2	2		0.358	0.358
		343	175	6	8	8	1.142	1.142	1.142
		344	334		2	2		0.537	0.537
			337		2	2		0.358	0.358
PQBP1	O60828	2	18	20	20	20	3.382	3.382	3.382
		18	2	20	20	20	3.382	3.382	3.382
		87	102	19	20	22	2.892	2.892	2.892
			109	8	15	15	1.258	1.258	1.258
			116	12	19	19	1.594	1.594	1.594
			123	9	11	11	1.164	1.164	1.164
			148		2	2		0.504	0.504
		90	102	17	22	23	1.105	1.105	1.105
			123	2	2	2	1.427	1.427	1.427
		102	87	19	20	22	2.892	2.892	2.892
			90	17	22	23	1.105	1.105	1.105
			109	28	29	29	2.833	2.833	2.833
			116	28	29	29	2.430	2.430	2.430
			123	14	14	14	1.556	1.556	1.556
			137		1	1		0.546	0.546
		109	87	8	15	15	1.258	1.258	1.258
			102	28	29	29	2.833	2.833	2.833
			116	9	12	12	1.997	1.997	1.997
			123	18	20	20	2.161	2.161	2.161
			148	5	6	6	1.609	1.609	1.609
		116	87	12	19	19	1.594	1.594	1.594
			102	28	29	29	2.430	2.430	2.430
			109	9	12	12	1.997	1.997	1.997
			123	46	46	46	3.450	3.450	3.450
			137	10	13	13	0.847	0.847	0.847
			148	47	50	50	3.263	3.263	3.263
			189	10	10	10	2.484	2.484	2.484
			192	5	5	5	2.460	2.460	2.460
			223		3	3		0.472	0.472
			228	10	13	13	1.027	1.027	1.027
			262	3	4	4	0.480	0.480	0.480
		123	87	9	11	11	1.164	1.164	1.164
			90	2	2	2	1.427	1.427	1.427
			102	14	14	14	1.556	1.556	1.556
			109	18	20	20	2.161	2.161	2.161
			116	46	46	46	3.450	3.450	3.450
			130	3	4	4	1.720	1.720	1.720
			137	12	14	14	0.941	0.941	0.941
			148	14	16	16	1.518	1.518	1.518
			167		2	2		0.399	0.399
			189	2	2	2	1.076	1.076	1.076
			228	3	3	3	1.007	1.007	1.007
		130	123	3	4	4	1.720	1.720	1.720
		137	102		1	1		0.546	0.546

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			116	10	13	13	0.847	0.847	0.847
			123	12	14	14	0.941	0.941	0.941
			148	37	42	42	1.315	1.315	1.315
		148	87		2	2		0.504	0.504
			109	5	6	6	1.609	1.609	1.609
			116	47	50	50	3.263	3.263	3.263
			123	14	16	16	1.518	1.518	1.518
			137	37	42	42	1.315	1.315	1.315
			167	16	19	19	1.136	1.136	1.136
			174	4	5	5	1.177	1.177	1.177
			189	4	5	5	1.218	1.218	1.218
			192	3	3	3	2.521	2.521	2.521
		167	123		2	2		0.399	0.399
			148	16	19	19	1.136	1.136	1.136
			174	2	5	5	0.480	0.480	0.480
			189	23	24	24	2.187	2.187	2.187
			192	11	11	11	1.776	1.776	1.776
			197	2	2	2	1.445	1.445	1.445
			223		3	3		0.451	0.451
			228		3	3		0.642	0.642
		174	148	4	5	5	1.177	1.177	1.177
			167	2	5	5	0.480	0.480	0.480
			189	12	13	13	1.412	1.412	1.412
			192	5	5	5	1.566	1.566	1.566
			197	2	2	3	1.094	1.094	1.094
		189	116	10	10	10	2.484	2.484	2.484
			123	2	2	2	1.076	1.076	1.076
			148	4	5	5	1.218	1.218	1.218
			167	23	24	24	2.187	2.187	2.187
			174	12	13	13	1.412	1.412	1.412
			192	116	117	117	3.077	3.077	3.077
			197	43	44	44	4.653	4.653	4.653
			223	17	20	20	2.410	2.410	2.410
			228	19	24	25	1.962	1.962	1.962
			262	3	3	4	0.843	0.843	0.843
		192	116	5	5	5	2.460	2.460	2.460
			148	3	3	3	2.521	2.521	2.521
			167	11	11	11	1.776	1.776	1.776
			174	5	5	5	1.566	1.566	1.566
			189	116	117	117	3.077	3.077	3.077
			197	4	4	4	2.587	2.587	2.587
			223	10	11	11	1.773	1.773	1.773
			228	19	20	20	1.792	1.792	1.792
			262	5	7	7	2.370	2.370	2.370
		197	167	2	2	2	1.445	1.445	1.445
			174	2	2	3	1.094	1.094	1.094
			189	43	44	44	4.653	4.653	4.653
			192	4	4	4	2.587	2.587	2.587
			223	11	12	12	1.362	1.362	1.362
			228	12	12	12	16.731	16.731	16.731
		223	116		3	3		0.472	0.472
			167		3	3		0.451	0.451
			189	17	20	20	2.410	2.410	2.410
			192	10	11	11	1.773	1.773	1.773
			197	11	12	12	1.362	1.362	1.362
			228	31	33	33	1.529	1.529	1.529
			262	2	5	7	0.494	0.494	0.494
		228	116	10	13	13	1.027	1.027	1.027
			123	3	3	3	1.007	1.007	1.007
			167		3	3		0.642	0.642
			189	19	24	25	1.962	1.962	1.962
			192	19	20	20	1.792	1.792	1.792
			197	12	12	12	16.731	16.731	16.731
			223	31	33	33	1.529	1.529	1.529
			262	25	34	35	1.880	1.880	1.880
		262	116	3	4	4	0.480	0.480	0.480
			189	3	3	4	0.843	0.843	0.843
			192	5	7	7	2.370	2.370	2.370
			223	2	5	7	0.494	0.494	0.494
			228	25	34	35	1.880	1.880	1.880
PRP17	O60508	177	203		2	3		0.601	0.601
		203	177		2	3		0.601	0.601
		228	333	24	27	27	2.744	2.744	2.744
		233	333	13	13	13	2.236	2.236	2.236
		238	312	13	31	34	2.437	2.437	2.437
			314	7	7	7	1.970	1.970	1.970
		289	312		2	2		0.776	0.776
			333	85	93	93	2.109	2.109	2.109
			551	63	64	64	5.238	5.238	5.238
		312	238	13	31	34	2.437	2.437	2.437
			289		2	2		0.776	0.776
		314	238	7	7	7	1.970	1.970	1.970
		333	228	24	27	27	2.744	2.744	2.744

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			233	13	13	13	2.236	2.236	2.236
			289	85	93	93	2.109	2.109	2.109
			374	1	1	1	1.471	1.471	1.471
		374	333	1	1	1	1.471	1.471	1.471
			398	26	27	27	2.362	2.362	2.362
		398	374	26	27	27	2.362	2.362	2.362
		531	547	6	6	6	1.746	1.746	1.746
		547	531	6	6	6	1.746	1.746	1.746
		551	289	63	64	64	5.238	5.238	5.238
PRP19	Q9UMS4	1	425		2	3		0.353	0.353
		64	76	3	3	3	3.200	3.200	3.200
		76	64	3	3	3	3.200	3.200	3.200
		122	135	27	27	27	7.016	7.016	7.016
			179	55	58	58	2.989	2.989	2.989
			192	71	71	71	3.369	3.369	3.369
			244	86	86	86	3.657	3.657	3.657
			261	40	41	41	3.177	3.177	3.177
			265	34	36	36	2.689	2.689	2.689
			266	57	59	59	3.747	3.747	3.747
			380	28	30	30	3.100	3.100	3.100
			423	1	1	1	1.254	1.254	1.254
			425	34	36	36	1.519	1.519	1.519
			428	7	7	7	2.460	2.460	2.460
		135	122	27	27	27	7.016	7.016	7.016
			135	4	4	6	2.352	2.352	2.352
			179	14	14	14	2.662	2.662	2.662
			192	19	19	19	15.970	15.970	15.970
			244	50	51	51	12.810	12.810	12.810
			261	11	12	12	12.714	12.714	12.714
			265	5	10	10	0.892	0.892	0.892
			266	12	12	12	8.408	8.408	8.408
			380	4	7	7	1.314	1.314	1.314
			423	14	14	14	2.329	2.329	2.329
			425	19	25	25	5.991	5.991	5.991
		179	122	55	58	58	2.989	2.989	2.989
			135	14	14	14	2.662	2.662	2.662
			192	183	195	195	5.759	5.759	5.759
			244	86	88	88	4.331	4.331	4.331
			261	6	6	6	2.990	2.990	2.990
			265	31	32	32	1.507	1.507	1.507
			266	19	19	19	4.044	4.044	4.044
			425	11	11	11	1.258	1.258	1.258
		189	200	72	143	154	1.525	1.525	1.525
			425	25	28	28	1.460	1.460	1.460
		192	122	71	71	71	3.369	3.369	3.369
			135	19	19	19	15.970	15.970	15.970
			179	183	195	195	5.759	5.759	5.759
			244	38	38	38	3.295	3.295	3.295
			261	10	10	10	2.775	2.775	2.775
			265	15	15	15	1.413	1.413	1.413
			266	12	12	12	4.817	4.817	4.817
			423	59	59	60	2.720	2.720	2.720
			425	53	71	71	2.482	2.482	2.482
			463	81	81	81	30.792	30.792	30.792
		200	189	72	143	154	1.525	1.525	1.525
			463	15	15	15	5.716	5.716	5.716
		206	244		2	2		0.717	0.717
		244	122	86	86	86	3.657	3.657	3.657
			135	50	51	51	12.810	12.810	12.810
			179	86	88	88	4.331	4.331	4.331
			192	38	38	38	3.295	3.295	3.295
			206		2	2		0.717	0.717
			244			6			0.213
			261	359	361	361	5.387	5.387	5.387
			265	802	874	877	10.209	10.209	10.209
			266	191	195	195	22.729	22.729	22.729
			380	8	8	8	1.356	1.356	1.356
			425	2	2	2	0.509	0.509	0.509
		261	122	40	41	41	3.177	3.177	3.177
			135	11	12	12	12.714	12.714	12.714
			179	6	6	6	2.990	2.990	2.990
			192	10	10	10	2.775	2.775	2.775
			244	359	361	361	5.387	5.387	5.387
			265		3	3		0.487	0.487
			266	17	17	17	2.313	2.313	2.313
		265	122	34	36	36	2.689	2.689	2.689
			135	5	10	10	0.892	0.892	0.892
			179	31	32	32	1.507	1.507	1.507
			192	15	15	15	1.413	1.413	1.413
			244	802	874	877	10.209	10.209	10.209
			261		3	3		0.487	0.487
			266	31	36	37	1.296	1.296	1.296
		266	122	57	59	59	3.747	3.747	3.747

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			135	12	12	12	8.408	8.408	8.408
			179	19	19	19	4.044	4.044	4.044
			192	12	12	12	4.817	4.817	4.817
			244	191	195	195	22.729	22.729	22.729
			261	17	17	17	2.313	2.313	2.313
			265	31	36	37	1.296	1.296	1.296
			266	16	18	18	1.357	1.357	1.357
		380	122	28	30	30	3.100	3.100	3.100
			135	4	7	7	1.314	1.314	1.314
			244	8	8	8	1.356	1.356	1.356
		417	423	2	2	2	0.469	0.469	0.469
			425	23	32	33	1.162	1.162	1.162
		423	122	1	1	1	1.254	1.254	1.254
			135	14	14	14	2.329	2.329	2.329
			192	59	59	60	2.720	2.720	2.720
			417	2	2	2	0.469	0.469	0.469
		425	1		2	3		0.353	0.353
			122	34	36	36	1.519	1.519	1.519
			135	19	25	25	5.991	5.991	5.991
			179	11	11	11	1.258	1.258	1.258
			189	25	28	28	1.460	1.460	1.460
			192	53	71	71	2.482	2.482	2.482
			244	2	2	2	0.509	0.509	0.509
			417	23	32	33	1.162	1.162	1.162
			463	2	2	2	0.717	0.717	0.717
		428	122	7	7	7	2.460	2.460	2.460
		463	192	81	81	81	30.792	30.792	30.792
			200	15	15	15	5.716	5.716	5.716
			425	2	2	2	0.717	0.717	0.717
PRP38	Q8NAV1	92	118	64	67	67	3.744	3.744	3.744
		97	131	67	68	68	2.837	2.837	2.837
		118	92	64	67	67	3.744	3.744	3.744
		131	97	67	68	68	2.837	2.837	2.837
PRP4B	Q13523	111	117	4	4	4	3.177	3.177	3.177
		117	111	4	4	4	3.177	3.177	3.177
		168	177	15	26	27	1.127	1.127	1.127
			181	3	6	6	1.924	1.924	1.924
		170	181	8	13	15	0.680	0.680	0.680
		177	168	15	26	27	1.127	1.127	1.127
		181	168	3	6	6	1.924	1.924	1.924
			170	8	13	15	0.680	0.680	0.680
			202		2	2		0.294	0.294
		198	200	2	3	3	1.076	1.076	1.076
			202	12	16	16	1.837	1.837	1.837
			209	4	5	5	0.772	0.772	0.772
		199	202	6	6	6	1.549	1.549	1.549
		200	198	2	3	3	1.076	1.076	1.076
		202	181		2	2		0.294	0.294
			198	12	16	16	1.837	1.837	1.837
			199	6	6	6	1.549	1.549	1.549
			211	22	23	23	1.839	1.839	1.839
		209	198	4	5	5	0.772	0.772	0.772
		211	202	22	23	23	1.839	1.839	1.839
		254	268	4	4	4	1.867	1.867	1.867
		262	990		2	2		0.152	0.152
		268	254	4	4	4	1.867	1.867	1.867
		393	625		2	2		0.571	0.571
		551	915	6	6	6	1.790	1.790	1.790
		593	747	10	10	11	1.921	1.921	1.921
		608	625	14	14	14	2.239	2.239	2.239
			626	19	20	20	3.950	3.950	3.950
			656	3	3	3	1.611	1.611	1.611
			780	61	67	67	2.625	2.625	2.625
			783	76	76	76	3.752	3.752	3.752
			790	13	13	13	2.596	2.596	2.596
			827	2	2	2	0.659	0.659	0.659
		625	393		2	2		0.571	0.571
			608	14	14	14	2.239	2.239	2.239
			656	2	3	3	1.098	1.098	1.098
			659		2	2		1.478	1.478
			783	17	17	17	3.014	3.014	3.014
			892	23	24	24	1.581	1.581	1.581
			936	3	3	3	1.545	1.545	1.545
		626	608	19	20	20	3.950	3.950	3.950
			656	5	5	5	1.217	1.217	1.217
			659	6	6	6	1.886	1.886	1.886
			780	5	5	5	1.856	1.856	1.856
			783	23	23	23	3.470	3.470	3.470
			892	27	27	27	2.783	2.783	2.783
			936	4	4	4	3.500	3.500	3.500
		656	608	3	3	3	1.611	1.611	1.611
			625	2	3	3	1.098	1.098	1.098
			626	5	5	5	1.217	1.217	1.217

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			685	8	8	8	1.271	1.271	1.271
			717	25	27	27	2.579	2.579	2.579
			727	7	9	9	1.073	1.073	1.073
			760	6	6	6	1.430	1.430	1.430
			780	18	18	18	1.609	1.609	1.609
			783	25	25	25	2.834	2.834	2.834
			817	3	4	4	1.275	1.275	1.275
			892	44	45	45	2.624	2.624	2.624
			936	6	7	8	3.261	3.261	3.261
		659	625		2	2		1.478	1.478
			626	6	6	6	1.886	1.886	1.886
			685	5	6	6	0.825	0.825	0.825
			717	6	6	6	1.341	1.341	1.341
			760	11	11	11	1.803	1.803	1.803
			780	16	16	16	2.571	2.571	2.571
			783	13	13	13	2.359	2.359	2.359
			892	37	38	38	2.129	2.129	2.129
			936	3	3	3	2.301	2.301	2.301
		685	656	8	8	8	1.271	1.271	1.271
			659	5	6	6	0.825	0.825	0.825
			737		2	2		0.341	0.341
			760	98	114	116	2.944	2.944	2.944
		717	656	25	27	27	2.579	2.579	2.579
			659	6	6	6	1.341	1.341	1.341
		727	656	7	9	9	1.073	1.073	1.073
			737	3	3	3	1.830	1.830	1.830
			738	22	22	22	2.495	2.495	2.495
		731	738	18	18	18	4.065	4.065	4.065
		737	685		2	2		0.341	0.341
			727	3	3	3	1.830	1.830	1.830
		738	727	22	22	22	2.495	2.495	2.495
			731	18	18	18	4.065	4.065	4.065
		747	593	10	10	11	1.921	1.921	1.921
			804	44	47	47	2.179	2.179	2.179
		760	656	6	6	6	1.430	1.430	1.430
			659	11	11	11	1.803	1.803	1.803
			685	98	114	116	2.944	2.944	2.944
		780	608	61	67	67	2.625	2.625	2.625
			626	5	5	5	1.856	1.856	1.856
			656	18	18	18	1.609	1.609	1.609
			659	16	16	16	2.571	2.571	2.571
			827	24	25	26	1.344	1.344	1.344
		783	608	76	76	76	3.752	3.752	3.752
			625	17	17	17	3.014	3.014	3.014
			626	23	23	23	3.470	3.470	3.470
			656	25	25	25	2.834	2.834	2.834
			659	13	13	13	2.359	2.359	2.359
			892	13	14	14	2.007	2.007	2.007
		790	608	13	13	13	2.596	2.596	2.596
			827	7	8	8	2.209	2.209	2.209
			972	2	2	2	2.587	2.587	2.587
			1006	33	33	33	4.463	4.463	4.463
		804	747	44	47	47	2.179	2.179	2.179
			817	6	7	7	2.415	2.415	2.415
			1006	3	3	3	1.948	1.948	1.948
		817	656	3	4	4	1.275	1.275	1.275
			804	6	7	7	2.415	2.415	2.415
		827	608	2	2	2	0.659	0.659	0.659
			780	24	25	26	1.344	1.344	1.344
			790	7	8	8	2.209	2.209	2.209
		864	911	2	2	2	1.437	1.437	1.437
			915	17	17	17	3.640	3.640	3.640
		892	625	23	24	24	1.581	1.581	1.581
			626	27	27	27	2.783	2.783	2.783
			656	44	45	45	2.624	2.624	2.624
			659	37	38	38	2.129	2.129	2.129
			783	13	14	14	2.007	2.007	2.007
			936	120	120	120	3.438	3.438	3.438
			942	35	35	35	3.135	3.135	3.135
		905	911	3	3	3	1.347	1.347	1.347
		907	990	10	10	10	2.138	2.138	2.138
		911	864	2	2	2	1.437	1.437	1.437
			905	3	3	3	1.347	1.347	1.347
			915	149	153	153	8.649	8.649	8.649
			990	140	151	152	2.829	2.829	2.829
		915	551	6	6	6	1.790	1.790	1.790
			864	17	17	17	3.640	3.640	3.640
			911	149	153	153	8.649	8.649	8.649
			990	104	106	106	5.628	5.628	5.628
		936	625	3	3	3	1.545	1.545	1.545
			626	4	4	4	3.500	3.500	3.500
			656	6	7	8	3.261	3.261	3.261
			659	3	3	3	2.301	2.301	2.301

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			892	120	120	120	3.438	3.438	3.438
			942	23	24	24	2.688	2.688	2.688
		942	892	35	35	35	3.135	3.135	3.135
			936	23	24	24	2.688	2.688	2.688
		972	790	2	2	2	2.587	2.587	2.587
		990	262		2	2		0.152	0.152
			907	10	10	10	2.138	2.138	2.138
			911	140	151	152	2.829	2.829	2.829
			915	104	106	106	5.628	5.628	5.628
		1006	790	33	33	33	4.463	4.463	4.463
			804	3	3	3	1.948	1.948	1.948
PRP8	Q6P2Q9	29	36	77	77	77	4.423	4.423	4.423
		33	36	10	10	10	2.368	2.368	2.368
		36	29	77	77	77	4.423	4.423	4.423
			33	10	10	10	2.368	2.368	2.368
		43	48	25	35	35	1.173	1.173	1.173
			50	22	23	23	2.397	2.397	2.397
		48	43	25	35	35	1.173	1.173	1.173
			50	10	16	17	0.907	0.907	0.907
			58		2	2		0.374	0.374
		50	43	22	23	23	2.397	2.397	2.397
			48	10	16	17	0.907	0.907	0.907
			480	2	2	2	2.068	2.068	2.068
		58	48		2	2		0.374	0.374
		85	93	2	2	2	0.599	0.599	0.599
			480		2	2		1.055	1.055
		93	85	2	2	2	0.599	0.599	0.599
			666	13	14	14	1.836	1.836	1.836
		101	744		2	2		0.163	0.163
			773	4	5	5	1.057	1.057	1.057
		278	449	3	3	3	1.372	1.372	1.372
		442	609	53	53	53	3.316	3.316	3.316
		449	278	3	3	3	1.372	1.372	1.372
		480	50	2	2	2	2.068	2.068	2.068
			85		2	2		1.055	1.055
		511	533	2	4	4	0.780	0.780	0.780
			666	27	29	29	2.780	2.780	2.780
			670	5	5	5	1.443	1.443	1.443
		533	511	2	4	4	0.780	0.780	0.780
		606	1838		2	2		0.385	0.385
		609	442	53	53	53	3.316	3.316	3.316
			623	4	4	4	2.672	2.672	2.672
		623	609	4	4	4	2.672	2.672	2.672
		666	93	13	14	14	1.836	1.836	1.836
			511	27	29	29	2.780	2.780	2.780
			674	6	6	6	1.973	1.973	1.973
		670	511	5	5	5	1.443	1.443	1.443
		674	666	6	6	6	1.973	1.973	1.973
			774	30	30	31	4.083	4.083	4.083
		721	746	59	61	61	3.080	3.080	3.080
			774	43	43	43	4.032	4.032	4.032
			1020	10	10	10	2.541	2.541	2.541
		727	744	2	3	3	1.038	1.038	1.038
		744	101		2	2		0.163	0.163
			727	2	3	3	1.038	1.038	1.038
		746	721	59	61	61	3.080	3.080	3.080
			774	102	105	105	3.541	3.541	3.541
		769	774	15	16	17	2.219	2.219	2.219
			853	82	89	89	3.821	3.821	3.821
			1020	5	5	5	1.332	1.332	1.332
			1262	89	90	90	2.299	2.299	2.299
			1449	100	100	100	3.261	3.261	3.261
			2140		2	2		0.352	0.352
		773	101	4	5	5	1.057	1.057	1.057
			853	21	29	30	1.038	1.038	1.038
			1020	46	52	52	1.462	1.462	1.462
			1262	20	20	21	1.300	1.300	1.300
			1449	22	24	24	1.537	1.537	1.537
		774	674	30	30	31	4.083	4.083	4.083
			721	43	43	43	4.032	4.032	4.032
			746	102	105	105	3.541	3.541	3.541
			769	15	16	17	2.219	2.219	2.219
			1020	134	139	140	3.037	3.037	3.037
		833	932	29	29	29	1.912	1.912	1.912
		853	769	82	89	89	3.821	3.821	3.821
			773	21	29	30	1.038	1.038	1.038
			1262	2	2	2	0.985	0.985	0.985
			1320	55	63	64	2.393	2.393	2.393
			1449	88	92	92	2.505	2.505	2.505
			1463	37	44	44	2.032	2.032	2.032
			1505	64	71	72	2.277	2.277	2.277
		882	892	26	27	27	3.186	3.186	3.186
		892	882	26	27	27	3.186	3.186	3.186

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		932	833	29	29	29	1.912	1.912	1.912
		983	1158	89	94	94	4.846	4.846	4.846
		987	1158	99	105	105	3.960	3.960	3.960
		1020	721	10	10	10	2.541	2.541	2.541
			769	5	5	5	1.332	1.332	1.332
			773	46	52	52	1.462	1.462	1.462
			774	134	139	140	3.037	3.037	3.037
		1131	1158	2	3	3	1.239	1.239	1.239
		1132	1344	2	2	2	2.554	2.554	2.554
			1463			2			0.113
			2113	3	7	7	0.449	0.449	0.449
		1144	1180	21	21	21	3.530	3.530	3.530
		1158	983	89	94	94	4.846	4.846	4.846
			987	99	105	105	3.960	3.960	3.960
			1131	2	3	3	1.239	1.239	1.239
		1180	1144	21	21	21	3.530	3.530	3.530
		1262	769	89	90	90	2.299	2.299	2.299
			773	20	20	21	1.300	1.300	1.300
			853	2	2	2	0.985	0.985	0.985
		1290	1449	2	3	3	0.567	0.567	0.567
		1320	853	55	63	64	2.393	2.393	2.393
			1463	80	85	86	3.076	3.076	3.076
		1344	1132	2	2	2	2.554	2.554	2.554
			1735	135	135	135	4.372	4.372	4.372
			1838	9	10	10	1.497	1.497	1.497
		1449	769	100	100	100	3.261	3.261	3.261
			773	22	24	24	1.537	1.537	1.537
			853	88	92	92	2.505	2.505	2.505
			1290	2	3	3	0.567	0.567	0.567
			1463	42	43	43	3.956	3.956	3.956
		1463	853	37	44	44	2.032	2.032	2.032
			1132			2			0.113
			1320	80	85	86	3.076	3.076	3.076
			1449	42	43	43	3.956	3.956	3.956
			1505	2	2	2	1.369	1.369	1.369
		1505	853	64	71	72	2.277	2.277	2.277
			1463	2	2	2	1.369	1.369	1.369
		1735	1344	135	135	135	4.372	4.372	4.372
			2244	66	67	67	2.214	2.214	2.214
			2249	61	62	62	3.134	3.134	3.134
		1801	1831	2	4	4	1.044	1.044	1.044
		1820	1840	24	24	24	3.458	3.458	3.458
		1831	1801	2	4	4	1.044	1.044	1.044
			1838	20	21	21	1.213	1.213	1.213
			1840	34	35	35	3.524	3.524	3.524
		1838	606		2	2		0.385	0.385
			1344	9	10	10	1.497	1.497	1.497
			1831	20	21	21	1.213	1.213	1.213
			1866	52	53	53	4.246	4.246	4.246
			2031	6	6	6	0.785	0.785	0.785
			2034	34	34	34	2.032	2.032	2.032
			2049	2	2	2	1.078	1.078	1.078
			2098	2	2	2	1.503	1.503	1.503
			2249	8	10	10	1.217	1.217	1.217
		1840	1820	24	24	24	3.458	3.458	3.458
			1831	34	35	35	3.524	3.524	3.524
			1993		3	3		0.282	0.282
		1859	1885	3	3	3	2.506	2.506	2.506
			2034	1	2	2	1.825	1.825	1.825
		1866	1838	52	53	53	4.246	4.246	4.246
			2034	6	6	6	4.164	4.164	4.164
		1885	1859	3	3	3	2.506	2.506	2.506
		1901	1955	1	1	1	1.256	1.256	1.256
		1955	1901	1	1	1	1.256	1.256	1.256
		1978	2034		4	4		1.646	1.646
			2249	5	6	6	14.929	14.929	14.929
		1984	2034	5	5	5	2.247	2.247	2.247
			2249	15	15	16	2.869	2.869	2.869
		1993	1840		3	3		0.282	0.282
		2031	1838	6	6	6	0.785	0.785	0.785
		2034	1838	34	34	34	2.032	2.032	2.032
			1859	1	2	2	1.825	1.825	1.825
			1866	6	6	6	4.164	4.164	4.164
			1978		4	4		1.646	1.646
			1984	5	5	5	2.247	2.247	2.247
			2049	12	12	12	3.082	3.082	3.082
			2070	3	3	3	2.736	2.736	2.736
		2049	1838	2	2	2	1.078	1.078	1.078
			2034	12	12	12	3.082	3.082	3.082
			2249	3	3	3	1.398	1.398	1.398
		2070	2034	3	3	3	2.736	2.736	2.736
		2098	1838	2	2	2	1.503	1.503	1.503
		2113	1132	3	7	7	0.449	0.449	0.449

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		2140	769		2	2		0.352	0.352
		2198	2244	5	6	7	1.346	1.346	1.346
		2244	1735	66	67	67	2.214	2.214	2.214
			2198	5	6	7	1.346	1.346	1.346
		2249	1735	61	62	62	3.134	3.134	3.134
			1838	8	10	10	1.217	1.217	1.217
			1978	5	6	6	14.929	14.929	14.929
			1984	15	15	16	2.869	2.869	2.869
			2049	3	3	3	1.398	1.398	1.398
RBM22	Q9NW64	76	109	2	3	3	0.568	0.568	0.568
			114		2	2		0.258	0.258
			139	4	6	6	0.799	0.799	0.799
		78	139	7	7	7	2.730	2.730	2.730
		104	114	26	26	26	3.521	3.521	3.521
			139	21	22	22	1.300	1.300	1.300
			149	48	48	48	2.925	2.925	2.925
			286	18	18	18	2.528	2.528	2.528
		109	76	2	3	3	0.568	0.568	0.568
			139	12	14	14	2.186	2.186	2.186
			149	12	13	13	1.352	1.352	1.352
			286	7	8	9	2.085	2.085	2.085
		114	76		2	2		0.258	0.258
			104	26	26	26	3.521	3.521	3.521
			139	19	19	20	2.363	2.363	2.363
			149	8	10	10	3.178	3.178	3.178
			286	19	20	20	3.425	3.425	3.425
		139	76	4	6	6	0.799	0.799	0.799
			78	7	7	7	2.730	2.730	2.730
			104	21	22	22	1.300	1.300	1.300
			109	12	14	14	2.186	2.186	2.186
			114	19	19	20	2.363	2.363	2.363
			229	3	4	4	2.232	2.232	2.232
			286	14	15	15	1.531	1.531	1.531
			290	3	4	4	1.314	1.314	1.314
		149	104	48	48	48	2.925	2.925	2.925
			109	12	13	13	1.352	1.352	1.352
			114	8	10	10	3.178	3.178	3.178
			158	5	9	9	1.131	1.131	1.131
			286	11	11	11	2.547	2.547	2.547
			290	8	9	9	1.891	1.891	1.891
			301	10	11	11	1.918	1.918	1.918
		158	149	5	9	9	1.131	1.131	1.131
			212		2	2		0.782	0.782
			290	4	5	5	1.299	1.299	1.299
		212	158		2	2		0.782	0.782
		229	139	3	4	4	2.232	2.232	2.232
		286	104	18	18	18	2.528	2.528	2.528
			109	7	8	9	2.085	2.085	2.085
			114	19	20	20	3.425	3.425	3.425
			139	14	15	15	1.531	1.531	1.531
			149	11	11	11	2.547	2.547	2.547
		290	139	3	4	4	1.314	1.314	1.314
			149	8	9	9	1.891	1.891	1.891
			158	4	5	5	1.299	1.299	1.299
		301	149	10	11	11	1.918	1.918	1.918
			313	6	6	6	1.197	1.197	1.197
			315	3	3	3	1.361	1.361	1.361
		313	301	6	6	6	1.197	1.197	1.197
		315	301	3	3	3	1.361	1.361	1.361
RED	Q13123	1	305			2			0.169
		30	38	13	16	16	2.303	2.303	2.303
		38	30	13	16	16	2.303	2.303	2.303
			54	4	4	4	1.643	1.643	1.643
		54	38	4	4	4	1.643	1.643	1.643
		112	142		2	2		0.388	0.388
		137	151	5	9	9	0.795	0.795	0.795
		142	112		2	2		0.388	0.388
			151	7	13	14	1.221	1.221	1.221
			164	2	3	4	0.642	0.642	0.642
		151	137	5	9	9	0.795	0.795	0.795
			142	7	13	14	1.221	1.221	1.221
			173	6	6	6	1.732	1.732	1.732
		164	142	2	3	4	0.642	0.642	0.642
		173	151	6	6	6	1.732	1.732	1.732
			181	11	11	11	1.881	1.881	1.881
		181	173	11	11	11	1.881	1.881	1.881
		191	197	25	27	27	2.493	2.493	2.493
			198	34	36	36	3.358	3.358	3.358
			205	25	26	26	2.053	2.053	2.053
			209	24	24	24	2.261	2.261	2.261
			222	2	7	7	0.632	0.632	0.632
		194	198	110	117	118	4.046	4.046	4.046
			205	41	48	49	1.721	1.721	1.721

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			209	43	48	48	2.397	2.397	2.397
			222	20	30	30	1.775	1.775	1.775
			224	9	11	11	1.588	1.588	1.588
			509		2	3		0.170	0.170
		197	191	25	27	27	2.493	2.493	2.493
			222	16	19	19	1.658	1.658	1.658
		198	191	34	36	36	3.358	3.358	3.358
			194	110	117	118	4.046	4.046	4.046
			222	72	75	75	2.429	2.429	2.429
			224	22	23	23	1.992	1.992	1.992
		205	191	25	26	26	2.053	2.053	2.053
			194	41	48	49	1.721	1.721	1.721
			222	101	116	117	2.370	2.370	2.370
			224	12	12	12	2.232	2.232	2.232
		209	191	24	24	24	2.261	2.261	2.261
			194	43	48	48	2.397	2.397	2.397
			222	5	6	6	1.110	1.110	1.110
		222	191	2	7	7	0.632	0.632	0.632
			194	20	30	30	1.775	1.775	1.775
			197	16	19	19	1.658	1.658	1.658
			198	72	75	75	2.429	2.429	2.429
			205	101	116	117	2.370	2.370	2.370
			209	5	6	6	1.110	1.110	1.110
		224	194	9	11	11	1.588	1.588	1.588
			198	22	23	23	1.992	1.992	1.992
			205	12	12	12	2.232	2.232	2.232
		281	303	11	11	11	2.292	2.292	2.292
			305	4	4	4	2.310	2.310	2.310
		303	281	11	11	11	2.292	2.292	2.292
		305	1			2			0.169
			281	4	4	4	2.310	2.310	2.310
		378	386	25	28	28	1.988	1.988	1.988
			388	4	4	4	0.942	0.942	0.942
			404	4	4	4	1.206	1.206	1.206
		379	386	10	19	19	1.173	1.173	1.173
			388	4	4	5	1.238	1.238	1.238
			398	5	6	6	0.832	0.832	0.832
		386	378	25	28	28	1.988	1.988	1.988
			379	10	19	19	1.173	1.173	1.173
			398	14	18	18	1.775	1.775	1.775
			404	17	17	17	1.992	1.992	1.992
			408	7	7	8	1.890	1.890	1.890
		388	378	4	4	4	0.942	0.942	0.942
			379	4	4	5	1.238	1.238	1.238
		398	379	5	6	6	0.832	0.832	0.832
			386	14	18	18	1.775	1.775	1.775
			408	23	24	24	2.835	2.835	2.835
			413	27	27	27	14.887	14.887	14.887
			427	6	6	6	1.698	1.698	1.698
			432	4	4	4	1.492	1.492	1.492
		404	378	4	4	4	1.206	1.206	1.206
			386	17	17	17	1.992	1.992	1.992
			413	29	30	30	4.196	4.196	4.196
			427	13	13	13	1.929	1.929	1.929
			432	4	4	4	2.410	2.410	2.410
			534	2	2	2	0.476	0.476	0.476
		408	386	7	7	8	1.890	1.890	1.890
			398	23	24	24	2.835	2.835	2.835
			427	13	13	13	1.124	1.124	1.124
			428	2	3	3	0.890	0.890	0.890
			432	7	10	10	1.322	1.322	1.322
		413	398	27	27	27	14.887	14.887	14.887
			404	29	30	30	4.196	4.196	4.196
		427	398	6	6	6	1.698	1.698	1.698
			404	13	13	13	1.929	1.929	1.929
			408	13	13	13	1.124	1.124	1.124
		428	408	2	3	3	0.890	0.890	0.890
		432	398	4	4	4	1.492	1.492	1.492
			404	4	4	4	2.410	2.410	2.410
			408	7	10	10	1.322	1.322	1.322
		474	496	20	20	20	1.813	1.813	1.813
			509	13	13	13	1.519	1.519	1.519
			520	10	10	11	1.233	1.233	1.233
		475	496	23	23	23	1.931	1.931	1.931
			501	16	16	16	3.532	3.532	3.532
			509	20	21	21	2.142	2.142	2.142
			520	8	8	8	3.436	3.436	3.436
			525	2	2	2	1.241	1.241	1.241
			534	2	2	2	3.476	3.476	3.476
		496	474	20	20	20	1.813	1.813	1.813
			475	23	23	23	1.931	1.931	1.931
			509	7	7	7	1.299	1.299	1.299
		501	475	16	16	16	3.532	3.532	3.532

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		509	194		2	3		0.170	0.170
			474	13	13	13	1.519	1.519	1.519
			475	20	21	21	2.142	2.142	2.142
			496	7	7	7	1.299	1.299	1.299
			520	34	34	34	2.936	2.936	2.936
			534	3	3	3	2.471	2.471	2.471
		520	474	10	10	11	1.233	1.233	1.233
			475	8	8	8	3.436	3.436	3.436
			509	34	34	34	2.936	2.936	2.936
			533	2	3	3	0.958	0.958	0.958
			534	29	29	29	4.860	4.860	4.860
		525	475	2	2	2	1.241	1.241	1.241
			533	7	10	11	0.716	0.716	0.716
		533	520	2	3	3	0.958	0.958	0.958
			525	7	10	11	0.716	0.716	0.716
			541	22	23	23	1.751	1.751	1.751
		534	404	2	2	2	0.476	0.476	0.476
			475	2	2	2	3.476	3.476	3.476
			509	3	3	3	2.471	2.471	2.471
			520	29	29	29	4.860	4.860	4.860
			544	6	6	6	3.360	3.360	3.360
			556	3	3	3	1.192	1.192	1.192
		541	533	22	23	23	1.751	1.751	1.751
			544	42	43	43	4.222	4.222	4.222
			553	28	30	30	3.682	3.682	3.682
			556	7	9	9	0.691	0.691	0.691
		544	534	6	6	6	3.360	3.360	3.360
			541	42	43	43	4.222	4.222	4.222
			556	7	8	8	1.184	1.184	1.184
		553	541	28	30	30	3.682	3.682	3.682
		556	534	3	3	3	1.192	1.192	1.192
			541	7	9	9	0.691	0.691	0.691
			544	7	8	8	1.184	1.184	1.184
SF3A1	Q15459	1	424		2	2		0.219	0.219
		2	30	6	7	7	3.171	3.171	3.171
			55	3	3	3	2.757	2.757	2.757
			102	15	15	15	2.253	2.253	2.253
			105	2	2	2	1.161	1.161	1.161
			115	16	16	16	2.956	2.956	2.956
			131	5	5	5	2.351	2.351	2.351
		20	102	2	2	2	2.436	2.436	2.436
		30	2	6	7	7	3.171	3.171	3.171
			97	15	17	17	1.213	1.213	1.213
			102	22	26	27	1.037	1.037	1.037
			105	14	21	21	2.030	2.030	2.030
			115	10	10	10	5.294	5.294	5.294
		37	97	14	14	14	1.298	1.298	1.298
			102	14	17	17	1.368	1.368	1.368
			105	3	3	3	1.637	1.637	1.637
			115	12	12	12	2.817	2.817	2.817
		55	2	3	3	3	2.757	2.757	2.757
			80	153	157	158	6.078	6.078	6.078
			97	84	85	85	3.144	3.144	3.144
			102	4	4	4	3.018	3.018	3.018
			105	127	129	129	3.230	3.230	3.230
			115	38	38	38	2.722	2.722	2.722
			131	7	7	7	2.132	2.132	2.132
		80	55	153	157	158	6.078	6.078	6.078
			115	31	31	31	5.066	5.066	5.066
			131	3	3	3	3.311	3.311	3.311
		97	30	15	17	17	1.213	1.213	1.213
			37	14	14	14	1.298	1.298	1.298
			55	84	85	85	3.144	3.144	3.144
			105	56	65	65	2.942	2.942	2.942
			115	8	11	12	1.638	1.638	1.638
		102	2	15	15	15	2.253	2.253	2.253
			20	2	2	2	2.436	2.436	2.436
			30	22	26	27	1.037	1.037	1.037
			37	14	17	17	1.368	1.368	1.368
			55	4	4	4	3.018	3.018	3.018
			115	44	45	45	1.993	1.993	1.993
			131	14	15	15	1.539	1.539	1.539
			188		3	4		0.353	0.353
		105	2	2	2	2	1.161	1.161	1.161
			30	14	21	21	2.030	2.030	2.030
			37	3	3	3	1.637	1.637	1.637
			55	127	129	129	3.230	3.230	3.230
			97	56	65	65	2.942	2.942	2.942
			131	3	3	3	1.509	1.509	1.509
		115	2	16	16	16	2.956	2.956	2.956
			30	10	10	10	5.294	5.294	5.294
			37	12	12	12	2.817	2.817	2.817
			55	38	38	38	2.722	2.722	2.722

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			80	31	31	31	5.066	5.066	5.066
			97	8	11	12	1.638	1.638	1.638
			102	44	45	45	1.993	1.993	1.993
		131	2	5	5	5	2.351	2.351	2.351
			55	7	7	7	2.132	2.132	2.132
			80	3	3	3	3.311	3.311	3.311
			102	14	15	15	1.539	1.539	1.539
			105	3	3	3	1.509	1.509	1.509
			223	2	3	3	1.193	1.193	1.193
			230	3	5	5	1.588	1.588	1.588
			231	25	25	25	2.504	2.504	2.504
			251	10	10	10	1.271	1.271	1.271
			258	4	4	4	1.632	1.632	1.632
			259	32	32	32	2.027	2.027	2.027
		144	251	30	32	32	1.678	1.678	1.678
			259	26	27	27	9.195	9.195	9.195
		188	102		3	4		0.353	0.353
		223	131	2	3	3	1.193	1.193	1.193
			230	32	34	34	1.705	1.705	1.705
			231	268	277	277	2.541	2.541	2.541
		228	231	47	47	47	2.972	2.972	2.972
		230	131	3	5	5	1.588	1.588	1.588
			223	32	34	34	1.705	1.705	1.705
		231	131	25	25	25	2.504	2.504	2.504
			223	268	277	277	2.541	2.541	2.541
			228	47	47	47	2.972	2.972	2.972
		251	131	10	10	10	1.271	1.271	1.271
			144	30	32	32	1.678	1.678	1.678
			258	146	146	146	4.306	4.306	4.306
			259	40	46	46	3.412	3.412	3.412
		258	131	4	4	4	1.632	1.632	1.632
			251	146	146	146	4.306	4.306	4.306
		259	131	32	32	32	2.027	2.027	2.027
			144	26	27	27	9.195	9.195	9.195
			251	40	46	46	3.412	3.412	3.412
		391	399	30	30	30	2.674	2.674	2.674
			419	45	52	53	2.195	2.195	2.195
			424	44	49	49	2.083	2.083	2.083
			449	6	6	6	2.085	2.085	2.085
		396	424	27	30	30	2.102	2.102	2.102
			449	2	2	2	2.132	2.132	2.132
		399	391	30	30	30	2.674	2.674	2.674
			424	76	80	80	3.466	3.466	3.466
			449	13	14	14	4.494	4.494	4.494
			499	2	2	2	4.232	4.232	4.232
		419	391	45	52	53	2.195	2.195	2.195
			449	9	9	9	2.412	2.412	2.412
			467	13	13	13	2.742	2.742	2.742
			486	5	6	6	2.719	2.719	2.719
			487	9	9	9	2.325	2.325	2.325
			499	2	2	2	2.809	2.809	2.809
			533	5	5	6	1.129	1.129	1.129
		424	1		2	2		0.219	0.219
			391	44	49	49	2.083	2.083	2.083
			396	27	30	30	2.102	2.102	2.102
			399	76	80	80	3.466	3.466	3.466
			449	52	52	52	3.223	3.223	3.223
			467	5	5	5	2.034	2.034	2.034
			486	7	7	7	2.528	2.528	2.528
			487	26	26	26	3.960	3.960	3.960
			499	20	20	20	2.961	2.961	2.961
			533	18	18	18	2.729	2.729	2.729
		449	391	6	6	6	2.085	2.085	2.085
			396	2	2	2	2.132	2.132	2.132
			399	13	14	14	4.494	4.494	4.494
			419	9	9	9	2.412	2.412	2.412
			424	52	52	52	3.223	3.223	3.223
			486	4	4	4	2.920	2.920	2.920
			487	8	8	8	2.604	2.604	2.604
			499	2	2	2	3.064	3.064	3.064
			533	7	7	7	1.635	1.635	1.635
		467	419	13	13	13	2.742	2.742	2.742
			424	5	5	5	2.034	2.034	2.034
			486	20	22	22	4.000	4.000	4.000
			487	11	11	11	4.564	4.564	4.564
		486	419	5	6	6	2.719	2.719	2.719
			424	7	7	7	2.528	2.528	2.528
			449	4	4	4	2.920	2.920	2.920
			467	20	22	22	4.000	4.000	4.000
			495	46	48	48	3.431	3.431	3.431
			499	50	51	51	23.334	23.334	23.334
			533	16	16	16	2.957	2.957	2.957
		487	419	9	9	9	2.325	2.325	2.325

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			424	26	26	26	3.960	3.960	3.960
			449	8	8	8	2.604	2.604	2.604
			467	11	11	11	4.564	4.564	4.564
			531	8	8	8	3.471	3.471	3.471
			533	22	22	22	3.433	3.433	3.433
		495	486	46	48	48	3.431	3.431	3.431
			533	13	14	14	2.777	2.777	2.777
		499	399	2	2	2	4.232	4.232	4.232
			419	2	2	2	2.809	2.809	2.809
			424	20	20	20	2.961	2.961	2.961
			449	2	2	2	3.064	3.064	3.064
			486	50	51	51	23.334	23.334	23.334
			531	9	9	9	3.232	3.232	3.232
			533	66	66	66	3.889	3.889	3.889
			542	4	5	5	0.638	0.638	0.638
		531	487	8	8	8	3.471	3.471	3.471
			499	9	9	9	3.232	3.232	3.232
		533	419	5	5	6	1.129	1.129	1.129
			424	18	18	18	2.729	2.729	2.729
			449	7	7	7	1.635	1.635	1.635
			486	16	16	16	2.957	2.957	2.957
			487	22	22	22	3.433	3.433	3.433
			495	13	14	14	2.777	2.777	2.777
			499	66	66	66	3.889	3.889	3.889
		542	499	4	5	5	0.638	0.638	0.638
		717	754	3	3	3	0.708	0.708	0.708
		754	717	3	3	3	0.708	0.708	0.708
SF3A2	Q15428	10	77	5	6	6	2.107	2.107	2.107
			88	3	3	3	1.025	1.025	1.025
			91	16	16	16	2.894	2.894	2.894
			101	2	2	2	0.723	0.723	0.723
		77	10	5	6	6	2.107	2.107	2.107
		88	10	3	3	3	1.025	1.025	1.025
		91	10	16	16	16	2.894	2.894	2.894
			103	2	2	2	1.002	1.002	1.002
		101	10	2	2	2	0.723	0.723	0.723
			108	43	46	46	2.539	2.539	2.539
		103	91	2	2	2	1.002	1.002	1.002
			108	31	33	33	2.910	2.910	2.910
		108	101	43	46	46	2.539	2.539	2.539
			103	31	33	33	2.910	2.910	2.910
		190	213	14	15	15	2.330	2.330	2.330
		194	216	7	7	7	2.554	2.554	2.554
		213	190	14	15	15	2.330	2.330	2.330
		216	194	7	7	7	2.554	2.554	2.554
SF3A3	Q12874	24	115	3	3	3	2.216	2.216	2.216
		89	97	54	59	59	1.817	1.817	1.817
		92	97	3	5	6	1.761	1.761	1.761
			463	10	11	11	1.097	1.097	1.097
		97	89	54	59	59	1.817	1.817	1.817
			92	3	5	6	1.761	1.761	1.761
			463	10	10	10	1.335	1.335	1.335
		115	24	3	3	3	2.216	2.216	2.216
		153	306	2	3	3	0.463	0.463	0.463
		175	306		4	4		0.395	0.395
			308	10	10	10	1.833	1.833	1.833
		264	303	26	26	26	2.626	2.626	2.626
			306	10	12	12	0.993	0.993	0.993
			308	5	5	5	2.010	2.010	2.010
		291	308	4	5	5	1.834	1.834	1.834
		303	264	26	26	26	2.626	2.626	2.626
			308	9	10	10	1.040	1.040	1.040
		306	153	2	3	3	0.463	0.463	0.463
			175		4	4		0.395	0.395
			264	10	12	12	0.993	0.993	0.993
		308	175	10	10	10	1.833	1.833	1.833
			264	5	5	5	2.010	2.010	2.010
			291	4	5	5	1.834	1.834	1.834
			303	9	10	10	1.040	1.040	1.040
		463	92	10	11	11	1.097	1.097	1.097
			97	10	10	10	1.335	1.335	1.335
		489	496		2	2		0.600	0.600
		496	489		2	2		0.600	0.600
SF3B1	O75533	3	21	22	22	22	5.588	5.588	5.588
			22	13	13	13	2.162	2.162	2.162
			81	6	6	6	1.369	1.369	1.369
			943	24	28	30	2.492	2.492	2.492
		6	21	12	14	14	2.284	2.284	2.284
			22	21	21	21	4.158	4.158	4.158
			858	12	12	12	3.801	3.801	3.801
		21	3	22	22	22	5.588	5.588	5.588
			6	12	14	14	2.284	2.284	2.284
			943	5	7	7	1.820	1.820	1.820

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		22	3	13	13	13	2.162	2.162	2.162
			6	21	21	21	4.158	4.158	4.158
			943	13	13	13	1.526	1.526	1.526
		81	3	6	6	6	1.369	1.369	1.369
			120	7	8	8	1.596	1.596	1.596
			163	2	2	2	1.268	1.268	1.268
		111	866	8	8	8	12.335	12.335	12.335
		120	81	7	8	8	1.596	1.596	1.596
			163	13	14	14	2.885	2.885	2.885
			177	7	7	7	1.725	1.725	1.725
		141	163	5	8	8	0.769	0.769	0.769
			175		3	3		0.384	0.384
			177	2	2	2	0.665	0.665	0.665
		146	163	2	2	2	1.995	1.995	1.995
			175	5	8	8	1.073	1.073	1.073
			177	5	5	5	2.185	2.185	2.185
		163	81	2	2	2	1.268	1.268	1.268
			120	13	14	14	2.885	2.885	2.885
			141	5	8	8	0.769	0.769	0.769
			146	2	2	2	1.995	1.995	1.995
			175	15	16	17	1.881	1.881	1.881
			177	7	8	8	1.210	1.210	1.210
			195	5	6	6	0.927	0.927	0.927
			214	11	11	11	3.388	3.388	3.388
			656	31	31	31	2.540	2.540	2.540
		175	141		3	3		0.384	0.384
			146	5	8	8	1.073	1.073	1.073
			163	15	16	17	1.881	1.881	1.881
			182	71	75	75	2.555	2.555	2.555
			195	9	11	11	1.966	1.966	1.966
			214	10	12	12	1.698	1.698	1.698
		177	120	7	7	7	1.725	1.725	1.725
			141	2	2	2	0.665	0.665	0.665
			146	5	5	5	2.185	2.185	2.185
			163	7	8	8	1.210	1.210	1.210
			195	11	13	13	1.283	1.283	1.283
			213	2	2	2	0.768	0.768	0.768
			214	4	4	4	2.467	2.467	2.467
		182	175	71	75	75	2.555	2.555	2.555
		195	163	5	6	6	0.927	0.927	0.927
			175	9	11	11	1.966	1.966	1.966
			177	11	13	13	1.283	1.283	1.283
			213	11	11	11	2.419	2.419	2.419
			214	31	36	36	2.641	2.641	2.641
			240	15	15	15	3.087	3.087	3.087
			252	2	2	2	0.920	0.920	0.920
		213	177	2	2	2	0.768	0.768	0.768
			195	11	11	11	2.419	2.419	2.419
			240	29	29	29	2.548	2.548	2.548
		214	163	11	11	11	3.388	3.388	3.388
			175	10	12	12	1.698	1.698	1.698
			177	4	4	4	2.467	2.467	2.467
			195	31	36	36	2.641	2.641	2.641
			240	40	42	42	4.631	4.631	4.631
			252	19	19	19	7.320	7.320	7.320
		240	195	15	15	15	3.087	3.087	3.087
			213	29	29	29	2.548	2.548	2.548
			214	40	42	42	4.631	4.631	4.631
			499		4	4		0.467	0.467
		252	195	2	2	2	0.920	0.920	0.920
			214	19	19	19	7.320	7.320	7.320
		298	333	2	2	2	1.038	1.038	1.038
		333	298	2	2	2	1.038	1.038	1.038
		493	496	20	26	26	1.866	1.866	1.866
			522	8	8	8	1.076	1.076	1.076
		496	493	20	26	26	1.866	1.866	1.866
			503	67	71	72	3.720	3.720	3.720
		499	240		4	4		0.467	0.467
		503	496	67	71	72	3.720	3.720	3.720
			522	38	48	48	2.326	2.326	2.326
		513	554	13	13	13	1.986	1.986	1.986
		522	493	8	8	8	1.076	1.076	1.076
			503	38	48	48	2.326	2.326	2.326
		554	513	13	13	13	1.986	1.986	1.986
		649	656	8	8	8	2.903	2.903	2.903
		653	656	28	28	28	3.546	3.546	3.546
		656	163	31	31	31	2.540	2.540	2.540
			649	8	8	8	2.903	2.903	2.903
			653	28	28	28	3.546	3.546	3.546
		666	700	6	6	6	2.829	2.829	2.829
			741	3	3	3	3.423	3.423	3.423
		700	666	6	6	6	2.829	2.829	2.829
			741	16	16	16	2.126	2.126	2.126

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		741	666	3	3	3	3.423	3.423	3.423
			700	16	16	16	2.126	2.126	2.126
			786	27	27	27	2.559	2.559	2.559
			790	30	30	30	2.765	2.765	2.765
		786	741	27	27	27	2.559	2.559	2.559
			793	15	16	16	1.329	1.329	1.329
		790	741	30	30	30	2.765	2.765	2.765
		793	786	15	16	16	1.329	1.329	1.329
		858	6	12	12	12	3.801	3.801	3.801
		866	111	8	8	8	12.335	12.335	12.335
		923	963	4	5	5	2.137	2.137	2.137
		943	3	24	28	30	2.492	2.492	2.492
			21	5	7	7	1.820	1.820	1.820
			22	13	13	13	1.526	1.526	1.526
		963	923	4	5	5	2.137	2.137	2.137
			1008	43	43	43	3.170	3.170	3.170
		1008	963	43	43	43	3.170	3.170	3.170
SF3B2	Q13435	10	77	2	2	2	4.448	4.448	4.448
		35	77	11	11	11	12.052	12.052	12.052
		77	10	2	2	2	4.448	4.448	4.448
			35	11	11	11	12.052	12.052	12.052
		148	165	25	26	26	3.804	3.804	3.804
			173	10	12	12	2.163	2.163	2.163
			182	2	4	4	0.660	0.660	0.660
			198	3	3	3	2.014	2.014	2.014
		165	148	25	26	26	3.804	3.804	3.804
			182	15	16	16	2.839	2.839	2.839
		173	148	10	12	12	2.163	2.163	2.163
		182	148	2	4	4	0.660	0.660	0.660
			165	15	16	16	2.839	2.839	2.839
		198	148	3	3	3	2.014	2.014	2.014
		268	280	94	127	129	1.725	1.725	1.725
		280	268	94	127	129	1.725	1.725	1.725
		323	335			2			0.148
		335	323			2			0.148
		387	394	17	17	17	3.629	3.629	3.629
			400	9	9	9	1.496	1.496	1.496
			403	9	9	9	2.562	2.562	2.562
			405	3	3	3	2.506	2.506	2.506
		394	387	17	17	17	3.629	3.629	3.629
			405	5	6	6	1.098	1.098	1.098
		400	387	9	9	9	1.496	1.496	1.496
			405	31	33	34	2.359	2.359	2.359
			409	9	10	10	1.332	1.332	1.332
			412	2	2	2	0.856	0.856	0.856
		401	405	26	26	26	1.480	1.480	1.480
			409	3	4	4	0.683	0.683	0.683
		403	387	9	9	9	2.562	2.562	2.562
			409	4	5	5	1.680	1.680	1.680
			420	3	3	3	2.064	2.064	2.064
		405	387	3	3	3	2.506	2.506	2.506
			394	5	6	6	1.098	1.098	1.098
			400	31	33	34	2.359	2.359	2.359
			401	26	26	26	1.480	1.480	1.480
			412	47	73	74	1.028	1.028	1.028
			420	6	8	8	2.450	2.450	2.450
		409	400	9	10	10	1.332	1.332	1.332
			401	3	4	4	0.683	0.683	0.683
			403	4	5	5	1.680	1.680	1.680
			412		3	3		0.226	0.226
			420	6	25	27	0.560	0.560	0.560
			556			2			0.242
		412	400	2	2	2	0.856	0.856	0.856
			405	47	73	74	1.028	1.028	1.028
			409		3	3		0.226	0.226
			605		2	2		0.419	0.419
			891		4	4		0.310	0.310
			894		5	5		0.558	0.558
		420	403	3	3	3	2.064	2.064	2.064
			405	6	8	8	2.450	2.450	2.450
			409	6	25	27	0.560	0.560	0.560
		529	570	38	38	38	3.544	3.544	3.544
		543	550	17	21	21	2.362	2.362	2.362
			552	7	9	9	1.562	1.562	1.562
			556	25	26	26	2.484	2.484	2.484
			560	45	45	45	1.869	1.869	1.869
			563	34	34	34	3.236	3.236	3.236
			570	6	6	6	0.982	0.982	0.982
			857	10	10	10	1.930	1.930	1.930
		547	552	19	23	24	1.522	1.522	1.522
			556	26	26	26	2.044	2.044	2.044
			560	25	26	26	1.293	1.293	1.293
			563	10	10	10	2.241	2.241	2.241

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			857	6	6	6	2.462	2.462	2.462
			870	11	11	11	2.145	2.145	2.145
		550	543	17	21	21	2.362	2.362	2.362
			556	14	14	14	2.139	2.139	2.139
			560	13	14	14	1.476	1.476	1.476
			870	32	33	33	1.286	1.286	1.286
			877	8	11	11	1.671	1.671	1.671
		552	543	7	9	9	1.562	1.562	1.562
			547	19	23	24	1.522	1.522	1.522
			556	15	16	17	1.227	1.227	1.227
			560	6	6	6	1.425	1.425	1.425
			563		2	2		0.468	0.468
			570	2	2	2	0.601	0.601	0.601
			870	11	12	12	0.921	0.921	0.921
		556	409			2			0.242
			543	25	26	26	2.484	2.484	2.484
			547	26	26	26	2.044	2.044	2.044
			550	14	14	14	2.139	2.139	2.139
			552	15	16	17	1.227	1.227	1.227
			563	23	23	23	2.517	2.517	2.517
			570	14	14	14	1.786	1.786	1.786
			870	11	11	11	1.866	1.866	1.866
		560	543	45	45	45	1.869	1.869	1.869
			547	25	26	26	1.293	1.293	1.293
			550	13	14	14	1.476	1.476	1.476
			552	6	6	6	1.425	1.425	1.425
			570	46	46	46	1.889	1.889	1.889
			857	8	8	8	1.035	1.035	1.035
		563	543	34	34	34	3.236	3.236	3.236
			547	10	10	10	2.241	2.241	2.241
			552		2	2		0.468	0.468
			556	23	23	23	2.517	2.517	2.517
		570	529	38	38	38	3.544	3.544	3.544
			543	6	6	6	0.982	0.982	0.982
			552	2	2	2	0.601	0.601	0.601
			556	14	14	14	1.786	1.786	1.786
			560	46	46	46	1.889	1.889	1.889
			815	3	3	3	2.907	2.907	2.907
		581	604	25	25	25	2.941	2.941	2.941
			605	15	15	15	3.408	3.408	3.408
		583	604	46	46	46	3.021	3.021	3.021
			605	32	32	32	3.580	3.580	3.580
		604	581	25	25	25	2.941	2.941	2.941
			583	46	46	46	3.021	3.021	3.021
		605	412		2	2		0.419	0.419
			581	15	15	15	3.408	3.408	3.408
			583	32	32	32	3.580	3.580	3.580
			649	17	17	17	18.369	18.369	18.369
		627	672	2	2	2	12.700	12.700	12.700
		649	605	17	17	17	18.369	18.369	18.369
		672	627	2	2	2	12.700	12.700	12.700
		770	815	6	6	6	10.981	10.981	10.981
		790	815	26	27	27	15.919	15.919	15.919
		815	570	3	3	3	2.907	2.907	2.907
			770	6	6	6	10.981	10.981	10.981
			790	26	27	27	15.919	15.919	15.919
			857	6	6	6	7.208	7.208	7.208
			870	4	4	4	2.008	2.008	2.008
			877	3	3	3	2.612	2.612	2.612
		843	877	6	7	7	2.947	2.947	2.947
		857	543	10	10	10	1.930	1.930	1.930
			547	6	6	6	2.462	2.462	2.462
			560	8	8	8	1.035	1.035	1.035
			815	6	6	6	7.208	7.208	7.208
			874			2			0.179
			875		5	6		0.372	0.372
			877	18	18	18	4.163	4.163	4.163
		870	547	11	11	11	2.145	2.145	2.145
			550	32	33	33	1.286	1.286	1.286
			552	11	12	12	0.921	0.921	0.921
			556	11	11	11	1.866	1.866	1.866
			815	4	4	4	2.008	2.008	2.008
			874		2	2		0.401	0.401
			877	98	105	105	3.362	3.362	3.362
			891	6	7	7	1.176	1.176	1.176
		874	857			2			0.179
			870		2	2		0.401	0.401
		875	857		5	6		0.372	0.372
		877	550	8	11	11	1.671	1.671	1.671
			815	3	3	3	2.612	2.612	2.612
			843	6	7	7	2.947	2.947	2.947
			857	18	18	18	4.163	4.163	4.163
			870	98	105	105	3.362	3.362	3.362

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			889		2	2		0.865	0.865
			891	40	42	42	2.585	2.585	2.585
			894	12	12	12	1.637	1.637	1.637
		889	877		2	2		0.865	0.865
		891	412		4	4		0.310	0.310
			870	6	7	7	1.176	1.176	1.176
			877	40	42	42	2.585	2.585	2.585
		894	412		5	5		0.558	0.558
			877	12	12	12	1.637	1.637	1.637
SF3B3	Q15393	96	650	2	2	2	2.145	2.145	2.145
			1189	21	21	21	2.005	2.005	2.005
		109	137	51	53	53	2.793	2.793	2.793
		137	109	51	53	53	2.793	2.793	2.793
		541	546	3	3	3	0.813	0.813	0.813
		546	541	3	3	3	0.813	0.813	0.813
		650	96	2	2	2	2.145	2.145	2.145
			1191	6	8	8	2.540	2.540	2.540
			1206	5	5	5	1.137	1.137	1.137
		911	942	12	12	12	9.887	9.887	9.887
		929	974	6	6	6	1.659	1.659	1.659
		942	911	12	12	12	9.887	9.887	9.887
		974	929	6	6	6	1.659	1.659	1.659
			980	9	9	9	1.906	1.906	1.906
		980	974	9	9	9	1.906	1.906	1.906
		984	1074	10	10	10	4.631	4.631	4.631
		1074	984	10	10	10	4.631	4.631	4.631
		1189	96	21	21	21	2.005	2.005	2.005
		1191	650	6	8	8	2.540	2.540	2.540
		1206	650	5	5	5	1.137	1.137	1.137
SF3B4	Q15427	23	82	17	18	18	1.607	1.607	1.607
		82	23	17	18	18	1.607	1.607	1.607
SF3B6	Q9Y3B4	7	41	12	13	13	2.214	2.214	2.214
			71	2	2	2	1.098	1.098	1.098
		41	7	12	13	13	2.214	2.214	2.214
		71	7	2	2	2	1.098	1.098	1.098
		100	105	27	27	27	1.476	1.476	1.476
			106	9	9	9	1.780	1.780	1.780
			125	3	3	3	2.124	2.124	2.124
		104	111	11	13	13	1.274	1.274	1.274
			116		2	2		0.386	0.386
		105	100	27	27	27	1.476	1.476	1.476
			114	1	1	1	0.911	0.911	0.911
			116	91	93	93	2.544	2.544	2.544
			125	4	4	4	1.371	1.371	1.371
		106	100	9	9	9	1.780	1.780	1.780
			114	13	15	15	1.554	1.554	1.554
			116	3	3	3	1.331	1.331	1.331
			125	4	4	4	1.700	1.700	1.700
		111	104	11	13	13	1.274	1.274	1.274
			116	43	43	43	1.857	1.857	1.857
		114	105	1	1	1	0.911	0.911	0.911
			106	13	15	15	1.554	1.554	1.554
		116	104		2	2		0.386	0.386
			105	91	93	93	2.544	2.544	2.544
			106	3	3	3	1.331	1.331	1.331
			111	43	43	43	1.857	1.857	1.857
		125	100	3	3	3	2.124	2.124	2.124
			105	4	4	4	1.371	1.371	1.371
			106	4	4	4	1.700	1.700	1.700
SKIP	Q13573	23	48	29	31	31	1.682	1.682	1.682
			81	8	8	8	2.803	2.803	2.803
			97	7	8	9	1.096	1.096	1.096
			110		2	2		0.472	0.472
		48	23	29	31	31	1.682	1.682	1.682
			97	2	2	2	3.392	3.392	3.392
			110	4	4	4	3.228	3.228	3.228
			193	3	3	3	2.306	2.306	2.306
		81	23	8	8	8	2.803	2.803	2.803
			97	36	36	36	3.764	3.764	3.764
			108	6	6	6	2.305	2.305	2.305
			110	14	14	14	3.442	3.442	3.442
			115	3	3	3	1.989	1.989	1.989
			122	6	6	6	3.235	3.235	3.235
		95	108	5	6	6	1.400	1.400	1.400
			110	3	3	3	2.579	2.579	2.579
		97	23	7	8	9	1.096	1.096	1.096
			48	2	2	2	3.392	3.392	3.392
			81	36	36	36	3.764	3.764	3.764
			108	36	42	43	1.536	1.536	1.536
			110	48	49	49	3.638	3.638	3.638
			115	23	25	25	2.697	2.697	2.697
			122	9	10	10	2.335	2.335	2.335
			153	3	3	3	1.744	1.744	1.744

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			193	7	7	7	1.757	1.757	1.757
		108	81	6	6	6	2.305	2.305	2.305
			95	5	6	6	1.400	1.400	1.400
			97	36	42	43	1.536	1.536	1.536
			115	15	16	16	1.332	1.332	1.332
			122	3	5	5	0.935	0.935	0.935
			153	2	2	2	0.810	0.810	0.810
		110	23		2	2		0.472	0.472
			48	4	4	4	3.228	3.228	3.228
			81	14	14	14	3.442	3.442	3.442
			95	3	3	3	2.579	2.579	2.579
			97	48	49	49	3.638	3.638	3.638
			122	33	34	34	2.778	2.778	2.778
		115	81	3	3	3	1.989	1.989	1.989
			97	23	25	25	2.697	2.697	2.697
			108	15	16	16	1.332	1.332	1.332
		122	81	6	6	6	3.235	3.235	3.235
			97	9	10	10	2.335	2.335	2.335
			108	3	5	5	0.935	0.935	0.935
			110	33	34	34	2.778	2.778	2.778
		141	153	5	5	5	1.189	1.189	1.189
		146	153	102	110	111	2.059	2.059	2.059
			158	2	2	2	0.793	0.793	0.793
			170	4	4	4	1.105	1.105	1.105
		153	97	3	3	3	1.744	1.744	1.744
			108	2	2	2	0.810	0.810	0.810
			141	5	5	5	1.189	1.189	1.189
			146	102	110	111	2.059	2.059	2.059
			170	49	49	49	2.362	2.362	2.362
			193	26	26	26	3.090	3.090	3.090
			217	12	12	12	1.847	1.847	1.847
		158	146	2	2	2	0.793	0.793	0.793
			170	11	11	11	2.027	2.027	2.027
			193	18	20	20	1.676	1.676	1.676
			204		3	3		0.457	0.457
			216	25	28	28	1.471	1.471	1.471
			217	34	36	36	2.345	2.345	2.345
		170	146	4	4	4	1.105	1.105	1.105
			153	49	49	49	2.362	2.362	2.362
			158	11	11	11	2.027	2.027	2.027
			193	27	28	28	3.005	3.005	3.005
			204	4	4	4	1.509	1.509	1.509
			213	2	2	2	0.611	0.611	0.611
			217	8	8	8	1.243	1.243	1.243
		193	48	3	3	3	2.306	2.306	2.306
			97	7	7	7	1.757	1.757	1.757
			153	26	26	26	3.090	3.090	3.090
			158	18	20	20	1.676	1.676	1.676
			170	27	28	28	3.005	3.005	3.005
			204	3	3	3	0.661	0.661	0.661
			217	2	4	4	1.649	1.649	1.649
		204	158		3	3		0.457	0.457
			170	4	4	4	1.509	1.509	1.509
			193	3	3	3	0.661	0.661	0.661
			217	4	4	4	1.439	1.439	1.439
		213	170	2	2	2	0.611	0.611	0.611
			217	26	26	26	4.954	4.954	4.954
		216	158	25	28	28	1.471	1.471	1.471
		217	153	12	12	12	1.847	1.847	1.847
			158	34	36	36	2.345	2.345	2.345
			170	8	8	8	1.243	1.243	1.243
			193	2	4	4	1.649	1.649	1.649
			204	4	4	4	1.439	1.439	1.439
			213	26	26	26	4.954	4.954	4.954
		246	266	18	26	26	0.966	0.966	0.966
		255	266	6	7	7	0.784	0.784	0.784
		266	246	18	26	26	0.966	0.966	0.966
			255	6	7	7	0.784	0.784	0.784
		311	317	3	3	3	2.150	2.150	2.150
			319	42	46	46	2.686	2.686	2.686
		315	323	16	16	16	1.485	1.485	1.485
		317	311	3	3	3	2.150	2.150	2.150
			323	3	3	3	0.785	0.785	0.785
		319	311	42	46	46	2.686	2.686	2.686
			330	3	3	3	1.643	1.643	1.643
		323	315	16	16	16	1.485	1.485	1.485
			317	3	3	3	0.785	0.785	0.785
			330	43	48	48	2.505	2.505	2.505
			339	2	4	4	0.702	0.702	0.702
		330	319	3	3	3	1.643	1.643	1.643
			323	43	48	48	2.505	2.505	2.505
			339	42	51	51	1.574	1.574	1.574
			344	12	18	18	1.533	1.533	1.533

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			379	4	4	4	0.683	0.683	0.683
		339	323	2	4	4	0.702	0.702	0.702
			330	42	51	51	1.574	1.574	1.574
			376	27	28	28	1.435	1.435	1.435
			379	47	48	48	2.228	2.228	2.228
			468	2	2	2	1.422	1.422	1.422
		344	330	12	18	18	1.533	1.533	1.533
			376	8	9	9	1.030	1.030	1.030
			379	19	19	19	1.529	1.529	1.529
		376	339	27	28	28	1.435	1.435	1.435
			344	8	9	9	1.030	1.030	1.030
			379	40	44	45	3.877	3.877	3.877
			416	3	4	4	0.465	0.465	0.465
			441	19	21	21	2.303	2.303	2.303
			452	2	5	5	0.592	0.592	0.592
			456	4	5	5	0.962	0.962	0.962
			468	7	7	7	2.348	2.348	2.348
			476	2	5	6	0.940	0.940	0.940
		379	330	4	4	4	0.683	0.683	0.683
			339	47	48	48	2.228	2.228	2.228
			344	19	19	19	1.529	1.529	1.529
			376	40	44	45	3.877	3.877	3.877
			416	4	5	5	0.706	0.706	0.706
			441	5	6	6	1.519	1.519	1.519
			452		4	5		0.775	0.775
			456	4	4	4	1.722	1.722	1.722
			468	3	3	3	3.411	3.411	3.411
			476	7	9	9	0.894	0.894	0.894
		416	376	3	4	4	0.465	0.465	0.465
			379	4	5	5	0.706	0.706	0.706
			441	22	22	22	3.233	3.233	3.233
			452	4	4	4	2.109	2.109	2.109
			456	3	3	3	3.067	3.067	3.067
			468		2	2		0.568	0.568
		441	376	19	21	21	2.303	2.303	2.303
			379	5	6	6	1.519	1.519	1.519
			416	22	22	22	3.233	3.233	3.233
			456	48	48	48	4.458	4.458	4.458
			468	27	28	28	1.888	1.888	1.888
			476	48	49	49	2.469	2.469	2.469
			515	9	9	9	1.469	1.469	1.469
		452	376	2	5	5	0.592	0.592	0.592
			379		4	5		0.775	0.775
			416	4	4	4	2.109	2.109	2.109
			468	54	69	71	1.158	1.158	1.158
			476	42	53	53	1.900	1.900	1.900
			515		6	7		0.325	0.325
			524		3	3		0.610	0.610
			531		3	3		0.901	0.901
		456	376	4	5	5	0.962	0.962	0.962
			379	4	4	4	1.722	1.722	1.722
			416	3	3	3	3.067	3.067	3.067
			441	48	48	48	4.458	4.458	4.458
			468	47	48	48	2.543	2.543	2.543
			476	19	21	21	1.322	1.322	1.322
			515	8	12	12	0.725	0.725	0.725
			524	4	5	5	0.747	0.747	0.747
		468	339	2	2	2	1.422	1.422	1.422
			376	7	7	7	2.348	2.348	2.348
			379	3	3	3	3.411	3.411	3.411
			416		2	2		0.568	0.568
			441	27	28	28	1.888	1.888	1.888
			452	54	69	71	1.158	1.158	1.158
			456	47	48	48	2.543	2.543	2.543
			476	103	127	129	1.437	1.437	1.437
			509	7	9	9	1.084	1.084	1.084
		476	376	2	5	6	0.940	0.940	0.940
			379	7	9	9	0.894	0.894	0.894
			441	48	49	49	2.469	2.469	2.469
			452	42	53	53	1.900	1.900	1.900
			456	19	21	21	1.322	1.322	1.322
			468	103	127	129	1.437	1.437	1.437
			503	4	4	4	1.369	1.369	1.369
			509	4	7	7	0.782	0.782	0.782
			515	13	25	26	0.719	0.719	0.719
			524	4	11	11	0.766	0.766	0.766
		503	476	4	4	4	1.369	1.369	1.369
			515	2	3	3	0.690	0.690	0.690
			524	4	7	7	0.705	0.705	0.705
		509	468	7	9	9	1.084	1.084	1.084
			476	4	7	7	0.782	0.782	0.782
			524	10	14	14	1.433	1.433	1.433
			531		3	3		1.022	1.022

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		515	441	9	9	9	1.469	1.469	1.469
			452		6	7		0.325	0.325
			456	8	12	12	0.725	0.725	0.725
			476	13	25	26	0.719	0.719	0.719
			503	2	3	3	0.690	0.690	0.690
			531	2	3	5	0.896	0.896	0.896
		524	452		3	3		0.610	0.610
			456	4	5	5	0.747	0.747	0.747
			476	4	11	11	0.766	0.766	0.766
			503	4	7	7	0.705	0.705	0.705
			509	10	14	14	1.433	1.433	1.433
		531	452		3	3		0.901	0.901
			509		3	3		1.022	1.022
			515	2	3	5	0.896	0.896	0.896
SmB	P14678	50	57	3	3	3	1.471	1.471	1.471
			64		3	3		0.373	0.373
			88	2	3	3	0.568	0.568	0.568
		52	57	2	3	3	0.600	0.600	0.600
		57	50	3	3	3	1.471	1.471	1.471
			52	2	3	3	0.600	0.600	0.600
			64		3	5		0.210	0.210
		64	50		3	3		0.373	0.373
			57		3	5		0.210	0.210
		88	50	2	3	3	0.568	0.568	0.568
SmD1	P62314	1	86	3	3	3	1.855	1.855	1.855
		86	1	3	3	3	1.855	1.855	1.855
SmD2	P62316	6	18	55	60	61	2.169	2.169	2.169
		18	6	55	60	61	2.169	2.169	2.169
		51	118	24	25	25	8.665	8.665	8.665
		71	118	13	13	13	1.690	1.690	1.690
		86	98		3	3		0.260	0.260
		88	98		6	8		0.344	0.344
			118	8	10	10	2.687	2.687	2.687
		92	98		17	20		0.391	0.391
		98	86		3	3		0.260	0.260
			88		6	8		0.344	0.344
			92		17	20		0.391	0.391
		118	51	24	25	25	8.665	8.665	8.665
			71	13	13	13	1.690	1.690	1.690
			88	8	10	10	2.687	2.687	2.687
SmD3	P62318	8	84		2	2		0.416	0.416
			87		2	2		0.201	0.201
		84	8		2	2		0.416	0.416
		87	8		2	2		0.201	0.201
SmE	P62304	67	72	4	4	4	1.916	1.916	1.916
		72	67	4	4	4	1.916	1.916	1.916
SmG	P62308	3	11		2	2		0.253	0.253
		10	15	9	10	10	0.937	0.937	0.937
		11	3		2	2		0.253	0.253
		15	10	9	10	10	0.937	0.937	0.937
SMU1	Q2TAY7	107	247	6	6	6	2.758	2.758	2.758
		170	190	55	57	57	2.869	2.869	2.869
			194	23	23	23	4.209	4.209	4.209
			512		3	3		0.360	0.360
		190	170	55	57	57	2.869	2.869	2.869
			200	13	13	13	1.573	1.573	1.573
			247	5	5	5	1.777	1.777	1.777
			292	6	6	6	1.689	1.689	1.689
			317	7	7	7	2.212	2.212	2.212
			337	13	13	13	2.112	2.112	2.112
			408	6	6	6	1.286	1.286	1.286
			512	25	28	29	3.322	3.322	3.322
		194	170	23	23	23	4.209	4.209	4.209
			292		2	2		2.341	2.341
			337	9	9	9	1.401	1.401	1.401
			408	10	10	10	3.445	3.445	3.445
			512	41	53	53	1.436	1.436	1.436
		200	190	13	13	13	1.573	1.573	1.573
			512	30	43	44	0.970	0.970	0.970
		210	247	77	78	78	2.806	2.806	2.806
			485	135	148	150	15.323	15.323	15.323
			509	119	125	126	6.908	6.908	6.908
		214	485	74	77	79	22.768	22.768	22.768
		247	107	6	6	6	2.758	2.758	2.758
			190	5	5	5	1.777	1.777	1.777
			210	77	78	78	2.806	2.806	2.806
			289	3	5	6	0.601	0.601	0.601
			308	2	2	2	0.980	0.980	0.980
		289	247	3	5	6	0.601	0.601	0.601
			512	4	4	4	2.380	2.380	2.380
		292	190	6	6	6	1.689	1.689	1.689
			194		2	2		2.341	2.341
			337	171	176	176	2.799	2.799	2.799

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		308	247	2	2	2	0.980	0.980	0.980
		317	190	7	7	7	2.212	2.212	2.212
			343	11	11	11	0.988	0.988	0.988
			379	21	22	22	3.645	3.645	3.645
			408	5	5	5	2.928	2.928	2.928
		337	190	13	13	13	2.112	2.112	2.112
			194	9	9	9	1.401	1.401	1.401
			292	171	176	176	2.799	2.799	2.799
			343	12	12	12	1.826	1.826	1.826
		340	379	3	3	3	1.711	1.711	1.711
		343	317	11	11	11	0.988	0.988	0.988
			337	12	12	12	1.826	1.826	1.826
			379	54	54	54	2.546	2.546	2.546
		374	388	10	11	11	1.491	1.491	1.491
		379	317	21	22	22	3.645	3.645	3.645
			340	3	3	3	1.711	1.711	1.711
			343	54	54	54	2.546	2.546	2.546
		388	374	10	11	11	1.491	1.491	1.491
		408	190	6	6	6	1.286	1.286	1.286
			194	10	10	10	3.445	3.445	3.445
			317	5	5	5	2.928	2.928	2.928
		485	210	135	148	150	15.323	15.323	15.323
			214	74	77	79	22.768	22.768	22.768
		509	210	119	125	126	6.908	6.908	6.908
		512	170		3	3		0.360	0.360
			190	25	28	29	3.322	3.322	3.322
			194	41	53	53	1.436	1.436	1.436
			200	30	43	44	0.970	0.970	0.970
			289	4	4	4	2.380	2.380	2.380
SNIP1	Q8TAD8	214	223	2	2	2	1.081	1.081	1.081
			301		2	2		0.486	0.486
		216	223	18	18	18	2.358	2.358	2.358
			301	15	15	15	1.335	1.335	1.335
		221	301	10	10	10	1.982	1.982	1.982
		223	214	2	2	2	1.081	1.081	1.081
			216	18	18	18	2.358	2.358	2.358
			265	3	3	3	2.697	2.697	2.697
			301	4	6	6	2.554	2.554	2.554
		245	301	106	108	109	3.723	3.723	3.723
		265	223	3	3	3	2.697	2.697	2.697
		301	214		2	2		0.486	0.486
			216	15	15	15	1.335	1.335	1.335
			221	10	10	10	1.982	1.982	1.982
			223	4	6	6	2.554	2.554	2.554
			245	106	108	109	3.723	3.723	3.723
		325	353		2	2		2.426	2.426
		342	353	11	15	15	1.635	1.635	1.635
		353	325		2	2		2.426	2.426
			342	11	15	15	1.635	1.635	1.635
SNU114	Q15029	95	602	3	3	3	2.589	2.589	2.589
		98	602	4	4	4	2.106	2.106	2.106
			646	7	7	7	3.103	3.103	3.103
		244	963			2			0.241
		352	355	66	68	68	2.340	2.340	2.340
			359	191	194	194	4.313	4.313	4.313
		355	352	66	68	68	2.340	2.340	2.340
			359	7	7	7	2.261	2.261	2.261
		359	352	191	194	194	4.313	4.313	4.313
			355	7	7	7	2.261	2.261	2.261
		602	95	3	3	3	2.589	2.589	2.589
			98	4	4	4	2.106	2.106	2.106
			609	3	3	3	1.931	1.931	1.931
		609	602	3	3	3	1.931	1.931	1.931
		646	98	7	7	7	3.103	3.103	3.103
		673	684	21	21	21	1.789	1.789	1.789
			790	30	30	30	4.858	4.858	4.858
		684	673	21	21	21	1.789	1.789	1.789
			963	3	3	3	2.055	2.055	2.055
		790	673	30	30	30	4.858	4.858	4.858
		914	951	11	11	11	11.140	11.140	11.140
			963	9	9	9	14.354	14.354	14.354
		951	914	11	11	11	11.140	11.140	11.140
		963	244			2			0.241
			684	3	3	3	2.055	2.055	2.055
			914	9	9	9	14.354	14.354	14.354
SNU17	Q9Y388	203	208	3	3	3	3.171	3.171	3.171
		208	203	3	3	3	3.171	3.171	3.171
			215	2	2	2	1.891	1.891	1.891
		215	208	2	2	2	1.891	1.891	1.891
			223	5	5	5	1.486	1.486	1.486
			226	6	7	7	1.560	1.560	1.560
		223	215	5	5	5	1.486	1.486	1.486
		226	215	6	7	7	1.560	1.560	1.560

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
SNU23	Q96NC0	8	15	86	88	88	3.718	3.718	3.718
			18	49	50	50	1.701	1.701	1.701
			23	5	6	6	1.681	1.681	1.681
			27	20	25	25	1.647	1.647	1.647
			36	11	11	11	2.247	2.247	2.247
			45	2	2	2	0.691	0.691	0.691
		15	8	86	88	88	3.718	3.718	3.718
			23	2	4	4	0.683	0.683	0.683
			27	22	25	25	2.874	2.874	2.874
		18	8	49	50	50	1.701	1.701	1.701
			27	8	13	13	1.471	1.471	1.471
		23	8	5	6	6	1.681	1.681	1.681
			15	2	4	4	0.683	0.683	0.683
		27	8	20	25	25	1.647	1.647	1.647
			15	22	25	25	2.874	2.874	2.874
			18	8	13	13	1.471	1.471	1.471
			36	8	10	11	1.165	1.165	1.165
			39	6	6	6	1.372	1.372	1.372
		36	8	11	11	11	2.247	2.247	2.247
			27	8	10	11	1.165	1.165	1.165
		39	27	6	6	6	1.372	1.372	1.372
		45	8	2	2	2	0.691	0.691	0.691
			55	4	4	4	1.056	1.056	1.056
		55	45	4	4	4	1.056	1.056	1.056
			64	2	3	3	1.092	1.092	1.092
			132	6	6	6	2.294	2.294	2.294
			136	25	27	27	1.402	1.402	1.402
		64	55	2	3	3	1.092	1.092	1.092
		123	130	9	10	10	2.168	2.168	2.168
		124	132	3	3	3	1.117	1.117	1.117
		130	123	9	10	10	2.168	2.168	2.168
		132	55	6	6	6	2.294	2.294	2.294
			124	3	3	3	1.117	1.117	1.117
			138	8	8	8	2.432	2.432	2.432
		136	55	25	27	27	1.402	1.402	1.402
		138	132	8	8	8	2.432	2.432	2.432
		155	160	1	2	3	0.418	0.418	0.418
		157	163	3	3	3	1.303	1.303	1.303
		160	155	1	2	3	0.418	0.418	0.418
			167		2	2		0.995	0.995
		163	157	3	3	3	1.303	1.303	1.303
		167	160		2	2		0.995	0.995
SPF27	O75934	151	158	14	14	14	2.693	2.693	2.693
		158	151	14	14	14	2.693	2.693	2.693
			168	18	18	18	4.826	4.826	4.826
		168	158	18	18	18	4.826	4.826	4.826
SRRM1	Q8IYB3	36	128	4	4	4	1.321	1.321	1.321
		37	78	3	3	3	2.236	2.236	2.236
		42	890	4	6	6	0.709	0.709	0.709
		54	128	90	100	101	2.301	2.301	2.301
			132	13	21	21	0.879	0.879	0.879
		78	37	3	3	3	2.236	2.236	2.236
		128	36	4	4	4	1.321	1.321	1.321
			54	90	100	101	2.301	2.301	2.301
			140	17	18	18	2.017	2.017	2.017
		132	54	13	21	21	0.879	0.879	0.879
			140	93	101	102	1.814	1.814	1.814
			145	4	4	4	1.835	1.835	1.835
		140	128	17	18	18	2.017	2.017	2.017
			132	93	101	102	1.814	1.814	1.814
			146	64	64	64	3.401	3.401	3.401
			151	26	28	28	1.849	1.849	1.849
			153	5	5	5	1.529	1.529	1.529
			156	3	3	3	1.677	1.677	1.677
			159	2	2	2	1.257	1.257	1.257
		145	132	4	4	4	1.835	1.835	1.835
			151	13	13	13	1.602	1.602	1.602
			153	4	4	4	1.210	1.210	1.210
			156	6	7	8	1.286	1.286	1.286
			159	9	11	11	2.275	2.275	2.275
		146	140	64	64	64	3.401	3.401	3.401
			159	4	4	4	2.304	2.304	2.304
		151	140	26	28	28	1.849	1.849	1.849
			145	13	13	13	1.602	1.602	1.602
			159	4	5	5	1.941	1.941	1.941
		153	140	5	5	5	1.529	1.529	1.529
			145	4	4	4	1.210	1.210	1.210
		156	140	3	3	3	1.677	1.677	1.677
			145	6	7	8	1.286	1.286	1.286
		159	140	2	2	2	1.257	1.257	1.257
			145	9	11	11	2.275	2.275	2.275
			146	4	4	4	2.304	2.304	2.304
			151	4	5	5	1.941	1.941	1.941

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		231	249	20	20	20	2.398	2.398	2.398
		249	231	20	20	20	2.398	2.398	2.398
		438	443	19	21	21	2.152	2.152	2.152
		443	438	19	21	21	2.152	2.152	2.152
		885	890	5	5	5	1.544	1.544	1.544
		890	42	4	6	6	0.709	0.709	0.709
			885	5	5	5	1.544	1.544	1.544
			897	35	36	36	2.308	2.308	2.308
		897	890	35	36	36	2.308	2.308	2.308
SRRT	Q9BXP5	153	176	22	22	22	2.796	2.796	2.796
			442	2	2	2	1.503	1.503	1.503
		170	687	13	13	13	1.767	1.767	1.767
		176	153	22	22	22	2.796	2.796	2.796
			200	2	6	6	0.888	0.888	0.888
			442		2	2		0.342	0.342
			476	3	3	3	1.759	1.759	1.759
			687	216	228	229	3.940	3.940	3.940
		192	739	3	3	3	0.909	0.909	0.909
		200	176	2	6	6	0.888	0.888	0.888
			413	7	7	7	2.381	2.381	2.381
			419	7	9	9	3.023	3.023	3.023
			442	77	80	80	3.005	3.005	3.005
			476	30	30	30	3.627	3.627	3.627
		208	442	7	7	7	0.754	0.754	0.754
		256	285	3	3	3	2.418	2.418	2.418
		279	286	17	18	18	3.152	3.152	3.152
			301	18	18	18	2.243	2.243	2.243
			403			2			0.334
		285	256	3	3	3	2.418	2.418	2.418
			301	57	62	65	2.956	2.956	2.956
			305	10	12	12	1.467	1.467	1.467
			308	2	4	4	0.685	0.685	0.685
			313	6	6	6	2.787	2.787	2.787
			327	2	3	3	0.734	0.734	0.734
			395	7	9	9	1.107	1.107	1.107
			397		2	2		0.671	0.671
			413	5	5	5	1.043	1.043	1.043
			442	4	7	7	0.827	0.827	0.827
			476	3	3	3	1.250	1.250	1.250
			607	4	7	7	0.661	0.661	0.661
		286	279	17	18	18	3.152	3.152	3.152
			301	30	30	30	2.641	2.641	2.641
			305	15	16	16	1.337	1.337	1.337
			309	3	4	4	1.273	1.273	1.273
			313	2	3	4	0.596	0.596	0.596
			413		2	2		0.585	0.585
			442	2	3	3	0.654	0.654	0.654
			476	2	2	2	0.892	0.892	0.892
			607	6	6	6	0.940	0.940	0.940
		301	279	18	18	18	2.243	2.243	2.243
			285	57	62	65	2.956	2.956	2.956
			286	30	30	30	2.641	2.641	2.641
			309	5	5	5	1.978	1.978	1.978
			313	3	3	3	1.399	1.399	1.399
		305	285	10	12	12	1.467	1.467	1.467
			286	15	16	16	1.337	1.337	1.337
			309	10	12	12	1.412	1.412	1.412
		308	285	2	4	4	0.685	0.685	0.685
			313	10	14	14	1.507	1.507	1.507
		309	286	3	4	4	1.273	1.273	1.273
			301	5	5	5	1.978	1.978	1.978
			305	10	12	12	1.412	1.412	1.412
		313	285	6	6	6	2.787	2.787	2.787
			286	2	3	4	0.596	0.596	0.596
			301	3	3	3	1.399	1.399	1.399
			308	10	14	14	1.507	1.507	1.507
		327	285	2	3	3	0.734	0.734	0.734
			338	2	2	2	1.686	1.686	1.686
		334	338	15	21	21	2.045	2.045	2.045
			343	3	4	4	0.447	0.447	0.447
		337	338	15	15	15	1.628	1.628	1.628
			343	17	21	22	1.040	1.040	1.040
			346		2	2		0.535	0.535
		338	327	2	2	2	1.686	1.686	1.686
			334	15	21	21	2.045	2.045	2.045
			337	15	15	15	1.628	1.628	1.628
		343	334	3	4	4	0.447	0.447	0.447
			337	17	21	22	1.040	1.040	1.040
		346	337		2	2		0.535	0.535
		395	285	7	9	9	1.107	1.107	1.107
			413	5	5	5	1.479	1.479	1.479
		397	285		2	2		0.671	0.671
			413	2	5	5	1.106	1.106	1.106

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			419		2	2		0.315	0.315
			442	2	4	4	0.500	0.500	0.500
		403	279			2			0.334
			413	11	25	25	1.112	1.112	1.112
			419		5	6		0.399	0.399
			442		3	3		0.302	0.302
		413	200	7	7	7	2.381	2.381	2.381
			285	5	5	5	1.043	1.043	1.043
			286		2	2		0.585	0.585
			395	5	5	5	1.479	1.479	1.479
			397	2	5	5	1.106	1.106	1.106
			403	11	25	25	1.112	1.112	1.112
			442	19	21	22	1.713	1.713	1.713
		419	200	7	9	9	3.023	3.023	3.023
			397		2	2		0.315	0.315
			403		5	6		0.399	0.399
			442	4	12	12	0.676	0.676	0.676
		442	153	2	2	2	1.503	1.503	1.503
			176		2	2		0.342	0.342
			200	77	80	80	3.005	3.005	3.005
			208	7	7	7	0.754	0.754	0.754
			285	4	7	7	0.827	0.827	0.827
			286	2	3	3	0.654	0.654	0.654
			397	2	4	4	0.500	0.500	0.500
			403		3	3		0.302	0.302
			413	19	21	22	1.713	1.713	1.713
			419	4	12	12	0.676	0.676	0.676
			476	15	16	16	2.498	2.498	2.498
			669	3	3	3	0.853	0.853	0.853
		476	176	3	3	3	1.759	1.759	1.759
			200	30	30	30	3.627	3.627	3.627
			285	3	3	3	1.250	1.250	1.250
			286	2	2	2	0.892	0.892	0.892
			442	15	16	16	2.498	2.498	2.498
			669	2	2	2	2.341	2.341	2.341
		514	522	2	2	2	1.880	1.880	1.880
			590	2	3	4	1.729	1.729	1.729
			604	4	4	4	12.406	12.406	12.406
		522	514	2	2	2	1.880	1.880	1.880
		590	514	2	3	4	1.729	1.729	1.729
			611		2	2		1.180	1.180
			664	3	3	3	4.173	4.173	4.173
		604	514	4	4	4	12.406	12.406	12.406
			611	108	110	110	5.223	5.223	5.223
		607	285	4	7	7	0.661	0.661	0.661
			286	6	6	6	0.940	0.940	0.940
		611	590		2	2		1.180	1.180
			604	108	110	110	5.223	5.223	5.223
		664	590	3	3	3	4.173	4.173	4.173
		669	442	3	3	3	0.853	0.853	0.853
			476	2	2	2	2.341	2.341	2.341
		687	170	13	13	13	1.767	1.767	1.767
			176	216	228	229	3.940	3.940	3.940
			691	4	4	4	2.670	2.670	2.670
		691	687	4	4	4	2.670	2.670	2.670
			739	5	5	5	3.119	3.119	3.119
		699	739	4	5	5	1.982	1.982	1.982
		710	720	4	4	4	1.041	1.041	1.041
			721	20	22	22	2.275	2.275	2.275
			723	11	12	12	1.493	1.493	1.493
		712	721	43	43	43	3.574	3.574	3.574
			723	25	26	26	2.838	2.838	2.838
		720	710	4	4	4	1.041	1.041	1.041
			723	15	15	15	2.025	2.025	2.025
			730	4	4	4	1.994	1.994	1.994
			735	18	19	19	1.431	1.431	1.431
		721	710	20	22	22	2.275	2.275	2.275
			712	43	43	43	3.574	3.574	3.574
			730	12	12	12	4.077	4.077	4.077
		723	710	11	12	12	1.493	1.493	1.493
			712	25	26	26	2.838	2.838	2.838
			720	15	15	15	2.025	2.025	2.025
			730	81	81	81	4.256	4.256	4.256
		730	720	4	4	4	1.994	1.994	1.994
			721	12	12	12	4.077	4.077	4.077
			723	81	81	81	4.256	4.256	4.256
			739	1	5	5	0.487	0.487	0.487
			744	27	27	27	3.625	3.625	3.625
		735	720	18	19	19	1.431	1.431	1.431
		739	192	3	3	3	0.909	0.909	0.909
			691	5	5	5	3.119	3.119	3.119
			699	4	5	5	1.982	1.982	1.982
			730	1	5	5	0.487	0.487	0.487

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		744	730	27	27	27	3.625	3.625	3.625
SRSF1	Q07955	30	138	43	43	43	2.567	2.567	2.567
			165	4	4	4	3.719	3.719	3.719
			174	15	15	15	3.320	3.320	3.320
			179	6	6	6	2.591	2.591	2.591
		38	138	330	333	333	2.918	2.918	2.918
			165	18	18	18	4.208	4.208	4.208
			174	22	22	22	3.275	3.275	3.275
			179	6	7	7	0.860	0.860	0.860
			193	27	27	27	2.226	2.226	2.226
		48	138	326	390	390	1.592	1.592	1.592
			165	55	60	61	1.994	1.994	1.994
			174	61	68	69	2.234	2.234	2.234
			179	21	30	30	1.116	1.116	1.116
			193	17	19	19	1.433	1.433	1.433
		138	30	43	43	43	2.567	2.567	2.567
			38	330	333	333	2.918	2.918	2.918
			48	326	390	390	1.592	1.592	1.592
		165	30	4	4	4	3.719	3.719	3.719
			38	18	18	18	4.208	4.208	4.208
			48	55	60	61	1.994	1.994	1.994
			174	39	40	40	3.200	3.200	3.200
		174	30	15	15	15	3.320	3.320	3.320
			38	22	22	22	3.275	3.275	3.275
			48	61	68	69	2.234	2.234	2.234
			165	39	40	40	3.200	3.200	3.200
		179	30	6	6	6	2.591	2.591	2.591
			38	6	7	7	0.860	0.860	0.860
			48	21	30	30	1.116	1.116	1.116
		193	38	27	27	27	2.226	2.226	2.226
			48	17	19	19	1.433	1.433	1.433
SYF1	Q9HCS7	2	45	4	4	4	2.027	2.027	2.027
			50	3	3	3	1.546	1.546	1.546
			62	3	3	3	2.048	2.048	2.048
		45	2	4	4	4	2.027	2.027	2.027
		50	2	3	3	3	1.546	1.546	1.546
			76	12	15	16	1.659	1.659	1.659
			83	31	34	34	2.236	2.236	2.236
		62	2	3	3	3	2.048	2.048	2.048
		76	50	12	15	16	1.659	1.659	1.659
			83	37	40	40	3.540	3.540	3.540
		83	50	31	34	34	2.236	2.236	2.236
			76	37	40	40	3.540	3.540	3.540
		393	485	5	5	5	4.014	4.014	4.014
		420	458	26	27	27	3.136	3.136	3.136
		423	458	41	42	42	3.135	3.135	3.135
		458	420	26	27	27	3.136	3.136	3.136
			423	41	42	42	3.135	3.135	3.135
		485	393	5	5	5	4.014	4.014	4.014
		549	590	5	6	6	2.881	2.881	2.881
		590	549	5	6	6	2.881	2.881	2.881
		708	721	106	107	107	4.241	4.241	4.241
			747	6	6	6	1.744	1.744	1.744
			770	22	22	22	3.485	3.485	3.485
			794	57	64	64	4.822	4.822	4.822
			853	3	3	3	1.379	1.379	1.379
		721	708	106	107	107	4.241	4.241	4.241
			794	3	4	4	1.799	1.799	1.799
		747	708	6	6	6	1.744	1.744	1.744
			794	2	2	2	1.013	1.013	1.013
		770	708	22	22	22	3.485	3.485	3.485
			794	2	2	2	1.555	1.555	1.555
		794	708	57	64	64	4.822	4.822	4.822
			721	3	4	4	1.799	1.799	1.799
			747	2	2	2	1.013	1.013	1.013
			770	2	2	2	1.555	1.555	1.555
		853	708	3	3	3	1.379	1.379	1.379
SYF3	Q9BZJ0	175	180	3	4	4	1.230	1.230	1.230
			182	10	11	11	1.950	1.950	1.950
		180	175	3	4	4	1.230	1.230	1.230
		182	175	10	11	11	1.950	1.950	1.950
		213	256	3	3	3	2.205	2.205	2.205
		224	256	16	17	17	3.349	3.349	3.349
		256	213	3	3	3	2.205	2.205	2.205
			224	16	17	17	3.349	3.349	3.349
		288	315	2	2	2	1.175	1.175	1.175
		315	288	2	2	2	1.175	1.175	1.175
		357	388	122	124	124	3.196	3.196	3.196
		381	388	10	11	11	1.890	1.890	1.890
		388	357	122	124	124	3.196	3.196	3.196
			381	10	11	11	1.890	1.890	1.890
		427	461	6	7	7	0.851	0.851	0.851
		437	445	32	34	34	3.452	3.452	3.452

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		445	437	32	34	34	3.452	3.452	3.452
			460	9	10	10	1.142	1.142	1.142
			475	28	29	29	4.229	4.229	4.229
			485	2	2	2	0.982	0.982	0.982
		460	445	9	10	10	1.142	1.142	1.142
			475	62	64	64	4.822	4.822	4.822
		461	427	6	7	7	0.851	0.851	0.851
		475	445	28	29	29	4.229	4.229	4.229
			460	62	64	64	4.822	4.822	4.822
		485	445	2	2	2	0.982	0.982	0.982
		527	531	2	2	2	0.736	0.736	0.736
		531	527	2	2	2	0.736	0.736	0.736
		568	602	21	23	23	2.060	2.060	2.060
		602	568	21	23	23	2.060	2.060	2.060
		730	740	19	20	20	1.210	1.210	1.210
			823	2	2	2	1.616	1.616	1.616
			827		2	3		0.237	0.237
		740	730	19	20	20	1.210	1.210	1.210
			776	10	10	10	1.814	1.814	1.814
			823	15	15	15	1.475	1.475	1.475
			827	2	3	3	0.474	0.474	0.474
		776	740	10	10	10	1.814	1.814	1.814
			832		3	3		0.190	0.190
		823	730	2	2	2	1.616	1.616	1.616
			740	15	15	15	1.475	1.475	1.475
			827	30	38	39	2.343	2.343	2.343
			832	6	27	31	1.064	1.064	1.064
		826	832		3	4		0.166	0.166
		827	730		2	3		0.237	0.237
			740	2	3	3	0.474	0.474	0.474
			823	30	38	39	2.343	2.343	2.343
		832	776		3	3		0.190	0.190
			823	6	27	31	1.064	1.064	1.064
			826		3	4		0.166	0.166
TCERG1	O14776	1	734			5			0.107
		146	161	10	10	10	2.214	2.214	2.214
			444	4	4	4	1.816	1.816	1.816
		161	146	10	10	10	2.214	2.214	2.214
		444	146	4	4	4	1.816	1.816	1.816
			459	5	5	5	1.907	1.907	1.907
		459	444	5	5	5	1.907	1.907	1.907
			468	15	15	15	1.126	1.126	1.126
		464	468	5	6	6	0.804	0.804	0.804
			478	3	3	3	2.232	2.232	2.232
		466	474	2	2	2	1.635	1.635	1.635
			478	6	7	7	1.514	1.514	1.514
		468	459	15	15	15	1.126	1.126	1.126
			464	5	6	6	0.804	0.804	0.804
			474	15	15	15	1.662	1.662	1.662
			478	20	28	29	4.811	4.811	4.811
		472	478	76	91	95	6.493	6.493	6.493
		474	466	2	2	2	1.635	1.635	1.635
			468	15	15	15	1.662	1.662	1.662
		478	464	3	3	3	2.232	2.232	2.232
			466	6	7	7	1.514	1.514	1.514
			468	20	28	29	4.811	4.811	4.811
			472	76	91	95	6.493	6.493	6.493
			500		7	8		0.493	0.493
			503		2	2		0.209	0.209
			507		2	2		0.127	0.127
			985	19	66	73	0.974	0.974	0.974
		495	503	284	387	395	14.529	14.529	14.529
			507	7	17	18	1.273	1.273	1.273
			516	6	16	16	2.580	2.580	2.580
			729			2			0.169
			734			2			0.163
		500	478		7	8		0.493	0.493
		503	478		2	2		0.209	0.209
			495	284	387	395	14.529	14.529	14.529
		507	478		2	2		0.127	0.127
			495	7	17	18	1.273	1.273	1.273
		516	495	6	16	16	2.580	2.580	2.580
			522	6	6	6	1.743	1.743	1.743
			585		2	2		1.099	1.099
		520	578	39	41	41	4.055	4.055	4.055
			585	8	9	9	1.681	1.681	1.681
		522	516	6	6	6	1.743	1.743	1.743
			578	5	5	5	3.265	3.265	3.265
			585	5	5	5	1.524	1.524	1.524
			724	2	2	2	0.709	0.709	0.709
		570	579	92	92	93	3.492	3.492	3.492
			585	23	25	25	1.848	1.848	1.848
		578	520	39	41	41	4.055	4.055	4.055

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			522	5	5	5	3.265	3.265	3.265
			585	34	35	35	1.798	1.798	1.798
		579	570	92	92	93	3.492	3.492	3.492
			640		4	7		0.275	0.275
		585	516		2	2		1.099	1.099
			520	8	9	9	1.681	1.681	1.681
			522	5	5	5	1.524	1.524	1.524
			570	23	25	25	1.848	1.848	1.848
			578	34	35	35	1.798	1.798	1.798
			599	4	4	4	1.354	1.354	1.354
		599	585	4	4	4	1.354	1.354	1.354
		626	634	2	2	2	0.977	0.977	0.977
		627	634	4	4	4	1.555	1.555	1.555
		634	626	2	2	2	0.977	0.977	0.977
			627	4	4	4	1.555	1.555	1.555
			649	2	2	2	0.682	0.682	0.682
			700		5	5		0.645	0.645
		640	579		4	7		0.275	0.275
		649	634	2	2	2	0.682	0.682	0.682
		667	711	155	160	160	3.302	3.302	3.302
		700	634		5	5		0.645	0.645
			703	103	103	103	4.908	4.908	4.908
			711	5	5	5	0.737	0.737	0.737
		703	700	103	103	103	4.908	4.908	4.908
		711	667	155	160	160	3.302	3.302	3.302
			700	5	5	5	0.737	0.737	0.737
		721	729	5	6	6	1.682	1.682	1.682
		722	729	11	17	18	1.207	1.207	1.207
			740			2			0.113
			761	1	1	1	0.863	0.863	0.863
			796	1	1	1	0.388	0.388	0.388
		724	522	2	2	2	0.709	0.709	0.709
			761	17	18	18	2.113	2.113	2.113
			796		8	9		1.473	1.473
		729	495			2			0.169
			721	5	6	6	1.682	1.682	1.682
			722	11	17	18	1.207	1.207	1.207
			765	9	10	10	0.899	0.899	0.899
			767	75	91	92	2.446	2.446	2.446
		734	1			5			0.107
			495			2			0.163
		740	722			2			0.113
			753	104	107	107	3.540	3.540	3.540
			756	23	25	25	2.935	2.935	2.935
			783	5	5	5	2.262	2.262	2.262
			785	20	20	20	2.421	2.421	2.421
		753	740	104	107	107	3.540	3.540	3.540
			761	15	16	16	1.610	1.610	1.610
		756	740	23	25	25	2.935	2.935	2.935
			761	60	61	61	4.649	4.649	4.649
			1016		2	2		0.312	0.312
		761	722	1	1	1	0.863	0.863	0.863
			724	17	18	18	2.113	2.113	2.113
			753	15	16	16	1.610	1.610	1.610
			756	60	61	61	4.649	4.649	4.649
		765	729	9	10	10	0.899	0.899	0.899
		767	729	75	91	92	2.446	2.446	2.446
			895			2			0.130
		782	796	2	2	2	1.734	1.734	1.734
		783	740	5	5	5	2.262	2.262	2.262
			796	7	7	7	1.694	1.694	1.694
		785	740	20	20	20	2.421	2.421	2.421
			796	174	183	184	3.265	3.265	3.265
			828	3	3	3	1.908	1.908	1.908
			846	20	20	20	1.626	1.626	1.626
			1016		2	2		0.519	0.519
		789	796	84	92	92	1.572	1.572	1.572
			846	6	8	8	1.165	1.165	1.165
		796	722	1	1	1	0.388	0.388	0.388
			724		8	9		1.473	1.473
			782	2	2	2	1.734	1.734	1.734
			783	7	7	7	1.694	1.694	1.694
			785	174	183	184	3.265	3.265	3.265
			789	84	92	92	1.572	1.572	1.572
			841	2	2	2	1.526	1.526	1.526
			846	88	89	89	2.672	2.672	2.672
			849	92	93	93	5.860	5.860	5.860
			855	3	3	3	2.033	2.033	2.033
		816	820	17	26	26	0.930	0.930	0.930
			857		2	2		0.781	0.781
			1016		3	3		0.319	0.319
		818	828	57	61	61	1.650	1.650	1.650
			841	7	7	7	1.621	1.621	1.621

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
			857	2	2	2	0.705	0.705	0.705
		820	816	17	26	26	0.930	0.930	0.930
			828	4	4	4	1.240	1.240	1.240
			855	6	7	7	0.481	0.481	0.481
			857	5	8	8	1.027	1.027	1.027
		828	785	3	3	3	1.908	1.908	1.908
			818	57	61	61	1.650	1.650	1.650
			820	4	4	4	1.240	1.240	1.240
			927	8	8	8	1.035	1.035	1.035
		841	796	2	2	2	1.526	1.526	1.526
			818	7	7	7	1.621	1.621	1.621
			849	24	24	24	2.873	2.873	2.873
		846	785	20	20	20	1.626	1.626	1.626
			789	6	8	8	1.165	1.165	1.165
			796	88	89	89	2.672	2.672	2.672
			855	9	9	9	1.673	1.673	1.673
			857	15	15	15	3.179	3.179	3.179
			927	2	2	2	1.069	1.069	1.069
		849	796	92	93	93	5.860	5.860	5.860
			841	24	24	24	2.873	2.873	2.873
		855	796	3	3	3	2.033	2.033	2.033
			820	6	7	7	0.481	0.481	0.481
			846	9	9	9	1.673	1.673	1.673
			927	11	12	12	1.171	1.171	1.171
		857	816		2	2		0.781	0.781
			818	2	2	2	0.705	0.705	0.705
			820	5	8	8	1.027	1.027	1.027
			846	15	15	15	3.179	3.179	3.179
		878	885	110	124	124	2.531	2.531	2.531
			895	3	4	4	1.211	1.211	1.211
			927	2	2	2	1.296	1.296	1.296
			942	7	14	15	0.706	0.706	0.706
		885	878	110	124	124	2.531	2.531	2.531
			895	10	14	14	1.182	1.182	1.182
			942	13	15	15	1.297	1.297	1.297
		895	767			2			0.130
			878	3	4	4	1.211	1.211	1.211
			885	10	14	14	1.182	1.182	1.182
		904	944	19	19	19	3.138	3.138	3.138
			955	20	20	20	2.726	2.726	2.726
		927	828	8	8	8	1.035	1.035	1.035
			846	2	2	2	1.069	1.069	1.069
			855	11	12	12	1.171	1.171	1.171
			878	2	2	2	1.296	1.296	1.296
			942	20	34	35	1.217	1.217	1.217
		942	878	7	14	15	0.706	0.706	0.706
			885	13	15	15	1.297	1.297	1.297
			927	20	34	35	1.217	1.217	1.217
		944	904	19	19	19	3.138	3.138	3.138
		955	904	20	20	20	2.726	2.726	2.726
		957	992	8	8	8	2.017	2.017	2.017
		978	982	22	29	29	1.323	1.323	1.323
			985	36	44	44	1.231	1.231	1.231
		982	978	22	29	29	1.323	1.323	1.323
		985	478	19	66	73	0.974	0.974	0.974
			978	36	44	44	1.231	1.231	1.231
		992	957	8	8	8	2.017	2.017	2.017
		1016	756		2	2		0.312	0.312
			785		2	2		0.519	0.519
			816		3	3		0.319	0.319
			1024	203	211	211	2.913	2.913	2.913
			1069	346	358	358	4.508	4.508	4.508
		1024	1016	203	211	211	2.913	2.913	2.913
		1027	1049	10	10	10	2.308	2.308	2.308
		1035	1049	2	2	2	0.918	0.918	0.918
		1049	1027	10	10	10	2.308	2.308	2.308
			1035	2	2	2	0.918	0.918	0.918
		1069	1016	346	358	358	4.508	4.508	4.508
U2-A'	P09661	3	56	14	14	14	3.777	3.777	3.777
		56	3	14	14	14	3.777	3.777	3.777
			179	17	19	19	3.533	3.533	3.533
		128	172	11	12	12	3.018	3.018	3.018
			179	60	61	61	1.970	1.970	1.970
			193	8	8	8	2.056	2.056	2.056
			221	1	1	1	1.036	1.036	1.036
		129	160	58	58	58	2.972	2.972	2.972
			163	15	15	15	5.315	5.315	5.315
			172	90	99	99	2.816	2.816	2.816
			179	188	189	190	4.380	4.380	4.380
			191	70	73	73	3.528	3.528	3.528
			192	12	12	12	2.323	2.323	2.323
			193	53	54	54	2.962	2.962	2.962
			205	53	54	54	2.256	2.256	2.256

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			221	80	81	81	3.263	3.263	3.263
		138	163	10	10	10	0.966	0.966	0.966
			172	73	77	77	3.973	3.973	3.973
			179	8	8	8	1.893	1.893	1.893
		160	129	58	58	58	2.972	2.972	2.972
			172	12	13	13	1.262	1.262	1.262
			179		2	2		0.872	0.872
			221	4	5	5	1.496	1.496	1.496
		163	129	15	15	15	5.315	5.315	5.315
			138	10	10	10	0.966	0.966	0.966
			172	30	34	35	2.004	2.004	2.004
			205		2	2		0.676	0.676
		172	128	11	12	12	3.018	3.018	3.018
			129	90	99	99	2.816	2.816	2.816
			138	73	77	77	3.973	3.973	3.973
			160	12	13	13	1.262	1.262	1.262
			163	30	34	35	2.004	2.004	2.004
			179	430	443	445	5.234	5.234	5.234
			191	62	67	67	2.368	2.368	2.368
			192	6	6	6	1.674	1.674	1.674
			193	45	45	45	3.941	3.941	3.941
			205	32	32	32	2.055	2.055	2.055
			221	55	55	55	4.443	4.443	4.443
		179	56	17	19	19	3.533	3.533	3.533
			128	60	61	61	1.970	1.970	1.970
			129	188	189	190	4.380	4.380	4.380
			138	8	8	8	1.893	1.893	1.893
			160		2	2		0.872	0.872
			172	430	443	445	5.234	5.234	5.234
			193	107	109	109	4.249	4.249	4.249
			205	30	30	30	5.153	5.153	5.153
			221	51	53	53	5.324	5.324	5.324
		191	129	70	73	73	3.528	3.528	3.528
			172	62	67	67	2.368	2.368	2.368
			193	72	73	73	3.849	3.849	3.849
			205	74	78	78	1.932	1.932	1.932
			221	23	26	26	1.553	1.553	1.553
		192	129	12	12	12	2.323	2.323	2.323
			172	6	6	6	1.674	1.674	1.674
			205	11	12	12	0.922	0.922	0.922
			221	1	1	1	0.516	0.516	0.516
		193	128	8	8	8	2.056	2.056	2.056
			129	53	54	54	2.962	2.962	2.962
			172	45	45	45	3.941	3.941	3.941
			179	107	109	109	4.249	4.249	4.249
			191	72	73	73	3.849	3.849	3.849
			205	2	4	4	0.833	0.833	0.833
			221	47	47	47	4.631	4.631	4.631
		205	129	53	54	54	2.256	2.256	2.256
			163		2	2		0.676	0.676
			172	32	32	32	2.055	2.055	2.055
			179	30	30	30	5.153	5.153	5.153
			191	74	78	78	1.932	1.932	1.932
			192	11	12	12	0.922	0.922	0.922
			193	2	4	4	0.833	0.833	0.833
			221	148	151	151	4.753	4.753	4.753
		221	128	1	1	1	1.036	1.036	1.036
			129	80	81	81	3.263	3.263	3.263
			160	4	5	5	1.496	1.496	1.496
			172	55	55	55	4.443	4.443	4.443
			179	51	53	53	5.324	5.324	5.324
			191	23	26	26	1.553	1.553	1.553
			192	1	1	1	0.516	0.516	0.516
			193	47	47	47	4.631	4.631	4.631
			205	148	151	151	4.753	4.753	4.753
U2-B"	P08579	1	111		2	2		0.206	0.206
		17	20			2			0.350
		20	17			2			0.350
		47	111	3	3	3	1.070	1.070	1.070
		57	93	72	75	75	3.367	3.367	3.367
			101	73	78	78	1.924	1.924	1.924
		85	111	6	6	6	3.160	3.160	3.160
		93	57	72	75	75	3.367	3.367	3.367
			101	44	48	48	3.933	3.933	3.933
			103	14	15	15	1.880	1.880	1.880
		101	57	73	78	78	1.924	1.924	1.924
			93	44	48	48	3.933	3.933	3.933
			108	2	2	2	0.877	0.877	0.877
			109		2	2		0.576	0.576
			111	4	5	5	1.321	1.321	1.321
		103	93	14	15	15	1.880	1.880	1.880
			109	2	2	2	0.848	0.848	0.848
			111	23	23	23	2.990	2.990	2.990

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		108	101	2	2	2	0.877	0.877	0.877
		109	101		2	2		0.576	0.576
			103	2	2	2	0.848	0.848	0.848
		111	1		2	2		0.206	0.206
			47	3	3	3	1.070	1.070	1.070
			85	6	6	6	3.160	3.160	3.160
			101	4	5	5	1.321	1.321	1.321
			103	23	23	23	2.990	2.990	2.990
U5-40K	Q96DI7	1	8	91	94	94	4.834	4.834	4.834
			18	46	51	52	3.907	3.907	3.907
			270	6	6	6	1.658	1.658	1.658
			275	86	87	87	3.528	3.528	3.528
			322	23	24	24	2.990	2.990	2.990
		6	8	137	140	142	2.844	2.844	2.844
			18	36	48	51	1.750	1.750	1.750
			270	9	11	11	0.931	0.931	0.931
			275	62	71	71	2.777	2.777	2.777
			322	9	14	15	1.663	1.663	1.663
		8	1	91	94	94	4.834	4.834	4.834
			6	137	140	142	2.844	2.844	2.844
			270	6	6	6	1.809	1.809	1.809
			275	66	66	66	4.785	4.785	4.785
			322	2	2	2	2.667	2.667	2.667
			349	4	4	4	3.251	3.251	3.251
		18	1	46	51	52	3.907	3.907	3.907
			6	36	48	51	1.750	1.750	1.750
			270	8	9	10	0.674	0.674	0.674
			275	126	135	136	4.576	4.576	4.576
			322	45	46	46	15.244	15.244	15.244
			349	8	12	12	2.838	2.838	2.838
		131	148	82	90	90	11.321	11.321	11.321
		148	131	82	90	90	11.321	11.321	11.321
			184	74	74	74	14.767	14.767	14.767
		184	148	74	74	74	14.767	14.767	14.767
		226	270	119	133	133	2.234	2.234	2.234
		270	1	6	6	6	1.658	1.658	1.658
			6	9	11	11	0.931	0.931	0.931
			8	6	6	6	1.809	1.809	1.809
			18	8	9	10	0.674	0.674	0.674
			226	119	133	133	2.234	2.234	2.234
			275	14	21	21	1.905	1.905	1.905
		275	1	86	87	87	3.528	3.528	3.528
			6	62	71	71	2.777	2.777	2.777
			8	66	66	66	4.785	4.785	4.785
			18	126	135	136	4.576	4.576	4.576
			270	14	21	21	1.905	1.905	1.905
		322	1	23	24	24	2.990	2.990	2.990
			6	9	14	15	1.663	1.663	1.663
			8	2	2	2	2.667	2.667	2.667
			18	45	46	46	15.244	15.244	15.244
		349	8	4	4	4	3.251	3.251	3.251
			18	8	12	12	2.838	2.838	2.838
WBP11	Q9Y2W2	46	98	33	33	33	1.999	1.999	1.999
		48	59		2	2		0.937	0.937
			81	5	5	5	0.958	0.958	0.958
		51	109	66	66	66	3.015	3.015	3.015
			117	6	6	6	0.924	0.924	0.924
		59	48		2	2		0.937	0.937
			81	2	2	2	0.966	0.966	0.966
			117	2	2	2	2.937	2.937	2.937
		81	48	5	5	5	0.958	0.958	0.958
			59	2	2	2	0.966	0.966	0.966
		98	46	33	33	33	1.999	1.999	1.999
		105	109	175	180	180	4.827	4.827	4.827
			599		2	2		0.281	0.281
		109	51	66	66	66	3.015	3.015	3.015
			105	175	180	180	4.827	4.827	4.827
			372		2	2		0.213	0.213
		117	51	6	6	6	0.924	0.924	0.924
			59	2	2	2	2.937	2.937	2.937
		168	218	7	7	7	3.202	3.202	3.202
		218	168	7	7	7	3.202	3.202	3.202
		302	316	12	19	19	1.328	1.328	1.328
		303	316	27	28	28	3.472	3.472	3.472
			322		3	4		1.111	1.111
		316	302	12	19	19	1.328	1.328	1.328
			303	27	28	28	3.472	3.472	3.472
			322	24	27	27	2.292	2.292	2.292
			325	14	14	14	1.621	1.621	1.621
		322	303		3	4		1.111	1.111
			316	24	27	27	2.292	2.292	2.292
		325	316	14	14	14	1.621	1.621	1.621
		372	109		2	2		0.213	0.213

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
		556	565	43	45	45	2.552	2.552	2.552
			572	16	17	17	2.536	2.536	2.536
			590	2	2	2	0.917	0.917	0.917
		557	590	9	9	9	2.093	2.093	2.093
			614	2	2	2	1.442	1.442	1.442
		565	556	43	45	45	2.552	2.552	2.552
			590	11	13	13	2.100	2.100	2.100
			614	4	4	4	2.175	2.175	2.175
		572	556	16	17	17	2.536	2.536	2.536
			590	42	48	48	2.624	2.624	2.624
			599	4	4	4	2.865	2.865	2.865
			610	4	4	4	2.320	2.320	2.320
			614	9	9	9	2.576	2.576	2.576
		590	556	2	2	2	0.917	0.917	0.917
			557	9	9	9	2.093	2.093	2.093
			565	11	13	13	2.100	2.100	2.100
			572	42	48	48	2.624	2.624	2.624
			599	95	95	95	4.091	4.091	4.091
			610	98	104	104	2.992	2.992	2.992
			614	36	36	36	2.399	2.399	2.399
		599	105		2	2		0.281	0.281
			572	4	4	4	2.865	2.865	2.865
			590	95	95	95	4.091	4.091	4.091
			614	48	49	49	3.666	3.666	3.666
		610	572	4	4	4	2.320	2.320	2.320
			590	98	104	104	2.992	2.992	2.992
			626	11	11	11	2.141	2.141	2.141
		614	557	2	2	2	1.442	1.442	1.442
			565	4	4	4	2.175	2.175	2.175
			572	9	9	9	2.576	2.576	2.576
			590	36	36	36	2.399	2.399	2.399
			599	48	49	49	3.666	3.666	3.666
			635	3	3	3	1.626	1.626	1.626
		626	610	11	11	11	2.141	2.141	2.141
		635	614	3	3	3	1.626	1.626	1.626
YBOX1	P67809	26	53	13	20	22	1.686	1.686	1.686
			58		2	2		0.896	0.896
			301		4	4		1.629	1.629
		53	26	13	20	22	1.686	1.686	1.686
		58	26		2	2		0.896	0.896
		296	304	25	30	31	1.220	1.220	1.220
		301	26		4	4		1.629	1.629
		304	296	25	30	31	1.220	1.220	1.220
ZC3H18	Q86VM9	149	168	21	22	22	16.541	16.541	16.541
			177	7	8	8	6.115	6.115	6.115
		158	168	113	123	124	11.070	11.070	11.070
			177	5	5	5	2.437	2.437	2.437
			184	2	3	3	0.537	0.537	0.537
			211		2	13		0.234	0.234
			500		2	3		0.153	0.153
		168	149	21	22	22	16.541	16.541	16.541
			158	113	123	124	11.070	11.070	11.070
			184	21	21	21	2.761	2.761	2.761
			188	13	13	13	2.775	2.775	2.775
		177	149	7	8	8	6.115	6.115	6.115
			158	5	5	5	2.437	2.437	2.437
			187	3	3	3	2.201	2.201	2.201
			188	42	51	52	3.867	3.867	3.867
		184	158	2	3	3	0.537	0.537	0.537
			168	21	21	21	2.761	2.761	2.761
			188	180	212	219	5.190	5.190	5.190
		186	188	47	53	55	6.102	6.102	6.102
		187	177	3	3	3	2.201	2.201	2.201
		188	168	13	13	13	2.775	2.775	2.775
			177	42	51	52	3.867	3.867	3.867
			184	180	212	219	5.190	5.190	5.190
			186	47	53	55	6.102	6.102	6.102
		211	158		2	13		0.234	0.234
			217		16	17		0.749	0.749
		217	211		16	17		0.749	0.749
		309	314	26	27	27	2.917	2.917	2.917
			319	9	9	9	2.449	2.449	2.449
		314	309	26	27	27	2.917	2.917	2.917
			319	26	26	26	2.899	2.899	2.899
			328	1	1	1	0.498	0.498	0.498
		319	309	9	9	9	2.449	2.449	2.449
			314	26	26	26	2.899	2.899	2.899
			343		4	4		0.392	0.392
		328	314	1	1	1	0.498	0.498	0.498
		343	319		4	4		0.392	0.392
			838		3	3		0.175	0.175
		450	510	2	2	2	1.017	1.017	1.017
		463	499	2	4	4	0.550	0.550	0.550

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		480	622	11	11	11	1.434	1.434	1.434
		484	499		2	2		0.338	0.338
			500		2	2		0.184	0.184
		499	463	2	4	4	0.550	0.550	0.550
			484		2	2		0.338	0.338
			510	3	4	4	0.941	0.941	0.941
			516	4	5	5	0.841	0.841	0.841
			622		3	3		0.203	0.203
		500	158		2	3		0.153	0.153
			484		2	2		0.184	0.184
			651	2	2	2	1.225	1.225	1.225
		510	450	2	2	2	1.017	1.017	1.017
			499	3	4	4	0.941	0.941	0.941
			516	42	46	46	1.636	1.636	1.636
		516	499	4	5	5	0.841	0.841	0.841
			510	42	46	46	1.636	1.636	1.636
		622	480	11	11	11	1.434	1.434	1.434
			499		3	3		0.203	0.203
			635	25	26	26	2.308	2.308	2.308
			638	67	68	68	3.345	3.345	3.345
			639	80	80	80	3.614	3.614	3.614
			651	62	64	64	2.503	2.503	2.503
			661	18	18	18	1.188	1.188	1.188
			827	4	4	4	0.731	0.731	0.731
		631	635	24	42	46	0.839	0.839	0.839
			638	109	115	115	3.333	3.333	3.333
			639	154	158	161	3.956	3.956	3.956
			651	35	50	53	1.517	1.517	1.517
			661	6	7	7	1.230	1.230	1.230
		635	622	25	26	26	2.308	2.308	2.308
			631	24	42	46	0.839	0.839	0.839
			639	12	14	14	1.925	1.925	1.925
			651	23	37	38	0.761	0.761	0.761
			661	8	11	11	0.701	0.701	0.701
		638	622	67	68	68	3.345	3.345	3.345
			631	109	115	115	3.333	3.333	3.333
			661	38	38	38	1.975	1.975	1.975
			880		2	2		0.364	0.364
			921	2	2	2	1.120	1.120	1.120
		639	622	80	80	80	3.614	3.614	3.614
			631	154	158	161	3.956	3.956	3.956
			635	12	14	14	1.925	1.925	1.925
			661	24	29	29	1.549	1.549	1.549
		651	500	2	2	2	1.225	1.225	1.225
			622	62	64	64	2.503	2.503	2.503
			631	35	50	53	1.517	1.517	1.517
			635	23	37	38	0.761	0.761	0.761
			780	9	9	9	0.732	0.732	0.732
			827		6	6		0.246	0.246
			880	9	9	9	1.608	1.608	1.608
			901	7	7	7	0.926	0.926	0.926
			908	4	6	6	1.105	1.105	1.105
			918	2	2	2	0.762	0.762	0.762
			921	2	3	3	0.717	0.717	0.717
			948		4	4		0.614	0.614
		661	622	18	18	18	1.188	1.188	1.188
			631	6	7	7	1.230	1.230	1.230
			635	8	11	11	0.701	0.701	0.701
			638	38	38	38	1.975	1.975	1.975
			639	24	29	29	1.549	1.549	1.549
			671	2	6	7	0.558	0.558	0.558
			880	2	2	2	0.602	0.602	0.602
		671	661	2	6	7	0.558	0.558	0.558
		758	761	4	5	5	1.375	1.375	1.375
		761	758	4	5	5	1.375	1.375	1.375
		766	780	5	12	12	0.966	0.966	0.966
		769	780	11	13	13	2.064	2.064	2.064
		780	651	9	9	9	0.732	0.732	0.732
			766	5	12	12	0.966	0.966	0.966
			769	11	13	13	2.064	2.064	2.064
			814	14	15	15	1.278	1.278	1.278
			817	9	12	12	1.649	1.649	1.649
		783	814	15	19	19	1.225	1.225	1.225
			817	2	2	2	1.391	1.391	1.391
			823	2	3	3	0.684	0.684	0.684
			827		2	2		0.707	0.707
			880	2	2	2	0.794	0.794	0.794
		789	817	32	32	32	2.229	2.229	2.229
			823	10	11	11	0.990	0.990	0.990
			827	10	11	11	0.985	0.985	0.985
			880	2	2	2	0.894	0.894	0.894
		814	780	14	15	15	1.278	1.278	1.278
			783	15	19	19	1.225	1.225	1.225

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	FDR1%	FDR3%	FDR5%
			823	16	20	20	0.894	0.894	0.894
			827	26	33	34	1.388	1.388	1.388
			861	2	3	4	1.111	1.111	1.111
			880	5	5	5	1.287	1.287	1.287
		817	780	9	12	12	1.649	1.649	1.649
			783	2	2	2	1.391	1.391	1.391
			789	32	32	32	2.229	2.229	2.229
			827	72	78	80	2.225	2.225	2.225
			861	3	3	3	1.821	1.821	1.821
		823	783	2	3	3	0.684	0.684	0.684
			789	10	11	11	0.990	0.990	0.990
			814	16	20	20	0.894	0.894	0.894
			861	3	3	3	2.393	2.393	2.393
		827	622	4	4	4	0.731	0.731	0.731
			651		6	6		0.246	0.246
			783		2	2		0.707	0.707
			789	10	11	11	0.985	0.985	0.985
			814	26	33	34	1.388	1.388	1.388
			817	72	78	80	2.225	2.225	2.225
			880	12	14	14	1.460	1.460	1.460
		838	343		3	3		0.175	0.175
		861	814	2	3	4	1.111	1.111	1.111
			817	3	3	3	1.821	1.821	1.821
			823	3	3	3	2.393	2.393	2.393
			880	36	39	39	3.341	3.341	3.341
			901	3	3	3	2.078	2.078	2.078
			908	7	7	7	1.893	1.893	1.893
			933	4	5	5	0.949	0.949	0.949
		880	638		2	2		0.364	0.364
			651	9	9	9	1.608	1.608	1.608
			661	2	2	2	0.602	0.602	0.602
			783	2	2	2	0.794	0.794	0.794
			789	2	2	2	0.894	0.894	0.894
			814	5	5	5	1.287	1.287	1.287
			827	12	14	14	1.460	1.460	1.460
			861	36	39	39	3.341	3.341	3.341
			901	47	48	48	2.242	2.242	2.242
			908	27	28	28	2.256	2.256	2.256
			918	17	20	20	2.766	2.766	2.766
			921	21	21	21	2.008	2.008	2.008
			933	22	23	23	1.685	1.685	1.685
			936	4	4	4	1.142	1.142	1.142
			948	6	6	6	1.585	1.585	1.585
		901	651	7	7	7	0.926	0.926	0.926
			861	3	3	3	2.078	2.078	2.078
			880	47	48	48	2.242	2.242	2.242
			918	41	42	42	2.204	2.204	2.204
			921	50	50	50	2.203	2.203	2.203
			933	26	28	28	1.629	1.629	1.629
			936	12	12	12	1.897	1.897	1.897
			948	8	8	8	1.292	1.292	1.292
		908	651	4	6	6	1.105	1.105	1.105
			861	7	7	7	1.893	1.893	1.893
			880	27	28	28	2.256	2.256	2.256
			921	51	51	51	3.438	3.438	3.438
			933	16	18	18	1.770	1.770	1.770
			936	10	11	11	1.755	1.755	1.755
			948	15	15	15	1.597	1.597	1.597
		918	651	2	2	2	0.762	0.762	0.762
			880	17	20	20	2.766	2.766	2.766
			901	41	42	42	2.204	2.204	2.204
			933	77	78	78	2.935	2.935	2.935
			936	20	20	20	2.685	2.685	2.685
		921	638	2	2	2	1.120	1.120	1.120
			651	2	3	3	0.717	0.717	0.717
			880	21	21	21	2.008	2.008	2.008
			901	50	50	50	2.203	2.203	2.203
			908	51	51	51	3.438	3.438	3.438
			933	114	118	118	3.182	3.182	3.182
			936	25	25	25	2.277	2.277	2.277
		933	861	4	5	5	0.949	0.949	0.949
			880	22	23	23	1.685	1.685	1.685
			901	26	28	28	1.629	1.629	1.629
			908	16	18	18	1.770	1.770	1.770
			918	77	78	78	2.935	2.935	2.935
			921	114	118	118	3.182	3.182	3.182
			948	3	5	5	1.294	1.294	1.294
		936	880	4	4	4	1.142	1.142	1.142
			901	12	12	12	1.897	1.897	1.897
			908	10	11	11	1.755	1.755	1.755
			918	20	20	20	2.685	2.685	2.685
			921	25	25	25	2.277	2.277	2.277
			948	13	13	13	1.626	1.626	1.626

Continued on next page

Table B.2 – *Continued from previous page*

Name	UniProt ID	Residue 1	Residue 2	FDR 1%	FDR 3%	FDR 5%	F*D*R1%	F*D*R3%	F*D*R5%
		948	651		4	4		0.614	0.614
			880	6	6	6	1.585	1.585	1.585
			901	8	8	8	1.292	1.292	1.292
			908	15	15	15	1.597	1.597	1.597
			933	3	5	5	1.294	1.294	1.294
			936	13	13	13	1.626	1.626	1.626
ZNF830	Q96NB3	2	13			2			0.109
		13	2			2			0.109
		44	86	4	5	5	1.062	1.062	1.062
		86	44	4	5	5	1.062	1.062	1.062
		102	104	20	21	21	2.669	2.669	2.669
			114		3	3		1.735	1.735
			117	8	8	8	2.097	2.097	2.097
		104	102	20	21	21	2.669	2.669	2.669
			117	20	20	20	4.242	4.242	4.242
			149	2	2	2	1.245	1.245	1.245
		114	102		3	3		1.735	1.735
			117	45	52	53	2.534	2.534	2.534
			140	5	6	6	1.236	1.236	1.236
		117	102	8	8	8	2.097	2.097	2.097
			104	20	20	20	4.242	4.242	4.242
			114	45	52	53	2.534	2.534	2.534
			140	4	5	5	0.872	0.872	0.872
		137	149	2	2	2	0.805	0.805	0.805
		140	114	5	6	6	1.236	1.236	1.236
			117	4	5	5	0.872	0.872	0.872
		149	104	2	2	2	1.245	1.245	1.245
			137	2	2	2	0.805	0.805	0.805
			190	3	3	3	1.733	1.733	1.733
		190	149	3	3	3	1.733	1.733	1.733
			216	4	4	4	1.077	1.077	1.077
		216	190	4	4	4	1.077	1.077	1.077
			234	5	6	6	1.232	1.232	1.232
		234	216	5	6	6	1.232	1.232	1.232
		329	333	9	10	10	1.841	1.841	1.841
		331	339	12	14	15	0.996	0.996	0.996
		333	329	9	10	10	1.841	1.841	1.841
		339	331	12	14	15	0.996	0.996	0.996

Data collection			
Microscope model	FEI Titan Krios I		
Voltage (kV)	300		
Additional hardware	CEOS Cs Corrector		
Detector	Falcon III		
Image acquisition mode	Linear (integration)		
Data collection software	MaxWare (Thermo Fisher)		
Magnification	59,000x		
Pixel size (Å)	1.16		
Electron exposure ($e^-/Å^2$)	45		
Frames per micrograph	20		
Electron exposure/micrograph (e-/$Å^2$)	2.25		
Exposure time per micrograph (sec)	1		
Defocus range (μm)	0.9 – 3.3		
Number of datasets	1		
Total number of micrographs collected	10000		
Micrographs used for processing	4676		
Reconstruction	pre-B^{act-1} (core)	pre-B^{act-2} (core)	pre-B^{act-2} (SF3B)
EMDB map entry	11694	11693	11696
PDB coordinate entry	7ABF	7AAV	7ABH
Image processing			
Symmetry imposed	C1	C1	C1
Initial particle images	503208	503208	503208
Final particle images	84539	39336	39336
Resolution (Å)			
- FSC 0.143	3.9	4.2	4.5
Local resolution range (Å)	3.4 – 6.3	3.5 – 6.3	3.9 – 9.2
Map sharpening *B* factor ($Å^2$)	-177	-186	-219
Model refinement statistics			
Model resolution (Å)	4.2	4.2	4.6
Model composition			
Non-hydrogen atoms	25835	25798	17922
Protein residues	3884	4151	3353
Nucleic acid residues	152	181	65
Ligands	3	3	0
Refinement method (PHENIX)	Real-space	Real-space	Real-space
RMS deviations			
Bond lengths (Å)	0.005	0.004	0.005
Bond angles (Å)	1.075	1.094	1.101
Validation			
MolProbity score	1.86	1.59	1.4
Clashscore	4.95	2.84	2.09
Poor rotamers (%)	0.19	0.18	0
C-beta deviations	0	0	0
Ramachandran plot			
Outliers (%)	0.08	0.05	0.25
Allowed (%)	11.82	8.88	6.02
Favored (%)	88.10	91.07	93.73
RNA validation			
Correct sugar puckers (%)	99.34	97.24	95.38
Good backbone conformations (%)	67.76	59.67	67.69
CaBLAM outliers (%)	6.4	6.7	1.19

Fig. B.1: Data collection and model refinement statistics. From [131]. Reprinted with permission from AAAS. Data acquisition was carried out by Dr. Karl Bertram (Department of Structural Dynamics, MPI-BPC).

Name	Component	UniProt ID	Chain (pre-B^{act-1})	Chain (pre-B^{act-2})	Domain/ Region	Total aa/ nt	Modelled (pre-B^{act-1})	Modelled (pre-B^{act-2})	Template	Method
U2 snRNP	A'	P09661	W	W		255	2-163	2-163	6QX9	docked/adjusted
	B''	P08579	B	B		225	3-94	3-94	6QX9	docked/adjusted
						-	149-225	149-225	6FF7	docked/adjusted
	SF3A1	Q15459	p	p		793	-	47-110	6FF7	docked/adjusted
							160-219	160-219	6QX9	docked/adjusted
	SF3A2	Q15428	F	F		464	45-93	19-93	6FF4	docked/adjusted
							-	114-209	6QX9	docked/adjusted
	SF3A3	Q12874	4	4		501	1-362	1-362	6QX9	docked/adjusted
							392-499	392-499	6FF7	docked/adjusted
	SmB	P14678	m	m		240	6-87	6-87	6QX9	docked/adjusted
	SmD3	P62318	l	l		126	2-84	2-84	6QX9	docked/adjusted
	SmD1	P62314	n	n		119	1-80	1-80	6QX9	docked/adjusted
	SmD2	P62316	h	h		118	13-78	13-78	6QX9	docked/adjusted
	SmF	P62306	i	i		86	4-75	4-75	6QX9	docked/adjusted
	SmE	P62304	j	j		92	11-91	11-91	6QX9	docked/adjusted
	SmG	P62308	k	k		76	4-76	4-76	6QX9	docked/adjusted
	U2 snRNA	NR_00271 6.3	2	2		187	3-15	3-31	6FF4	docked/adjusted
							32-64	32-73	5Z58	docked/adjusted
							79-85	79-85	5Z58	docked/adjusted
							97-184	97-184	6Y5Q	docked/adjusted
	SF3B1	O75533	u	u		1304	465-1304	118-1304	5Z58	docked/adjusted
	SF3B2	Q13435	T	T		895	458-545	458-545	6FF4	docked/adjusted
							563-601	563-601	6FF4	docked/adjusted
							602	602	-	de novo
							603-689	603-698	5Z58	docked/adjusted
	SF3B3	Q15393	E	E		1222	1-1222	1-1222	5IFE	docked/adjusted
	SF3B4	Q15427	w	w	RRM	424	12-89	12-89	5Z58	docked/adjusted
	SF3B5	Q9BWJ5	x	x		86	1-79	1-79	6FF4	docked/adjusted
	SF3B6	Q9Y3B4	z	z		125	12-119	12-119	5Z58	docked/adjusted
	PHF5A	Q7RTV0	y	y		110	2-101	2-101	6FF4	docked/adjusted
U5 snRNP	U5-40K	Q96DI7	D	D		357	56-357	56-357	3JCR	docked/adjusted
	BRR2	O75643	s	s		2136	404-2125	404-2125	4KIT	docked/adjusted
	PRP8	Q6P2Q9	A	A	NTD	2335	63-658	63-658	5O9Z	docked/adjusted
					HB		677-799	677-799	5O9Z	docked/adjusted
					RT/En		800-1503	800-1503	5O9Z	docked/adjusted
							1528-1755	1528-1755	5O9Z	docked/adjusted
					RH		1770-2011	1769-2016	5O9Z	docked/adjusted
							2055-2068	2055-2068	-	de novo
					Jab1		2069-2335	2069-2318	4KIT	docked/adjusted
	SNU114	Q15029	r	r		972	114-957	63-957	6FF4	docked/adjusted
	SmB	P14678	f	f		240	7-86	7-86	3JCR	docked/adjusted
	SmD3	P62318	e	e		126	4-81	4-81	3JCR	docked/adjusted
	SmD1	P62314	g	g		119	2-95	2-95	3JCR	docked/adjusted
	SmD2	P62316	a	a		118	26-114	26-114	3JCR	docked/adjusted
	SmF	P62306	b	b		86	4-76	4-76	3JCR	docked/adjusted
	SmE	P62304	c	c		92	15-92	15-92	3JCR	docked/adjusted
	SmG	P62308	d	d		76	4-72	4-72	3JCR	docked/adjusted
	U5 snRNA	X01691.1	5	5		116	3-116	3-116	5O9Z	docked/adjusted
U4/U6 snRNP	U6 snRNA	NR_00439 4.1	6	6		106	1-46	1-46	5Z58	docked/adjusted
							-	47-51	6FF4	rebuilt
							-	52-57	6FF4	docked/adjusted
							58-76	60-75	6FF7	docked/rebuilt
							83-95	78-95	6FF4	docked/adjusted
	LSm2	Q9Y333	V	-		95	1-90	-	5O9Z	docked/adjusted
	LSm3	P62310	9	-		102	16-98	-	5O9Z	docked/adjusted
	LSm4	Q9Y4Z0	C	-		139	1-80	-	5O9Z	docked/adjusted
	LSm5	Q9Y4Y9	H	-		91	10-85	-	5O9Z	docked/adjusted
	LSm6	P62312	J	-		80	8-76	-	5O9Z	docked/adjusted
	LSm7	Q9UK45	A2	-		103	11-89	-	5O9Z	docked/adjusted
	LSm8	O95777	A3	-		96	4-73	-	5O9Z	docked/adjusted

Name	Component	UniProt ID	Chain (pre-B^{act-1})	Chain (pre-B^{act-2})	Domain /Region	Total aa/ nt	Modelled (pre-B^{act-1})	Modelled (pre-B^{act-2})	Template	Method
B-specific	PRP38	Q8NAV1	I	I		312	1-176	1-176	5O9Z	docked/adjusted
	SNU23	Q96NC0	N	N		199	79-134	79-134	5O9Z	docked/adjusted
	MFAP1	P55081	K	K		439	271-313	271-313	5O9Z	docked/adjusted
							314-393	314-393	-	de novo
	UBL5	Q9BZL1	q	q		73	1-73	1-73	6AHD	docked/adjusted
	WBP11	Q9Y2W2	X	-	-	-	24-59	-	SwissModel	docked
hNTC	CDC5L	Q99459	L	L		802	8-110	8-110	6FF4	docked/adjusted
								-	147-159	SpliProt3D	
	PLRG1	O43660	G	G		514	181-500	181-500	6FF4	docked/adjusted
	CWC15	Q9P013	R	R		229	218-226	186-229	6FF4	docked/adjusted
	CTNNBL1	Q8WYA6	-	S		571	-	209-563	4HM9	docked/adjusted
IBC	AQR	O60306	-	U		1485	-	19-1381	5MQF	docked/adjusted
	SYF1	Q9HCS7	-	M		855	-	33-738	5MQF	docked/adjusted
	PPIE	Q9UNP9	-	o		301	-	1-82	6FF7	docked/adjusted
							-	141-160	6FF7	docked/adjusted
							-	166-218	6FF7	docked/adjusted
							-	228-301	6FF7	docked/adjusted
hNTR	SYF3	Q9BZJ0	-	O		848	-	215-467	6FF4	docked/adjusted
	SKIP	Q13573	v	v		536	137-162	129-162	6FF4	docked/adjusted
							237-249	176-255	6FF4	docked/adjusted
							282-313	277-334	6FF4	docked/adjusted
							-	405-438	6FF4	docked/adjusted
	RBM22	Q9NW64	P	P		420	15-220	15-200	6FF4	docked/adjusted
							-	201-308	6FF4	docked/adjusted
	BUD31	P41223	Q	Q		144	3-140	3-140	6FF4	docked/adjusted
Splicing	PPIL2	Q13356	-	t	U box	35-108	-	43-93	SpliProt3D	docked/adjusted
							-	116-159	SpliProt3D	docked/adjusted
					PPIase	278-433	-	280-435	SpliProt3D	docked/adjusted
	SRRM1	Q8IYB3	Y	Y		904	32-126	32-126	6FF4	docked/adjusted
	PRP17	O60508	-	8			-	114-131	6FF4	docked/adjusted
RES	SNU17	Q9Y388	--	1		322	-	2-123	6FF4	docked/adjusted
	SNIP	Q8TAD8	0	0		396	221-370	221-370	6FF7	docked/adjusted
	BUD13	Q9BRD0	--	3		619	-	563-619	6FF4	docked/adjusted
Other	TCERG1	O14776-1	A4	-	FF1-3	1098	664-845	-	SpliProt3D	docked/adjusted
				-	CC		856-893	-	SpliProt3D	docked/adjusted
				-	FF4-6		896-1080	-	4FQG	docked/adjusted
	CBP80	Q09161	A5	-		790	26-790	-	1N52	docked/adjusted
	CBP20	P52298	A1	-		156	6-153	-	1N52	docked/adjusted
	MINX M3	-	Z	Z		324	50-78	50-78	5Z58	docked/adjusted
							141-158	141-158	6FF4	docked/adjusted
							-	159-168	5Z58	docked/adjusted
	SRSF1	Q07955-1	A6	-	RRM2	248	122-195	-	2O3D	docked/adjusted
	KIN17	O60870-1	-	7		393	-	60-142	2CKK	docked/adjusted
							-	147-179	ModBase	docked/adjusted
							-	274-393	2V1N	docked/adjusted
Ligands	GTP		r	r					5Z58	docked/adjusted
	Mg^{2+}		r	r					6FF4	docked/adjusted
	I6P		A	A					6FF4	docked/adjusted
	GTG		A1	A1					1N52	docked

Fig. B.2: Modeling summary for pre-B^{act-1} and pre-B^{act-2}. From [131]. Reprinted with permission from AAAS. Model building was overseen by Dr. Berthold Kastner (Department of Cellular Biochemistry, MPI-BPC) (see 2.2.10).

C References

[1] Crick, F. Central dogma of molecular biology. *Nature* 227, 561–563 (1970). http://doi.org/10.1038/227561a0.

[2] Sharp, P. A. The discovery of split genes and RNA splicing. In *Trends in Biochemical Sciences*, vol. 30, 279–281 (Elsevier, 2005). http://doi.org/10.1016/j.tibs.2005.04.002.

[3] Gilbert, W. Why genes in pieces? (1978). http://doi.org/10.1038/271501a0.

[4] Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5 ends of adenovirus 2 messenger RNA. *Cell* (1977). http://doi.org/10.1016/0092-8674(77)90180-5.

[5] Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA* (adenovirus 2 mRNA processing/5' tails on mRNAs/electron microscopy of mRNA-DNA hybrids). *Biochemistry* 74, 3171–3175 (1977).

[6] Crick, F. Split genes and RNA splicing (1979). http://doi.org/10.1126/science.373120.

[7] Ruis, B. L., Kivens, W. J. & Siliciano, P. G. The interaction between the first and last intron nucleotides in the second step of pre-mrna splicing is independent of other conserved intron nucleotides. *Nucleic Acids Research* 22, 5190–5195 (1994). http://doi.org/10.1093/nar/22.24.5190.

[8] Burset, M., Seledtsov, I. A. & Solovyev, V. V. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research* 28, 4364–4375 (2000). http://doi.org/10.1093/nar/28.21.4364.

[9] Tarn, W. Y. & Steitz, J. A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* 84, 801–811 (1996). http://doi.org/10.1016/S0092-8674(00)81057-0.

[10] Tarn, W. Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273, 1824–1832 (1996). http://doi.org/10.1126/science.273.5283.1824.

[11] Domdey, H. *et al.* Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell* 39, 611–621 (1984). http://doi.org/10.1016/0092-8674(84)90468-9.

[12] Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* 225, 898–903 (1984). http://doi.org/10.1126/science.6206566.

[13] Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* 38, 317–331 (1984). http://doi.org/10.1016/0092-8674(84)90553-1.

[14] Rodriguez, J. R., Pikielny, C. W. & Rosbash, M. In vivo characterization of yeast mRNA processing intermediates. *Cell* 39, 603–610 (1984). http://doi.org/10.1016/0092-8674(84)90467-7.

[15] 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). http://doi.org/10.1261/rna.039024.113.

[16] Costa, M., Walbott, H., Monachello, D., Westhof, E. & Michel, F. Crystal structures of a group II intron lariat primed for reverse splicing. *Science* 354, aaf9258–aaf9258 (2016). http://doi.org/10.1126/science.aaf9258.

[17] Heath, C. G., Viphakone, N. & Wilson, S. A. The role of TREX in gene expression and disease (2016). http://doi.org/10.1042/BCJ20160010.

[18] Alt, F. W. *et al.* Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3 ends. *Cell* 20, 293–301 (1980). http://doi.org/10.1016/0092-8674(80)90615-7.

[19] Early, P. *et al.* Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* 20, 313–319 (1980). http://doi.org/10.1016/0092-8674(80)90617-0.

[20] Schmucker, D. *et al.* Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671–684 (2000). http://doi.org/10.1016/S0092-8674(00)80878-8.

[21] Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing (2010). http://doi.org/10.1038/nature08909.

[22] Black, D. L. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry* 72, 291–336 (2003). http://doi.org/10.1146/annurev.biochem.72.121801.161720.

[23] Wang, Z. & Burge, C. B. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code (2008). http://doi.org/10.1261/rna.876308.

[24] Juneau, K., Nislow, C. & Davis, R. W. Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. *Genetics* 183, 185–194 (2009). http://doi.org/10.1534/genetics.109.105155.

[25] Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E. & Vilardell, J. Deciphering 3'ss Selection in the Yeast Genome Reveals an RNA Thermosensor that Mediates Alternative Splicing. *Molecular Cell* 43, 1033–1039 (2011). http://doi.org/10.1016/j.molcel.2011.07.030.

[26] Mishra, S. K. *et al.* Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. *Nature* 474, 173–180 (2011). http://doi.org/10.1038/nature10143.

[27] Aslanzadeh, V., Huang, Y., Sanguinetti, G. & Beggs, J. D. Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. *Genome Research* 28, 203–213 (2018). http://doi.org/10.1101/gr.225615.117.

[28] Pertea, M. *et al.* CHESS: A new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biology* 19, 1–14 (2018). http://doi.org/10.1186/s13059-018-1590-2.

[29] Chen, J. & Weiss, W. A. Alternative splicing in cancer: Implications for biology and therapy. *Oncogene* 34, 1–14 (2015). http://doi.org/10.1038/onc.2013.570.

[30] Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nature Reviews Molecular Cell Biology* 10, 741–754 (2009). http://doi.org/10.1038/nrm2777.

[31] Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nature Reviews Genetics* 17, 19–32 (2016). http://doi.org/10.1038/nrg.2015.3.

[32] Dreyfuss, G. hnRNP proteins and the biogenesis of mRNA. In *Annu. Rev. Biochem.* (1993). http://doi.org/10.1109/IEEM.2013.6962450.

[33] Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. SR proteins: A conserved family of pre-mRNA splicing factors. *Genes and Development* 6, 837–847 (1992). http://doi.org/10.1101/gad.6.5.837.

[34] Lee, Y. & Rio, D. C. Mechanisms and regulation of alternative Pre-mRNA splicing (2015). http://doi.org/10.1146/annurev-biochem-060614-034316.

[35] Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors (2012). http://doi.org/10.1002/wrna.100.

[36] Martinez-Contreras, R. *et al.* hnRNP proteins and splicing control. (2007). http://doi.org/10.1007/978-0-387-77374-2_8.

[37] Conway, G., Wooley, J., Bibring, T. & LESTOURGEONl, W. M. Ribonucleoproteins Package 700 Nucleotides of Pre-mRNA into a Repeating Array of Regular Particles. Tech. Rep. 7 (1988).

[38] Choi, Y. D., Grabowski, P. J., Sharp, P. A. & Dreyfuss, G. Heterogeneous nuclear ribonucleoproteins: Role in RNA splicing. *Science* 231, 1534–1539 (1986). http://doi.org/10.1126/science.3952495.

[39] Wang, Y., Ma, M., Xiao, X. & Wang, Z. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nature Structural and Molecular Biology* 19, 1044–1053 (2012). http://doi.org/10.1038/nsmb.2377.

[40] Lührmann, R., Kastner, B. & Bach, M. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing (1990). http://doi.org/10.1016/0167-4781(90)90001-I.

[41] Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* 3, a003707 (2011). http://doi.org/10.1101/cshperspect.a003707.

[42] Kunkel, G. R., Maser, R. L., Calvet, J. P. & Pederson, T. U6 small nuclear RNA is transcribed by RNA polymerase III. *Proceedings of the National Academy of Sciences of the United States of America* 83, 8575–8579 (1986). http://doi.org/10.1073/pnas.83.22.8575.

[43] Hamm, J. & Mattaj, I. W. Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* 63, 109–118 (1990). http://doi.org/10.1016/0092-8674(90)90292-M.

[44] Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation (1997). http://doi.org/10.1101/gad.11.21.2755.

[45] Vankan, P., Mcguigan, C. & Mattaj, W. Domains of U4 and U6 snRNAs required for snRNP assembly and splicing complementation in Xenopus oocytes. Tech. Rep. 10 (1990).

[46] Pessa, H. K. *et al.* Minor spliceosome components are predominantly localized in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 105, 8655–8660 (2008). http://doi.org/10.1073/pnas.0803646105.

[47] Branlant, C. *et al.* U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *The EMBO Journal* 1, 1259–1265 (1982). http://doi.org/10.1002/j.1460-2075.1982.tb00022.x.

[48] Raker, V. A., Plessel, G. & Luhrmann1, R. The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. Tech. Rep. 9 (1996).

[49] Urlaub, H., Raker, V. A., Kostka, S. & Lührmann, R. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO Journal* 20, 187–196 (2001). http://doi.org/10.1093/emboj/20.1.187.

[50] Yuo, C.-Y., Ares, M. & Weiner, A. M. Sequences Required for 3' End Formation of Human U2 Small Nuclear RNA. Tech. Rep. (1985).

[51] Mattaj, I. W. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* 46, 905–911 (1986). http://doi.org/10.1016/0092-8674(86)90072-3.

[52] Neuman de Vegvar, H. E. & Dahlberg, J. E. Nucleocytoplasmic transport and processing of small nuclear RNA precursors. *Molecular and Cellular Biology* 10, 3365–3375 (1990). http://doi.org/10.1128/mcb.10.7.3365.

[53] Hamm, J., Darzynkiewicz, E., Tahara, S. M. & Mattaj, I. W. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* 62, 569–577 (1990). http://doi.org/10.1016/0092-8674(90)90021-6.

[54] Fischer, U. & Lührmann, R. An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science* 249, 786–790 (1990). http://doi.org/10.1126/science.2143847.

[55] Hetzer, M. & Mattaj, I. W. An ATP-dependent, Ran-independent mechanism for nuclear import of the U1A and U2B' spliceosome proteins. *Journal of Cell Biology* 148, 293–303 (2000). http://doi.org/10.1083/jcb.148.2.293.

[56] Patel, S. B. & Bellini, M. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Research* 36, 6482–6493 (2008). http://doi.org/10.1093/nar/gkn658.

[57] Achsel, T. *et al.* 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 Journal* 18, 5789–5802 (1999). http://doi.org/10.1093/emboj/18.20.5789.

[58] Reddy, R. & Busch, H. Small Nuclear RNAs: RNA Sequences, Structure, and Modifications. In *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*, 1–37 (Springer Berlin Heidelberg, 1988). http://doi.org/10.1007/978-3-642-73020-7_1.

[59] Bohnsack, M. T. & Sloan, K. E. Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biological Chemistry* 399, 1265–1276 (2018). http://doi.org/10.1515/hsz-2018-0205.

[60] Yu, Y. T., Shu, M. D. & Steitz, J. A. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO Journal* 17, 5783–5795 (1998). http://doi.org/10.1093/emboj/17.19.5783.

[61] Newby, M. I. & Greenbaum, N. L. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nature Structural Biology* 9, 958–965 (2002). http://doi.org/10.1038/nsb873.

[62] Rueter, S. M., Dawson, T. R. & Emeson, R. B. Regulation of alternative splicing by RNA editing. *Nature* 399, 75–80 (1999). http://doi.org/10.1038/19992.

[63] Melcher, T. *et al.* A mammalian RNA editing enzyme. *Nature* 379, 460–464 (1996). http://doi.org/10.1038/379460a0.

[64] Tarn, W. Y. Site-specific substitution of inosine at the terminal positions of a pre-mRNA intron: Implications for the configuration of the terminal base interaction. *Biochimie* 78, 1057–1065 (1996). http://doi.org/10.1016/S0300-9084(97)86730-2.

[65] Gerber, A., O'Connell, M. A. & Keller, W. Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. *RNA* 3, 453–463 (1997).

[66] Mendel, M. *et al.* Article Splice site m 6 A methylation prevents binding of U2AF35 to inhibit RNA splicing Splice site m 6 A methylation prevents binding of U2AF35 to inhibit RNA splicing 1–18 (2021). http://doi.org/10.1016/j.cell.2021.03.062.

[67] McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–360 (1997). http://doi.org/10.1038/385357a0.

[68] Carrillo Oesterreich, F. *et al.* Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* 165, 372–381 (2016). http://doi.org/10.1016/j.cell.2016.02.045.

[69] Zhuang, Y. & Weiner, A. M. A compensatory base change in U1 snRNA suppresses a 5 splice site mutation. *Cell* 46, 827–835 (1986). http://doi.org/10.1016/0092-8674(86)90064-4.

[70] Ruskin, B., Zamore, P. D. & Green, M. R. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52, 207–219 (1988). http://doi.org/10.1016/0092-8674(88)90509-0.

[71] Berg, M. G. *et al.* U1 snRNP determines mRNA length and regulates isoform expression. *Cell* (2012). http://doi.org/10.1016/j.cell.2012.05.029.

[72] Venters, C. C., Oh, J.-m., Di, C., So, B. R. & Dreyfuss, G. U1 snRNP Telescripting : Suppression of Premature Transcription Termination in Introns as a New Layer of Gene Regulation (2019). http://doi.org/10.1101/cshperspect.a032235.

[73] Zhang, S. *et al.* Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* 371, 305–309 (2021). http://doi.org/10.1126/science.abf1870.

[74] Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182–185 (2002). http://doi.org/10.1038/nature01031.

[75] Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Research* 12, 1231–1245 (2002). http://doi.org/10.1101/gr.473902.

[76] Kornblihtt, A. R., De La Mata, M., Fededa, J. P., Muñoz, M. J. & Nogués, G. Multiple links between transcription and splicing (2004). http://doi.org/10.1261/rna.7100104.

[77] McKay, S. L. & Johnson, T. L. A bird's-eye view of post-translational modifications in the spliceosome and their roles in spliceosome dynamics (2010). http://doi.org/10.1039/c002828b.

[78] Mermoud, J. E., Cohen, P. T. & Lamond, A. I. Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO Journal* 13, 5679–5688 (1994). http://doi.org/10.1002/j.1460-2075.1994.tb06906.x.

[79] Duncan, P. I., Stojdl, D. F., Marius, R. M. & Bell, J. C. In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Molecular and Cellular Biology* 17, 5996–6001 (1997). http://doi.org/10.1128/mcb.17.10.5996.

[80] Manley, J. L. & Tacke, R. SR proteins and splicing control (1996). http://doi.org/10.1101/gad.10.13.1569.

[81] Wang, C. *et al.* Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes and Development* 12, 1409–1414 (1998). http://doi.org/10.1101/gad.12.10.1409.

[82] Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840 (2009). http://doi.org/10.1126/science.1175371.

[83] Kuhn, A. N., Van Santen, M. A., Schwienhorst, A., Urlaub, H. & Lührmann, R. Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA* 15, 153–175 (2009). http://doi.org/10.1261/rna.1332609.

[84] Bellare, P. *et al.* A role for ubiquitin in the spliceosome assembly pathway. *Nature Structural and Molecular Biology* 15, 444–451 (2008). http://doi.org/10.1038/nsmb.1401.

[85] Michaud, S. & Reed, R. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes and Development* 5, 2534–2546 (1991). http://doi.org/10.1101/gad.5.12b.2534.

[86] Mount, S. M., Pettersson, L., Hinterberger, M., Karmas, A. & Steitz, J. A. The Ul Small Nuclear RNA-Protein Complex Selectively Binds a 5' Splice Site In Vitro. Tech. Rep. (1983).

[87] Query, C. C., McCaw, P. S. & Sharp, P. A. A minimal spliceosomal complex A recognizes the branch site and polypyrimidine tract. *Molecular and Cellular Biology* 17, 2944–2953 (1997). http://doi.org/10.1128/mcb.17.5.2944.

[88] Pruzan, R., Furneaux, H., Lassota, P., Hong, G. Y. & Hurwitz, J. Assemblage of the prespliceosome complex with separated fractions isolated from HeLa cells. *Journal of Biological Chemistry* 265, 2804–2813 (1990).

[89] Boesler, C. *et al.* ARTICLE A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nature Communications* 7 (2016). http://doi.org/10.1038/ncomms11997.

[90] Staley, J. P. & Guthrie, C. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Molecular Cell* 3, 55–64 (1999). http://doi.org/10.1016/S1097-2765(00)80174-4.

[91] Fortner, D. M., Troy, R. G. & Brow^, D. A. A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. Tech. Rep. (1994).

[92] Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Current Biology* 8, 847–855 (1998). http://doi.org/10.1016/S0960-9822(07)00345-4.

[93] Mefford, M. A. & Staley, J. P. Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* 15, 1386–1397 (2009). http://doi.org/10.1261/rna.1582609.

[94] Agafonov, D. E. *et al.* Semiquantitative Proteomic Analysis of the Human Spliceosome via a Novel Two-Dimensional Gel Electrophoresis Method. *Molecular and Cellular Biology* 31, 2667–2682 (2011). http://doi.org/10.1128/MCB.05266-11.

[95] Sidarovich, A. *et al.* Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. *eLife* 6 (2017). http://doi.org/10.7554/eLife.23533.

[96] Warkocki, Z. *et al.* Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nature Structural & Molecular Biology* 16 (2009). http://doi.org/10.1038/nsmb.1729.

[97] Makarov, E. M. *et al.* Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205–2208 (2002). http://doi.org/10.1126/science.1077783.

[98] Bertram, K. *et al.* Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* 170, 701–713.e11 (2017). http://doi.org/10.1016/j.cell.2017.07.011.

[99] Haselbach, D. *et al.* Structure and Conformational Dynamics of the Human Spliceosomal B Structure and Conformational Dynamics of the Human Spliceosomal B act Complex. *Cell* 172, 1–11 (2018). http://doi.org/10.1016/j.cell.2018.01.010.

[100] Wan, R., Bai, R., Yan, C., Lei, J. & Shi, Y. Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching. *Cell* 177, 339–351.e13 (2019). http://doi.org/10.1016/j.cell.2019.02.006.

[101] 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). http://doi.org/10.1038/nature06842.

[102] Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* 349, 494–499 (1991). http://doi.org/10.1038/349494a0.

[103] 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). http://doi.org/10.1038/349487a0.

[104] Arenas, J. E. & Abelson, J. N. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proceedings of the National Academy of Sciences of the United States of America* 94, 11798–802 (1997).

[105] Yan, C., Wan, R. & Shi, Y. Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome. *Cold Spring Harbor Perspectives in Biology* 11, a032409 (2019). http://doi.org/10.1101/cshperspect.a032409.

[106] Pyle, A. M. Group II Intron Self-Splicing. *Annual Review of Biophysics* 45, 183–205 (2016). http://doi.org/10.1146/annurev-biophys-062215-011149.

[107] Linder, P. *et al.* Birth of the D-E-A-D box (1989). http://doi.org/10.1038/337121a0.

[108] Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. The DEAD-box protein family of RNA helicases (2006). http://doi.org/10.1016/j.gene.2005.10.019.

[109] Liu, Y. C. & Cheng, S. C. Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing (2015). http://doi.org/10.1186/s12929-015-0161-z.

[110] O'Day, C. L., Dalbadie-McFarland, G. & Abelson, J. The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *Journal of Biological Chemistry* 271, 33261–33267 (1996). http://doi.org/10.1074/jbc.271.52.33261.

[111] Fleckner, J., Zhang, M., Valcárcel, J. & Green, M. R. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes and Development* 11, 1864–1872 (1997). http://doi.org/10.1101/gad.11.14.1864.

[112] Zhang, Z. *et al.* Molecular architecture of the human 17S U2 snRNP. *Nature* 583, 310–313 (2020). http://doi.org/10.1038/s41586-020-2344-3.

[113] Boesler, C. *et al.* A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nature Communications* 7, 1–12 (2016). http://doi.org/10.1038/ncomms11997.

[114] Teigelkamp, S., Mundt, C., Achsel, T., Will, C. L. & Lührmann, R. The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *RNA* 3, 1313–1326 (1997).

[115] Madhani, H. D. & Guthrie, C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71, 803–817 (1992). http://doi.org/10.1016/0092-8674(92)90556-R.

[116] Bao, P., Höbartner, C., Hartmuth, K. & Lührmann, R. Yeast Prp2 liberates the 5 splice site and the branch site adenosine for catalysis of pre-mRNA splicing. *RNA* 23, 1770–1779 (2017). http://doi.org/10.1261/rna.063115.117.

[117] Tseng, C. K. & Cheng, S. C. Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science* 320, 1782–1784 (2008). http://doi.org/10.1126/science.1158993.

[118] Bertram, K. *et al.* Structural Insights into the Roles of Metazoan-Specific Splicing Factors in the Human Step 1 Spliceosome. *Molecular Cell* 80, 127–139.e6 (2020). http://doi.org/10.1016/j.molcel.2020.09.012.

[119] Fica, S. M., Oubridge, C., Wilkinson, M. E., Newman, A. J. & Nagai, K. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* 363, 710–714 (2019). http://doi.org/10.1126/science.aaw5569.

[120] Zhang, X. *et al.* Structures of the human spliceosomes before and after release of the ligated exon. *Cell Research* 29, 274–285 (2019). http://doi.org/10.1038/s41422-019-0143-x.

[121] Arenas, J. E. & Abelson, J. N. Prp43: An RNA helicase-like factor involved in spliceosome dissassembly. *Proceedings of the National Academy of Sciences of the United States of America* 94, 11798–11802 (1997). http://doi.org/10.1073/pnas.94.22.11798.

[122] Kastner, B., Will, C. L., Stark, H. & Lührmann, R. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes (2019). http://doi.org/10.1101/cshperspect.a032417.

[123] Mozaffari-Jovin, S. *et al.* The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes and Development* 26, 2422–2434 (2012). http://doi.org/10.1101/gad.200949.112.

[124] Nguyen, T. H. D. *et al.* Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure* 21, 910–919 (2013). http://doi.org/10.1016/j.str.2013.04.017.

[125] Mozaffari-Jovin, S. *et al.* Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* 341, 80–84 (2013). http://doi.org/10.1126/science.1237515.

[126] Henning, L. M. *et al.* A new role for FBP21 as regulator of Brr2 helicase activity. *Nucleic Acids Research* 45, 7922–7937 (2017). http://doi.org/10.1093/nar/gkx535.

[127] Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229–234 (2013). http://doi.org/10.1038/nature12734.

[128] Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nature Structural and Molecular Biology* 21, 464–471 (2014). http://doi.org/10.1038/nsmb.2815.

[129] Anokhina, M. *et al.* RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO Journal* 32, 2804–2818 (2013). http://doi.org/10.1038/emboj.2013.198.

[130] Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA (phosphoryl transfer mechanism/ribozyme/group I splicing/spliceosome/group H splicing). Tech. Rep. (1993).

[131] Townsend, C. *et al.* Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. *Science* 370 (2020). http://doi.org/10.1126/science.abc3753.

[132] Chan, R. T. *et al.* Structural basis for the second step of group II intron splicing. *Nature Communications* 9, 1–10 (2018). http://doi.org/10.1038/s41467-018-06678-0.

[133] Lambowitz, A. M. & Zimmerly, S. Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology* 3, 1–19 (2011). http://doi.org/10.1101/cshperspect.a003616.

[134] Lambowitz, A. M. & Belfort, M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. In *Mobile DNA III*, vol. 3, 1209–1236 (American Society of Microbiology, 2015). http://doi.org/10.1128/microbiolspec.mdna3-0050-2014.

[135] Michel, F., Kazuhiko, U. & Haruo, O. Comparative and functional anatomy of group II catalytic introns - a review. *Gene* 82, 5–30 (1989). http://doi.org/10.1016/0378-1119(89)90026-7.

[136] Pyle, A. M. The tertiary structure of group II introns: Implications for biological function and evolution (2010). http://doi.org/10.3109/10409231003796523.

[137] Jacquier, A. & Michel, F. Base-pairing interactions involving the 5 and 3-terminal nucleotides of group II self-splicing introns. *Journal of Molecular Biology* 213, 437–447 (1990). http://doi.org/10.1016/S0022-2836(05)80206-2.

[138] De Lencastre, A. & Pyle, A. M. Three essential and conserved regions of the group II intron are proximal to the 5-splice site. *RNA* 14, 11–24 (2008). http://doi.org/10.1261/rna.774008.

[139] Madhani, H. D. snRNA catalysts in the spliceosome's ancient core. *Cell* 155, 1213–1215 (2013). http://doi.org/10.1016/j.cell.2013.11.022.

[140] Nguyen, T. H. D. *et al.* CryoEM structures of two spliceosomal complexes: Starter and dessert at the spliceosome feast. *Current Opinion in Structural Biology* 36, 48–57 (2016). http://doi.org/10.1016/j.sbi.2015.12.005.

[141] Gordon, P. M., Fong, R. & Piccirilli, J. A. A Second Divalent Metal Ion in the Group II Intron Reaction Center. *Chemistry and Biology* 14, 607–612 (2007). http://doi.org/10.1016/j.chembiol.2007.05.008.

[142] Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. *Cell* 151, 497–507 (2012). http://doi.org/10.1016/j.cell.2012.09.033.

[143] Keating, K. S., Toor, N., Perlman, P. S. & Pyle, A. M. A structural analysis of the group II intron active site and implications for the spliceosome (2010). http://doi.org/10.1261/rna.1791310.

[144] Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* 320, 77–82 (2008). http://doi.org/10.1126/science.1153803.

[145] Grainger, R. J. & Beggs, J. D. Prp8 protein: At the heart of the spliceosome (2005). http://doi.org/10.1261/rna.2220705.

[146] Hoddges, P. E., Jackson, S. P., Brown, J. D. & Beggs, J. D. Extraordinary sequence conservation of the PRP8 splicing factor. *Yeast* 11, 337–342 (1995). http://doi.org/10.1002/yea.320110406.

[147] Fabrizio, P. *et al.* The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Molecular Cell* 36, 593–608 (2009). http://doi.org/10.1016/j.molcel.2009.09.040.

[148] Dubochet, J. & McDowall, A. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *Journal of Microscopy* 124, 3–4 (1981). http://doi.org/10.1111/j.1365-2818.1981.tb02483.x.

[149] Adrian, M., Dubochet, J., Lepault, J. & McDowall, A. W. Cryo-electron microscopy of viruses. *Nature* (1984). http://doi.org/10.1038/308032a0.

[150] Unwin, P. N. T. Phase contrast and interference microscopy with the electron microscope. *Phil. Trans. Roy. Soc. Lond. B* 261, 95–104 (1971).

[151] Van Heel, M. *et al.* Single-particle electron cryo-microscopy: Towards atomic resolution. *Quarterly Reviews of Biophysics* 33, 307–369 (2000). http://doi.org/10.1017/S0033583500003644.

[152] Frank, J. J. *Three-dimensional electron microscopy of macromolecular assemblies : visualization of biological molecules in their native state* (Oxford University Press, 2006).

[153] Wade, R. A brief look at imaging and contrast transfer. *Ultramicroscopy* 46, 145–156 (1992). http://doi.org/10.1016/0304-3991(92)90011-8.

[154] Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics* 28, 171–193 (1995).

[155] Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* 177, 630–637 (2012). http://doi.org/10.1016/J.JSB.2012.02.003.

[156] Naydenova, K., Jia, P. & Russo, C. J. Cryo-EM with sub–1 Å specimen movement. *Science* 370, 223–226 (2020). http://doi.org/10.1126/science.abb7927.

[157] Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* 10, 584–590 (2013). http://doi.org/10.1038/nmeth.2472.

[158] Stark, H., Zemlin, F. & Boettcher, C. Electron radiation damage to protein crystals of bacteriorhodopsin at different temperatures. *Ultramicroscopy* 63, 75–79 (1996). http://doi.org/10.1016/0304-3991(96)00045-9.

[159] Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* 4, e06980 (2015). http://doi.org/10.7554/eLife.06980.

[160] Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Publishing Group* 14 (2017). http://doi.org/10.1038/nmeth.4193.

[161] van Heel, M. & Frank, J. Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy* 6, 187–94 (1981).

[162] DE ROSIER, D. J. & KLUG, A. Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature* 217, 130–134 (1968). http://doi.org/10.1038/217130a0.

[163] Van Heel, M. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* 21, 111–123 (1987). http://doi.org/10.1016/0304-3991(87)90078-7.

[164] Sigworth, F. J. Principles of cryo-EM single-particle image processing. *Microscopy (Oxford, England)* 65, 57–67 (2016). http://doi.org/10.1093/jmicro/dfv370.

[165] Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human g-secretase. *eLife* 4, e11182 (2015). http://doi.org/10.7554/eLife.11182.

[166] Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* 7, 1–18 (2018). http://doi.org/10.7554/eLife.36861.

[167] Haselbach, D. *et al.* Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nature Communications* 8, 1–8 (2017). http://doi.org/10.1038/ncomms15578.

[168] W Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* 9 (2012). http://doi.org/10.1038/nmeth.2115.

[169] Rosenthal, P. B. & Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology* 333, 721–745 (2003). http://doi.org/10.1016/J.JMB.2003.07.013.

[170] van Heel, M. & Schatz, M. Fourier shell correlation threshold criteria. *Journal of Structural Biology* 151, 250–262 (2005). http://doi.org/10.1016/J.JSB.2005.05.009.

[171] Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* 11, 63–65 (2014). http://doi.org/10.1038/nmeth.2727.

[172] Topf, M. & Sali, A. Combining electron microscopy and comparative protein structure modeling. *Current Opinion in Structural Biology* 15, 578–585 (2005). http://doi.org/10.1016/j.sbi.2005.08.001.

[173] Urlaub, H., Hartmuth, K. & Lührmann, R. A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods* 26, 170–181 (2002). http://doi.org/10.1016/S1046-2023(02)00020-8.

[174] Schmidt, C. & Urlaub, H. Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies. *Current Opinion in Structural Biology* 46, 157–168 (2017). http://doi.org/10.1016/j.sbi.2017.10.005.

[175] Murshudov, G. Refinement of Atomic Structures Against cryo-EM Maps. *Methods in Enzymology* 579, 277–305 (2016). http://doi.org/10.1016/BS.MIE.2016.05.033.

[176] Beckers, M., Mann, D. & Sachse, C. Structural interpretation of cryo-EM image reconstructions. *Progress in Biophysics and Molecular Biology* 160, 26–36 (2021). http://doi.org/10.1016/j.pbiomolbio.2020.07.004.

[177] Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 resolution. *Nature* 458, 475–480 (2009). http://doi.org/10.1038/nature07851.

[178] Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M. C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO Journal* 29, 4172–4184 (2010). http://doi.org/10.1038/emboj.2010.295.

[179] Kondo, Y., Oubridge, C., van Roon, A. M. M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *eLife* 4, 1–19 (2015). http://doi.org/10.7554/eLife.04986.

[180] 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). http://doi.org/10.1038/nature11843.

[181] Santos, K. F. *et al.* Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. *Proceedings of the National Academy of Sciences of the United States of America* 109, 17418–17423 (2012). http://doi.org/10.1073/pnas.1208098109.

[182] Kühlbrandt, W. The resolution revolution (2014). http://doi.org/10.1126/science.1251652.

[183] Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* 587, 152–156 (2020). http://doi.org/10.1038/s41586-020-2829-0.

[184] Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* 587, 157–161 (2020). http://doi.org/10.1038/s41586-020-2833-4.

[185] Li, X. *et al.* A unified mechanism for intron and exon definition and back-splicing. *Nature* 573, 375–380 (2019). http://doi.org/10.1038/s41586-019-1523-6.

[186] Plaschka, C., Newman, A. J. & Nagai, K. Structural Basis of Nuclear pre-mRNA Splicing : Lessons from Yeast 1–21 (2019). http://doi.org/10.1101/cshperspect.a032391.

[187] Brown, J. D. & Beggs, J. D. Roles of PRP8 protein in the assembly of splicing complexes. *EMBO Journal* 11, 3721–3729 (1992). http://doi.org/10.1002/j.1460-2075.1992.tb05457.x.

[188] Pena, V., Rozov, A., Fabrizio, P., Lührmann, R. & Wahl, M. C. Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO Journal* 27, 2929–2940 (2008). http://doi.org/10.1038/emboj.2008.209.

[189] Whittaker, E. & Beggs, J. D. The yeast PRP8 protein interacts directly with pre-mRNA. *Nucleic Acids Research* 19, 5483–5489 (1991). http://doi.org/10.1093/nar/19.20.5483.

[190] Maeder, C., Kutach, A. K. & Guthrie, C. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nature Structural and Molecular Biology* 16, 42–48 (2009). http://doi.org/10.1038/nsmb.1535.

[191] Mozaffari-Jovin, S. *et al.* Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* 341, 80–84 (2013). http://doi.org/10.1126/science.1237515.

[192] Ritchie, D. B. *et al.* Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nature Structural and Molecular Biology* 15, 1199–1205 (2008). http://doi.org/10.1038/nsmb.1505.

[193] Mozaffari-Jovin, S. *et al.* The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes and Development* 26, 2422–2434 (2012). http://doi.org/10.1101/gad.200949.112.

[194] Mayerle, M. *et al.* Structural toggle in the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. *Proceedings of the National Academy of Sciences of the United States of America* 114, 4739–4744 (2017). http://doi.org/10.1073/pnas.1701462114.

[195] Agafonov, D. E. *et al.* Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* 351 (2016). http://doi.org/10.1126/science.aad2085.

[196] Bertram, K. *et al.* Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* 170, 701–713.e11 (2017). http://doi.org/10.1016/J.CELL.2017.07.011.

[197] Bertram, K. *et al.* Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature Publishing Group* 542 (2017). http://doi.org/10.1038/nature21079.

[198] Haselbach, D. *et al.* Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. *Cell* 172, 454–464.e11 (2018). http://doi.org/10.1016/j.cell.2018.01.010.

[199] Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structure of a human catalytic step I spliceosome. *Science* 359, 537–545 (2018). http://doi.org/10.1126/science.aar6401.

[200] Jia, X. & Sun, C. SURVEY AND SUMMARY Structural dynamics of the N-terminal domain and the Switch loop of Prp8 during spliceosome assembly and activation. *Nucleic Acids Research* 46, 3833–3840 (2018). http://doi.org/10.1093/nar/gky242.

[201] Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5' splice site transfer for human spliceosome activation. *Science* 364, 362–367 (2019). http://doi.org/10.1126/science.aax3289.

[202] Agafonov, D. E. *et al.* Semiquantitative Proteomic Analysis of the Human Spliceosome via a Novel Two-Dimensional Gel Electrophoresis Method. *Molecular and Cellular Biology* 31, 2667–2682 (2011). http://doi.org/10.1128/mcb.05266-11.

[203] Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Research* (2018). http://doi.org/10.1038/s41422-018-0094-7.

[204] Schütze, T. *et al.* Multiple protein-protein interactions converging on the Prp38 protein during activation of the human spliceosome. *Rna* 22, 265–277 (2016). http://doi.org/10.1261/rna.054296.115.

[205] Keiper, S. *et al.* Smu1 and RED are required for activation of spliceosomal B complexes assembled on short introns. *Nature Communications* 10, 1–15 (2019). http://doi.org/10.1038/s41467-019-11293-8.

[206] Dziembowski, A. *et al.* Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO Journal* 23, 4847–4856 (2004). http://doi.org/10.1038/sj.emboj.7600482.

[207] Wysoczanski, P. *et al.* Cooperative structure of the heterotrimeric pre-mRNA retention and splicing complex. *Nature Structural and Molecular Biology* 21, 911–918 (2014). http://doi.org/10.1038/nsmb.2889.

[208] Ohrt, T. *et al.* Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* 18, 1244–1256 (2012). http://doi.org/10.1261/rna.033316.112.

[209] Schneider, C. *et al.* Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3' Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes. *PLoS Genetics* 11, 1–27 (2015). http://doi.org/10.1371/journal.pgen.1005539.

[210] Gottschalk, A., Bartels, C., Neubauer, G., Lührmann, R. & Fabrizio, P. A Novel Yeast U2 snRNP Protein, Snu17p, Is Required for the First Catalytic Step of Splicing and for Progression of Spliceosome Assembly. *Molecular and Cellular Biology* 21, 3037–3046 (2001). http://doi.org/10.1128/mcb.21.9.3037-3046.2001.

[211] Bao, P., Höbartner, C., Hartmuth, K. & Lührmann, R. Yeast Prp2 liberates the 5 splice site and the branch site adenosine for catalysis of pre-mRNA splicing. *RNA* 23, 1770–1779 (2017). http://doi.org/10.1261/rna.063115.117.

[212] Tarn, W. Y. *et al.* Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO Journal* 13, 2421–2431 (1994). http://doi.org/10.1002/j.1460-2075.1994.tb06527.x.

[213] Makarova, O. V. *et al.* A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO Journal* 23, 2381–2391 (2004). http://doi.org/10.1038/sj.emboj.7600241.

[214] Chan, S. P., Kao, D. I., Tsai, W. Y. & Cheng, S. C. The Prp19p-associated complex in spliceosome activation. *Science* 302, 279–282 (2003). http://doi.org/10.1126/science.1086602.

[215] Grote, M. *et al.* Molecular Architecture of the Human Prp19/CDC5L Complex. *Molecular and Cellular Biology* 30, 2105–2119 (2010). http://doi.org/10.1128/mcb.01505-09.

[216] Burns, C. G., Ohi, R., Krainer, A. R. & Gould, K. L. Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America* 96, 13789–13794 (1999). http://doi.org/10.1073/pnas.96.24.13789.

[217] Huang, X., Wang, G., Wu, Y. & Du, Z. The structure of full-length human CTNNBL1 reveals a distinct member of the armadillo-repeat protein family. *Acta Crystallographica Section D: Biological Crystallography* 69, 1598–1608 (2013). http://doi.org/10.1107/S0907444913011360.

[218] Ahn, J. W., Kim, S., Kim, E. J., Kim, Y. J. & Kim, K. J. Structural insights into the novel ARM-repeat protein CTNNBL1 and its association with the hPrp19-CDC5L complex. *Acta Crystallographica Section D: Biological Crystallography* 70, 780–788 (2014). http://doi.org/10.1107/S139900471303318X.

[219] Hogg, R., McGrail, J. C. & O'Keefe, R. T. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochemical Society Transactions* 38, 1110–1115 (2010). http://doi.org/10.1042/BST0381110.

[220] Zhang, X. *et al.* An Atomic Structure of the Human Spliceosome. *Cell* 169, 918–929.e14 (2017). http://doi.org/10.1016/J.CELL.2017.04.033.

[221] McGrail, J. C., Krause, A. & O'Keefe, R. T. The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. *Nucleic Acids Research* 37, 4205–4217 (2009). http://doi.org/10.1093/nar/gkp341.

[222] Xu, C. *et al.* Solution structure of human peptidyl prolyl isomerase-like protein 1 and insights into its interaction with SKIP. *Journal of Biological Chemistry* 281, 15900–15908 (2006). http://doi.org/10.1074/jbc.M511155200.

[223] Wang, X. *et al.* A large intrinsically disordered region in SKIP and its disorder-order transition induced by PPIL1 binding revealed by NMR. *Journal of Biological Chemistry* 285, 4951–4963 (2010). http://doi.org/10.1074/jbc.M109.087528.

[224] Li, Y. *et al.* The SNW Domain of SKIP Is Required for Its Integration into the Spliceosome and Its Interaction with the Paf1 Complex in Arabidopsis. *Molecular Plant* 9, 1040–1050 (2016). http://doi.org/10.1016/j.molp.2016.04.011.

[225] De, I. *et al.* The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nature Structural & Molecular Biology* 22, 138–144 (2015). http://doi.org/10.1038/nsmb.2951.

[226] Nakatsu, Y. *et al.* XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *Journal of Biological Chemistry* 275, 34931–34937 (2000). http://doi.org/10.1074/jbc.M004936200.

[227] Chen, C. H. *et al.* Functional and physical interactions between components of the Prp19p-associated complex. *Nucleic Acids Research* 30, 1029–1037 (2002). http://doi.org/10.1093/nar/30.4.1029.

[228] Mi, H., Kops, O., Zimmermann, E., Jäschke, A. & Tropschug, M. A nuclear RNA-binding cyclophilin in human T cells. *FEBS Letters* 398, 201–205 (1996). http://doi.org/10.1016/S0014-5793(96)01248-3.

[229] Villa, T. & Guthrie, C. The Isy1p component of the NineTeen Complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes and Development* 19, 1894–1904 (2005). http://doi.org/10.1101/gad.1336305.

[230] Hatakeyama, S., Matsumoto, M., Yada, M. & Nakayama, K. I. Interaction of U-box-type ubiquitin-protein ligases (E3s) with molecular chaperones. *Genes to Cells* 9, 533–548 (2004). http://doi.org/10.1111/j.1356-9597.2004.00742.x.

[231] Hatakeyama, S., Yada, M., Matsumoto, M., Ishida, N. & Nakayama, K. I. U Box Proteins as a New Family of Ubiquitin-Protein Ligases. *Journal of Biological Chemistry* 276, 33111–33120 (2001). http://doi.org/10.1074/jbc.M102755200.

[232] Busetto, V. *et al.* Structural and functional insights into CWC27/CWC22 heterodimer linking the exon junction complex to spliceosomes. *Nucleic acids research* 48, 5670–5683 (2020). http://doi.org/10.1093/nar/gkaa267.

[233] Brickner, J. R. *et al.* A ubiquitin-dependent signalling axis specific for ALKBH-mediated DNA dealkylation repair. *Nature* 551, 389–393 (2017). http://doi.org/10.1038/nature24484.

[234] Blencowe, B. J., Issner, R., Nickerson, J. A. & Sharp, P. A. A coactivator of pre-mRNA splicing. *Genes and Development* 12, 996–1009 (1998). http://doi.org/10.1101/gad.12.7.996.

[235] Eldridge, A. G., Li, Y., Sharp, P. A. & Blencowe, B. J. The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proceedings of the National Academy of Sciences of the United States of America* 96, 6125–6130 (1999). http://doi.org/10.1073/pnas.96.11.6125.

[236] Tomsic, J. *et al.* A germline mutation in SRRM2, a splicing factor gene, is implicated in papillary thyroid carcinoma predisposition. *Scientific reports* 5, 10566 (2015). http://doi.org/10.1038/srep10566.

[237] Zhang, X. *et al.* Structure of the human activated spliceosome in three conformational states. *Cell Research* 28, 307–322 (2018). http://doi.org/10.1038/cr.2018.14.

[238] Bai, R. *et al.* Mechanism of spliceosome remodeling by the ATPase/helicase Prp2 and its coactivator Spp2. *Science* 371 (2021). http://doi.org/10.1126/science.abe8863.

[239] Valadkhan, S. & Manley, J. L. Splicing-related catalysis by protein-free snRNAS. *Nature* 413, 701–707 (2001). http://doi.org/10.1038/35099500.

[240] Valadkhan, S. & Manley, J. L. Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA* 9, 892–904 (2003). http://doi.org/10.1261/rna.5440303.

[241] Valadkhan, S., Mohammadi, A., Wachtel, C. & Manley, J. L. Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA* 13, 2300–2311 (2007). http://doi.org/10.1261/rna.626207.

[242] Valadkhan, S., Mohammadi, A., Jaladat, Y. & Geisler, S. Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proceedings of the National Academy of Sciences of the United States of America* 106, 11901–11906 (2009). http://doi.org/10.1073/pnas.0902020106.

[243] Saha, D., Khandelia, P., O'Keefe, R. T. & Vijayraghavan, U. Saccharomyces cerevisiae NineTeen Complex (NTC)-associated factor Bud31/Ycr063w assembles on precatalytic spliceosomes and improves first and second step pre-mRNA splicing efficiency. *Journal of Biological Chemistry* 287, 5390–5399 (2012). http://doi.org/10.1074/jbc.M111.298547.

[244] Bao, P., Will, C. L., Urlaub, H., Boon, K. L. & Lührmann, R. The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated Bactcomplex. *Genes and Development* 31, 2416–2429 (2017). http://doi.org/10.1101/gad.308163.117.

[245] Tarn, W. Y., Lee, K. R. & Cheng, S. C. The yeast PRP19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the pre-mRNA and prior to formation of the functional spliceosome. *Molecular and Cellular Biology* 13, 1883–1891 (1993). http://doi.org/10.1128/mcb.13.3.1883.

[246] Shi, Y. The Spliceosome: A Protein-Directed Metalloribozyme (2017). http://doi.org/10.1016/j.jmb.2017.07.010.

[247] Plaschka, C., Lin, P.-C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* 546, 617 (2017). http://doi.org/10.1038/nature22799.

[248] Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Åresolution. *Science (New York, N.Y.)* 353, 904–911 (2016). http://doi.org/10.1126/science.aag0291.

[249] Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nature Communications* 7 (2016). http://doi.org/10.1038/ncomms10615.

[250] Effenberger, K. A., Urabe, V. K. & Jurica, M. S. Modulating splicing with small molecular inhibitors of the spliceosome. *Wiley Interdisciplinary Reviews: RNA* 8 (2017). http://doi.org/10.1002/wrna.1381.

[251] Bonnal, S., Vigevani, L. & Valcárcel, J. The spliceosome as a target of novel antitumour drugs. *Nature Reviews Drug Discovery* 11, 847–859 (2012). http://doi.org/10.1038/nrd3823.

[252] Cretu, C. *et al.* Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds. *Molecular Cell* 70, 265–273.e8 (2018). http://doi.org/10.1016/j.molcel.2018.03.011.

[253] Mermoud, J. E., Cohen, P. & Lamond, A. I. Ser/thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Research* 20, 5263–5269 (1992). http://doi.org/10.1093/nar/20.20.5263.

[254] Pilch, B. *et al.* Specific inhibition of serine- and arginine-rich splicing factors phosphorylation, spliceosome assembly, and splicing by the antitumor drug NB-506. *Cancer Research* 61, 6876–6884 (2001).

[255] Berg, M. G. *et al.* A Quantitative High-Throughput In Vitro Splicing Assay Identifies Inhibitors of Spliceosome Catalysis. *Molecular and Cellular Biology* 32, 1271–1283 (2012). http://doi.org/10.1128/mcb.05788-11.

[256] Sidarovich, A. Identification and characterization of small molecule inhibitors of pre-mRNA splicing that block spliceosome assembly at novel stages (2015).

[257] Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* 60, 2126–2132 (2004). http://doi.org/10.1107/S0907444904019158.

[258] Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D: Structural Biology* 74, 814–840 (2018). http://doi.org/10.1107/S2059798318009324.

[259] Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 14, 290–296 (2017). http://doi.org/10.1038/nmeth.4169.

[260] Zhang, K. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* 193, 1–12 (2016). http://doi.org/10.1016/J.JSB.2015.11.003.

[261] Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567 (1999). http://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

[262] Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* 66, 12–21 (2010). http://doi.org/10.1107/S0907444909042073.

[263] Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology* 74, 531–544 (2018). http://doi.org/10.1107/S2059798318006551.

[264] Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nature Methods* 9, 904–906 (2012). http://doi.org/10.1038/nmeth.2099.

[265] Chen, Z. L. *et al.* A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nature Communications* 10, 1–12 (2019). http://doi.org/10.1038/s41467-019-11337-z.

[266] Burley, S. K. *et al.* Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Research* 47, D520–D528 (2019). http://doi.org/10.1093/nar/gky949.

[267] Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2. *eLife* 5 (2016). http://doi.org/10.7554/eLife.18722.

[268] Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7 (2018). http://doi.org/10.7554/eLife.42166.

[269] Korneta, I., Magnus, M. & Bujnicki, J. M. Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Research* 40, 7046–7065 (2012). http://doi.org/10.1093/nar/gks347.

[270] Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* 46, W296–W303 (2018). http://doi.org/10.1093/nar/gky427.

[271] Pettersen, E. F. *et al.* UCSF Chimera–A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25, 1605–1612 (2004). http://doi.org/10.1002/jcc.20084.

[272] Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* 27, 14–25 (2018). http://doi.org/10.1002/pro.3235.

[273] The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research* 47, D506–D515 (2019). http://doi.org/10.1093/nar/gky1049.

[274] 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 Research* 11, 1475–1489 (1983). http://doi.org/10.1093/nar/11.5.1475.

[275] Das, R. & Reed, R. Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* 5, 1504–1508 (1999). http://doi.org/10.1017/S1355838299991501.

[276] Agafonov, D. E. *et al.* Semiquantitative Proteomic Analysis of the Human Spliceosome via a Novel Two-Dimensional Gel Electrophoresis Method. *Molecular and Cellular Biology* (2011). http://doi.org/10.1128/MCB.05266-11.

[277] Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W. S. & Lührmann, R. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO Journal* 16, 4092–4106 (1997). http://doi.org/10.1093/emboj/16.13.4092.

[278] Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature Methods* 5, 53–55 (2008). http://doi.org/10.1038/nmeth1139.

[279] Lu, M. *et al.* Crystal Structure of the Three Tandem FF Domains of the Transcription Elongation Regulator CA150. *Journal of Molecular Biology* 393, 397–408 (2009). http://doi.org/10.1016/j.jmb.2009.07.086.

[280] Izaurralde, E. *et al.* A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* 78, 657–668 (1994). http://doi.org/10.1016/0092-8674(94)90530-4.

[281] Görnemann, J., Kotovic, K. M., Hujer, K. & Neugebauer, K. M. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Molecular Cell* 19, 53–63 (2005). http://doi.org/10.1016/j.molcel.2005.05.007.

[282] Pabis, M. *et al.* The nuclear cap-binding complex interacts with the U4/U6 · U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *Rna* 19, 1054–1063 (2013). http://doi.org/10.1261/rna.037069.112.

[283] Llorian, M., Beullens, M., Andrés, I., Ortiz, J. M. & Bollen, M. SIPP1, a novel pre-mRNa splicing factor and interactor of protein phosphatase-1. *Biochemical Journal* 378, 229–238 (2004). http://doi.org/10.1042/BJ20030950.

[284] Komuro, A., Saeki, M. & Kato, S. Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-rich motif containing glycine and arginine. *Journal of Biological Chemistry* 274, 36513–36519 (1999). http://doi.org/10.1074/jbc.274.51.36513.

[285] Masson, C. *et al.* Global genome repair is required to activate KIN17, a UVC-responsive gene involved in DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* 100, 616–621 (2003). http://doi.org/10.1073/pnas.0236176100.

[286] Miccoli, L. *et al.* The Human Stress-Activated Protein kin17 Belongs to the Multiprotein DNA Replication Complex and Associates In Vivo with Mammalian Replication Origins. *Molecular and Cellular Biology* 25, 3814–3830 (2005). http://doi.org/10.1128/mcb.25.9.3814-3830.2005.

[287] Carlier, L. *et al.* Solution structure of the region 51-160 of human KIN17 reveals an atypical winged helix domain. *Protein Science* 16, 2750–2755 (2007). http://doi.org/10.1110/ps.073079107.

[288] Schellenberg, M. J. *et al.* A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nature Structural and Molecular Biology* 20, 728–734 (2013). http://doi.org/10.1038/nsmb.2556.

[289] Carty, S. M., Goldstrohm, A. C., Suñé, C., Garcia-Blanco, M. A. & Greenleaf, A. L. Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* 97, 9015–9020 (2000). http://doi.org/10.1073/pnas.160266597.

[290] Goldstrohm, A. C., Albrecht, T. R., Suñé, C., Bedford, M. T. & Garcia-Blanco, M. A. The Transcription Elongation Factor CA150 Interacts with RNA Polymerase II and the Pre-mRNA Splicing Factor SF1. *Molecular and Cellular Biology* 21, 7617–7628 (2001). http://doi.org/10.1128/mcb.21.22.7617-7628.2001.

[291] Sánchez-Álvarez, M., Goldstrohm, A. C., Garcia-Blanco, M. A. & Suñé, C. Human Transcription Elongation Factor CA150 Localizes to Splicing Factor-Rich Nuclear Speckles and Assembles Transcription and Splicing Components into Complexes through Its Amino and Carboxyl Regions. *Molecular and Cellular Biology* 26, 4998–5014 (2006). http://doi.org/10.1128/mcb.01991-05.

[292] Waragai, M. *et al.* PQBP-1/Npw38, a nuclear protein binding to the polyglutamine tract, interacts with U5-15kD/dim1p via the carboxyl-terminal domain. *Biochemical and Biophysical Research Communications* 273, 592–595 (2000). http://doi.org/10.1006/bbrc.2000.2992.

[293] Zhang, Y. Z. *et al.* Evidence that Dim1 associates with proteins involved in pre-mRNA splicing, and delineation of residues essential for Dim1 interactions with hnRNP F and Npw38/PQBP-1. *Gene* 257, 33–43 (2000). http://doi.org/10.1016/S0378-1119(00)00372-3.

[294] Okazawa, H. *et al.* Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. *Neuron* 34, 701–713 (2002). http://doi.org/10.1016/S0896-6273(02)00697-9.

[295] Nabeshima, Y., Mizuguchi, M., Kajiyama, A. & Okazawa, H. Segmental isotope-labeling of the intrinsically disordered protein PQBP1. *FEBS Letters* 588, 4583–4589 (2014). http://doi.org/10.1016/j.febslet.2014.10.028.

[296] Kohtz, J. D. *et al.* Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368, 119–124 (1994). http://doi.org/10.1038/368119a0.

[297] Sánchez-Hernández, N. *et al.* The FF4 and FF5 domains of transcription elongation regulator 1 (TCERG1) target proteins to the periphery of speckles. *Journal of Biological Chemistry* 287, 17789–17800 (2012). http://doi.org/10.1074/jbc.M111.304782.

[298] Misteli, T. RNA splicing: What has phosphorylation got to do with it? (1999). http://doi.org/10.1016/S0960-9822(99)80128-6.

[299] Cho, S. *et al.* Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences of the United States of America* 108, 8233–8238 (2011). http://doi.org/10.1073/pnas.1017700108.

[300] Sánchez-Hernández, N. *et al.* The in vivo dynamics of TCERG1, a factor that couples transcriptional elongation with splicing. *Rna* 22, 571–582 (2016). http://doi.org/10.1261/rna.052795.115.

[301] Ulrich, A. K., Schulz, J. F., Kamprad, A., Schütze, T. & Wahl, M. C. Structural Basis for the Functional Coupling of the Alternative Splicing Factors Smu1 and RED. *Structure* 24, 762–773 (2016). http://doi.org/10.1016/j.str.2016.03.016.

[302] Huang, X. *et al.* Structure and function of the two tandem WW domains of the pre-mRNA splicing factor FBP21 (Formin-binding protein 21). *Journal of Biological Chemistry* 284, 25375–25382 (2009). http://doi.org/10.1074/jbc.M109.024828.

[303] Xie, J., Beickman, K., Otte, E. & Rymond, B. C. Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO Journal* 17, 2938–2946 (1998). http://doi.org/10.1093/emboj/17.10.2938.

[304] Chung, S. *et al.* Crooked neck is a component of the human spliceosome and implicated in the splicing process. *Biochimica et Biophysica Acta - Gene Structure and Expression* 1576, 287–297 (2002). http://doi.org/10.1016/S0167-4781(02)00368-8.

[305] Davis, T. L. *et al.* Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. *PLoS Biology* 8 (2010). http://doi.org/10.1371/journal.pbio.1000439.

[306] Rajiv, C. & Davis, T. L. Structural and functional insights into human nuclear cyclophilins (2018). http://doi.org/10.3390/biom8040161.

[307] Adams, B. M., Coates, M. N., Jackson, S. R. E., Jurica, M. S. & Davis, T. L. Nuclear cyclophilins affect spliceosome assembly and function in vitro. *Biochemical Journal* 469, 223–233 (2015). http://doi.org/10.1042/BJ20150396.

[308] le Maire, A. *et al.* A Tandem of SH3-like Domains Participates in RNA Binding in KIN17, a Human Protein Activated in Response to Genotoxics. *Journal of Molecular Biology* 364, 764–776 (2006). http://doi.org/10.1016/j.jmb.2006.09.033.

[309] Kim, S. H. & Lin, R. J. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Molecular and Cellular Biology* 16, 6810–6819 (1996). http://doi.org/10.1128/mcb.16.12.6810.

[310] Bao, P., Will, C. L., Urlaub, H., Boon, K. L. & Lührmann, R. The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated Bact complex. *Genes and Development* 31, 2416–2429 (2017). http://doi.org/10.1101/gad.308163.117.

[311] Ajuh, P., Sleeman, J., Chusainow, J. & Lamond, A. I. A Direct Interaction between the Carboxyl-terminal Region of CDC5L and the WD40 Domain of PLRG1 Is Essential for Pre-mRNA Splicing. *Journal of Biological Chemistry* 276, 42370–42381 (2001). http://doi.org/10.1074/jbc.M105453200.

[312] Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. The WD repeat: A common architecture for diverse functions. *Trends in Biochemical Sciences* 24, 181–185 (1999). http://doi.org/10.1016/S0968-0004(99)01384-5.

[313] Van Maldegem, F. *et al.* CTNNBL1 facilitates the association of CWC15 with CDC5L and is required to maintain the abundance of the Prp19 spliceosomal complex. *Nucleic Acids Research* 43, 7058–7069 (2015). http://doi.org/10.1093/nar/gkv643.

[314] 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 (2010). http://doi.org/10.1371/journal.pone.0010013.

[315] Lindsey, L. A. & Garcia-Blanco, M. A. Functional conservation of the human homolog of the yeast pre-mRNA splicing factor Prp17p. *Journal of Biological Chemistry* 273, 32771–32775 (1998). http://doi.org/10.1074/jbc.273.49.32771.

[316] Sapra, A. K., Arava, Y., Khandelia, P. & Vijayraghavan, U. Genome-wide analysis of Pre-mRNA splicing: Intron features govern the requirement for the second-step factor, Prp17 in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Journal of Biological Chemistry* 279, 52437–52446 (2004). http://doi.org/10.1074/jbc.M408815200.

[317] Yehuda, S. B. *et al.* Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40 yeast gene involved in splicing and cell cycle control. *Rna* 4, 1304–1312 (1998). http://doi.org/10.1017/S1355838298980712.

[318] Chai, G. *et al.* Mutations in Spliceosomal Genes PPIL1 and PRP17 Cause Neurodegenerative Pontocerebellar Hypoplasia with Microcephaly. *Neuron* 109, 241–256.e9 (2021). http://doi.org/10.1016/j.neuron.2020.10.035.

[319] Reiter, N. J., Blad, H., Abildgaard, F. & Butcher, S. E. Dynamics in the U6 RNA intramolecular stem-loop: A base flipping conformational change. *Biochemistry* 43, 13739–13747 (2004). http://doi.org/10.1021/bi048815y.

[320] Blad, H., Reiter, N. J., Abildgaard, F., Markley, J. L. & Butcher, S. E. Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. *Journal of Molecular Biology* 353, 540–555 (2005). http://doi.org/10.1016/j.jmb.2005.08.030.

[321] Venditti, V., Clos, L., Niccolai, N. & Butcher, S. E. Minimum-Energy Path for a U6 RNA Conformational Change Involving Protonation, Base-Pair Rearrangement and Base Flipping. *Journal of Molecular Biology* 391, 894–905 (2009). http://doi.org/10.1016/j.jmb.2009.07.003.

[322] Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* 466, 329–333 (2010). http://doi.org/10.1038/nature09206.

[323] Hoskins, A. A., Rodgers, M. L., Friedman, L. J., Gelles, J. & Moore, M. J. Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. *eLife* 5 (2016). http://doi.org/10.7554/eLife.14166.

[324] Hoskins, A. a. *et al.* Ordered and Dynamic Assembly of Single Spliceosomes. *Science* 331, 1289–1296 (2011).

[325] Shcherbakova, I. *et al.* Alternative Spliceosome Assembly Pathways Revealed by Single-Molecule Fluorescence Microscopy. *Cell Reports* 5, 151–165 (2013). http://doi.org/10.1016/j.celrep.2013.08.026.

[326] David, C. J., Boyne, A. R., Millhouse, S. R. & Manley, J. L. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes and Development* 25, 972–982 (2011). http://doi.org/10.1101/gad.2038011.

[327] Lazo, J. S. *et al.* Identification of a potent and selective pharmacophore for Cdc25 dual specificity phosphatase inhibitors. *Molecular Pharmacology* 61, 720–728 (2002). http://doi.org/10.1124/mol.61.4.720.

[328] Dulyaninova, N. G. *et al.* Cysteine 81 is critical for the interaction of s100a4 and myosin-IIA. *Biochemistry* 50, 7218–7227 (2011). http://doi.org/10.1021/bi200853y.

[329] Nilsson, I. & Hoffmann, I. Cell cycle regulation by the Cdc25 phosphatase family. *Progress in cell cycle research* 4, 107–114 (2000). http://doi.org/10.1007/978-1-4615-4253-7_10.

[330] You, Y. J., Zheng, X. G., Yong, K. & Ahn, B. Z. Naphthazarin derivatives: Synthesis, cytotoxic mechanism and evaluation of antitumor activity. *Archives of Pharmacal Research* 21, 595–598 (1998). http://doi.org/10.1007/BF02975381.

[331] Valente, C. *et al.* The 1,4-naphthoquinone scaffold in the design of cysteine protease inhibitors. *Bioorganic and Medicinal Chemistry* 15, 5340–5350 (2007). http://doi.org/10.1016/j.bmc.2007.04.068.

[332] Soares, K. M. *et al.* Profiling the NIH small molecule repository for compounds that generate H2O2 by redox cycling in reducing environments. *Assay and Drug Development Technologies* 8, 152–174 (2010). http://doi.org/10.1089/adt.2009.0247.

[333] Effenberger, K. A. *et al.* A high-throughput splicing assay identifies new classes of inhibitors of human and yeast spliceosomes. *Journal of Biomolecular Screening* 18, 1110–1120 (2013). http://doi.org/10.1177/1087057113493117.

[334] Wan, L., Ottinger, E., Cho, S. & Dreyfuss, G. Inactivation of the SMN Complex by Oxidative Stress. *Molecular Cell* 31, 244–254 (2008). http://doi.org/10.1016/j.molcel.2008.06.004.

[335] Wilkinson, M. E., Fica, S. M., Galej, W. P. & Nagai, K. Structural basis for conformational equilibrium of the catalytic spliceosome. *Molecular Cell* 81, 1439–1452.e9 (2021). http://doi.org/10.1016/j.molcel.2021.02.021.

[336] Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods* 17, 1214–1221 (2020). http://doi.org/10.1038/s41592-020-00990-8.

[337] Wu, Z. *et al.* Deep manifold learning reveals hidden dynamics of proteasome autoregulation. *arXiv* (2020).

[338] Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* 18, 176–185 (2021). http://doi.org/10.1038/s41592-020-01049-4.

[339] Schliep, J. E. *Structural Characterization of the Eukaryotic Translation Initiation by Electron submitted by Electron Cryo-Microscopy.* Ph.D. thesis (2018).

[340] Singh, K. *et al.* Discovery of a Regulatory Subunit of the Yeast Fatty Acid Synthase. *Cell* 180, 1130–1143.e20 (2020). http://doi.org/10.1016/j.cell.2020.02.034.