Structural and functional studies of the spliceosomal RNP remodeling enzyme Brr2

Dissertation
for the award of the degree
"Doctor of Philosophy"
Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

submitted by
Karine Santos
from São José do Rio Preto, Brazil

Göttingen, 2012

Thesis committee

Prof. Dr. Markus C. Wahl (reviewer)
Freie University Berlin
Institute for Chemistry and Biochemistry
Department of Structural Biochemistry
Berlin, Germany

Prof. Dr. Ralf Ficner (reviewer)
Georg August University Göttingen
Institute for Microbiology and Genetics
Department of Molecular Structural Biology
Göttingen, Germany

Prof. Dr. Detlef Doenecke
Georg August University Göttingen
Institute for Biochemistry and Molecular Cell Biology
Department of Molecular Biology
Göttingen, Germany

Prof. Dr. Dirk Fasshauer
University of Lausanne
Faculty of Biology and Medicine
Department of Cellular Biology and Morphology
Lausanne, Switzerland

Prof. Dr. Marina Rodnina
Georg August University Göttingen
Max Planck Institute for Biophysical Chemistry
Department of Physical Biochemistry
Göttingen, Germany

Prof. Dr. Kai Tittmann
Georg August University Göttingen
Albrecht-von-Haller Institute
Department of Bioanalytics
Göttingen, Germany

Date of oral examination: 20.11.2012

Affidavit

I hereby declare that my thesis entitled "Structural and functional studies of the spliceosomal RNP remodeling enzyme Brr2" has been written independently and with no other sources and aids than quoted. This thesis (wholly or in part) has not been submitted elsewhere for any academic award or qualification.

Karine Santos

September, 2012

Göttingen

Related publications

Parts of this thesis were published in:

Weber, G., Cristao, V.F., de L Alves, F., Santos, K.F., Holton, N., Rappsilber, J., Beggs, J.D., Wahl, M.C. (2011) Mechanism for Aar2p function as a U5 snRNP assembly factor. *Genes Dev.* 25(15):1601-12. DOI: 10.1101/gad.635911.

Santos, K. F., Mozaffari-Jovin, S., Weber, G., Pena, V., Lührmann, R., Wahl, M. C. (2012) Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. *Proc. Natl. Acad. Sci. USA*, 109(43):17418-23. DOI: 10.1073/pnas.1208098109.

Mozaffari-Jovin, S., Santos, K.F., Hsiao, H.H., Urlaub, H., Wahl, M.C., Lührmann, R. (2012) The Prp8 RNase H-like domain inhibits Brr-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes Dev.* 26(21):2422-34. DOI: 10.1101/gad.200949.112.

Weber, G., Cristao, V.F., Santos, K.F., Jovin, S.M., Heroven, A.C., Holton, N., Lührmann, R., Beggs, J.D., Wahl, M.C. (2013) Structural Basis for dual roles of Aar2p in U5 snRNP assembly. *Genes Dev.* 27(5):525-40. DOI: 10.1101/gad.213207.113.

Mozaffari-Jovin, S.*, Wandersleben, T.*, Santos, K.F.*, Will, C., Lührmann, R., Wahl, M.C. (2013) Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science.* DOI:10.1126/science.1237515.
*These authors contributed equally to this work.

Abstract

The spliceosome is a highly dynamic, multi-MDa eukaryotic RNA-protein (RNP) machinery that catalyzes precursor messenger RNA (pre-mRNA) splicing. Pre-mRNA splicing entails the excision of non-coding introns and the joining of the neighboring coding exons *via* two consecutive transesterification reactions. For each catalytic cycle, a spliceosome is assembled on a substrate pre-mRNA by the stepwise recruitment of five small nuclear (sn) ribonucleoproteins (RNPs) and numerous non-snRNP factors. In contrast to ribosomal subunits, for example, none of the snRNPs or non-snRNP complexes contain a pre-formed catalytic center for the splicing catalysis. Instead, the active center of the spliceosome is formed anew during each spliceosome assembly cycle.

Spliceosome assembly occurs stepwise *via* several discrete intermediates that have been experimentally defined. During the transition from one assembly stage to the next, the spliceosome undergoes profound compositional and conformational remodeling. These remodeling events are driven and controlled by eight highly conserved Superfamily (SF) 2 RNP remodeling enzymes. In particular, an initial assembly containing all snRNPs (the so-called B complex) is still catalytically inactive and requires major rearrangements of its RNA-RNA, RNA-protein and protein-protein interaction networks in order to produce a catalytically competent spliceosome. Spliceosome catalytic activation requires a 650 kDa sub-complex that is part of the U5 snRNP and is composed of a large scaffolding protein, Prp8, a G-protein, Snu114, and a Ski-2 RNA helicase, Brr2. The molecular mechanisms underlying spliceosome catalytic activation are poorly understood. To elucidate the architecture of the complex formed by Prp8, Snu114 and Brr2 we aimed at recombinant reconstitution of this micro-machinery. We managed to successfully co-express human (h) Prp8, hSnu114 and hBrr2 in insect cells. However, we were not able to co-purify all the components. Only hBrr2 could be efficiently purified, indicating that under these working conditions, hBrr2 did not stably interact with hPrp8 and hSnu114. Using an ortholog screening approach, we tried to co-express the yeast (y) Prp8-ySnu114 sub-complex. Although yPrp8 was poorly expressed, we were able to co-purify small amounts of yPrp8 with ySnu114. Although these two proteins form a stable complex that can be purified, size exclusion chromatography revealed that the complex was possibly aggregated and unsuitable for further structural analysis since it migrated in the void volume of the column.

Abstract

While our initial strategy of co-expressing Prp8, Snu114 and Brr2 as a complex failed, we have succeeded in isolating human Brr2. Brr2 is an essential RNA helicase needed for U4/U6 di-snRNP disruption during spliceosome catalytic activation. Brr2 is also the only spliceosomal helicase that is permanently associated with the spliceosome and thus requires faithful regulation. Concomitantly, Brr2 represents a unique subclass of SF2 nucleic acid helicases, containing tandem helicase cassettes. Presently, the mechanistic and regulatory consequences of this unconventional architecture are unknown. Henceforth, we then aimed at producing highly purified and homogeneous human and yeast Brr2 for further structural and functional investigations.

Full length human and yeast Brr2 could be expressed and purified to near homogeneity. Both enzymes were active in ATP-dependent U4/U6 duplex unwinding but failed to crystallize. In order to remove putatively flexible regions that may hinder crystallization, we treated hBrr2 and yBrr2 with proteases, several of which gave rise to a protease-resistant ca. 200 kDa fragment encompassing the two helicase cassettes. One of the six truncated hBrr2 proteins, whose borders were designed based on the proteolysis experiments, crystallized readily and the crystals diffracted to 2.65 Å resolution. Elucidation of the crystal structure and biochemical analyses showed that in hBrr2 two ring-like helicase cassettes intimately interact and functionally cooperate. Only the N-terminal cassette harbored ATPase and helicase activities in isolation. Structural comparisons and mutational analyses suggested that the N-terminal cassette of hBrr2 threads single-stranded RNA through a central tunnel and across a helix-loop-helix domain during duplex unwinding. While the C-terminal cassette did not seem to engage RNA in this fashion, it bound ATP and it strongly stimulated the N-terminal helicase. Stimulation depended on two inter-cassette communication lines, disruption of which affected ATPase and helicase activities in diverse ways. Additionally, mutations at the C-terminal ATP pocket affected the crosstalk between the two cassettes, suggesting that ATP binding may induce a specific C-terminal cassette conformation that solidifies important inter-cassette contacts. Using pre-steady state kinetics, we also probed the nucleotide binding preferences and worked out possible nucleotide binding mechanisms of either cassette, confirming that the C-terminal cassette strongly binds ATP in solution. Taken together, our results revealed the structural and functional interplay between two helicase cassettes in a tandem SF2 enzyme and suggested how Brr2 interactors may exploit the C-terminal cassette as a "remote control" to regulate the N-terminal helicase of the enzyme.

1. Introduction

1.1 pre-mRNA splicing

The entirety of an organism's hereditary information is contained in the genome. Functionally, the genome can be split into repetitive, regulatory and gene sequences. A gene is a sequence of deoxyribonucleic acid (DNA) that encodes for a ribonucleic acid (RNA). In protein-coding genes, the RNA instructs production of protein.

The process by which a gene gives rise to a protein is called gene expression. For most bacterial genes, gene expression consists of two steps. The first step is called transcription, when RNA, which is a copy of one strand of the DNA, is produced. These RNA molecules can play regulatory roles and can also carry the information into the next step. The latter RNA molecules are called messenger RNAs (mRNAs). In the second step, proteins are synthesized by ribosomes, using mRNA molecules as templates. The sequence of an mRNA is read in triplets (codons) to provide the series of amino acids that make the corresponding protein.

An mRNA includes a sequence of nucleotides (nt) that corresponds to the sequence of amino acids in the protein. This part of the mRNA is called the coding region. Furthermore, the mRNA includes additional sequences on either end. The 5' untranslated region (UTR) is called the leader and the 3' UTR region is called the trailer. UTRs usually contain regulatory sequences controlling stability and translational activity of the transcripts.

Considering that prokaryotic organisms possess only a single compartment, transcription and translation occur in the same place and are tightly coupled. In eukaryotes, however, mRNA molecules are first transcribed as primary transcripts in the nucleus and must be transported to the cytoplasm in order to be translated.

Typically, eukaryotic RNA transcripts undergo several steps of processing before a mature mRNA is exported from the nucleus and is suitable for translation (Fig. 1.1). The maturation of a nuclear mRNAs starts with co-transcriptional capping of its 5' end. A 7-methylguanosine is linked by an unusual 5'-5' triphosphate bond to the ribose at the 5' end of the transcript (Shatkin, 1976; Shatkin and Manley, 2000). This m_7G cap (m_7GpppN) is involved in mRNA export from the nucleus, initiation of protein synthesis and stabilization of mRNA (McCracken *et al.*, 1997). Additionally, the 3' end of the mRNA is modified by a process called polyadenylation, in which the growing transcript is cleaved at a specific site

and a polyadenine (poly(A)) tail (100-200 adenine nt) is added by a poly(A) polymerase (Colgan and Manley, 1997).

Furthermore, most genes in higher Eukarya (and few genes in Bacteria and Archaea) contain additional sequences that lie within the coding region that do not code for protein. These intervening, non-coding, sequences of the genes are called introns and the coding regions are called exons. The initial transcripts of such genes are called precursor mRNAs (pre-mRNAs) and require special processing in which the introns are removed from the nascent transcript and the adjacent exons are joined to obtain an mRNA that has a continuous open reading frame (ORF).

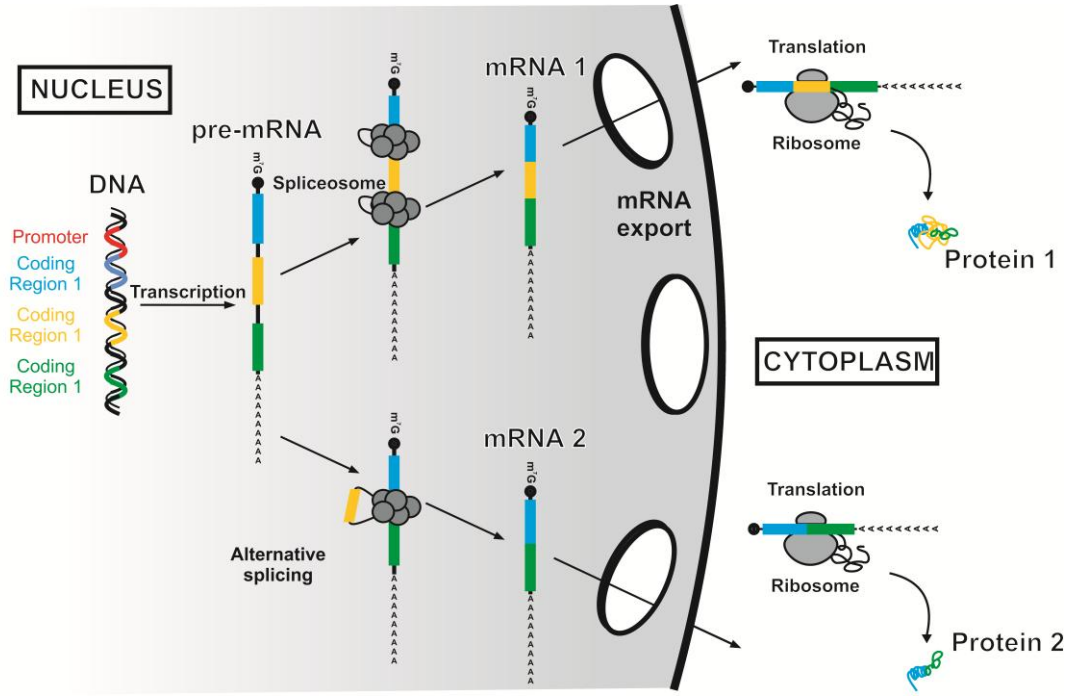


Figure 1.1: Gene expression in eukaryotes. Following transcription, the pre-mRNA, containing exons (boxes) and introns (lines), is subjected to processing events such as capping, polyadenylation and splicing before it is exported to the cytoplasm, where translation takes place.

There are four major classes of introns that are recognized based on their splicing mechanism:

(I) Group I self-splicing introns are widespread in protist nuclear rDNA, fungal mitochondria, some bacteria and bacteriophages. These group I intron RNAs self-splice utilizing a distinctive two-step pathway that relies on an external guanosine nucleotide as a cofactor. Within group I introns, open reading frames are often found encoding homing endonucleases that promote intron mobility at the DNA level into intronless cognate sites (Haugen *et al.*, 2005).

(II) Group II self-splicing introns are phylogenetically unrelated to group I introns and are found in some bacteria and organellar genomes of plants, fungi, protists and some animals. Most group II intron RNAs consist of two basic components, a self-splicing ribozyme and an ORF for expression of a multifunctional protein (maturase) that facilitates splicing. These maturases contain endonuclease and reverse transcriptase domains that play crucial roles in homing and retrotransposition. Group II introns self-splice through a mechanism that is different from group I introns but similar to nuclear pre-mRNA splicing using the 2'-OH of an adenosine nucleotide within the intron as a nucleophile (Chan *et al.*, 2012; Toor *et al.*, 2008a).

(III) The transfer RNA (tRNA) introns are found in eukaryotic nuclei and in Archaea and are enzymatically removed by a cut-and-rejoin mechanism that requires adenosine triphosphate (ATP) and an endonuclease (Abelson *et al.*, 1998).

(IV) Since this thesis deals with nuclear pre-mRNA splicing, the fourth class of introns present in nuclear pre-mRNAs will be described in more detail in the following sections.

The process of nuclear pre-mRNA splicing is highly conserved from yeast to humans and is carried out by the spliceosome. Whereas nearly all human genes contain introns, they are present in only 3.8% of all genes of the yeast *Saccharomyces cerevisiae* (Lopez and Séraphin, 1999; Barrass and Beggs, 2003). Furthermore, the small percentage of intron-containing genes in yeast generally harbors only one intron of relatively short size (approximately 100-400 nt). In contrast, human pre-mRNAs contain on average of 7.8 introns varying from 100 to 500.000 nucleotides (Ast, 2004; Rowen *et al.*, 2002). The exons, however, have a rather fixed length of only approximately 120 nucleotides.

Interestingly, the human genome consists of 20,000 – 25,000 genes although the number of translated products is much larger (80,000 – 120,000 different proteins) (Liang *et al.*, 2000; Yura *et al.*, 2006; Nilsen and Graveley, 2010). Possibly, this discrepancy can be explained by alternative splicing, which is thought to occur in up to 90% of all human genes (Wang *et al.*, 2008; Luco *et al.*, 2011). Alternative splicing is a mechanism that generates variable forms of mRNA from a single pre-mRNA species due to differential incorporation of exons or retention of introns into the final transcript (Black, 2003). This finally leads to protein isoforms that differ in their amino acid sequence and, therefore, in their chemical and biological properties. Many alternative splicing events are regulated at the early phases of splicing by positive and negative regulators. These regulators bind to *cis*-acting sequences (exonic or intronic splicing enhancers – ESEs or ISEs; or inhibitors – ESIs or ISIs) and, thereby, influence the binding of splicing factors to the pre-mRNA.

Mutations affecting splicing are usually deleterious. The majority of these mutations are single base substitutions at the junctions between introns and exons. They may cause an exon to be left out of the product, cause an intron to be included or make splicing occur at an aberrant site. The most common outcome is to introduce a termination codon that results in truncation of the protein sequence. Anomalously processed or mutated mRNAs are eliminated via survey mechanisms, like nonsense-mediated decay, No-Go decay, non-stop decay or ribosome extension decay (Doma and Parker, 2006; Frischmeyer *et al.*, 2002; Inada and Aiba, 2005; Isken and Maquat, 2007; Kong and Liebhaber, 2007; Maquat *et al.*, 2001; van Hoof *et al.*, 2002).

Once the mRNA is in the cytoplasm, translation takes place. Translation is accomplished by a complex apparatus that includes protein and RNA components, the ribosome. The process of recognizing which amino acid corresponds to a particular nucleotide triplet requires an intermediate tRNA. There is at least one tRNA species for every amino acid.

1.1.1 The mechanism of nuclear pre-mRNA splicing

Nuclear pre-mRNA splicing requires a large number of *trans*-acting factors that aid proper splice site selection as well as pre-mRNA folding, bringing together the sites at which ligation of exons will occur. These factors are brought together in a stepwise manner to form a dynamic macromolecular machine called the spliceosome.

The spliceosome assembles from 5 small uridine-rich nuclear RNAs (U1, U2, U4, U5 and U6 snRNAs) organized in ribonucleoprotein complexes (snRNPs) and a plethora of non-snRNP proteins (Will and Lührmann, 2006; Shatkin, 1976; Shatkin and Manley, 2000; Jurica and Moore, 2003). In metazoans, about 1% of all introns (called U12-type introns in contrast to U2-type introns of the major spliceosome) are spliced by a distinct spliceosome, called the minor spliceosome or U12-dependent spliceosome, comprising U11, U12, U4atac, U6atac snRNPs, which are functional analogs of U1, U2, U4, U6 snRNPS of the major spliceosome, respectively (Patel and Steitz, 2003; Will and Lührmann, 2005). The U5 snRNP is shared by both spliceosomes. The spliceosome assembles *de novo* on each intron of the pre-mRNA and catalyzes two transesterification reactions which are required for excision of introns and ligation of exons.

Chemically, the splicing process seems very simple. It involves a two step reaction which produces an excised intron and ligated exons. However, the sites at which the mRNA is cleaved to splice out the intron(s) must be precisely selected since an error of one nucleotide shifts the reading frame and results in a completely different protein product. Correct splice

site (SS) selection is a major challenge to the splicing machinery, especially in higher eukaryotes, considering that introns have variable sequences and lengths as well as low information content defining exon-intron boundaries. Nevertheless, a set of specific sequences required for splicing can, in most cases, be identified in introns and at the intron/exon boundaries (Fig. 1.2) (Stephens and Schneider, 1992). The 5' SS defines the 5' end of the intron and, in yeast, it is composed of 5'-R|**GUAUGU**-3' (Lopez and Séraphin, 1999) ('|' defines the exon-intron boundary, 'R' is a purine base and nucleotides in bold indicate at least 90% conservation among yeast introns). In higher eukaryotes, the 5' SS is characterized by the consensus sequence 5'-AG|**GU**RAGU-3' (bold letters indicate invariable nucleotides). The highly conserved yeast branch point (BP) sequence 5'-UACUAAC-3' contains a conserved adenosine (underlined) which is essential for catalysis of the first step of nuclear pre-mRNA splicing. The BP adenosine is located 10 to 155 nucleotides upstream of the yeast 3' SS (Spingola *et al.*, 1999). In human, the BP sequence is normally located 18-40 nucleotides upstream of the 3' SS and comprises a degenerate sequence 5'-YNCURAC-3' ('Y' is a pyrimidine base, 'N' is any nucleotide) (Reed, 1989; Wahl *et al.*, 2009; Zhang, 1998). A 10-15 nucleotide long pyrimidine rich sequence, the polypyrimidine tract (PPT), can often be found in higher Eukarya introns one to five nucleotide upstream of the 3' SS. The PPT is essential for splicing in humans possibly due to the less conserved BP sequence in human introns. The 3' SS follows the sequence 5'-YAG|-3' in most of the introns in vertebrates. The invariant AG dinucleotide defines the end of the intron (Reed, 1989). In addition, splicing is modulated by ESEs/ISEs and ESSs/ISSs, which are short sequences within the pre-mRNA recruiting regulatory proteins that either repress or stimulate spliceosome assembly. These sequences are not only important for constitutive splicing but also play a crucial role in alternative splicing as mentioned above (Black, 2003; Matlin *et al.*, 2005).

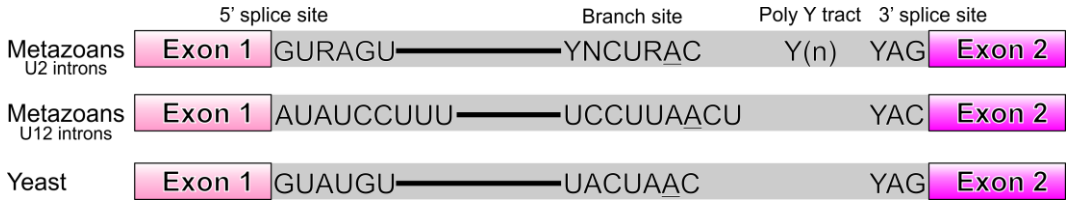


Figure 1.2: Conserved sequence elements found in introns from metazoans and budding yeast (*S. cerevisiae*). The 5' and 3' exons are shown as boxes. The branch point adenosine is underlined. "Y" – pyrimidine and "R" – purine. The polypyrimidine tract is indicated by "Y(n)".

In metazoans and plants, a separate, less abundant class of introns exists harboring different consensus sequences (Jackson, 1991). In these U12-type introns, the sequences 5'-|AUAUCCUUU-3' and 5'-YAC|-3' represent highly conserved elements at the 5' and 3' SS,

respectively. Additionally, the U12-type introns lack the PPT and show a higher level of conservation of the 5' SS and BP sequence. The first identified introns of this class had 5' AT (AU for RNA) and 3' AC di-nucleotides instead of the highly conserved GT-AG (GU-AG) present at the 5' and 3' ends of U2-type introns. Consequently, they were originally called 'ATAC'-introns. As previously mentioned, these introns are removed by the minor spliceosome (Burge *et al.*, 1999).

Regardless of the spliceosome type, introns are removed by a conserved mechanism involving two sequential S_N2-type transesterification reactions (Fig. 1.3) (Query *et al.*, 1994; Will and Lührmann, 2006). First, the oxygen of the 2' OH group of the BP adenosine makes a nucleophilic attack at the phosphodiester bond of the 5' SS exon-intron boundary. This leads to the formation of a free 3' OH group at the 3' terminal nucleotide of the 5' exon and the formation of 5'-2' phosphodiester bond between the 5' SS guanosine and the BP adenosine. The result is a free 5' exon and a lariat intermediate containing the intron and the downstream exon. In the second step, the 3' OH group of the 5' exon attacks the phosphodiester bond at the 3' SS, thereby joining 5' and 3' exons and excising the intron as a lariat. Subsequently, the lariat intron is debranched and typically degraded, but can also be a source of regulatory RNAs (Carthew and Sontheimer, 2009; Voinnet, 2009), whereas the mRNA is exported from the nucleus into the cytoplasm (Brow, 2002).

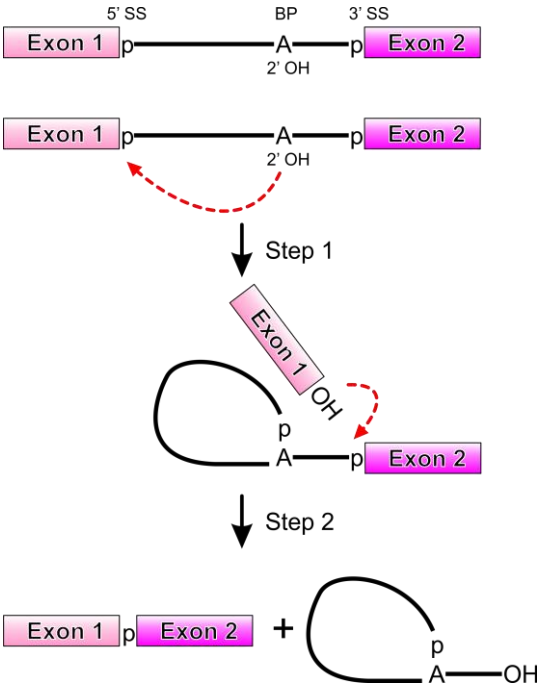


Figure 1.3: Schematic representation of the two-step mechanism of pre-mRNA splicing. Boxes and solid lines represent the exons and the intron, respectively. The branch site adenosine is indicated by the letter "A" and the phosphate groups by the letters "p" at the 5' and 3' splice sites. The red arrows indicate the nucleophilic attacks at the phosphodiester bond at the 5' and 3' splice sites during step 1 and 2.

1.1.2 Components of the spliceosome

The major building blocks of both major and minor spliceosomes are snRNPs. Each snRNP is comprised of one (or two in case of U4/U6 snRNP) U snRNA molecule, seven common Sm or Sm-like (LSm) proteins, the latter in case of U6 and U6atac, and a variable number of snRNP-specific proteins (Will and Lührmann, 2006). The four major snRNPs (U1, U2, U5, U4/U6) are highly conserved in eukaryotes.

The five U snRNAs are numbered according to the order of their discovery: U1, U2, U4, U5 and U6 snRNA (remark: U3 is involved in ribosomal RNA processing). U4 and U6 snRNAs possess a large region of sequence complementarity and base pair *via* stem I and stem II through 22 Watson-Crick base pair interactions and, therefore, are usually found as a U4/U6 di-snRNP (Guthrie and Patterson, 1988). The major and minor spliceosomal snRNPs differ either with respect to their snRNAs and their specific proteins (U1 and U2 in comparison to U11/U12 snRNPs) or only with respect to their snRNAs (U4/U6 in comparison to U4atac/U6atac snRNPs).

All U snRNAs, except U6 and U6atac, are transcribed by RNA polymerase II, receive a 7-methyl-guanylate cap and are exported from the nucleus (Patel and Bellini, 2008). In the cytoplasm, a set of seven Sm proteins (namely B/B', D1, D2, D3, E, F and G) form a ring like structure (Sm ring) on a conserved uridine rich patch on the U snRNA, called the Sm-site (Achsel *et al.*, 1999; Kambach *et al.*, 1999; Kiss, 2004; Pomeranz Krummel *et al.*, 2009; Weber *et al.*, 2010; Leung *et al.*, 2011). The Sm core assembly is mediated by two coordinated multifactorial assembly machineries: the protein arginine methyltransferase 5 (PRMT5) complex and the survival of motor neurons (SMN) complex (Battle *et al.*, 2006; Fischer *et al*,. 2011). After assembly of the Sm core RNPs, the snRNA m_7GpppN caps are hypermethylated by a methyltransferase (Girard *et al.*, 2008) and converted to a 2,2,7-trimethylguanosine ($m_3^{2,2,7}$GpppN) cap, and the premature particles are returned to the nucleus. Constituting a unique feature of U snRNAs, the $m_3^{2,2,7}$G cap has been successfully employed for immunoaffinity-based purification of snRNPs (Bringmann *et al.*, 1983). Final stages of snRNP assembly take place in the nuclear Cajal bodies, where the particle-specific proteins are thought to be added (Schaffert *et al.*, 2004; Stanek and Neugebauer, 2006). Before association with the particle-specific proteins, many nucleotides of each U snRNA are modified by pseudouridylation and 2'-O-methylation (Jády and Kiss, 2001; Kiss, 2004).

Unlike other U snRNAs, the U6 and U6atac snRNA are transcribed by RNA polymerase III and are processed in the nucleus where they acquire a γ-monomethyl cap structure. U6 and

U6atac lack a Sm site and, instead, associate with a set of seven Sm-like proteins (LSm2-8) which bind to the 3' end of U6/U6atac snRNAs, before they interact with the U4/U4atac snRNP to form the di-snRNP (Achsel *et al.*, 1999; Séraphin, 1995).

As mentioned above, every snRNP contains its own specific set of proteins (Fig. 1.4). Considering that all of the identified yeast splicing factors are conserved in metazoans, we mainly refer to the human nomenclature henceforth. The 12S U1 snRNP contains only three additional factors, U1-70K, U1-A and U1-C (Will and Lührmann, 2006). The 17S U2 snRNP consists of U2-A', U2-B'' and the heteromeric complexes SF3a (consisting of the proteins SF3a120, 66 and 60, named according to their apparent molecular weight) and SF3b (SF3b155, 145, 130, 49, p14a/p14, 14b and 10). Additionally, several factors (i.e., U2AF65, U2AF35, SPF31, PUF60, CHERP, hPrp5, hPrp43, SPF30, SPF45 and SR140) have been identified in purified 17S U2 snRNPs, but seem to be loosely bound since they dissociate at salt concentrations higher than 250 mM (Will *et al.*, 2002).

The 13S U4/U6 di-snRNP contains a set of five specific proteins: hPrp3, hPrp31, hPrp4, CypH and 15.5K (hSnu13) (Behrens and Lührmann, 1991; Lauber *et al.*, 1997). The 20S U5 snRNP consists of eight additional factors: hPrp8, hBrr2, hSnu114, hPrp6, hPrp28, 52K (hLin1), 40K and hDib1, (Bach *et al.*, 1989).

Under physiological conditions, the 13S U4/U6 di-snRNP and the 20S U5 snRNP assemble a U4/U6-U5 tri-snRNP in the Cajal bodies (Schaffert *et al.*, 2004) before incorporation into spliceosomesl. This 25S U4/U6.U5 tri-snRNP recruits three additional proteins: hSnu66, hSad1 and 27K (Behrens and Lührmann, 1991). The U5-52K protein is absent in this complex, but is supposed to play a role during assembly of the U4/U6.U5 tri-snRNP (Laggerbauer *et al.*, 2005).

In yeast, a cytoplasmic precursor of U5 snRNP – containing only the U5 snRNA, the Sm proteins, and the U5-specific proteins Prp8p and Snu114p – has been characterized. Pre-U5 snRNP lacks the Brr2 helicase and instead includes the Aar2 protein (Gottschalk *et al.*, 2001; Boon *et al.*, 2007). Recent studies have shown that Aar2 and Brr2 bind, respectively, to an RNase H-like domain and a Jab1/MPN-like (Jab1) domain that lie next to each other in the C-terminal region of the Prp8 protein. Moreover, binding of Aar2 to the Prp8 RNase H-like domain sequestered the Jab1 domain by packaging the two Prp8 domains in a manner incompatible with Brr2 binding (Weber *et al.*, 2011). Furthermore, Aar2 was shown to be phosphorylated at five sites *in vivo* and a phospho-mimetic mutation of Aar2 (S253E) diminished binding to the RNase H-like domain *in vitro* and abolished Aar2-Prp8 interaction in extracts, allowing increased association of Brr2 with Prp8 (Weber *et al.*, 2011).

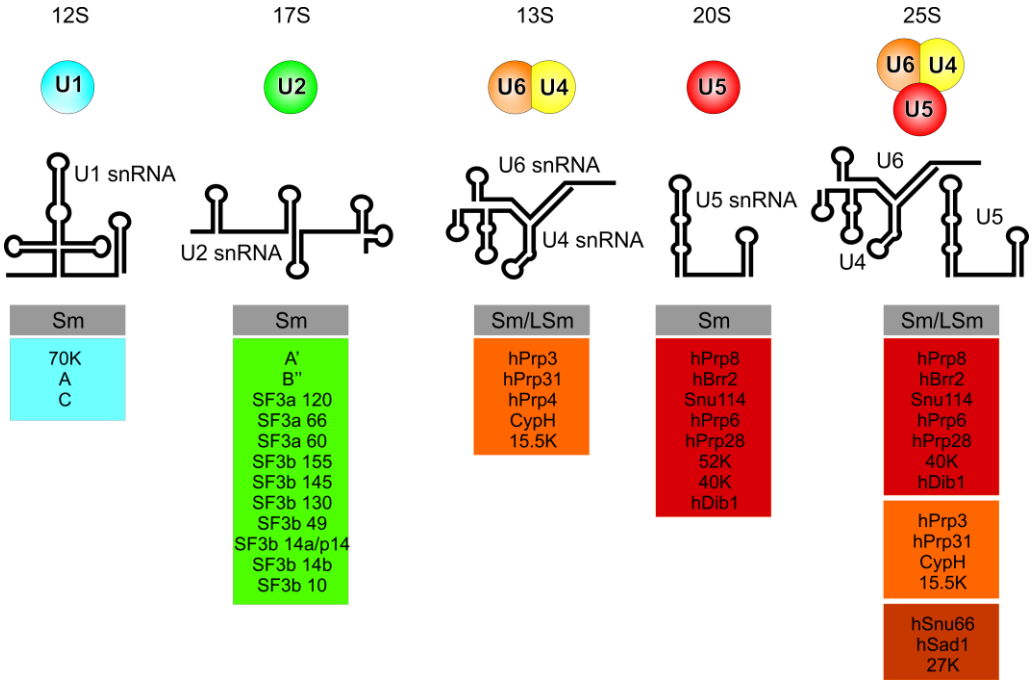


Figure 1.4: Protein composition and snRNA secondary structures of the human major spliceosomal snRNPs. The seven Sm (B/B', D3, D2, D1, E, F and G) or LSm (LSm 2-8) proteins are indicated by "Sm" or "LSm" in a gray box. The colored boxes list the specific proteins associated with each snRNP. The tri-snRNP contains two sets of Sm proteins and one set of LSm proteins.

Important bridging factors between U4/U6 di-snRNP and the U5 snRNP are U4/U6-specific hPrp31 and U5-specific hPrp6 and depletion of one of these two factors abolishes U4/U6.U5 tri-snRNP formation (Makarova *et al.*, 2002; Schaffert *et al.*, 2004). hPrp31 is known to further interact via its NOP domain with 15.5K and the 5' stem loop of U4 snRNA (Liu *et al.*, 2007). The other bridging protein hPrp6 has been shown to associate with the U5 snRNP components hPrp8, hBrr2 and hSnu114, but not with the U5 snRNA (Liu *et al.*, 2006; Makarov *et al.*, 2000).

A considerable number of factors are not stably attached to any U snRNA and are recruited to the spliceosome either individually or as part of a pre-assembled complex. These proteins or complexes are referred to as non-snRNP splicing factors and will be described briefly.

Many spliceosomal SF2 helicases are thought to play an essential role in each step of spliceosome assembly by rearranging inter- and intra-molecular RNA structures or by dissociating RNA-protein complexes using energy from ATP hydrolysis (reviewed in Rocak and Linder, 2004). RNA helicases such as UAP56, hPrp5, hPrp2, hPrp16, hPrp22 and hPrp43, are non-snRNP factors and are only transiently associated with the spliceosome, in contrast to

hBrr2 and hPrp28 helicases which are integral components of the U5 snRNP and U4/U6.U5 tri-snRNP.

SR (serine-arginine) proteins are known to act as activators of constitutive splicing and modulators of alternative splicing (reviewed in Graveley, 2000). SR proteins are typically characterized by the presence of one or more amino-terminal RNA recognition motifs (RRM) and an arginine/serine rich (RS) carboxy-terminal domain (Long and Caceres, 2009). The RRMs are involved in binding to ESEs, whereas the RS domain is engaged in protein-protein as well as protein-RNA interactions, which facilitate the recruitment of spliceosomal factors (Wu and Maniatis, 1993).

The human Prp19/CDC5L complex (NineTeen Complex or NTC – in yeast) is recruited to the spliceosome prior to the first step of splicing and plays an important role during spliceosome activation apparently after U4/U6.U5 tri-snRNP association (Wahl *et al.*, 2009; Makarov *et al.*, 2002; Makarova *et al.*, 2004). The hPrp19/CDC5L complex acts subsequent to the U4 dissociation, stabilizing the association of U5 and U6 with the activated spliceosome and specifying the proper interaction of U5 and U6 with the pre-mRNA prior to step 1 (Chan and Cheng, 2005; Chan *et al.*, 2003).

The RES (pre-mRNA REtention and Splicing) complex, which consists of three factors well characterized in yeast, Snu17, Bud13 and Pml1, was shown to influence the efficiency of the splicing process. This non-snRNP complex proved to be essential in yeast for the retention of unspliced pre-mRNAs in the nucleus (Dziembowski *et al.*, 2004).

It is noteworthy that in contrast to, for example, ribosomal subunits, none of the snRNPs as well as non-snRNP complexes contain a pre-formed catalytic center for the splicing transesterification reactions. Actually, the active center of the spliceosome is formed anew during each spliceosome assembly cycle.

1.1.3 Stepwise assembly of the spliceosome

In order to splice out the introns and ligate the adjacent exons, the 5' SS, BP and 3' SS have to be brought into close proximity. Self-splicing group II introns spontaneously adopt a three-dimensional fold that provides an active site where the reactive groups of the introns are juxtaposed (Toor *et al.*, 2008b; Chan *et al.*, 2012). In contrast, considering the limited information contained in the nuclear introns and the fact that the consensus sequences defining the 5' SS, BP and 3' SS are highly degenerated in metazoans, nuclear introns do not contain sufficient secondary and tertiary structure information to assemble in a productive fold that leads to splicing. As a result, the efficient folding of nuclear pre-mRNA introns in a

manner conductive to splicing depends on many *trans*-acting factors that are brought together with the pre-mRNA to form the spliceosome. For each round of splicing, the spliceosome is assembled *de novo* and undergoes several rearrangements of its components generating well defined intermediate complexes that can be isolated *in vitro* (reviewed in Will and Lührmann, 2011; Wahl *et al.*, 2009).

When introns do not exceed 200-250 nt, the spliceosome assembles across the intron (Fig. 1.5) (Fox-Walsh *et al.*, 2005). In major (U2-dependent) spliceosomes, the assembly starts with the ATP-independent recognition of the 5' splice site by U1 snRNP with the 5' end of the U1 snRNA base-pairing with the 5' splice site of the intron (Ruby and Abelson, 1988; Seraphin and Rosbash, 1989). Also, in the early assembly phase of the spliceosome, SF1/BBP protein and the 65 kDa subunit of the U2 auxiliary factor (U2AF) bind to the BPS and to the PPT, respectively (Berglund *et al.*, 1998). Additionally, the 35 kDa subunit of U2AF binds to the AG dinucleotide of the 3' SS and, together, all these interactions yield the spliceosomal E complex (Hong *et al*., 1997; Das *et al*., 2000).

Subsequent to E complex formation, U2 snRNA engages in base-pairing interactions with the BPS in an ATP-dependent manner, assisted by UAP56 and hPrp5 helicases (Xu *et al.*, 1996; Zhang, 2001; Fleckner *et al.*, 1997). In the short U2-BPS duplex, the branch site adenosine is bulged out offering its 2'-OH as a nucleophile for the first catalytic step (Query *et al.*, 1994). This base-pairing is stabilized by SF3a and SF3b protein complexes from U2 snRNP and by the RS domain of U2AF. Association of U2 snRNP leads to the dissociation of SF1/BBP from the BPS and results in A complex formation (Lim and Hertel, 2004).

In the next step, the pre-formed U4/U6.U5 tri-snRNP particle is recruited to the spliceosome, forming the B complex (Cheng and Abelson, 1987). Although it contains all snRNPs, the B complex is catalytically inactive and requires major compositional and conformational rearrangements. During spliceosome activation, U1 and U4 snRNPs dissociate from the spliceosome, giving rise to the activated spliceosome (B^{act} complex). hPrp28 and hBrr2 helicases are involved in disrupting the base-pair interactions of U1 and the 5' SS and of U4/U6 di-snRNA, respectively (Staley and Guthrie, 1999; Laggerbauer *et al.*, 1998; Raghunathan and Guthrie, 1998). Concomitantly, the 5' end of U6 snRNA substitutes U1 and base pairs with the 5' SS. Extensive base-pairing network is formed between U2 and U6, bringing together the 5' SS and the BP sequence for the first step of splicing. Additionally, a central region of U6 snRNA forms an intramolecular stem-loop structure (U6-ISL) that seems to be crucial for catalysis. U5 snRNA also interacts with nucleotides near the 5' SS.

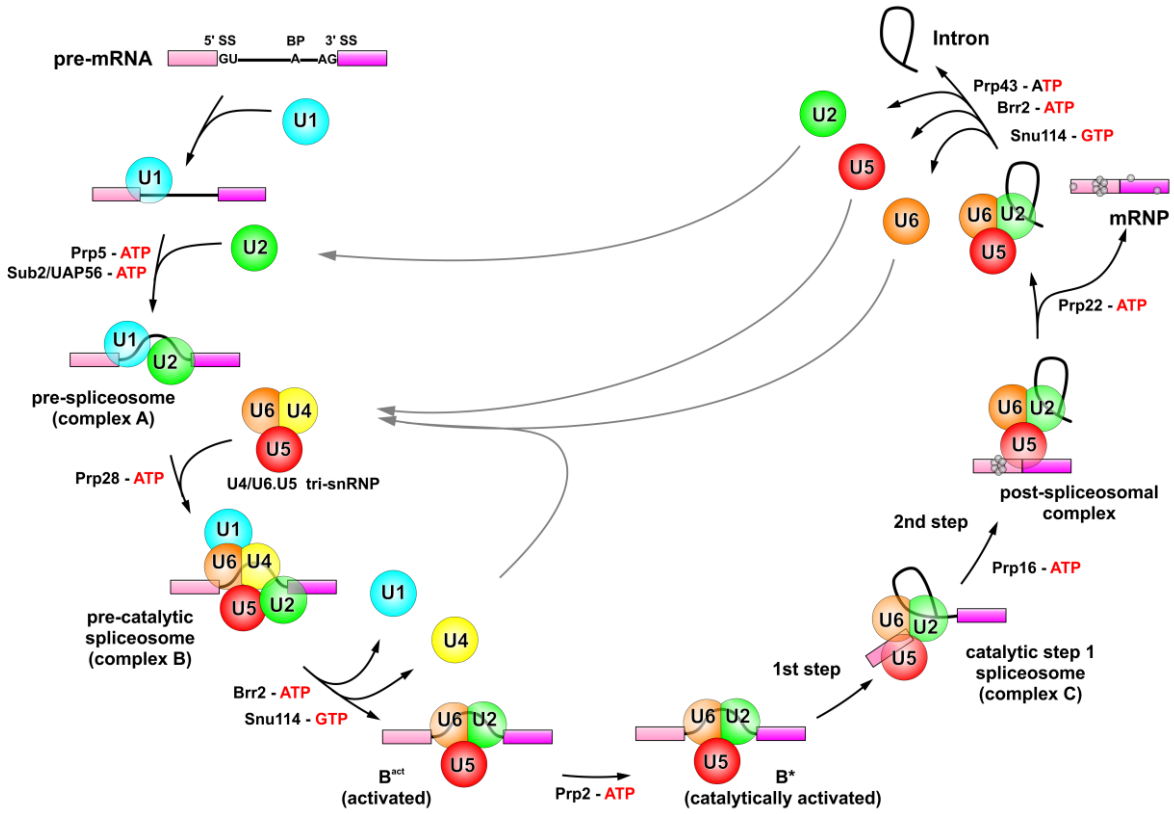


Figure 1.5: Cross-intron assembly and disassembly of the major spliceosome. Only the stepwise interactions of the spliceosomal snRNPs (colored circles) but not those of the non-snRNP factors are shown. The spliceosomal complexes are named according to the metazoan nomenclature. Exons and introns are represented by boxes and lines, respectively. The stages at which remodeling takes place driven by SF2 RNA helicases and the GTPase Snu114 are indicated.

At this stage, Prp2 plays a role in reorganizing the spliceosome, generating the B* complex (a catalytically activated spliceosome), which catalyses the first step of splicing (Kim and Lin, 1993; Warkocki *et al.*, 2009). This yields the C complex. Prior to the second catalytic step, the spliceosome is remodeled again by the Prp16 helicase possibly to reposition the splicing intermediates (Schwer and Guthrie, 1992). Also before step 2, U5 contacts exon nucleotides downstream of the 3' SS and aligns 5' and 3' exons for the second catalytic step. All these events lead to the catalysis of step 2 (reviewed by Umen and Guthrie, 1995; Smith *et al.*, 2008). Finally, the exon junction complex (EJC) is deposited 20 to 25 nt upstream of the exon-exon junction (Le Hir *et al.*, 2000; Bono *et al.*, 2006). The mRNA is then released in the form of an mRNP and transported out of the nucleus (Le Hir *et al.*, 2000; Bono and Gehring, 2011). At the same time, the post-spliceosomal complex (Makarov *et al.*, 2002) is disassembled and the snRNPs are recycled to take part in subsequent splicing events. The released lariat intron (Martin *et al.*, 2002) is debranched by Dbr1 and typically degraded (Chapman and Boeke, 1991).

Alternative spliceosome assembly pathways exist in metazoans, whose mRNAs contain multiple extremely large introns, from several hundred to several thousand nt (Deutsch and Long, 1999) and rather short exons. When intron length exceeds 250 nt, spliceosomal components assemble across an exon, a process called exon definition (Berget, 1995). During exon definition, U1 snRNP binds to the 5' SS downstream of an exon and stimulates the association of U2AF with the PPT and the 3' SS upstream of the same exon. Then, U2 snRNP is recruited to the BPS also upstream of the exon and ESEs recruit proteins of the SR family which stabilize the exon-defined complex (Hoffman and Grabowski, 1992; Reed, 2000). In a subsequent step, these cross-exon interactions must be substituted by cross-intron interactions. However, this process is poorly understood. It is suggested that exon exclusion and skipping during alternative splicing events occurs during the transition from a cross-exon to a cross-intron complex (Sharma *et al.*, 2008).

1.1.4 Rearrangements of the RNA-RNA network during the splicing cycle

An extensive RNA-RNA network is formed and significantly restructured during spliceosome assembly and activation (Fig. 1.6) (reviewed by Will and Lührmann, 2011; Nilsen, 1998).

In the pre-catalytic B spliceosome, the U1 snRNA base pairs with the 5' SS of the pre-mRNA (Zhuang and Weiner, 1986; Siliciano *et al.*, 1991; Heinrichs *et al.*, 1990) and U2 snRNA base pairs with the BPS (Wu and Manley, 1989; Parker *et al.*, 1987; Gozani *et al.*, 1998). In the latter duplex, the BP adenosine, that performs the nucleophilic attack in the first catalytic step, is bulged out. Furthermore, the U2 snRNA forms an internal stem-loop (SL I), while its 5' end base-pairs with the 3' region of the U6 snRNA forming the so-called helix II (Madhani and Guthrie, 1992). Simultaneously, nucleotides of U6 snRNA that are essential components of the spliceosome active site are extensively base-paired with the U4 snRNA forming stem I and stem II (Rinke *et al.*, 1985). Finally, the conserved U5 snRNA loop 1 has been shown to be in the direct vicinity of exon 1 by cross-linking experiments (Newman *et al.*, 1995; Wyatt *et al.*, 1992).

During the activation of the spliceosome, a dramatic remodeling of the RNA-RNA network positions the reactive groups of the pre-mRNA (i.e. the 5' SS and the BP adenosine) in a proper spatial organization for the first step of splicing (Fig. 1.6). The initial U1 snRNA base-pairing to the pre-mRNA is disrupted by Prp28 (Staley and Guthrie, 1999) and replaced by base-pairing between the pre-mRNA and the conserved ACAGA-box sequence in the 5' region of the U6 snRNA (Sawa and Abelson, 1992; Wassarman and Steitz, 1992).

Additionally, the extensive and thermodynamically stable U4/U6 base-pairing is unwound by Brr2 (Laggerbauer *et al.*, 1998; Raghunathan and Guthrie, 1998; Staley and Guthrie, 1999). Once the U4 snRNP has dissociated, short U2/U6 duplexes are formed and U6 snRNA rearranges from an inert to a catalytically active conformation, forming, in particular, the conserved internal stem-loop (ISL) structure which is involved in metal ion binding (Wolff and Bindereif, 1993; McManus *et al.*, 2007; Yean *et al.*, 2000). The initial delivery of U6 in an inactive conformation prevents the splicing cycle to proceed until the reactive groups of the pre-mRNA are properly positioned.

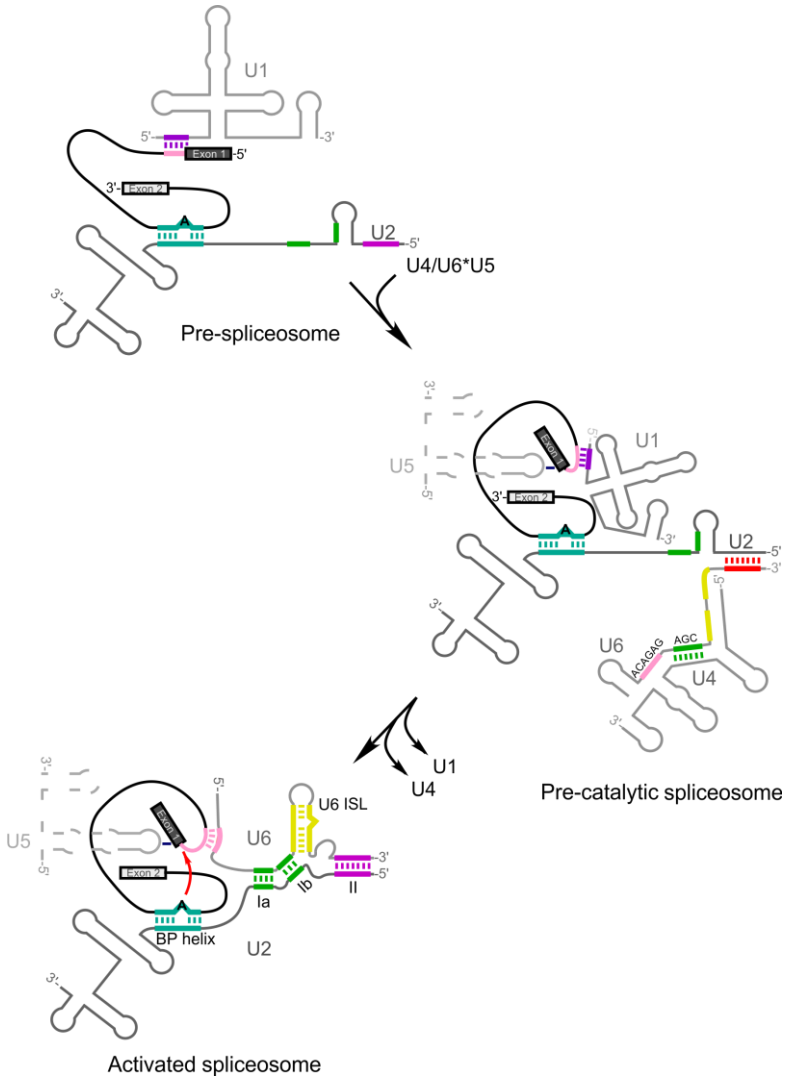


Figure 1.6: Dynamic network of RNA-RNA interactions in the spliceosome. Exons are indicated as boxes and intron sequences by a black line. snRNAs are shown schematically (secondary structure as observed in mammals) in gray, with regions engaging in base paring interactions highlighted in color.

The RNA-RNA network established by U2 and U6 snRNAs plays a crucial role in catalysis. Intriguingly, there are many similarities between the spliceosome and self-splicing group II introns. Both perform splicing through an identical two-step mechanism and the

splicing reaction requires divalent cations (Yean *et al.*, 2000; Sontheimer, 2001). In addition, there are significant similarities in the sequence and secondary structure of snRNAs and some domains of group II introns. The catalytically critical Domain V of group II introns resembles the ISL in U6 snRNA. The structures of Domain V and U6-ISL contain a bulge, known to bind a divalent metal ion (Yean *et al.*, 2000; Seetharaman *et al.*, 2006), and a conserved catalytic triad, AGC, at their 5' ends. Besides, Domain VI of group II introns contains an equivalent of the structure formed between U2 snRNA and the BPS, with the adenosine nucleotide, which carries the nucleophile of the first step, bulged out. Additionally, three sequences in Domain I of group II introns bind exonic sequences in a manner similar to U5 snRNA and another subdomain of Domain I recognizes the 5' SS and is considered a functional counterpart of U1 snRNA. Such parallels lead to the hypothesis that snRNAs might be evolutionary descendants of group II introns and, therefore, have a catalytic role during splicing (reviewed in Valadkhan, 2010).

In agreement with the proposal of an RNA-based catalysis, U6 and U2 snRNAs can catalyze a two-step reaction with short RNA substrates that remotely resemble the one catalyzed by the spliceosome (Valadkhan *et al.*, 2007, 2009). Additionally, structural studies of self-splicing group II introns place the elements equivalent to the U6-ISL and ACAGA-box as key active site components (Toor *et al.*, 2008a; Chan *et al.*, 2012). However, most similarities between group II introns and the spliceosome are related to the chemistry of splicing and the RNA components. Considering that the spliceosome is a protein-rich RNP, it is difficult to decide if the parallels between group II introns and the spliceosome truly reflect a common ancestry or if they are just a result of convergent evolution impelled by the need to catalyze the same reaction. However, recent studies have shown that Prp8, the largest and most conserved protein in the spliceosome, shares similarities to group II intron maturases (Dlakić and Mushegian, 2011). As mentioned above in section 1.1, maturases promote splicing by inducing and stabilizing the catalytically active RNA structure. The self-splicing process is very inefficient in the absence of maturases as is the splicing-like reaction catalyzed by a protein-free RNA duplex similar to U2/U6. The Prp8 protein interacts with all the reactive groups in the intron, 3' SS, 5' SS and BPS, and with U5 and U6 snRNAs, thus constituting a major scaffold and regulator of the RNA-RNA network in the spliceosome (reviewed in Grainger and Beggs, 2005). Therefore, Prp8 and maturases seem to play similar roles in their corresponding systems, providing additional evidence for a common evolutionary ancestry of group II introns and the spliceosome.

1.1.5 Compositional changes within the spliceosome

Despite the indications that pre-mRNA splicing catalysis is at least partially RNA-based, the spliceosome, unlike group II introns, requires a plethora of protein factors to assemble the introns and the snRNAs in a catalytic structure. These proteins play crucial roles in SS recognition; facilitate dynamics of RNA-RNA and RNA-protein interactions and ensure the proper arrangement of the catalytic centers of the spliceosome.

The protein composition of affinity purified spliceosomal complexes stalled at certain assembly stages has been determined by mass spectrometry. These studies demonstrated that the composition of the spliceosome is highly dynamic with remarkable exchanges of proteins from the assembly stage, throughout activation and disassembly (Makarov *et al.*, 2002; Makarova *et al.*, 2004; Behzadnia *et al.*, 2007; Bessonov *et al.*, 2008, 2010; Deckert *et al.*, 2006).

Besides the human and *D. melanogaster* spliceosomes, the protein composition of affinity purified, *in vitro* assembled *S. cerevisiae* spliceosomal complexes have been recently determined (Fabrizio *et al.*, 2009). It can be noted that the yeast splicing machinery contains the evolutionary conserved core of spliceosomal proteins required for constitutive splicing. Additional proteins found in higher Eukarya spliceosomes are mainly implicated in alternative splicing, a process mostly absent in yeast.

A dramatic exchange of proteins occurs during spliceosome assembly and activation. Interestingly, the proteins involved in dissociation/recruitment during B complex to C complex transitions are homologous in yeast and metazoans indicating that not only the proteins but also the compositional dynamics of the splicing machinery are evolutionarily conserved (Fabrizio *et al.*, 2009; Bessonov *et al.*, 2008, 2010).

Proteomic analysis of human spliceosomal A complex revealed that it consists of ten A complex-specific proteins besides U1 and U2 snRNPs (Behzadnia *et al.*, 2007). These non-snRNPs leave the spliceosome during A to B complex transition while ~60 other proteins are recruited. Thus, the B complex contains U1 and U2 snRNPs, the U4/U6.U5 tri-snRNP plus 35 non-snRNP proteins, including the hPrp19/CDC5L and RES complexes and a group of B complex-specific proteins (Deckert *et al.*, 2006). In the conversion from B to B^{act} complex, all U1 and U4/U6 proteins are lost. In contrast, several proteins are recruited to or become more stably associated with the spliceosome. All hPrp19/CDC5L complex proteins as well as related proteins are more abundant in B^{act}. The presence of hPrp2 in the purified B^{act} complexes indicates that these complexes have not yet undergone catalytic activation, as Prp2

is known to dissociate from the spliceosome after catalyzing the remodeling step that yields a catalytically active B* complex (Kim and Lin, 1996). In the transition from B^{act} to C complex, the new additions to the spliceosome mainly consist of C-complex specific proteins and the so-called step II factors, which are proteins known to function prior to or during the second transesterification reaction. Additionally, SF2 helicases and peptidyl-prolyl isomerases (PPIases) are recruited to the C complex, potentially playing a role in RNP remodeling at this stage of splicing. Furthermore, SF3a and SF3b proteins are specifically destabilized from the human spliceosome during B^{act} to C complex transition. Finally, members of the exon junction complex (EJC), which are important for mRNA translation, are recruited at this stage (Bessonov *et al.*, 2008, 2010).

In yeast, drastic rearrangements occur in the transition from B to B^{act} complex (Fig. 1.7), the latter representing the spliceosome prior to the final catalytic activation mediated by Prp2. U1 snRNP is released as well as the U4 snRNA and all the U4/U6 associated proteins. At this stage, 12 B^{act} proteins are recruited that may be either involved in establishing/stabilizing U2/U6 base-pairing (such as Ecm2 and Cwc2) or promoting step 1 (Prp2, Spp2 and Yju2). Comparison of B^{act} with C complexes (Fig. 1.7) shows that key proteins that are required at later stages of splicing are recruited during this transition. At least nine proteins, mainly step 2 factors, as well as the trimeric disassembly NTR complex join the spliceosome during C complex formation (Fabrizio *et al.*, 2009). Due to the limited number of proteins recruited during B^{act} to C complex transition, it has been possible to investigate the requirements of some of these factors for step 1 or 2 by complementing purified spliceosomes ($B^{act\Delta prp2}$), which were stalled before the catalytic activation step mediated by Prp2, with purified recombinant splicing factors (Warkocki *et al.*, 2009). The ability to rescue both steps of splicing in yeast using purified components could possibly help to elucidate the role of some RNA helicases implicated in RNP remodeling during activation.

It is important to mention that not only the protein composition during the splicing cycle changes but also the extent to which these proteins are modified. Several spliceosomal proteins are post-translationally modified prior to or during their involvement in the splicing cycle (Mathew *et al.*, 2008; Soulard *et al.*, 1993). Post-translational modification patterns constitute a code for recruitment of mRNA processing factors once they generate structural-mediated transitions that provide new interaction platforms. Consequently, these modifications play regulatory roles in the progression of splicing (Wahl *et al.*, 2009).

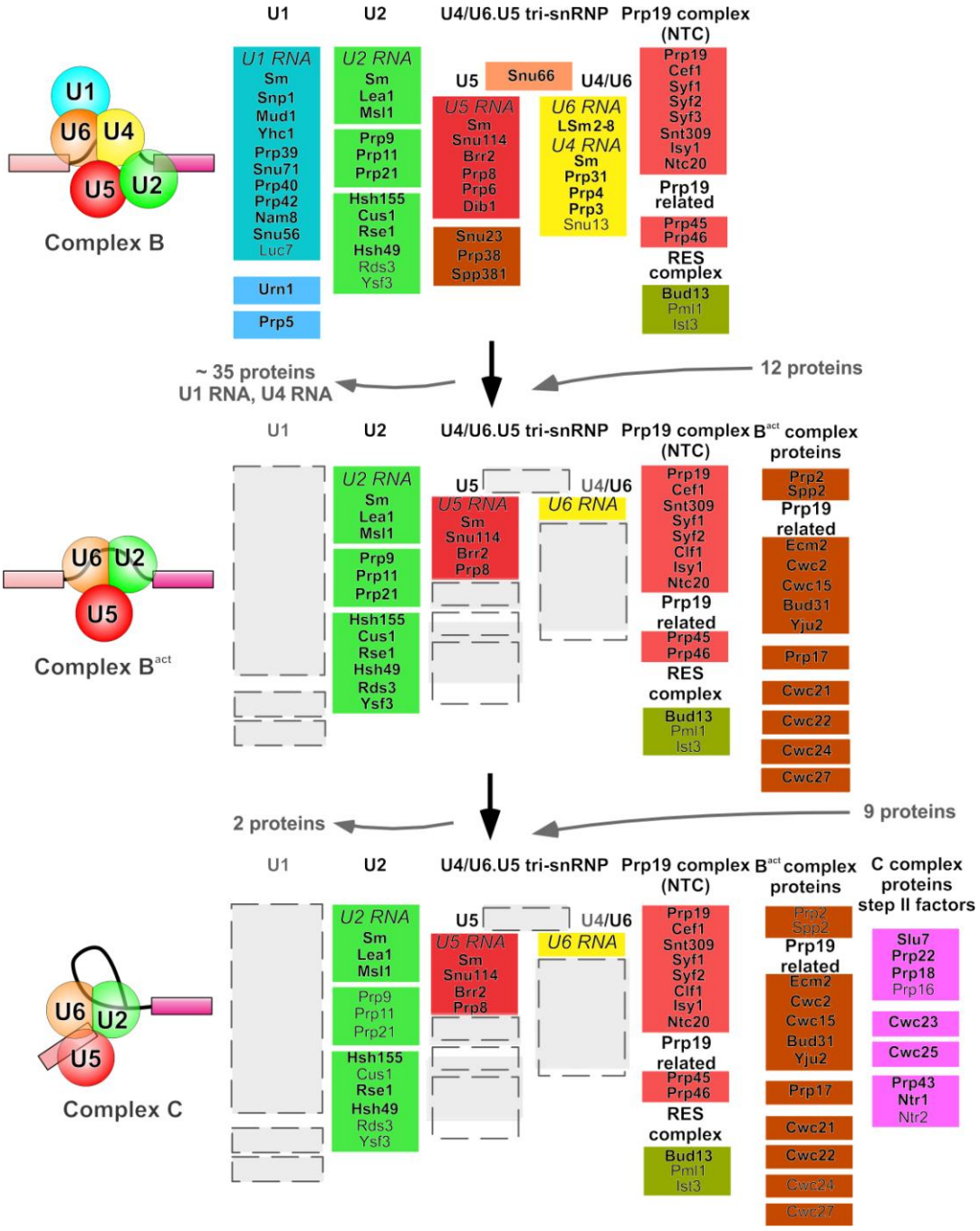


Figure 1.7: Compositional dynamics of the yeast spliceosome. The protein composition of B, B^{act} and C complexes of *S. cerevisiae*, identified by mass spectrometry, are shown. Proteins are grouped according to their function or association with a snRNP. The relative abundance of the proteins is indicated by the light (substoichiometric) or dark (stoichiometric) lettering. Considering that the compositional dynamics of the splicing machinery are evolutionarily conserved between yeast and human, only the representative scheme of the compositional dynamics of the yeast spliceosome is shown for simplicity.

Several enzymes responsible for introduction or removal of post-translational modifications are found in the spliceosomal complexes, such as SR protein kinases 1 and 2, Prp4 kinase and Clk/Sty kinase. In general, phosphorylation predominantly occurs during spliceosome assembly and activation, whereas dephosphorylation is more prominent during

catalysis and disassembly. Furthermore, spliceosomal proteins seem to undergo other types of modifications such as ubiquitination (Bellare *et al.*, 2008) and acetylation (Kuhn *et al.*, 2009).

1.2 Helicases: enzymatic motors of the spliceosome

The sequential rearrangements in the RNA-RNA and RNA-protein networks during assembly, activation and catalysis of the spliceosome are mainly driven by RNA unwindases/RNPases (Staley and Guthrie, 1998). The energy of nucleotide triphosphate hydrolysis is harnessed by these enzymes and is coupled to structural/compositional rearrangements, such as remodeling of snRNA conformations and positioning of splice sites at the catalytic center, at several steps of the splicing cycle. The activity of spliceosomal helicases must be highly coordinated; a task carried out by other spliceosomal proteins and modulated, in some cases, by post-translational modifications. These RNA helicases also govern timing and fidelity of splicing by facilitating the discard of aberrant/nonproductive splicing intermediates (Smith *et al.*, 2008).

1.2.1 RNA helicases

Helicases are highly conserved enzymes that bind or remodel DNA/RNA or protein-nucleic acid complexes in an energy-dependent manner utilizing the energy of hydrolysis of nucleotide triphosphate (Cordin *et al.*, 2006). DNA helicases unwind double-stranded DNA regions and are involved in DNA replication, recombination and repair as well as overall genome stability (Singleton *et al.*, 2007). RNA helicases are found in all kingdoms of life, where they are the largest group of enzymes involved in RNA metabolism (Anantharaman *et al.*, 2002). Many RNA helicases play critical roles during conformational changes of RNA and RNP complexes (Linder, 2006), in displacing proteins from RNA-protein complexes (Fairman *et al.*, 2004), in mediating RNA annealing or unwinding (Yang and Jankowsky, 2005; Halls *et al.*, 2007; Rajkowitsch and Schroeder, 2007), as well as during timing and proofreading events (Tanner and Linder, 2001; Staley and Guthrie, 1998). Although most RNA helicases receive their helicase designation based on highly conserved sequence motifs (Fairman-Williams *et al.*, 2010; Jankowsky *et al.*, 2011), ATP-driven RNA duplex unwinding has been shown only for a subset of proteins *in vitro* (Jankowsky *et al.*, 2005; Linder, 2006; Tanner and Linder, 2001). Therefore, it is assumed that most of the proteins categorized as RNA helicases are presumably capable to separate RNA duplexes if a suitable substrate is provided.

Both DNA and RNA helicases fall into two categories, those that form oligomeric (mostly hexameric) rings and those that do not. Based on sequence and comparative structural and functional analyses, all helicases are classified into six superfamilies (SFs). The ring-forming helicases comprise SFs 3 to 6 and the non-ring forming ones comprise SFs 1 and 2. All eukaryotic RNA helicases belong to SFs 1 and 2. Ring-shaped RNA helicases are found in bacteria and viruses (Jankowsky, 2011).

SF1 and SF2 include DNA and RNA helicases that contain a structurally conserved helicase core formed by two highly similar α-β RecA-like domains. Both SF1 and SF2 helicases share 12 characteristic sequence motifs at defined positions in the helicase core and seem to act as monomers or dimers (Jankowsky, 2011; Fairman-Williams *et al.*, 2010). However, not all motifs are present in each helicase family. The highest level of conservation across both SFs is seen in the residues that coordinate binding and hydrolysis of the triphosphate (motifs I, II and VI). These residues are located in the cleft between the two conserved RecA-like domains. Interestingly, the spatial arrangement displayed by these residues is highly conserved in other NTPases (Walker *et al.*, 1982; Fairman-Williams *et al.*, 2010).

Most RNA helicases are members of SF2. RNA helicases are found in five families belonging to SF2 (Ski2-like, RIG-I-like, DEAD-box, DEAH/RHA and NS3/NPH-II) and in one family of SF1 (Upf1-like) (Fairman-Williams *et al.*, 2010). In RNA helicases of SF1 and SF2, the two RecA-like domains are separated by a linker that allows movement of the domains relative to each other upon substrate binding. The binding of a single-stranded RNA substrate to the core of these helicases is achieved when residues of motifs Ia, Ib and Ic in the RecA-1 domain and of motifs IV, IVa, V and Vb in the RecA-2 interact with the phosphate backbone of the oligonucleotide. Concomitant binding of RNA substrate and ATP brings together the two RecA domains resulting in the formation of an NTP binding pocket competent for hydrolysis (Weir *et al.*, 2010; Montpetit *et al.*, 2011; Sengoku *et al.*, 2006). For the enzymes of all these families, NTP binding is executed in a similar manner with residues from motifs I, II and VI contacting the phosphates and the coordinated magnesium ion, while residues from motif VI are coordinating the ribose (reviewed by Cordin *et al.*, 2012).

In addition to the helicase core region which contains the conserved motifs, most RNA helicases contain variable N- and/or C-terminal extensions. These amino- and carboxyl-terminal extensions are highly variable in size and composition. This modular organization suggests that the core region functions as an NTP-dependent motor while the terminal extensions might confer substrate specificity, include protein or RNA binding motifs and/or

might direct the protein to a specific subcellular localization (Wang and Guthrie, 1998). Consistent with the critical roles in establishing physiological specificity for individual enzymes, C-terminal and N-terminal accessory domains are usually not conserved within a family. However, recent studies have revealed some degree of structural conservation of the C-terminal extension in the Ski2-like and DEAH families (He *et al.*, 2010; Büttner *et al.*, 2007). Additionally, most helicases that unwind duplexes with defined polarity have functionally important accessory domains located on top of the nucleic acid binding site on the helicase core. This arrangement encloses the bound nucleic acid strand to some extent, possibly facilitating directional translocation upon which polar unwinding is based (Fairman-Williams *et al.*, 2010).

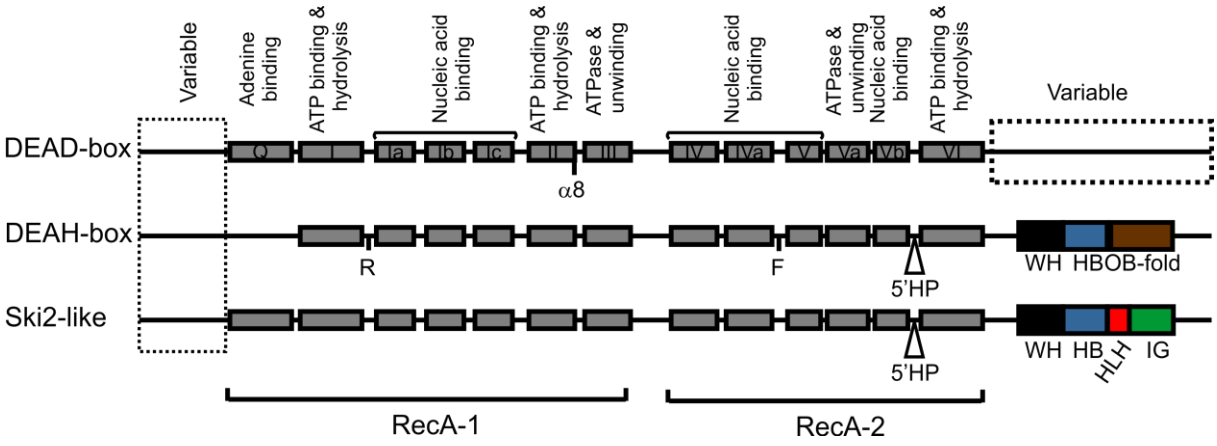


Figure 1.8: The conserved motifs of SF2 helicases. The motifs and functionally important features of DEAD-box, DEAH/RHA and Ski2-like helicases are represented. The length of the boxes is not to scale. All the characteristic sequence motifs summed to comparative structural and functional analyses determine the family to which a helicase belongs. For the Ski2-like family, the N-terminal cassette of Brr2 is represented.

Since all spliceosomal helicases belong to the Ski2-like, DEAD-box and DEAH/RHA families of SF2, a brief description of these families will follow.

1.2.1.1 DEAD-box family

The DEAD-box proteins were first identified by sequence alignments of eight homologues of yeast eIF4A translation initiation factor that highlighted the presence of several conserved motifs in the RecA-like domains (Linder and Slonimski, 1989). The name of the family derives from the amino acid sequence D-E-A-D (Asp-Glu-Ala-Asp) of its Walker B motif (motif II). The DEAD-box family is the largest within SF2 and appears to contain exclusively RNA helicases (Tanner *et al.*, 2003; Fairman-Williams *et al.*, 2010). DEAD-box helicases only hydrolyze ATP since they possess an adenine recognition motif (Q-motif) that interacts specifically with the adenine base. Furthermore, an interaction between an α-helix in the end of motif II and an arginine of motif V controls RNA binding. Simultaneous ATP and RNA

binding by DEAD-box helicases results in local duplex separation. Subsequently, ATP hydrolysis leads to RNA release. Interestingly, removal of stabilizing proteins can bypass the requirement of these helicases in the spliceosome (Kistler and Guthrie, 2001; Perriman and Ares, 2007; Chen *et al.*, 2001). Consequently, it is plausible that ATP hydrolysis by spliceosomal DEAD-box helicases may be required to time conformational switching events rather than to force conformational rearrangements through mechanical movement (Cordin *et al.*, 2012).

1.2.1.2 DEAH/RHA family

DEAH/RHA helicases hydrolyze any nucleotide tri-phosphate *in vitro* (Fairman-Williams *et al.*, 2010). In this family, the base of the bound nucleotide is twisted by 150º compared to the position adopted by the base in DEAD and Ski2-like helicases (He *et al.*, 2010; Walbott *et al.*, 2010; Pyle, 2008). Consequently, the base is sandwiched non-specifically between the two RecA-like domains. These helicases also possess a long β-hairpin located between motifs Va and VI which is proposed to control access to the RNA binding pocket in an ATP-dependent manner (He *et al.*, 2010). When ADP is bound, RNA binding is prevented. Additionally, other structural features such as winged helix, OB-fold and helical bundle domains cooperate with the β-hairpin to modulate RNA binding. Alternatively, this β-hairpin might be functionally similar to the strand separating device in the Ski2-like Hel308 DNA helicase. Interestingly, the presence of this β-hairpin correlates with polar unwinding, observed in DEAH/RHA, NS3/NPH-II and Ski2-like.

1.2.1.3 Ski2-like family

Ski2-like, as DEAD-box helicases, selectively hydrolyze ATP due to the presence of the Q-motif (Fairman-Williams *et al.*, 2010). In this family, nucleic acid substrates preferentially bind with 3' to 5' directionality across the first and second RecA-like domains, which could explain the 3' to 5' unwinding directionally observed in these enzymes (Hopfner and Michaelis, 2007). In particular, 3' to 5' translocation has been shown for Hel308 (Guy and Bolt, 2005), Mer3/HFM1 (Nakagawa and Kolodner, 2002) and Mtr4 (Bernstein *et al.*, 2008). Ski2-like helicases also contain a β-hairpin which is shorter compared to the one from DEAH-box helicases and is probably not involved in controlling RNA binding. This β-hairpin is proposed to act as a duplex separation device which takes apart the duplex strands when one strand of the RNA substrate is pulled in a 3' to 5' direction (Büttner *et al.*, 2007).

1.2.2 Spliceosomal RNA helicases

SF2 RNA helicases are probably the main driving force behind the extensive structural remodeling of the spliceosome, conferring speed, accuracy and directionality (Staley and Guthrie, 1998). They act at specific steps of splicing to facilitate transactions between mutually exclusive RNA-RNA or RNA-protein interactions. These helicases displace interaction partners (RNA or proteins) from particular regions of the pre-mRNA and/or snRNAs and, consequently, liberate RNA regions allowing them to engage in new base pairing interactions or bind a protein (Pyle, 2008; Staley and Guthrie, 1998).

Of the eight spliceosomal SF2 RNA helicases, three belong to the DEAD-box family (Sub2/UAP56, Prp5 and Prp28), four to the DEAH/RHA family (Prp2, Prp16, Prp22 and Prp43) and one to the Ski2-like family (Brr2). All of them are conserved between yeast and humans.

The DEAD-box proteins Prp5 and Sub2/UAP56 are required for pre-spliceosome formation. Sub2/UAP56 facilitates U2 addition to the spliceosome by displacing U2AF and/or BPP from the BPS (Kistler and Guthrie, 2001). Prp5 is implicated in remodeling of U2 snRNP, thereby facilitating U2 snRNA binding to the BPS. Prp5 may also play a role in proofreading the stability of the U2/BPS duplex (Xu and Query, 2007). Prp28 catalyzes the exchange of U1 for U6 at the 5' SS (Staley and Guthrie, 1999) during the transition from B to B^{act} complexes. Prp28 actively dissociates U1-C from the 5' SS, which is the protein that stabilizes the U1/5' SS base pairing interaction (Chen *et al.*, 2001). In humans, phosphorylation of Prp28 is required for its association with the tri-snRNP and the subsequent integration of this snRNP during B complex formation (Mathew *et al.*, 2008). It is presently not known if this post-translational modification regulates the function of Prp28. Thus, Sub2/UAP56, Prp5 and Prp28 seem to catalyze RNA-protein rearrangements in the spliceosome rather than unwind RNA duplexes. This is in agreement with the fact that, in the spliceosome, the targets of these DEAD-box helicases are unstable RNA structures requiring stabilization by proteins which are then displaced as a consequence of the helicase activity (Kistler and Guthrie, 2001; Perriman and Ares, 2007; Chen *et al.*, 2001; Staley and Guthrie, 1999).

The Ski2-like helicase Brr2 catalyzes a crucial step in spliceosome activation, the unwinding of the U4/U6 di-snRNA. Since the Brr2 helicase is the topic of this thesis, it will be described in more detail in the next section.

The DEAH/RHA protein Prp2 plays a role before step 1 of splicing promoting a poorly understood remodeling that converts B^{act} into B* complex (Warkocki *et al.*, 2009). The association of Prp2 with the spliceosome is transient and Prp2 is recruited through the interaction of its C-terminal region with Spp2, a G-patch protein (Roy *et al.*, 1995; Last *et al.*, 1987). *In vitro* biochemical experiments with purified Prp2 showed that ATP hydrolysis is stimulated by single-stranded RNA but not by DNA or RNA duplexes (Kim *et al.*, 1992). In contrast to the other DEAH/RHA spliceosomal helicases, Prp2 seems not to unwind RNA duplexes *in vitro* (Kim and Lin, 1996; Wang and Guthrie, 1998; Wagner *et al.*, 1998; Tanaka *et al.*, 2007; Martin *et al.*, 2002). Prp16 promotes conformational rearrangements in the spliceosome required for step 2 of splicing. However, the precise nature of such remodeling is unclear. In yeast, Prp16 interacts genetically with Prp8 (Query and Konarska, 2004) and U6 snRNA (Madhani and Guthrie, 1992) suggesting that it may act on a structure containing these components. Prp16 unwinds double-stranded RNA *in vitro* (Schwer and Guthrie, 1992; Wang and Guthrie, 1998) and also joins the spliceosome transiently. After NTP-mediated activity, Prp16 dissociates from the spliceosome (Schwer and Guthrie, 1992). Prp16 also regulates fidelity of BPS recognition, promoting the discard of aberrant lariat intermediates (Burgess and Guthrie, 1993). Another DEAH/RHA helicase, Prp22 not only functions subsequent to Prp16 in an ATP-dependent fashion in the catalysis of step 2 but it is also required for the ATP-independent release of the mRNA from the spliceosome. It is suggested that Prp22 displaces U5 snRNP from the mRNA by disrupting Prp8 and U5 snRNA interactions with the exon, leading to mRNA release (Aronova *et al.*, 2007; Schwer, 2008; Schwer and Gross, 1998). Prp22 is also involved in maintaining fidelity of exon ligation by repressing the splicing of aberrant splicing intermediates (Mayas *et al.*, 2006). Prp43 is implicated in the release of the excised intron from the post-splicing complex, which is accompanied by release of U2, U5 and U6. In yeast, Prp43 forms a complex with Ntr1 and Ntr2 which recruit Prp43 to the spliceosome and are further required for spliceosome disassembly (Tsai *et al.*, 2007). Additionally, Ntr1 stimulates the helicase activity of Prp43 (Tanaka *et al.*, 2007). In contrast to the DEAD-box family, DEAH/RHA and Ski2-like helicases can load on the RNA substrate through a 3' single stranded overhang and, by means of an ATP-dependent ratcheting mechanism, pull the substrate strand into the RNA binding groove. Therefore, these DEAH/RHA and Ski2-like helicases are involved in remodeling the extensive RNA-RNA and RNA-protein networks of a fully assembled spliceosome fulfilling the requirement for *bona fide* helicases actively rearranging RNA structures.

Furthermore, the regulatory mechanisms controlling the ATPase/helicase/RNPase activities of spliceosomal helicases are largely unknown. Most of these RNA helicases join the spliceosome transiently at the stage at which their activities are required. Therefore, the sequential recruitment of these helicases may be explained by overlapping binding sites and/or by conformational rearrangements elicited by one helicase that could activate or create the substrate for the following one. Additionally, specific partners mediate the recruitment of a helicase, such as Spp2 recruiting Prp2 (Roy *et al.*, 1995), and/or modulate its activity, as for Ntr1 and Prp43 (Tsai *et al.*, 2007). Interestingly, DEAH/RHA and Ski2-like helicases share C-terminal structural features such as the winged helix and helical bundle domains. Since these domains participate in RNA binding and contact the RecA domains, binding of an effector protein to such structures might remodel the active site and affect RNA binding and/or ATP hydrolysis. Along the same line, Prp2, Prp16, Prp22 and Prp43 exhibit high sequence conservation within their C-terminal regions, which comprise a winged helix (WH), a helix-loop-helix (HLH) and an oligonucleotide/oligosaccharide-binding (OB fold) domain (Arenas and Abelson, 1997; Burgess *et al.*, 1990; Chen and Lin, 1990; Company *et al.*, 1991; Kudlinzki *et al.*, 2009), also suggesting a conserved regulatory mechanism. Although DEAD-box proteins possess poorly conserved N- and C-terminal domains, recent reports indicate that such helicases can bind effector proteins through a conserved surface in their RecA-2 domain (Montpetit *et al.*, 2011; Oberer *et al.*, 2005; Tritschler *et al.*, 2009).

As mentioned previously, Prp5, Prp16 and Prp22 are considered important guarantors of splicing fidelity. It has been suggested that the function of these helicases is similar to the role played by the elongation factor eEF1A/EF-Tu during translation (Burgess *et al.*, 1990). The kinetic proofreading model for splicing fidelity suggests that the rate of ATP hydrolysis by an RNA helicase acts as a timer and influences the equilibrium between two competing events: stalling of the spliceosome or splicing catalysis (Burgess and Guthrie, 1993). The stalled spliceosome can either reorganize and proceed in the splicing cycle or be discarded. Prp43 is then implicated in the disassembly of such rejected spliceosomal complexes (Koodathingal *et al.*, 2010; Pandit *et al.*, 2006; Mayas *et al.*, 2010). However, the mechanisms underlying kinetic proofreading and spliceosome rejection are poorly understood.

1.2.3 Brr2: a helicase essential for spliceosome catalytic activation

Brr2 is an integral component of the U5 snRNP (Lauber *et al.*, 1996; Noble and Guthrie, 1996) and enters the spliceosome as part of the U4/U6.U5 tri-snRNP. Thereafter, Brr2 remains stably associated with the catalytic core of the spliceosome (Bessonov *et al.*, 2008).

Brr2 is much larger than other spliceosomal helicases (ca. 245 kDa for the human enzyme) and is the only member of the Ski2-like family in the spliceosome. Sequence analysis revealed that the protein comprises ca. 450 amino acid N-terminal region, predicted to be largely unstructured with exception to the region corresponding to residues 258-338 that are predicted to adopt a PWI-like helical bundle fold (Korneta *et al.*, 2012). This motif has also been encountered in the N-terminus of other spliceosomal helicases such as Prp2 and Prp22 and seems to be implicated in nucleic acid binding. The N-terminal extension is followed by a tandem repeat of helicase units, herein referred to as "cassettes". Both cassettes are envisaged to contain dual RecA-like ATPase domains followed by a domain that bears resemblance to the Sec63 subunit of the protein translocation apparatus of the endoplasmic reticulum, henceforth referred to as Sec63 units (Fig. 1.9) (Ponting, 2000). Although RecA-like domains are common to all helicases, only Brr2, the Ski2-like helicase 1 (Slh1p, an RNA helicase involved in antiviral defense (Martegani *et al.*, 1997), ASCC3 (DNA helicase of the activating signal cointegrator complex involved in genome maintenance) (Dango *et al.*, 2011) and Hfm1p/Hfm1 DNA helicase (Ponting, 2000) are predicted to contain Sec63 units.

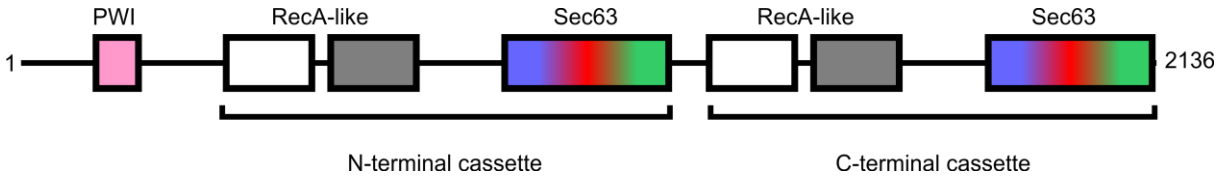


Figure 1.9: Special structure organization of Brr2. Sequence analysis predicts that Brr2 is comprised of an N-terminal region thought to be mostly unstructured, with exception of the PWI domain, and a tandem repeat of helicase cassettes. Each cassette (represented as N-terminal and C-terminal cassettes) harbors typical dual RecA-like domains followed by a Sec63 homology unit. The numbers indicated refer to the human sequence.

The dual cassette organization of Brr2 is not found in any other spliceosomal helicase and is shared by only few other known enzymes, including the RNA helicase Slh1p (Martegani *et al.*, 1997) and the ASCC3 DNA helicase (Dango *et al.*, 2011). The conserved domain structure and sequence suggest that the two cassettes of Brr2 originated from a gene duplication event and subsequently diverged. Genetic analyses suggested that only the ATPase and helicase activities of the N-terminal cassette of Brr2 are required for splicing (Kim and Rossi, 1999). Nevertheless, the C-terminal cassette is essential for yeast viability (Zhang *et al.*, 2009), represents a versatile protein-protein interaction platform (Liu *et al.*, 2006; van Nues and Beggs, 2001), may have retained nucleotide binding capacity (Hahn and Beggs, 2010) and may influence the activity of the N-terminal helicase (Zhang *et al.*, 2009; Hahn and Beggs, 2010).

Catalytic activation of the spliceosome involves the unwinding of the U4 and U6 snRNAs, which are extensively base-paired *via* two regions (stems 1 and 2) when delivered to the spliceosome in the framework of the U4/U6.U5 tri-snRNP. Brr2 is thought to elicit these rearrangements since the enzyme unwinds U4/U6 duplexes *in vitro* (Laggerbauer *et al.*, 1998; Raghunathan and Guthrie, 1998) and Brr2 mutations interfere with catalytic activation (Noble and Guthrie, 1996; Kim and Rossi, 1999; Zhao *et al.*, 2009). Furthermore, Brr2 and its U4/U6 substrate are pre-assembled in the U4/U6-U5 tri-snRNP before incorporation into the spliceosome. However, U4/U6 dissociation must be delayed until splice sites have been reliably located during spliceosome assembly. Additionally, unlike other spliceosomal helicases, Brr2 is stably associated with the spliceosome throughout the splicing cycle and is required again during spliceosome disassembly (Small *et al.*, 2006). Thus, Brr2 has to be strictly controlled to prevent premature unwinding of U4/U6 in the tri-snRNP and to allow its repeated on- and off-switching during splicing.

At least some regulation of Brr2 activity is exerted by two other U5 snRNP proteins, Prp8 (Maeder *et al.*, 2009) and Snu114 (Small *et al.*, 2006). A C-terminal fragment of Prp8, comprising an RNase H like and a Jab1/MPN domain, has been shown to interact with Brr2 and stimulate its helicase activity (Maeder *et al.*, 2009). In yeast, Prp8 is reversibly ubiquitinylated and its ubiquitinylated form represses Brr2 activity (Bellare *et al.*, 2008). Furthermore, Jab1/MPN domain of Prp8 binds ubiquitin (Bellare *et al.*, 2006), implying that upon deubiquitinylation of Prp8, this domain could be liberated to stimulate Brr2 activity. These findings suggest that this post-translational modification plays a role in regulating Brr2 activity. The GTPase Snu114, homologous to the ribosomal elongation factor eEF2 that catalyzes structural rearrangements in the ribosome during translocation, also seems to be required during catalytic activation. It was shown that GDP-bound Snu114 impairs Brr2 helicase activity while the GTP-bound form seems to elicit helicase activity (Small *et al.*, 2006). Snu114 and Brr2 interact with the same C-terminal region of Prp8 (Liu *et al.*, 2006) implying that Prp8 may function as a platform coordinating the activities of both proteins. Additional support for this idea comes from studies showing that Prp8, Sn114 and Brr2 form a stable salt-resistant complex (Achsel *et al.*, 1998) and may act as a functional unit within the spliceosome.

A crystal structure of the archaeal DNA helicase Hel308 in complex with DNA (Büttner *et al.*, 2007) has shown that Ski2-like helicases bind one strand of a nucleic acid duplex in a central tunnel surrounded by both RecA domains, a winged helix (WH) domain and a seven helix bundle (HB) domain. Unwinding was suggested to involve translocation of the enzyme

on the bound strand in 3'-to-5' direction with an extended ß-hairpin (the separator loop - SL), inserted between helicase motifs Vb and VI of the second RecA-like domain, positioned between the two substrate strands. This element contains aromatic residues that stack with the DNA bases in the structure and was suggested to work as a strand separation device in Hel308. A long helix (ratchet helix) of the HB domain, which employs positively charged and aromatic residues to contact the substrate, was thought to function as a ratchet allowing the helicase to processively unwind DNA. The encircled DNA strand was seen to exit across a C-terminal helix-loop-helix domain (HLH) (Büttner *et al.*, 2007), which is required for coupling ATPase and helicase activities (Richards *et al.*, 2008; Woodman *et al.*, 2007).

Structural analyses of the C-terminal Sec63 units of yeast (Zhang *et al.*, 2009; Pena *et al.*, 2009) and human (PDB ID 2Q0Z) Brr2 have shown that these elements contain Hel308-like HB and HLH domains followed by a Brr2-specific immunoglobulin-like (IG) domain and have led to the proposal that Brr2 comprises two expanded Hel308-like helicase cassettes. In these studies, a model of the N- and C-terminal cassettes of Brr2 was devised showing that the Sec63 unit is preceded by a WH domain and that both cassettes could be modeled into a similar circular domain arrangement. Additionally, the model predicted many mechanistic features associated with the N-terminal cassette that are crucial for Brr2 activity (Pena *et al.*, 2009). In the devised model of the N-terminal cassette, an equivalent element of the separator loop in Hel308 also contains aromatic residues that, when mutated, did not support cell growth suggesting that this element may act as a strand separator in Brr2 as well. Furthermore, the modeling exercise predicted that the N-terminal HB domain form the roof of the single stranded nucleic acid binding tunnel. As in Hel308, aromatic and positively charged side chains protrude from the under-side of the ratchet helix and are expected to contact the RNA substrate. Mutation of these residues gave rise to cellular defects and affected *in vivo* splicing efficiency. These results support a role for the HB domain of the N-terminal Sec63 unit in RNA binding and possibly as a ratchet coupling conformational changes to nucleic acid translocation. Finally, in order for the HB domain to act as a ratchet, it has to be functionally connected to the RecA-like domains. In Hel308, the winged helix domain connects the second RecA-like domain with the HB domain through a hydrophobic interface and residues in this interface are conserved in Brr2. Mutations affecting the hydrophobic interface residues in Brr2 led to growth defects as well as pre-mRNA splicing deficiency. Therefore, these results are consistent with the idea that the WH domain acts as a positioning device for the Sec63 unit relative to the catalytic RecA-domains.

Furthermore, in analogy to Hel308, Brr2 is also thought to translocate in 3'-to-5' direction on one of the substrate strands. However, in the U4/U6 di-snRNP both 3'-ends are sequestered in a stem-loop structure and/or are occluded by bound Sm/LSm proteins (Achsel *et al.*, 1999; Leung *et al.*, 2011). Thus, it is presently unclear how Brr2 can engage its U4/U6 substrate.

Regardless of the numerous studies on Brr2, the overall organization of this helicase and the nature of the putative communication between its cassettes are unknown. The unusual architecture of Brr2 is likely instrumental for its unique functions and may form the basis for the required regulation of the enzyme.

1.2.4 Brr2 and Renititis pigmentosa

Renititis pigmentosa (RP) is a progressive rod-cone dystrophy that leads to night blindness, gradual reduction of the peripheral visual field and, in severe cases, to total blindness. The incidence of RP is approximately 1 in 4000 individuals, summing up to a total of 1.5 million affected people. The disease is genetically diverse and can be inherited as an autosomal-dominant (adRP) (30 - 40% of the cases), autosomal-recessive (arRP) (50 - 60%), X-linked (xl-RP) (5 - 15%) and, in rare cases, as a non-Mendelian trait (Hartong et al., 2006). To date, 23 genes have been associated to adRP. Most of these genes are predominantly expressed in the retina and are determinant for the survival and function of the retinal cells (Swarrop and Zack, 2002). Surprisingly, five genes related to adRP code for ubiquitously expressed pre-mRNA splicing factors: hPrp8 (RP13 form of adRP) (McKie *et al.*, 2001), hPrp31 (RP11) (Vithana *et al.*, 2001), hPrp3 (RP18) (Chakarova *et al.*, 2002), hBrr2 (RP33) (Zhao *et al.*, 2006, 2009) and PAP-1 (RP9) (Li *et al.*, 2010; Benaglio *et al.*, 2011).

Considering the widespread expression of these splicing factors, it is intriguing that mutations in these proteins only affect retinal cells. A possible explanation rests on the fact that the retina normally expresses up to seven fold more major snRNAs and approximately twice as many minor snRNAs compared to other human tissues. The high level of snRNA expression in the retina is due to the highest amount of processed mRNAs in these cells within the whole body. Therefore, the deficiency in spliceosomal components is more deleterious for the retina than for any other human tissue, implying a possible molecular link between RP-proteins and the pathogenesis of RP (Tanackovic *et al.*, 2011). Additionally, since adRP patients carry heterozygous mutations, the wild type copy of the affected genes is always expressed and compensates for the effects of the defective proteins in other cell types.

The RP33 form is of great interest for this thesis since all mutations map to the N-terminal helicase cassette of Brr2. As mentioned above, the ATPase and helicase activities of the N-

terminal cassette are required for splicing (Kim and Rossi, 1999). One set of affected residues (R681C, R681H, V683L, Y689C) lies in between the N-terminal RecA domains of Brr2. Another set of mutations maps to the predicted ratchet helix of the N-terminal HB domain (S1087L, R1090L) and corresponding changes in yBrr2 were detrimental to U4/U6 unwinding and splicing (Zhao *et al.*, 2009; Zhang *et al.*, 2009; Pena *et al.*, 2009). Therefore, we can hypothesize that these mutations in the N-terminal cassette could possibly impair RNA binding and/or unwinding.

1.2.5 Structural studies of spliceosomes and their components

Obtaining structural information at the atomic level of a highly dynamic and complex machinery such as the spliceosome is extremely challenging. The amounts, for example, of affinity-purified spliceosomes stalled at a particular stage that can presently be obtained allow investigations by electron cryo-microscopy (cryo-EM) at intermediate to low resolution (Stark and Lührmann, 2006; Alcid and Jurica, 2008; Jurica *et al.*, 2004). Interpretation of such structures is difficult. Immuno-labeling of specific components allows the assignment of selected subunits to structural elements (Häcker *et al.*, 2008)

High resolution structural analysis by X-ray crystallography and/or NMR is currently possible for individual spliceosomal proteins/protein fragments or sub-complexes, typically produced recombinantly or, more rarely, natively purified material. The crystal structures of the an RNA recognition motif of the U1-A protein alone and in complex with stem-loop II of U1 snRNA (Oubridge *et al.*, 1994), the U2-A'/B" heterodimer bound to U2 snRNA stem-loop IV (Price *et al.*, 1998), the 15.5K protein complexed with the 5' SL of U4 snRNA (Vidovic *et al.*, 2000) and two heteromeric Sm protein dimers (Kambach *et al.*, 1999) were the pioneer structures of spliceosomal components. Further progress towards elucidating the structures of proteins and RNA fragments involved in the splice site recognition (rev. in Ritchie *et al.*, 2009) includes the structure of SF1 bound to a branch site RNA sequence (Liu *et al.*, 2001), U2AF65 bound to a polypyrimidine tract (Sickmier *et al.*, 2006), an RRM of U2AF65 bound to an amino-terminal SF1 peptide (Selenko *et al.*, 2003), a heterodimer of U2AF65 and U2AF35 (Kielkopf *et al.*, 2001), and SF3b14a/p14 bound to a SF3b155 peptide (Schellenberg *et al.*, 2006; Spadaccini *et al.*, 2006). More recently, the crystal structure of the U4 snRNA stem-loop complexed with the 15.5K protein and a fragment of the hPrp31 was reported (Liu *et al.*, 2007) as well as the structure of the complex between 15.5K and the GYF domain of 52K (Nielsen *et al.*, 2007), followed by crystal structures of the carboxy terminus of the *S. cerevisiae* (Pena *et al.*, 2007) and *C. elegans* (Zhang *et al.*, 2007) Prp8 revealing a Jab1/MPN

domain with insertions and appendices and with an impaired metal ion binding site. Furthermore, three groups independently solved the structure of a ~250 amino acid RNase H-like domain, located just upstream of the Jab1/MPN domain, of human and/or yeast Prp8, which encompasses amino acids that contact the 5' SS of the pre-mRNA (Pena *et al.*, 2008; Ritchie *et al.*, 2008; Yang *et al.*, 2008). Two independent groups have solved the crystal structure of the C-terminal Sec63 unit of *S. cerevisiae* Brr2 (Pena *et al.*, 2009; Zhang *et al.*, 2009). Further achievements include the structure of Prp43 (Walbott *et al.*, 2010), the structure of the complex of a fragment of the Aar2 protein with the RNase H-like domain of Prp8 (Weber *et al.*, 2011), the structure of the C-terminal domain of human Prp22 (Kudlinzki *et al.*, 2009) and finally the structure of the Cwc2 protein, solved independently by two groups (Lu *et al.*, 2012; Schmitzová *et al.*, 2012). An important milestone on the way to elucidating the spliceosomal structural organization was reached recently with the crystal structures of U1 snRNP recombinantly assembled (Pomeranz Krummel *et al.*, 2009) and natively purified (Weber *et al.*, 2010) and of a recombinant U4 snRNP core (Leung *et al.*, 2011). While they will not provide a complete picture of entire spliceosomal assemblies, such structures of spliceosomal components/complexes are very useful for guiding functional studies by biochemical, biophysical or genetics methods.

The two approaches, global structural analysis by cryo-EM and high resolution partial structures by crystallography or NMR, have also been successfully combined. Spatial arrangements of snRNP components have been elucidated by fitting partial high resolution structures into intermediate resolution cryo-EM maps, thereby devising quasi-atomic models (Stark *et al.*, 2001; Golas *et al.*, 2003, 2005). In a similar fashion, quasi-atomic models of entire spliceosomes may become available in the future.

1.3 Aim of this study

A hallmark of the spliceosome is that it is put together anew on a substrate pre-mRNA for each splicing event in a stepwise fashion. Transitions from one assembly intermediate to the next are invariably associated with profound compositional and conformational rearrangements. The most dramatic rearrangements ensue during the conversion of an initial assembly, which contains all snRNPs but is still catalytically inactive, to an activated spliceosome that can carry out the splicing transesterification reactions. Spliceosome catalytic activation requires a 650 kDa RNP remodeling machinery, composed of the U5 snRNP-specific proteins Prp8, Snu114 and Brr2. The three proteins are large, multi-domain factors (Prp8 ca. 280 kDa, Snu114 ca. 115 kDa, Brr2 ca. 250 kDa). Prp8 serves as a regulatory

scaffold of this machinery and is also considered the master regulator of the entire spliceosome (GRAINGER and BEGGS, 2005). Snu114 is a complex G-protein that bears close resemblance to the eukaryotic translation elongation factor 2 (eEF2) and the bacterial elongation factor G (EF-G). Due to its resemblance of the ribosomal translocases, Snu114 has been suggested to also act as a motor protein during spliceosome assembly (Bartels *et al.*, 2002). More recently, it has been suggested to also function as an on/off switch for the activity of Brr2 (Small *et al.*, 2006). Brr2 is a Ski2-like RNA helicase that functions as an RNP remodeling enzyme in the spliceosome under tight control of Prp8 and Snu114 (Small *et al.*, 2006; Maeder *et al.*, 2009; Wahl *et al.*, 2009). During catalytic activation, it is thought to unwind the U4/U6 di-snRNA duplex and dispose of all U4/U6 associated proteins (Raghunathan and Guthrie, 1998; van Nues and Beggs, 2001). Additionally, the same machinery is thought to be also involved in the ordered disassembly of the spliceosome (Small *et al.*, 2006; Staley and Guthrie, 1998).

At the start of this thesis, very little was known about the molecular mechanisms by which Prp8, Snu114 and Brr2 accomplish spliceosome remodeling and how these events are regulated. In particular, very little experimental structural information was available on this machinery. Thus, the overall goal of my thesis was to advance our understanding of the Prp8-Snu114-Brr2 RNP remodeling machinery by elucidating molecular mechanisms *via* a structural biochemical approach. To this end, I pursued the following specific aims:

1. Establishment of an efficient recombinant production pipeline for Prp8, Snu114, Brr2 or complexes of these proteins, using proteins from different organisms to increase chances of success.
2. Establishment of efficient purification protocols for recombinantly produced proteins or complexes that yield sufficient amounts of highly homogeneous material for further structural and functional studies.
3. Experimental definition of stable folding units within factors that can be produced and purified in suitable form but do not crystallize.
4. Crystallization and crystal structure analysis of proteins, folding units or complexes that can be produced in suitable amount and homogeneity.
5. Structure-based functional analyses, such as targeted mutagenesis followed by enzymatic or binding assays *in vitro*.

I started my work with attempts at the co-production of the ternary Prp8-Snu114-Brr2 complexes from human and yeast and then zoomed in on the analysis of a ca. 200 kDa fragment of human Brr2, encompassing both of its helicase cassettes. For the latter fragment, high resolution crystal structures were elucidated and a comprehensive *in vitro* functional analysis was conducted.

2. Materials and methods

2.1 Materials

2.1.1 Chemicals

Standard chemicals that are not listed here were purchased from Sigma-Aldrich, Merck or Fluka.

Table 2.1: Chemicals

Chemical	Supplier
1 kb DNA ladder, 2-log DNA ladder	New England Biolabs, Germany
1,4-Dithiothreitol (DTT)	Roth, Germany
2-Mercaptoethanol (β-ME)	Roth, Germany
2-Propanol	Merck, Germany
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Roth, Germany
Acetic acid	Merck, Germany
Acetone	Merck, Germany
Acrylamide solutions:	Roth, Germany
Rotiphorese Gel 40 (38% acrylamide, 2% B *N*,*N'*-methylene-*bis*-acrylamide)	
Rotiphorese Gel 30 (29.2% acrylamide, 0.8% *N*,*N'*-methylene-*bis*-acrylamide)	
Rotiphorese Gel A (30% acrylamide)	
Rotiphorese Gel B (2% *N*,*N'*-methylene-*bis*-acrylamide)	
Agarose (electrophoresis grade)	Invitrogen, Germany
Ammonium peroxodisulfate (APS)	Merck, Germany
Ammonium sulfate	Fluka, Switzerland
Ampicillin, sodium salt	Fluka, Switzerland
Bluo-Gal	Invitrogen, Germany
Boric acid	Merck, Germany
Bovine serum albumin	Sigma-Aldrich, Germany
Bradford assay reagent	Bio-Rad, Germany
Bromophenol blue, sodium salt	Merck, Germany
Calcium chloride dihydrate	Merck, Germany
Chloramphenicol	Roche, Germany
Complete-EDTA free protease inhibitor	Roche, Germany

Chemical	Supplier
Coomassie brilliant blue G-250	Serva, Germany
Coomassie brilliant blue R-250	Serva, Germany
Dimethylsulfoxide (DMSO)	Roth, Germany
Ethanol	Merck, Germany
Ethidium bromide solution (10mg/ml)	Roth, Germany
Ethylene glycol	Serva, Germany
Ethylenediaminetetraacetic acid, disodium salt dihydrate (EDTA)	Roth, Germany
Express Five® SFM	Invitrogen, Germany
Formaldehyde	Merck, Germany
Gel filtration standard	Bio-Rad, Germany
Glycerol	Merck, Germany
Guanidine hydrochloride	Roth, Germany
Heparin, sodium salt	Sigma-Aldrich, Germany
Hydrochloric acid (HCl)	Merck, Germany
Igepal^{R} CA-630 (NP-40)	Sigma-Aldrich, Germany
Imidazole	Merck, Germany
Isopropyl-β-D-1-thiogalactopyranoside (IPTG)	Roth, Germany
Izit protein crystal stain	Hampton research, USA
Kanamycine sulfate	Roth, Germany
LB medium	Q-Bio-gene, USA
LB-agar medium	Q-Bio-gene, USA
L-glutamine	Invitrogen, Germany
Lithium chloride	Fluka, Switzerland
Magnesium chloride hexahydrate	Fluka, Switzerland
Methanol	Merck, Germany
Milk powder, dry, instant	Heirler, Germany
N,N,N',N'- Tetramethylethylenediamine (TEMED)	Sigma-Aldrich, Germany
N-Z-Amine AS Casein enzymatic hydrolysate	Sigma-Aldrich, Germany
Phenylmethylsulfonyl fluoride (PMSF)	Roche, Germany
Piperazine-N,N′-bis-(2-ethanesulfonic acid) (PIPES)	Roth, Germany
Polyethylene glycol 200-35000	Fluka, Germany
Polyethylene glycol 3350	Sigma-Aldrich, Germany
Ponceau S	Serva, Germany
Potassium chloride	Merck, Germany

Chemical	Supplier
Potassium hydroxide	Merck, Germany
Protein molecular weight marker (unstained, pre-stained)	Bio-rad, Germany
Sf-900™ III SFM	Invitrogen, Germany
Silver nitrate	Merck, Germany
Sodium azide	Fluka, Switzerland
Sodium carbonate	Merck, Germany
Sodium chloride	Merck, Germany
Sodium dodecylsulfate (SDS)	Serva, Germany
Sodium hydroxide	Merck, Germany
SYPRO Orange Protein Stain	Invitrogen, USA
Tetracyclin	Fluka, Switzerland
Trifluoroacetic acid	Fluka, Switzerland
Tris-(hydroxymethyl)-aminomethane (TRIS)	VWR International, Germany
Triton X-100	Merck, Germany
Tween-20	Roth, Germany
Urea	Merck, Germany
X-tremeGene 9 Transfection Reagent	Roche, Germany
Xylene cyanol FF	Sigma-Aldrich, Germany
Yeast extract powder	Roth, Germany

2.1.2 Buffer solutions and media components

Buffers were prepared with Milli-Q water and filter-sterilized (0.22 μm) or autoclaved. The pH of buffers was adjusted by the addition of buffered stock solutions or by titration with 37% HCl or 10 M NaOH if not stated otherwise.

Table 2.2: Buffers, solutions and media components

Buffers	Composition
5×T7 buffer	1 M HEPES-KOH, pH 7.5, 160 mM $MgCl_2$, 10 mM spermidine, 200 mM DTT
10×TBE	0.89 M Tris pH 8.0, 0.89 M Boric Acid, 25 mM EDTA
10×TBS buffer	200 mM Tris pH 7.6, 1.37 M NaCl
5×SDS loading buffer	250 mM Tris pH 6.8, 8% (w/v) SDS, 10% (v/v) β-ME, 30% (v/v) Glycerol, 0.02% (w/v) Bromophenol blue
4×agarose gel loading buffer	1×TBE, 30% (v/v) Glycerol, 0.05% (w/v) Xylene cyanol, 0.05% (w/v) Bromophenol blue
10×PBS	1.35 M NaCl, 25 mM KCl, 100 mM NaH_2PO_4 17 mM KH_2PO_4 pH 7.4

Buffers	Composition
10×SDS running buffer	250 mM Tris pH 6.8, 2 mM Glycine, 1% (w/v) SDS
Coomassie staining solution	0.025% (w/v) Coomassie (R250); 0.025% (w/v) Coomassie (G250); 30% (v/v) isopropanol; 7.5% (v/v) acetic acid
Destaining solution I	40% (v/v) ethanol, 20% (v/v) acetic acid
Destaining solution II	10% (v/v) acetic acid
Stacking gel buffer	0.5 M Tris-HCl pH 6.8, 0.4% (w/v) SDS
Resolving gel buffer	1.5 M Tris pH 8.8, 0.4% (w/v) SDS
SLAB4	50 mM Tris, 380 mM glycine, 0.1% (w/v) SDS
Blotting buffer	1.5 l SLAB4, 0.6 l methanol, 0.9 l deionized water
Blocking buffer	1×PBS, 0.1% (v/v) Tween-20, 5% (w/v) milk powder
Washing buffer I	1×PBS, 0.1% (v/v) Tween-20, 1% (w/v) milk powder
Washing buffer II	1×PBS, 0.1% (v/v) Tween-20
Primary antibody solution	1×PBS, 0.1% (v/v) Tween-20, 1% (w/v) milk powder, primary antibody
Secondary antibody solution	1×PBS, 0.1% (v/v) Tween-20, 1% (w/v) milk powder, secondary antibody
TB buffer	10 mM PIPES pH 6.7, 250 mM KCl, 55 mM $MnCl_2$, 15 mM $CaCl_2$
SOB medium	2% (w/v) Tryptone, 0.55% (w/v) yeast extract, 10 mM NaCl, 10 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$
SOC medium	2% (w/v) Tryptone, 0.55% (w/v) yeast extract, 10 mM NaCl, 10 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose
ZY	10 g N-Z-amine AS, 5 g yeast extract, 950 ml of deionized water
ZYM-5052	950 ml ZY media supplemented with 2 ml $MgSO_4$ (1 M), 200 µl trace metals (1000×solution), 20 ml of 50×5052, 20 ml of 50×M
50×M	1.25 M Na_2HPO_4, 1.25 M KH_2PO_4, 2.5 M NH_4Cl, 0.25 M Na_2SO_4
50×5052	25% glycerol, 2.5% glucose, 10% α-lactose
1000×Trace Metals	50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, 2 mM Na_2SeO_3, 2 mM H_3BO_3 in ~60 mM HCl

2.1.3 Consumables

Table 2.3 lists the items used routinely in this thesis.

Table 2.3: Consumables

Item	Supplier
Acupuncture needle	Moxom Medical, Germany
Amersham Hyperfilm ECL	GE Healthcare, Germany
Amersham Hybond-P PVDF membrane	GE Healthcare, Germany
Autoradiography films BioMax MR	Kodak, USA

Item	Supplier
Concentrators (Amicon Ultra)	Millipore, USA
CrysChem plates, 24 well, hanging drop	Hampton Research, USA
CrysChem plates, 24 well, sitting drop	Hampton Research, USA
Crystallization plates MRC, 96 well, sitting drop	Molecular Dimensions, UK
Dialysis membranes	Spectra/Por, USA
Electroporation cuvettes	Bio-Rad, Germany
Eppendorf safe-lock micro test tubes	Eppendorf, Germany
Fluotrac-600, 96-well plates	Greiner-Bio-One, Germany
Gloves, Dermatril	KCL, Germany
Linbro plates, 24 well, hanging drop	Jena Biosciences, Germany
Linbro plates, 24 well, sitting drop	Jena Biosciences, Germany
Needles	Henke Sass Wolf, Germany
Parafilm	Pechiney Plastic Packaging, USA
Slide-A-lyzer	Pierce, USA
Sterile filters 0.22 µm, 0.45 µm	Sarstedt, Germany
Surgical blades	Martin, Germany
Syringes	Braun, Germany
Tubes (5 ml, 10 ml, 15 ml and 50 ml)	Greiner-Bio-One, Germany
Whatman 3MM paper	Whatman plc, UK
Weighting dishes	Roth, Germany

2.1.4 Chromatographic resins and columns

Table 2.4: Chromatographic resins and columns

Matrix	Supplier
Ni-NTA agarose	Qiagen, Germany
HisTrap FF Crude, MonoQ 5/50 GL, MonoQ HR 10/10, Superdex 200 26/60, Superdex 200 16/60, Superdex 200 10/300, Superdex 200 PC 3.2/30, HiPrep Heparin FF	GE Healthcare, Germany
PD-10 columns, ProbeQuant G-25/G50 columns	GE Healthcare, Germany

2.1.5 Molecular biology kits

Table 2.5: Commercial molecular biological kits

Kit	Supplier
BCA protein assay kit	Pierce, USA
QIAquick PCR Purification Kit, QIAprep spin miniprep Kit, QIAquick Gel Extraction Kit	Qiagen, Germany
Amersham ECL Western Blot Detection Reagent	GE Healthcare, Germany
QuikChange II XL Site-Directed Mutagenesis Kit	Agilent Technologies, Germany
PureLink™ HiPure Plasmid FP (Filter and Precipitator) Maxiprep Kit	Invitrogen, Germany

2.1.6 Nucleotides

Table 2.6: Nucleotides

Nucleotides	Supplier
Deoxynucleoside-5′-triphosphate (dNTPs, 100mM each) (dATP, dCTP, dGTP, dTTP)	New England Biolabs, Germany
2′/3′-O-(N-methylantranyloil)-nucleotides (*mant*-ATP, *mant*-ATPγS, *mant*-ADP, *mant*-GTP, *mant*-GTPγS, *mant*-GDP)	Jena Biosciences, Germany
TNP-nucleotides (TNP-ATP, TNP-CTP, TNP-UTP)	Invitrogen, Germany
ATP, ADP, AMP-PNP, AMP-PCP	Sigma-Aldrich, Germany
Nucleoside-5′-Triphosphate (rNTPs, 100mM) (ATP, CTP, GTP, UTP)	Jena Biosciences, Germany
[γ-^{32}P] dATP (6000 Ci/mmol; 10 mCi/ml)	Perkin Elmer, Germany

2.1.7 Crystallization screens and kits

The crystallization screen solutions were transferred to 96-well MRC crystallization plates for automated pipetting using a Cartesian robot. All heavy atom compounds used in this work were part of the "Heavy Atom Screen Kit" from Hampton Research or the Tantalum Cluster Derivatization Kit from Jena Bioscience. The heavy metals were dissolved in reservoir solution and utilized for the preparation of heavy-atom derivatives for structure determination of hBrr2 by X-ray analysis.

Table 2.7: Crystallization screens

Screen	Supplier
Additives, Index, SaltRX, Silver Bullets, Silver Bullets Bio, PCT (pre-crystallization screen), Heavy atom screen Au, Pt, Hg, M1, M2	Hampton Research, USA

Screen	Supplier
Classics Suite, Classics Lite Suite, Classics II Suite, Cryos Suite, PEGs Suite, AmSO4 Suite, MPD Suite, Anions Suite, Cations Suite, pHClear Suite, pHClear II Suite, MBClass Suite, MBClass II Suite, Protein Complex Suite, PEGs II Suite, ComPAS Suite, PACT Suite, Nucleix Suite, JCSG+ Suite	Qiagen, Germany
JBS Rainbow, Tantalum Cluster Derivatization Kit	Jena Bioscience, Germany
Wizard I+II, Wizard III+ IV	Emerald BioSystems, USA

2.1.8 Instrumentation

Table 2.8: Devices

Device	Manufacturer
Äkta Explorer, Purifier, Prime, Micro	GE Healthcare, Germany
Allegra X-15R	Beckman Coulter, Germany
Autoclaves	H+P Labortechnik, Germany
Avanti J-26 XP	Beckman Coulter, Germany
Axiovert100	Zeiss, Germany
Beamline 14.1	HZB, Berlin, Germany
Beamline 14.2	HZB, Berlin, Germany
Beamline 14.3	HZB, Berlin, Germany
Beamline PXII X10SA	SLS, PSI, Villigen, Switzerland
Beamline P14, Petra III	DESY, Hamburg, Germany
Biofuge fresco, Biofuge pico	Heraeus, Germany
Cartesian crystallization robot, 4 channels	Zinsser Analytic, Germany
Cartesian crystallization robot, 8 channels	Digilab, USA
CASY TT Counter	Innovatis, Germany
Cryofuge 6000i	Heraeus, Germany
Electroporator EasyjecT Prima	Equibio, England
Elexsys 500 EPR spectrometer	Bruker ASX B.V., Netherlands
Filter KV 408	Schott, Germany
Gel documentation system	Bio-Rad, Germany
Geldryer model 583	Bio-Rad, Germany
Gel electrophoresis equipment	Bio-Rad, Germany
Glass-ware	VWR International, Germany
Heating blocks	Eppendorf, Germany
HT multitron culture shaker	Infors, Switzerland

Device	Manufacturer
Ice machine	Ziegra, UK
Jasco J-810 CD Spectropolarimeter	Jasco, Inc., USA
4800 MALDI ToF/ToF	Applied Biosciences/MDS Sciex, Germany
QToF Ultima	Waters, UK
LTQ-Orbitrap XL	Thermo Scientific, Germany
Kodak X-Omat 2000 Processor	Kodak, USA
Milli-Q synthesis A10	Millipore, USA
Magnetic stirrer	IKA, Germany
Megafuge 1.0R, swing out rotor type 2704	Heraeus, Germany
Micro fluidizer M110S	Micro fluidics, USA
Microliter syringes	Hamilton, Switzerland
Microwave oven	Bosch
12-way multichannel pipette	Eppendorf, Germany
Mx3000P QPCR Systems	Agilent Technologies, Germany
NanoDrop 2000 Spectrometer	Thermo Fisher Scientific, USA
PCR cycler T-Professional	Biometra, Germany
PCR cycler Peqstar 2x gradient	PeqLab, Germany
pH-meter, Orion 2-Star	Thermo Fisher Scientific, USA
pH-meter, Professional Meter PP-20	Sartorius, Germany
Phosphorimager Typhoon 8600	GE Healthcare, Germany
Photometer DU 530	Beckmann, Germany
Pipettes, Pipetman	Eppendorf, Germany
Rocking platform	Biometra, Germany
Scintillation counter LS 1701/TRI-CARB 2100TR	Beckman/Packard, USA
Power supplies	Bio-Rad, Germany
Quartz Cuvette 1 mm, 110-QS	Hellma, Germany
Scales BP4100	Sartorius, Germany
Scales XS4002S DeltaRange^{R}	Mettler Toledo, Germany
Scales XS205 DualRange^{R}	Mettler Toledo, Germany
SMART	GE Healthcare, Germany
Sonifier II 250 Digi	Heinemann, Germany
Sonopuls Ultrasonic Homogenizer HD 3100	Bandelin, Germany
Sorvall Evolution RC, Sorvall SA-300 rotor, Sorvall SLC-6000 rotor, Sorvall SS-34 rotor, Sorvall HB-6 rotor	Thermo Fisher Scientific, USA

Device	Manufacturer
Spectrophotometer Ultrospec 3000 pro	GE Healthcare, Germany
Speed vac concentrator 5301	Eppendorf, Germany
VICTOR™ X3 multilabel plate reader	Perkin-Elmer, Germany
Vortex Genie 2	Scientific Industries, USA
Table centrifuge 5415R	Eppendorf, Germany
Tecan Evo liquid dispensing robot	Tecan Group, Germany
Thermomixer comfort	Eppendorf, Germany
Trans-Blot Cell	Bio-Rad, Germany
Tunair flasks	Sigma-Aldrich, Germany
SX-20MV stopped-flow apparatus	Applied Photophysics, Leatherhead, UK

2.1.9 Enzymes and proteins

Table 2.9: List of enzymes, proteins and enzyme inhibitors used in this study

Enzyme, proteins and peptides	Supplier
Bovine serum albumin (BSA)	Sigma-Aldrich, Germany
Bovine serum albumin (BSA) solution (10mg/ml)	New England Biolabs, Germany
Carboxypeptidase Y	Sigma-Aldrich, Germany
Chymotrypsin	Roche, Germany
Cloned *Pfu* DNA polymerase (10 U/µl)	Agilent Technologies, Germany
DNase I	Roche, Germany
Elastase	Sigma-Aldrich, Germany
Endoproteinase Lys-C	Sigma-Aldrich, Germany
Endoproteinase Asp-N	Sigma-Aldrich, Germany
Endoproteinase Glu-C	Sigma-Aldrich, Germany
Endoproteinase Arg-C	Sigma-Aldrich, Germany
Herculase enhanced DNA polymerase	Agilent Technologies, Germany
Lysozyme	Merck, Germany
Peroxidase-AffiniPure Goat Anti-Rabbit IgG	Jackson ImmunoResearch Laboratories, USA
Rabbit IgG against hBrr2	I. Oechsner, MPI-BPC, Göttingen
Proteinase K	Sigma-Aldrich, Germany
Restriction endonucleases	New England Biolabs, Germany
RNase A, T1	Ambion, Germany
RNasin	Promega, Germany

Enzyme, proteins and peptides	Supplier
Subtilisin	Sigma-Aldrich, Germany
TEV Protease	Home-made, recombinant
T4 DNA ligase	New England Biolabs, Germany
T4 polynucleotide kinase	New England Biolabs, Germany
T7 RNA polymerase	Self-made, recombinant
Thermolysin	Sigma-Aldrich, Germany
Trypsin	Roche, Germany

2.1.10 DNA oligonucleotides

All DNA oligonucleotides used in this work were purchased from Eurofins MWG-Operon, Germany. The oligonucleotides were dissolved in autoclaved Milli-Q water to a stock concentration of 100 pmol/µl and stored at -20°C.

Table 2.10: DNA oligonucleotides

Name	Used for	Sequence (5' → 3')
Prp8ct_1_for	Cloning in pETM11 *via Bsa*I	ATATATGGTCTCACATGAAGGCCAACCCCG CCCTGTACG
Prp8ct_2_for	Cloning in pETM11 *via Bsa*I	ATATATGGTCTCACATGAACTCCTCTAACTA CGCCGAGC
Prp8ct_rev	Cloning in pETM11 *via Not*I	ATACATTTATGCGGCCGCTACTTGTCGTCGT CGTCCTTGTAGTCG
Prp8ct_GEX1_for	Cloning in pGEX6P1 *via BamH*I	ATATATGGATCCATGAAGGCCAACCCCGCC CTGTACG
Prp8ct_GEX2_for	Cloning in pGEX6P1 *via BamH*I	ATATATGGATCCAACTCCTCTAACTACGCC GAGC
HsBrr2FL_for	Cloning in pFL *via BamH*I	CCGGAATTCATGGCTGACGTGACCGCTCG
HsBrr2FL_rev	Cloning in pFL *via Hind*III	TTAATTAAAAGCTTTTAGTCAGAGTCGGAG TCAGTCTCAGC
Brr2N_Ntrun1_For	Cloning in pFL10 His *via Eco*RI	CCGGAATTCGACCTGGACCAGGGTGGCGAG GC
Brr2N_Rev	Cloning in pFL10 His *via Hind*III	TTAATTAAAAGCTTTTAGTCAGAGTCGGAG TCAGTCTCAGC
Brr2N_Ctrun_Rev	Cloning in pFL10 His *via Hind*III	TTAATTAAAAGCTTTTAAGCTTCTTTCACGT CAACGG
Brr2N_Ntrun2_For	Cloning in pFL10 His *via Eco*RI	CCGGAATTCTCCGAGGAACAGCTGCTGCCC
Brr2C_Ntrun1_For	Cloning in pFL *via Eco*RI	CCGGAATTCATGGACCTGGACCAGGGTGGC GAGGC

Name	Used for	Sequence (5' → 3')
Brr2C_His_Rev	Cloning in pFL *via Hind*III	TTAATTAAAAGCTTTTAATGGTGATGGTGGTGGTGGTCAGAGTCGG
Brr2C_Ntrun2_For	Cloning in pFL *via Eco*RI	CCGGAATTCATGTCCGAGGAACAGCTGCTGCCC
10HisNterminus BamHI /EcoRI	Modifying pFL in pFL 10 His *via BamH*I and *Eco*RI	CGCGGATCCATGAAACATCACCATCACCATCACCATCACCATCACCCCATGAGCGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAGGGCGCCGAATTC
HSN_cas_For	Cloning in pFL10 His *via Eco*RI	CCGGAATTCGACCTGGACCAGGGTGGCGAGG
HSN_cas_Rev	Cloning in pFL10 His *via Hind*III	TTAATTAAAAGCTTTTACTTGTCCTGGTACAGGGACTCG
HSC_cas_For1	Cloning in pFL10 His *via Eco*RI	CCGGAATTCCTGCCCGTGTCCTTCCGTCACC
HSC_cas_Rev	Cloning in pFL10 His *via Hind*III	TTAATTAAAAGCTTTTAGTCAGAGTCGGAGTCAGTCTCAGC
ScBrr2_for	Cloning in pFL10 His *via Sac*I	ATATATGAGCTCTAATGACCGAACACGAAACCAAGGACAAGG
ScBrr2_rev	Cloning in pFL10 His *via Hind*III	TTAATTAAAAGCTTTTACTTGACGTTGATCTCGAAAGACAGTTCC
C_cas_For2	Cloning in pFL10 His *via Sal*I	ATATATGTCGACAGTCTGAAGTGTTCGAATTCAAGACC
C_cas_Rev	Cloning in pFL10 His *via Xba*I	TTAATTAATCTAGATTACTTGACGGTGATCTCGAAAGACAGTTCC
Mut1_1F	Mutagenesis of hBrr2^{HR,RP} R1133/K1134E	GCCCCCTGCGCCAGTTCGAGGAACTGCCCGAGGAAGTGG
Mut1_1R	Mutagenesis of hBrr2^{HR,RP} R1133/K1134E	CCACTTCCTCGGGCAGTTCCTCGAACTGGCGCAGGGGGC
Mut1_2F	Mutagenesis of hBrr2^{HR,RP} R1176E	GGCAAGACCATCCACGAGTACGTGCACCTGTTCC
Mut1_2R	Mutagenesis of hBrr2^{HR,RP} R1176E	GGAACAGGTGCACGTACTCGTGGATGGTCTTGCC
Mut2F	Mutagenesis of hBrr2^{HR,RP} deletion of the first HLH domain	GCCCCCTGCGCCAGTTCTCCTCCGCTTACGTGCACCTGTTCCCC
Mut2R	Mutagenesis of hBrr2^{HR,RP} deletion of the first HLH domain	GGGGAACAGGTGCACGTAAGCGGAGGAGAACTGGCGCAGGGGGC
Mut4F	Mutagenesis of hBrr2^{HR,RP} Q1332R	CCTTCTTCAACCCCATCCGTACCCAGGTGTTCAACACC
Mut4R	Mutagenesis of hBrr2^{HR,RP} Q1332R	GGTGTTGAACACCTGGGTACGGATGGGGTTGAAGAAGG

Name	Used for	Sequence (5' → 3')
Mut4R	Mutagenesis of hBrr2^{HR,RP} Q1332R	GGTGTTGAACACCTGGGTACGGATGGGGTTGAAGAAGG
Mut5_3F	Mutagenesis of hBrr2^{HR,RP} G1355Q/K1356E	GGTGCTCCTACCGGTTCCCAGGAGACCATCTGCGCTGAG
Mut5_3R	Mutagenesis of hBrr2^{HR,RP} G1355Q/K1356E	CTCAGCGCAGATGGTCTCCTGGGAACCGGTAGGAGCACC
Mut8_1F	Mutagenesis of hBrr2^{HR,RP} I1290/L1291/P1292A	CCCGTGTCCTTCCGTCACCTGGCTGCTGCCGAGAAGTACCCCCCTCCC
Mut8_1F	Mutagenesis of hBrr2^{HR,RP} I1290/L1291/P1292A	CCCGTGTCCTTCCGTCACCTGGCTGCTGCCGAGAAGTACCCCCCTCCC
Mut8_1R	Mutagenesis of hBrr2^{HR,RP} I1290/L1291/P1292A	GGGAGGGGGGTACTTCTCGGCAGCAGCCAGGTGACGGAAGGACACGGG
Mut8_2F	Mutagenesis of hBrr2^{HR,RP} L1307/P1308/V1309A	GAGCTGCTGGACCTGCAACCCGCTGCCGCTTCCGCTCTGCGTAACTCCGC
Mut8_2R	Mutagenesis of hBrr2^{HR,RP} L1307/P1308/V1309A	GCGGAGTTACGCAGAGCGGAAGCGGCAGCGGGTTGCAGGTCCAGCAGCTC
Mut8_3F	Mutagenesis of hBrr2^{HR,RP} P1296/P1297/P1298A	CTGCCCGAGAAGTACGCTGCTGCCACCGAGCTGCTGG
Mut8_3R	Mutagenesis of hBrr2^{HR,RP} P1296/P1297/P1298A	CCAGCAGCTCGGTGGCAGCAGCGTACTTCTCGGGCAG
Mut8_4F	Mutagenesis of hBrr2^{HR,RP} Y1295A	ATCCTGCCCGAGAAGGCTCCCCCTCCCACCGAG
Mut8_4R	Mutagenesis of hBrr2^{HR,RP} Y1295A	CTCGGTGGGAGGGGAGCCTTCTCGGGCAGGAT
Mut6F	Mutagenesis of hBrr2^{HR,RP} close the separator loop of the C-cassette	CATCATGGACACTCAGTCCGACTACCCCATCTACG
Mut6R	Mutagenesis of hBrr2^{HR,RP} close the separator loop of the C-cassette	CGTAGATGGGGTAGTCGGACTGAGTGTCCATGATG
Mut7a_F	Mutagenesis of hBrr2^{HR,RP} R637A	GGTGGCTCGTGCTATCGCTAACATCGAGATGACC
Mut7a_R	Mutagenesis of hBrr2^{HR,RP} R637A	GGTCATCTCGATGTTAGCGATAGCACGAGCCACC
Mut7b_F	Mutagenesis of hBrr2^{HR,RP} T1578A	GCTATCGACATCCTGGCTACCTGCGCTGCTGACATCC
Mut7b_R	Mutagenesis of hBrr2^{HR,RP} T1578A	GGATGTCAGCAGCGCAGGTAGCCAGGATGTCGATAGC
Mut7c_F	Mutagenesis of hBrr2^{HR,RP} R603A	CGTAAGGGTGGAGAGGCTACCTACACCCAGCTGG
Mut7c_R	Mutagenesis of hBrr2^{HR,RP} R603A	CCAGCTGGGTGTAGGTAGCCTCTCCACCCTTACG

Name	Used for	Sequence (5' → 3')
Mut7d_F	Mutagenesis of hBrr2^{HR,RP} K1544A	CGTCTGCTGTCTATGGCCGCTCCCGTGTACC ACGCTATCACC
Mut7d_R	Mutagenesis of hBrr2^{HR,RP} K1544A	GGTGATAGCGTGGTACACGGGAGCGGCCAT AGACAGCAGACG
Mut7e_F	Mutagenesis of hBrr2^{HR,RP} H1548A	TGGCCAAGCCCGTGTACGCTGCTATCACCA AGCACTCC
Mut7e_R	Mutagenesis of hBrr2^{HR,RP} H1548A	GGAGTGCTTGGTGATAGCAGCGTACACGGG CTTGGCCA
Mut9_1F	Mutagenesis of hBrr2^{HR,RP} R1090L	ACCCAGCTGGCTGGTCTGCTGATGCGTGCT ATCTTCG
Mut9_1R	Mutagenesis of hBrr2^{HR,RP} R1090L	CGAAGATAGCACGCATCAGCAGACCAGCCA GCTGGGT
Mut9_2F	Mutagenesis of hBrr2^{HR,RP}, N-cassette^{RP} and Brr2^{RP} L1087S	GGTGTACGTGACCCAGTCCGCTGGTCGCCT GATGC
Mut9_2R	Mutagenesis of hBrr2^{HR,RP}, N-cassette^{RP} and Brr2^{RP} L1087S	GCATCAGGCGACCAGCGGACTGGGTCACGT ACACC

2.1.11 Plasmids

Table 2.11: List of plasmids

Name	Description	Reference
pETM11	Vector for expression of genes with a TEV protease-cleavable N-terminal His_6 tag in *E. coli*; Kan^R	EMBL, Heidelberg
pGEX6P1	Vector for expression of genes with a PreScission protease-cleavable N-terminal GST tag in *E. coli*; Amp^R	GE Healthcare
pFL	MultiBac acceptor vector for bacmid preparation; Amp^R and Gen^R	Dr. I. Berger, EMBL, Grenoble
pFL-10His	MultiBac acceptor vector for bacmid preparation; MCSII was modified to code for an N-terminal His_{10} and a TEV protease cleavage site; Amp^R and Gen^R	This study
pFL hBrr2^{RP}	C-His_6; full length hBrr2^{RP} cloned *via BamH*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP}	N-His_{10}; TEV protease cleavage site; hBrr2$^{395-2129}$ Truncation 1 cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{395-2136,RP}$	N-His_{10}; TEV protease cleavage site; hBrr2 Truncation 2 cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{458-2129,RP}$	N-His_{10}; TEV protease cleavage site; hBrr2 Truncation 3 cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{458-2136,RP}$	N-His_{10}; TEV protease cleavage site; hBrr2 Truncation 4 cloned *via EcoR*I and *Hind*III	This study

Name	Description	Reference
pFL hBrr2$^{395\text{-}2136,RP}$	C-His$_6$; hBrr2 Truncation 5 cloned *via EcoR*I and *Hind*III	This study
pFL hBrr2$^{458\text{-}2136,RP}$	C-His$_6$; hBrr2 Truncation 6 cloned *via EcoR*I and *Hind*III	This study
pFL-10His HsN^{RP}	N-His$_{10}$; TEV protease cleavage site; hBrr2$^{395\text{-}1324}$ (N-cassette) cloned *via EcoR*I and *Hind*III	This study
pFL-10His HsC	N-His$_{10}$; TEV protease cleavage site; hBrr2$^{1282\text{-}2136}$ (C-cassette) cloned *via EcoR*I and *Hind*III	This study
pFL yBrr2	N-His$_6$; full length yBrr2 cloned *via Sac*I and *Hind*III	This study
pFL-10His T2	N-His$_{10}$; TEV protease cleavage site; yBrr2$^{271\text{-}2163}$	Christian Becke, FU-Berlin
pFL-10His ScC	N-His$_{10}$; TEV protease cleavage site; hBrr2$^{1339\text{-}2163}$ (C-cassette) cloned *via Sal*I and *Xba*I	This study
pFL-10His hBrr2$^{HR,RP,R1133/K1134E}$	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,R1176E}	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{HR,RP,\Delta HLH}$	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,Q1332R}	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{HR,RP,G1355Q/K1356E}$	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{HR,RP,I1290/L1291/P1292A}$	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{HR,RP,L1307/P1308/V1309A}$	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{HR,RP,P1296/P1297/P1298A}$	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,Y1295A}	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2$^{HR,RP,\Delta loop}$	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,R637A}	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,T1578A}	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,R603A}	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,K1544A}	N-His$_{10}$; TEV protease cleavage site; cloned *via EcoR*I and *Hind*III	This study

Name	Description	Reference
pFL-10His hBrr2^{HR,RP,H1548A}	N-His$_{10}$; TEV protease cleavage site; cloned *via Eco*RI and *Hind*III	This study
pFL-10His hBrr2^{HR,RP,R1090L}	N-His$_{10}$; TEV protease cleavage site; cloned *via Eco*RI and *Hind*III	This study
pFL-10His hBrr2^{HR,WT}	N-His$_{10}$; TEV protease cleavage site; hBrr2 wild type cloned *via Eco*RI and *Hind*III	This study
pFL hBrr2^{WT}	C-His$_6$; full length hBrr2 wild type cloned *via BamH*I and *Hind*III	This study
pFL-10His HsN^{WT}	N-His$_{10}$; TEV protease cleavage site; hBrr2$^{395-1324}$ (N-cassette) wild type cloned *via Eco*RI	This study
pETM11 Prp8ct$^{1806-2413}$	N-His$_6$; TEV protease cleavage site; *via Bsa*I and *Not*I	This study
pETM11 Prp8ct$^{1836-2413}$	N-His$_6$; TEV protease cleavage site; *via Bsa*I and *Not*I	This study
pGEX6P1 Prp8ct$^{1806-2413}$	N-GST; TEV protease cleavage site; *via BamH*I and *Not*I	This study
pGEX6P1 Prp8ct$^{1836-2413}$	N-GST; TEV protease cleavage site; *via BamH*I and *Not*I	This study
pETM11 Prp8ct$^{1836-2398}$	N-His$_6$; TEV protease cleavage site; *via Bsa*I and *Not*I	Dr. Gert Weber, FU-Berlin

2.1.12 Insect cell lines

Sf9 and Sf21 cells were used to isolate and propagate recombinant baculovirus stocks. High Five^{TM} cells were used for large scale expression of recombinant proteins. All cell lines above are adapted to serum-free suspension culture.

Table 2.12: Insect cell lines

Cell line	Description	Supplier
Sf9 cells	clonal isolate derived from the parental *Spodoptera frugiperda* (Fall Armyworm) cell line IPLB-Sf21-AE. Originated at the USDA Insect Pathology Laboratory (Vaughn *et al.*, 1977)	Invitrogen, Germany
Sf21 cells	isolated from *Spodoptera frugiperda* (Fall Armyworm) pupal ovarian tissue (Vaughn *et al.*, 1977)	Invitrogen, Germany
High Five^{TM} cells	BTI-TN-5B1-4 was developed by the Boyce Thompson Institute for Plant Research, Ithaca, NY and originated from a clonal isolate derived from the ovarian cells of the cabbage looper, *Trichoplusia ni* (Wickham *et al.*, 1992)	Invitrogen, Germany

2.1.13 Bacterial strains

Escherichia coli strains XL1 blue and XL2 blue were used for routine cloning applications of plasmid DNA. BL21 (DE3) "Rosetta2" strain was utilized for production of protein from

target genes cloned in T7-driven expression vectors. DH10MultiBacY strain was used for recombinant bacmid preparation.

Table 2.13: Bacterial strains

Strain	Description	Reference
XL1 blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*q*ZΔM15* Tn*10* (Tet^R)]	Agilent, Germany
XL2 blue	*endA1 supE44 thi-1 hsdR17 recA1 gyrA96 relA1 lac* [F´ *proAB lacI*q*ZΔM15* Tn*10* (Tet^R) *Amy* (Cam^R)]	Agilent, Germany
BL21 (DE3) "Rosetta2"	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* pRARE2 (Cam^R)	Novagen, USA
DH10 MultiBacY	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K λ– *rps*L *nup*G / bMON14272 Δ(*chiA, v-cath*) / pMON7124	Dr. I. Berger, EMBL, Grenoble

2.1.14 Software

Table 2.14: Software

Software	Reference
CNS	(Brünger *et al.*, 1998)
Collaborative Computational Project Number 4 (CCP4i) program suite	(Potterton *et al.*, 2003)
Coot	(Emsley and Cowtan, 2004; Emsley *et al.*, 2010)
Corel Draw	Corel Corporation, USA
GraphPad Prism	GraphPad Software, Inc., USA
HKL2000	(Minor *et al.*, 2006)
Pro-Data Viewer	Applied Photophysics Ltd.
Pymol	Schrödinger LLC, USA
Sharp/Autosharp	(Bricogne *et al.*, 2003)
ShelX suite	(Sheldrick, 2008)
XDS	(Kabsch, 2010)

2.2 Methods

Unless otherwise stated, all molecular biological methods were performed as described in (Sambrook *et al.*, 1989). Commercial kits were used according to the instructions of the manufacturer.

Special attention is given to some methods used extensively during this thesis. In such cases, the technique will be first introduced in more details followed by a description of how it was specifically applied to this work.

2.2.1 Nucleic acid methods

2.2.1.1 Determination of nucleic acid concentration

To determine the concentration of nucleic acids, the light absorption of an aqueous solution was measured at the wavelength of 260 nm using a Nanodrop spectrophotometer. The concentration was then calculated using theoretical absorption values at 260 nm (as described in Sambrook, 1989).

- double-stranded DNA 1 OD_{260} = 50 µg/ml
- single-stranded DNA 1 OD_{260} = 33 µg/ml
- single-stranded RNA 1 OD_{260} = 40 µg/ml

2.2.1.2 Agarose gel electrophoresis for DNA

To separate nucleic acids, agarose gel electrophoresis was used for both analytical visualization and purification of preparative amounts of DNA. Depending on the size of the DNA fragments, the gels were prepared with varying agarose concentrations of 0.8–1.5% in 1×TBE buffer. Commercial DNA ladders were used as marker on each gel. Before casting, ethidium bromide was added to the final concentration of 0.7 µg/ml in gel solution. Samples were mixed with DNA loading dye. Gels were run at a constant voltage of 100 V in 1×TBE buffer. DNA bands were visualized by UV illumination at 254 nm.

2.2.1.3 DNA purification using agarose gel electrophoresis

DNA bands were illuminated at a wavelength of 254 nm and excised from the gel with a sterile razor blade. DNA was extracted from the gel using QIAquick gel extraction kit following the instruction of the manufacturer.

2.2.1.4 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used for amplification of genes or gene fragments. Depending on the target gene, cloned *Pfu* polymerase or Herculase enhanced polymerase

were used according to the instructions of manufacturers. Typical PCR conditions are shown below.

Table 2.15: Typical conditions for PCR

Parameter	Typical condition
10×polymerase reaction buffer	1×
dNTP mixture (10 mM of each dNTP)	0.2 mM
DNA template (plasmid DNA)	10-50 ng
Primers	0.3 µM each
DNA polymerase	0.1 – 0.2 U/µl
DMSO	0 – 4%
Denaturing temperature	98°C
Annealing temperature	T_m-5°C
Extension time	0.5 – 2 min per kb
Extension temperature	72°C
Number of cycles	30

2.2.1.5 Site-directed mutagenesis

Desired mutations were introduced using the QuikChange II XL Site-Directed Mutagenesis Kit according to the manufacturer's instructions. The resulting clones were verified by DNA sequencing.

2.2.1.6 Restriction digestion and ligation of DNA

Restriction digestions were carried out to generate compatible ends in vectors and PCR products for subsequent ligation reactions. Buffers and temperatures were chosen according to the manufacturer's instruction. The PCR amplified fragments were purified before digestion using a QIAquick PCR purification kit. A typical digestion reaction contained 2–4 µg DNA and 3 units of restriction enzyme per µg of DNA in a total volume of 50 µl.

For ligation, the digested DNA was resolved on a preparative agarose gel and extracted as explained in sections 2.2.1.2-3. Ligation reactions typically contained 100 ng of plasmid DNA, three to five fold molar excess of the insert DNA, 800 units of T4 DNA ligase in 20 µl reaction volume. Reaction mixtures were incubated at 16°C overnight or at 25°C for 1 h.

2.2.1.7 Plasmid isolation from *Escherichia coli* cells

A single colony from an overnight grown LB-agar plate was used to inoculate LB medium. Cells were grown in LB medium overnight at 37°C. Plasmid purification was carried out using Mini- or Maxiprep kits, according to the manufacturer's instructions.

2.2.1.8 Plasmid verification

All cloned plasmids were verified for the presence of the correct insert by PCR or analytical restriction digestion. Sequences of the inserts which showed the correct size in agarose gel analysis were verified by DNA sequencing (Seqlab, Göttingen).

2.2.1.9 Radioactive labeling of the 5'-end of RNA oligonucleotides

For electrophoretic mobility shift assays, RNA oligonucleotides were radioactively labeled at their 5'-end using γ-[^{32}P]-ATP and T4-polynucleotide kinase (PNK).

A typical labeling reaction contained 10 pmol of the RNA oligo, 1 µl of 10×PNK buffer (New England Biolabs), 1 µl of T4 PNK (10 U/µl) and 5 µl of γ-[^{32}P]-ATP (6000 Ci/mmol, 10 mCi/ml). The reaction mixture was incubated at 37°C for 1 h. After incubation, the reaction volume was adjusted to 50 µl with ddH_2O and loaded on a ProbeQuant G-50 or G-25 column to purify the radioactively labeled oligonucleotide from free γ-[^{32}P]-ATP.

Synthethic RNA oligonucleotides from Dharmacon (Bonn, Germany) were deprotected before labeling. The 2'-protecting groups were removed by addition of 2'-deprotection buffer (100 mM acetic acid, adjusted to pH 3.8 with TEMED) and incubation at 60°C for 30 min. Samples were then dried in a Speed-Vac and the RNA pellets were resuspended in ddH_2O.

2.2.2 Cell and cell culture methods

2.2.2.1 *Escherichia coli* strains and cultivation

Escherichia coli (E. coli) cells were grown in liquid medium or on agar plates supplemented with adequate antibiotics in the following concentrations:

- ampicillin 100 µg/ml,
- chloramphenicol 34 µg/ml,
- tetracycline 15 µg/ml,
- gentamicin 7 µg/ml,
- kanamycin 50 µg/ml.

2.2.2.2 Preparation of chemically competent *Escherichia coli* cells

Chemically competent cells were prepared using the DMSO-method (Inoue *et al.*, 1990). For preparation of chemically competent cells, *E. coli* cells were grown in SOB medium with

appropriate antibiotics at 18°C to reach an OD_{600} of 0.5−0.6. Cells were collected in sterile bottles by centrifugation at 3000×g for 10 min at 4°C. The cell pellet was resuspended in ice-cold TB buffer and harvested by centrifugation (3000×g, 10 min, 4°C). This step was repeated twice and the final cell pellet was resuspended in ice-cold sterile TB buffer, containing 7% (v/v) DMSO.

The competent cells were aliquoted in cold Eppendorf tubes, flash-frozen in liquid nitrogen and stored at -80°C.

2.2.2.3 Preparation of electro-competent *Escherichia coli* cells

To prepare electro-competent bacterial cells, *E. coli* cells were grown in 500 ml LB medium to reach an OD_{600} of 0.4−0.5. Cells were harvested in sterile bottles by centrifugation at 4000×g for 15 min at 4°C. The cell pellet was washed twice with 500 ml ice-cold sterile H_2O and twice with 200 ml ice-cold sterile 10% glycerol. Cells were finally resuspended in 5 ml ice-cold sterile 10% glycerol, aliquoted in cold Eppendorf tubes, flash-frozen in liquid nitrogen and stored at -80°C.

2.2.2.4 Transformation of *Escherichia coli* cells

For electroporation, 50−100 ng DNA were mixed with 50 µl electro-competent *E. coli* cells on ice. The mixture was transferred to an ice-cold electroporation cuvette (2 mm electrode gap) and subjected to a 4.8 ms pulse of 2.5 kV. Cells were collected in 1 ml of SOC medium and incubated for 1 to 4 h at 37°C in a shaker. Subsequently, cells were harvested by 3 min centrifugation at 1500×g. The cell pellet was resuspended in 150 µl of LB medium and streaked out on an agar-plate containing the selective antibiotics.

For chemical transformation, 100−200 ng of plasmid DNA was mixed with 100 µl of chemically competent *E. coli* cells and incubated for 30 min on ice. Ice-cold cells were then heat-shocked for 45 sec at 42°C and cooled on ice for 3 min. Cells were mixed with 1 ml of SOC medium and incubated at 37°C for 1 h in a shaker. The cells were collected and selected on LB-agar plates supplemented with appropriate antibiotics.

2.2.2.5 Protein expression in *Escherichia coli*

For protein expression, cells were grown in ZYM-5052 (50 mM phosphate) auto-inducing complex medium (Studier, 2005). TunAir flasks were used to maintain adequate aeration in preparative protein production setups. Optical densities of cultures were measured in plastic cuvettes with 1 cm path-length in an Ultrospec 3000 pro spectral photometer at 600 nm wavelength using the respective plain media as reference.

Complex media containing enzymatic digests of casein and yeast extract are extensively used since they support growth of a wide range of *E. coli* strains, with different nutritional requirements, and cultures typically grow several times faster than in simple mineral salts media supplemented with glucose as the only carbon source. However, due to small amounts of lactose, inducing activity is fairly common in complex media.

Auto-induction depends on mechanisms bacteria use to regulate the use of carbon sources present in the growth medium. Lactose is prevented from inducing production of target protein by compounds that can be depleted during growth, such as glucose, glycerol and amino acids. Consequently, if glucose is present in the medium, it prevents the uptake of lactose. Ideally, the expression strain would grow to saturation in auto-inducing media, when depletion of inhibitory factors would allow cells to take up lactose and convert it to allolactose, the natural inducer of the *lac* operon. Induction unblocks both the *lac*UV5 and T7*lac* promoters of T7 RNA polymerase and target protein, respectively, and leads to large scale expression levels in highly saturated cultures.

2.2.2.6 Baculovirus expression vector system

Many eukaryotic proteins fail to be produced properly in *E. coli* due to protein size limitations or particular requirements for chaperone systems or post-translational modifications that *E. coli* cannot support. Overproduction of such proteins demands a eukaryotic expression system. Baculovirus expression vector system (BEVS) is robust and well suited for producing eukaryotic proteins as it provides a number of advantages:

- Baculoviruses do not replicate in eukaryotic cells besides their insect cell hosts which is associated with low safety requirements;
- Due to the flexibility of the *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) envelope, large DNA insertions can be accommodated in its 130 kb dsDNA genome and, consequently, large proteins often authentically processed can be produced;
- Insect cell cultures can be grown in regular shaker flasks;
- Protein yields are comparable to *E. coli.*

Baculovirus gene expression takes advantage of the viral life cycle. A well-studied baculovirus, the AcMNPV, has three classes of genes that are expressed in a chronologically regulated, sequential manner (Smith *et al.*, 1983; Pennock *et al.*, 1984). The first class comprises the early genes, controlled by host-like promoter and transcribed by the host transcriptional machinery (Friesen, 1997). Once the viral DNA is replicated, the late genes are

expressed, such as the p10-coding gene, which demand virus-encoded transcriptional machinery (Lu and Miller, 1997; Passarelli and Guarino, 2007). In the end of the infectious cycle, the very late genes are expressed which code for several proteins including polyhedrin (polh). Polyhedrin is one of the most abundantly expressed proteins and is involved in the formation of occlusion bodies in the nuclei of insect cells infected with wild type baculovirus. In tissue culture, these late and very late proteins (such as p10 and polh) are non-essential for viral replication and their genes can be exchanged with other genes to create a recombinant baculovirus. This circumstance was exploited for producing the first recombinant baculoviruses in 1983 (Smith *et al.*, 1983) by standard homologous recombination using transfer plasmids carrying the foreign genes.

2.2.2.6.1 The MultiBac system

The MultiBac system (Trowitzsch *et al.*, 2010) is based on an *Ac*MNPV baculovirus genome derived from the Tn7-based BAC variant described by Luckow and Summers, 1988. The MultiBac baculoviral genome is propagated as a bacterial artificial chromosome (BAC) in *E. coli* cells and utilizes a Tn7 attachment site embedded in a *LacZα* gene for integrating foreign genes *via* transfer plasmids (Acceptor plasmids) into the baculoviral genome. Successful integration of the expression cassette leads to disruption of the *LacZα* gene and enables the selection of positive clones by blue/white screening.

The MultiBac baculovirus also contains modifications to reduce proteolytic breakdown of the overproduced proteins. Baculoviral genes *v-cath* and *chiA*, coding for a cathepsin protease and a chitinase, were substituted by a LoxP imperfect inverted repeat introduced together with an enhanced YFP protein coding gene. The resulting BAC is called EMBacY.

Two families of modular transfer plasmids denominated Acceptors (pFL and pKL) and Donors (pUCDM, pIDC, pIDS, pIDK and pSPL) are currently used to deliver the heterologous genes, under control of p10 or polh promoters, to the BAC. Both Acceptor and Donor plasmids harbor a LoxP site that can be used to combine one Acceptor with one or more Donor plasmids through cre-recombination before integration into the EMBacY. This strategy is widely used in case of multiprotein complex expression. The transposition involving the Acceptor plasmid or Acceptor/Donor fusions and the EMBacY occurs *in vivo* in *E. coli* cells, called DH10MultiBacY, tailored for this purpose. These cells provide a helper plasmid encoding the Tn7 transposon complex for accessing the Tn7 site on the EMBacY. For virus production, the composite EMBacY is isolated from DH10MultiBacY and used to transfect insect cells.

The presence of YFP gene in the EMBacY serves the purpose of directly observing virus performance by using a fluorescence spectrophotometer (Trowitzsch *et al.*, 2010). YFP is under the control of a very late promoter (polh) as are, typically, the heterologous genes. It is observed that when YFP expression reaches a plateau, expression of other heterologous proteins under the same promoter (as well as p10 promoter) also reach their peak production. Consequently, it is possible to follow protein production levels by monitoring YFP expression.

2.2.2.6.2 Insect cell cultivation

All handling of insect cell lines was carried out under sterile conditions in a laminar flow hood. Sf9, Sf21 and High Five™ cell stocks (25 ml) were maintained in 250 ml shaker flasks at 27°C while shaking at 80 rpm. Sf9 and Sf21 cells were cultured in SF900 III SFM and High Five™ cells in Express Five® SFM medium supplemented with L-glutamine prior to use. Population doubling time for insect cells vary depending on growth conditions:

- Sf9 and Sf21 cells double approximately every 24 hours;
- High Five™ cells double every 18-24 hours.

The suspension cultures were subcultured every day to a density of 0.5×10^6 cells/ml to maintain log phase growth and maximum viability. Cell viability was assessed every day by cell counting using a Casy TT cell counter. Cell count using Casy TT is performed by suspending 50 µl of cells in 10 ml CASY®ton, an electrolyte developed specifically for cell counting. The solution (400 µl) is then aspirated through a precision measuring capillary (150 µm) of defined geometry at a constant flow speed. During the three measuring cycles, a pulsed low voltage field with 1 MHz is applied to the measuring capillary via two platinum electrodes. During the passage of the cells through the measuring capillary, they displace a quantity of electrolyte corresponding to their volume. Since intact cells can generally be considered insulators, an increased level of resistance is achieved over the measuring pore. This resistance corresponds to the dimension for the volume of the cells. In contrast, dead cells, whose membrane no longer acts as an electrical barrier, are recorded by the size of their cell nucleus. The size range used to calculate the measuring results for Sf9 and Sf21 was 10.70–40.00 µm and for High Five™ was 13.00–40.00 µm. The normalization cut off determines the size range for which all percentage calculations indicate 100% (% viability, % counts or % volume). The normalization cut off for Sf9 and Sf21 was 6.80–40.00 µm and for High Five™ was 7.80–40.00 µm. The lower range of the normalization cut off defines the size range for determining debris.

2.2.2.6.3 Bacmid preparation, transfection and virus amplification

Recombinant bacmid was prepared using the MultiBac system as described by Fitzgerald *et al.*, 2006.

The genes of hBrr2 as well as mutants and fragments were cloned in a pFL vector or in a modified version called pFL10His and transformed to electrocompetent DH10MultiBacY cells. Following electroporation, cells were incubated at 37°C for 4 h in SOC medium and plated on LB-agar media containing 50 μg/ml kanamycin, 7 μg/ml gentamicin, 10 μg/ml tetracycline, 100 μg/ml ampicillin, 100 μg/ml Bluo-Gal and 1 mM IPTG and incubated overnight.

The Cre-lox recombination process was used to join Acceptor and Donor plasmids to create constructs for co-expression of the human Prp8-Snu114-Brr2 and the yeast Prp8-Snu114 proteins. For this purpose, we have cloned hPrp8 in a Donor plasmid (pIDS), hBrr2 with a C-terminal His_6-tag in an Acceptor plasmid (pFL) and hSnu114 with a C-terminal strepII-tag in another Donor plasmid (pIDK) while yPrp8 with a C-terminal FLAG tag was cloned in an Acceptor plasmid (pFL) and ySnu114 in a Donor plasmid (pIDS).

For the reaction, donor and acceptor plasmids were mixed in equal amounts (1-2 μg each) in the Cre reaction buffer and Cre recombinase was added. The reactions were incubated at 37°C for 1h. The electrocompetent DH10MultiBacY cells were, then, transformed with the reaction mixtures. After electroporation, SOC media was added and cells were incubated at 37°C for 4h. The cells were plated on LB-agar media containing the desired combination of antibiotics and incubated overnight.

The transformants were verified by an antibiotic challenge. For this purpose, the colonies from the plate were picked and inoculated into LB without antibiotics. After incubation at 37°C for about 1h, aliquots of pre-incubated cell cultures were added to the wells of a 96-well plate containing the desired antibiotics in different combinations. The plate was left shaking at 37°C overnight. The cultures which demonstrated resistance to all of the antibiotics were selected for amplification of the recombination product. Finally, all Acceptor/Donor assemblies were verified by digestion to confirm the presence of the desired genes.

The expression constructs (Acceptor plasmid or Acceptor/Donor fusions) were integrated *via* Tn7 transposition into the EMBacY during the DH10MultiBacY transformation procedure. The Tn7 transposition site is embedded in a *lacZα* gene allowing the selection of positive EMBacY recombinants by blue/white screening. White positive clones were re-streaked on a new LB-agar plate containing antibiotics as well as IPTG and BluoGal. After the second round of blue-white screening, the single white colonies were inoculated into LB

media containing the desired antibiotics and cultivated O.N. with vigorous shaking. Recombinant EMBacYs were isolated from the bacterial hosts using QIAprep miniprep kit, omitting the column purification step.

The recombinant bacmids were further precipitated by addition of isopropanol and pelleted by centrifugation in a table-top centrifuge at 17000×g for 30 min. The pellets were washed with 70% ethanol and the dried bacmid pellets were dissolved in 40 µl ddH_2O and mixed with 200 µl of SF-900 III medium in a sterile laminar flow hood. The transfection reagent, X-tremeGene 9, was diluted in the same medium (1:10, 10 µl X-tremeGene 9 in 100 µl medium) and mixed with the bacmid solution. The mixture was incubated for 1 h and distributed equally in two wells of a 6-well plate, each containing 3 ml of Sf9 cells (0.3×10^6 cells/ml).

To obtain initial virus (V_0 generation), Sf9 cells grown as adhesive culture in 6-well plates were transfected with composite EMBacY BACs. The efficiency of transfection was monitored by eYFP fluorescence using an Axiovert 40 CFL microscope equipped with an HBO illuminator and a proper filter set for eYFP fluorescence visualization. The V_0 generation was harvested 60 h post-transfection and immediately used to infect a 25 ml suspension culture of Sf21 cells in an Erlenmeyer shaker flask for further virus amplification (V_1 generation). It is important to maintain a low multiplicity of infection (MOI – number of infectious virus particles per cell in a cell culture) during virus production and amplification. Low MOI regime is the best way to avoid detrimental gene deletions which can occur during virus amplification, adversely affecting protein yields (Braunagel *et al.*, 1998). Low MOI is ascertained by allowing one doubling time of the Sf21 cells after addition of V_0 (Bieniossek *et al.*, 2008). Infected Sf21 were diluted to a cell count below 10^6 cells/ml every 24 h until cell proliferation arrest (typically 2–4 days). After proliferation arrest, 10^6 cells were sampled from the infected culture every 12 h for cell counting and diameter determination using a Casy TT cell counter. For YFP fluorescence signal measurements (performed in a FLUOTRAC-600 black 96 well polystyrene plates using a Victor X3 multilabel plate reader), the cells were resuspended in 500 µl PBS and lysed by sonication. The amplified virus (V_1 generation) was harvested by centrifugation and the supernatant containing the virus 60 h after cell proliferation arrest was stored at 4°C. Usually, the same volume of fresh media was replenished to the culture. YFP signal from the 10^6 cells sampled every 12 h continued to be recorded until it reached a plateau. At this point, cells were harvested and protein production was analyzed by SDS-PAGE using a fraction of the samples used for YFP fluorescence measurements.

2.2.2.6.4 Protein expression in insect cells

To determine the appropriate concentration of V_1 to be used for large scale expression, 3 flasks each containing 25 ml of High Five^{TM} at 0.5×10^6 cells/ml were infected with varying volumes of V_1, such as 30 µl, 60 µl and 90 µl. Cell counting and diameter determination was performed every 24h as described above. The amount of V_1 that allowed one round of cell division and yielded high protein production was chosen for scalling up the expression.

For large scale expression, 400 ml of High Five^{TM} cells kept in suspension in an Erlenmeyer shaker flasks at 0.5×10^6 cells/ml were infected with 0.4 to 1 ml of V_1 virus. Samples of 10^6 cells were taken from the infected culture every 12 h for cell counting and diameter determination as described previously. The samples were further used for YFP fluorescence signal measurements as described above and for protein production analysis by SDS-PAGE. The infected cells were harvested when the YFP fluorescence signal reached a plateau (typically 2–3 days) or before the cell viability dropped below 90%.

2.2.3 Protein methods

2.2.3.1 Determination of protein concentration

Protein concentrations were determined using either the Bradford assay (Bradford, 1976) the BCA assay (Smith *et al.*, 1985) or by measuring the absorbance of denatured protein at the wavelength of 280 nm. The Bradford and BCA assays were used according to the manufacturer's instructions. When using the absorbance at 280 nm, the proteins were denatured in 20 mM phosphate buffer, pH 6.5, 6.0 M guanidinium hydrochloride and the theoretical extinction coefficients of the proteins were used to calculate the concentration (Walker, 2005).

2.2.3.2 Protein precipitation

Protein samples were precipitated to reduce the volume before running the gel. 5 volumes of ice-cold acetone were added to one volume of protein solution. The sample was rapidly mixed and incubated overnight at -20°C. The mixture was centrifuged for 15 min at 4°C in a tabletop centrifuge at 17000×g, leftover acetone was evaporated and the pellet was resuspended in SDS loading buffer.

2.2.3.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

To analyse proteins, SDS-polyacrylamide gels were used. According to the protein sizes to be separated, 10–12% (final acrylamide solution) gels were prepared and run vertically in SDS-PAGE running buffer. SDS-PAGE was basically performed as described elsewhere

(Laemmli, 1970; Weber *et al.*, 1972). In brief, a typical gel contained a stacking part, consisting of 1×stacking gel buffer and 16% (v/v) Rotiphorese Gel 30 solution, polymerized with 0.3% (w/v) APS and 0.03% (v/v) TEMED, and a separating part, consisting of 1×resolving gel buffer and 25–40% (v/v) Rotiphorese Gel 30 solution, polymerized with 0.5% (w/v) APS and 0.1% (v/v) TEMED. Protein samples were denatured in protein loading buffer and heated to 95°C for 5 min. Electrophoresis was typically stopped when the bromophenol blue border reached the bottom of the gel. To visualize the protein bands, SDS-PAGE gels were first fixed for 30 min with a solution containing 4.7% (v/v) formaldehyde and 50% (v/v) methanol. Subsequently, the gels were incubated in Coomassie staining solution for 5-15 minutes and destained by the sequential addition of destaining solution I until bands became visible. The gel was destained further with destaining solution II (Fairbanks *et al.*, 1971).

2.2.3.4 Silver staining of protein gels

For silver staining of protein gels, the gel was first fixed in 50% (v/v) methanol and 5% (v/v) acetic acid solution for 20 min and then washed with 50% (v/v) ethanol, 30% (v/v) ethanol and ddH_2O water for 10 min each. The gel was further treated with 1.25 mM $Na_2S_2O_3$ for exactly 60 sec and immediately rinsed thrice with ddH_2O to remove the thiosulfate from the gel surface. For staining, the gel was incubated in a solution containing 6 mM $AgNO_3$ and 0.0185% (v/v) formaldehyde for 20 min followed by three times washing with ddH_2O. For stain developing, the gel was rinsed with 0.19 M Na_2CO_3, 0.026% (v/v) formaldehyde. The solution was replaced when it turned yellow. Development continued until protein or RNA bands were visible. The developing procedure was stopped with cold 5% (v/v) acetic acid solution.

2.2.3.5 Western Blotting

Proteins were separated by 10–12% SDS PAGE and electro-blotted on a PVDF transfer membrane in a Trans-Blot Electrophoretic Transfer Cell according to the membrane manufacturer's instruction. A pre-stained MW standard was applied onto one lane to assess the transfer efficiency. The gel, membrane and filter paper were soaked in the blotting buffer prior to the transfer. Special care was taken to remove air bubbles trapped between the layers of the assembly. The transfer was performed in blotting buffer for 2 h at 70 V, 4°C. To decrease non-specific binding of antibodies, the membrane was blocked in blocking buffer overnight at 4°C. To detect hBrr2, the membrane was first incubated with a primary antibody solution for 1 h at room temperature and washed three times 15 min each wash with washing buffer I at room temperature. Subsequently, the membrane was incubated with goat

horseradish peroxidase-conjugated anti-rabbit anti-rabbit IgG (secondary antibody solution) for 1 h at room temperature and then washed three times for 15 min each wash with washing buffer II. Proteins were detected by enhanced chemiluminescence using an ECL kit according to the manufacturer's instructions. The luminescence was detected by exposure on the ECL Hyperfilm. The film was developed on an X-Omat 2000 developer.

2.2.3.6 Purification of human and yeast Brr2

If not mentioned otherwise, the same purification protocol was used for all human and yeast Brr2 constructs. The High Five^{TM} cell pellet was resuspended in 50 mM HEPES, pH 7.5, 600 mM NaCl, 2 mM β-mercaptoethanol, 0.05% NP40, 10% glycerol, 10 mM imidazole, supplemented with EDTA-free protease inhibitors and lysed by sonication using a Sonopuls Ultrasonic Homogenizer HD 3100. Cell debris was removed by centrifugation and the soluble extract was filtered using a 0.44 µm filter. The target was captured on a 5 ml HisTrap FF column and eluted with a linear gradient from 10 to 250 mM imidazole. The His-tag was cleaved with TEV protease during overnight dialysis at 4ºC against 50 mM HEPES, pH 7.5, 600 mM NaCl, 2 mM β-mercaptoethanol, 10% (v/v) glycerol and 15 mM imidazole. The cleaved protein was again loaded on a 5 ml HisTrap FF column to remove the His-tagged protease, uncut protein and cleaved His-tag. The flow-through containing the protein of interest was diluted to a final concentration of 80 mM sodium chloride and loaded on a MonoQ 10/100 GL column equilibrated with 25 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM β-mercaptoethanol. The protein was eluted with a linear 50 to 600 mM sodium chloride gradient and further purified by gel filtration on a 26/60 Superdex 200 column in 10 mM Tris-HCl, pH 7.5, 200 mM NaCl, 2 mM DTT. In case of yeast Brr2, after the recycling step, the protein was diluted to a final concentration of 80 mM sodium chloride and loaded on a HiPrep Heparin FF column equilibrated with 25 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM β-mercaptoethanol. The protein was eluted with a linear 50 to 600 mM sodium chloride gradient and further purified by gel filtration on a 26/60 Superdex 200 column in 10 mM Tris-HCl, pH 7.5, 200 mM NaCl, 2 mM DTT. For the purification of the N-terminal cassette construct and hBrr2^{HR} mutants, all solutions used were buffered at pH 8.0. All proteins produced for activity assays retained an N-terminal His-tag since the TEV cleavage and HisTrap FF recycling steps were omitted.

2.2.3.7 Expression and purification of C-terminal domain of Prp8 from *S. cerevisiae*

Prp8ct$^{1836-2398}$ was expressed in *E. coli* Rosetta 2(DE3). Cells were grown in autoinducing medium (Studier, 2005) to an OD_{600} of 0.3 at 37°C, shifted to 16°C for 60 h and harvested by centrifugation. Cell pellets were resuspended in 25 mM HEPES, pH 7.5, 500 mM NaCl, 0.5 mM DTT, 0.5% NP40, 5% (v/v) glycerol, supplemented with EDTA-free protease inhibitors, and lysed by five passes through a microfluidizer. The cell lysate was clarified by centrifugation and then loaded onto a 5 ml HisTrap FF column equilibrated in resuspension buffer without detergent. Protein was eluted with a linear gradient from 10 mM to 500 mM imidazol. Fractions containing the protein of interest were pooled, mixed with TEV protease and dialyzed over night at 4ºC against 25 mM HEPES, pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 0.5 mM DTT. The digested sample was again loaded on a 5 ml HisTrap FF column equilibrated in resuspension buffer (without detergent) in order to remove TEV protease, uncut protein and the cleaved His_6-tag. The flow-through containing Prp8ct$^{1836-2398}$ was diluted to a final concentration of 30 mM sodium chloride before being applied to a MonoQ HR 5/5 column equilibrated with 25 mM Tris-HCl pH 8.0, 30 mM sodium chloride and 1 mM DTT. Protein was eluted in a 50 ml linear gradient from 30 mM to 1.5 M sodium chloride. The fractions containing Prp8ct$^{1836-2398}$ were pooled and injected directly onto a 26/60 Superdex 200 column equilibrated in 10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM DTT.

2.2.3.8 Limited proteolysis

Limited proteolysis experiments can inform on conformational features of proteins. In a number of studies, it has been demonstrated that the sites of limited proteolysis along the polypeptide chain of a protein are characterized by enhanced backbone flexibility, implying that proteolytic probes can pinpoint the sites of local unfolding in a protein chain (Fontana and Laureto, 1997; Fontana *et al.*, 2004). This means that only flexible regions of the target protein such as disordered termini, exposed loops, or flexible domain linkers can be cleaved. Removing such flexible parts from a protein generates more compact and conformationally homogeneous molecules or compact single domains. These proteolytic fragments or domains of a protein may crystallize more readily or form better diffracting crystals than the intact protein.

Proteolysis of a protein substrate can occur only if the polypeptide chain can bind and adapt to the specific stereochemistry of the protease active site. However, since the active sites of the proteases have not been designed by nature to fit a specific sequence and fixed

stereochemistry of a stretch of at least 8–10 amino acid residues of a particular protein, an induced-fit mechanism of adaptation of the protein substrate to the active site of the protease is required for binding and formation of the transition state of the hydrolytic reaction (Herschlag, 1988). Therefore, the native rigid structure of a globular protein generally is not a good substrate for a protease, as documented by the fact that folded proteins under physiological conditions are rather resistant to proteolytic degradation. Nevertheless, native globular proteins can be attacked by proteases given the fact that it is not a static entity, but, instead, a dynamic system capable of fluctuations around its average native state. Indeed, crystallographers analyze this protein mobility in terms of *B*-factors for both side chain and Cα-backbone. The main chain *B*-factor is a measure of the fluctuations of the molecule energy around the potential minimum. These fluctuations can be visualised as the displacement of the polypeptide chain from its native structure. Naturally, these displacements can lead to local unfolding. It can be envisaged that the higher energy, locally unfolded states are those required for a native protein to be attacked by a proteolytic enzyme. Evidence for this mechanism of local unfolding required for limited proteolysis has been provided by demonstrating a close correspondence between sites of limited proteolysis and sites of higher backbone displacements or areas that are poorly resolved in the electron density map of proteins which have their 3D structures elucidated (Fontana *et al.*, 1986). Therefore, it was concluded that limited proteolysis of globular proteins preferably occurs at flexible loops or mobile and disordered stretches of amino acid residues (for example, N- or C-terminal extensions or conformationally unstable regions often found between domains). In particular, chain segments in a regular secondary structure (such as helices) are generally not good sites for limited proteolysis. Based on the fact that sites for limited proteolysis require large conformational changes (local unfolding), a possible explanation that rigid elements of secondary structure are not easily hydrolysed by proteases is that it would be energetically disadvantageous. If proteolysis is occurring at the center of a helical segment, likely the helix is fully destroyed and, consequently, all hydrogen bonds, which cooperatively stabilize it, are broken. On the other hand, a peptide bond fission at a disordered flexible site likely does not change much the energetic state of that site, since the peptide hydrolysis can easily be compensated by some hydrogen bonds with water.

It is plausible to suggest that limited proteolysis also requires that a specific chain segment of the folded protein substrate is sufficiently exposed to bind at the active site of the protease. Although the notion of exposure/protrusion/accessibility is a required property, it is clearly not sufficient to explain the selective hydrolysis of just one or few peptide bonds, since it is

evident that there are many exposed sites (the whole protein surface) which could be a target of proteolysis. Instead, enhanced chain flexibility appears to be the key feature of the site(s) of limited proteolysis (Fontana and Laureto, 1997).

Limited proteolysis has been widely used to define the boundaries of single domains or a set of tightly interacting domains in a molecule by trimming its flexible and unstructured parts and, thus, increasing its propensity to crystallization. The increased likelihood of a domain or a smaller set of domains to yield structural information can be explained, in a way, by the observation that large proteins are often composed of many individual domains. The conformational heterogeneity that results from motion between such domains is a severe impediment to crystallization.

2.2.3.8.1 Practical approach to limited proteolysis

The generation of domains by limited proteolysis relies directly on the tertiary structure of the protein under investigation and provides much firmer evidence for their existence than that provided by sequence homology and secondary structure predictions.

In practice, limited proteolysis is achieved by dilution of the proteases sufficiently so that they will only digest the most accessible and flexible regions of the protein substrate leaving the domains intact. Initially, the protein substrate should be digested with various proteases to establish which conditions are optimal for generating a protease-resistant domain. Three parameters are routinely varied: enzyme/substrate ratio, temperature and time of digestion. At first, it is common practice to perform the experiment for 30 min at room temperature using up to 6 proteases at 3 different dilutions each. More specifically, to determine the appropriate enzyme/substrate ratio for a particular protease, the same amount of substrate is digested with at least three different dilutions of the protease. The appearance of a discrete band, on SDS-PAGE, resistant to further degradation (even if only transiently), is evidence for the existence of a domain.

For optimization of the experiment, variation of reaction time, for example taking samples from the reaction mixture at regular time intervals for SDS-PAGE analysis, at different temperatures is used. The conditions which yield a clear band of the protease resistant fragment on SDS-PAGE are further analyzed.

Once a domain has been observed after the proteolysis experiment, N-terminal sequencing and mass spectrometry analysis of the band corresponding to the fragment on SDS-PAGE should enable unambiguous identification of the domain sequence. This information can be exploited for subsequent cloning, expression, purification and crystallization of the respective fragment/domain.

Scaling up the digestion is not generally a problem and can be a very suitable alternative to cloning the stable fragment/domain. It is possible that the fragment of a protein that would have the best crystallization properties cannnot be expressed in recombinant form. Indeed, it has been observed that unstructured N- or C-terminal extensions are sometimes required for recombinant expression. Limited proteolysis in large scale also appears advantageous in cases when the product of the proteolysis is a nicked protein (the non-covalent complex of usually two protein fragments).

Usually, digestion of a large quantity (several milligrams) of a specific protein is achieved under optimized conditions and the reaction is stopped by addition of an appropriate inhibitor. The stable fragment is, then, further purified and used for crystallization trials.

Note that instead of choosing the large scale proteolysis approach, *in situ* proteolysis can be performed by adding a protease digestion step to crystallization trials (Dong *et al.*, 2007).

2.2.3.8.2 Limited proteolysis of yeast and human Brr2

To search for stable fragments, full-length human and yeast Brr2 were treated with various proteases. For each reaction, 9 µg of protein were incubated with increasing amounts (0.004, 0.04 and 0.4 µg) of protease in the absence or presence of ATP/$MgCl_2$ at 20ºC for 30 min in buffer containing 10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM DTT. The reactions were stopped by addition of 2 µl PMSF (saturated solution in isopropanol) and 10 µl SDS-PAGE loading buffer. Half of each sample was separated by SDS-PAGE and bands were analyzed by tryptic mass spectrometric fingerprinting (Facility for Mass Spectrometry, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany). The remainder of the sample was separated by SDS-PAGE, blotted on a PVDF membrane, stained with Ponceau S and stable fragments were subjected to N-terminal sequencing (Microchemistry Core Facility, Max-Planck-Institute for Biochemistry, Martinsried, Germany).

2.2.3.9 Analytical gel filtration analysis

The hBrr2 constructs were analyzed by analytical size exclusion (gel filtration) chromatography on a Superdex 200 PC3.2 column in 10 mM Tris-HCl, pH 7.5, 200 mM NaCl, 2 mM DTT at a flow rate of 70 µl/min. For analysis of complex formation, proteins were mixed in defined ratios in a total volume of 60 µl and incubated for 30 min on ice prior to the chromatography. Fractions were concentrated and analyzed by SDS-PAGE.

2.2.3.10 Circular dichroism spectroscopy

Plane polarised light consists of two circularly polarised components of equal magnitude, one rotating counter-clockwise (left handed) and the other clockwise (right handed). Circular dichroism (CD) refers to the differential absorption of these two components. This effect will occur when a chromophore is chiral (optically active) by reason of its structure (for example, a carbon atom with four different substituents or a disulphide bond which is chiral because of the dihedral angles of the C-S-S-C atoms), or by being covalently linked to a chiral center or by being placed in an asymmetric environment by virtue of the three-dimensional structure adopted by the molecule. In practice, the plane polarized radiation is split into its two circularly polarized components by passage through a modulator. If one of the components is absorbed by the sample to a greater extent than the other after passage through the sample, the resultant radiation would be elliptically polarized. The CD spectropolarimeter does not recombine the two components but rather detects them separately. The dichroism at a given wavelength is then expressed as either the difference in absorbance of the two components or as the ellipticity in degrees (θ). A CD spectrum is obtained when the dichroism is measured as a function of the wavelength.

CD measurements were carried out on a J-810 spectropolarimeter in a quartz cuvette (0.1 cm path length). Far-UV spectra (190–240 nm) were recorded with proteins at 0.1 mg/ml in 50 mM potassium phosphate, pH 7.6, at 10°C with a scan rate of 100 nm/min and a data pitch of 0.2 nm. Ten consecutive scans were accumulated and averaged. Data were baseline corrected and secondary structure contents were calculated using software provided by the spectropolarimeter manufacturer. Melting profiles were recorded between 10 and 100°C with the proteins at 0.25 mg/ml in the same buffer plus 0.5 M urea by monitoring the CD signal at 220 nm. The temperature was ramped at 1°C per minute with data sampling at every 0.1°C.

2.2.3.11 Differential scanning fluorimetry (DSF)

Identifying conditions in which protein samples are stable over long periods of time, are prevented to aggregate or denature is extremely helpful for experiments involving analytical and biophysical techniques that require high protein concentrations.

Many factors can influence protein stability such as: buffers (chemical composition as well as pH), salts, detergents whose interactions with the protein are non-specific. Ligands, which bind proteins at specific sites, may also produce noticeable effects as protein stabilizers. The stability of a protein is characterized the Gibbs free energy of unfolding, ΔG_u, which is temperature-dependent. The stability of most proteins decreases as the temperature

increases, meaning that ΔG_u decreases and reaches zero at the equilibrium point where the concentrations of folded and unfolded protein are equal. At this equilibrium point, the temperature is called melting temperature (T_m).

DSF monitors the thermal unfolding of proteins in the presence of a fluorescent dye that is highly fluorescent in non-polar environment compared to aqueous solution where the fluorescence is quenched. To date, SYPRO orange is the dye possessing the most favorable properties for DSF owing to its signal to noise ratio. Additionally, the relatively long wavelength of excitation of SYPRO orange (near 500 nm) decreases the likelihood that buffer components would interfere with the optical properties of the dye. Practically, DSF is performed using a real-time PCR instrument and monitors the fluorescence intensity of the dye as a function of the temperature. Upon protein unfolding by thermal denaturation, the dye preferentially binds to the now exposed hydrophobic patches of the protein and the fluorescence intensity increases. This generates a sigmoidal curve that can be described as a two-state transition and the T_m can be extracted from the curve by determining the first derivative (Niesen *et al.*, 2007).

DSF is an excellent method to screen for conditions/compounds that stabilize proteins due to the small amounts and low protein concentrations required as well as for providing an easy readout for identification of such stabilizing conditions. Basically, the T_m value of the protein under each condition of the screen needs to be compared with the reference T_m, which is generally provided by protein in purification buffer.

DSF experiments were done in a 96-well plate in a plate reader combined with a thermocycler (Stratagene Mx3005P). In order to determine a suitable protein concentration, purified protein was diluted to series of concentrations varying from 1 µM to 10 µM in the final purification buffer supplemented with 10×SYPRO orange (1:500 dilution of the stock) in a total volume of 20 µl and pipetted into a 96-well plate. The temperature was increased from 25°C to 95°C and the fluorescence emission was monitored in steps of 1°C/min with hold steps of 30 sec between reads. The fluorescence intensity was then plotted as a function of temperature. The protein concentration that showed a clear sigmoidal curve was chosen for buffer optimization and compound screens. Different buffer compositions and compounds were tested for their stabilizing effect on the protein at the defined protein concentration. The sigmoidal curve from each condition was normalized and corrected for the background signal of the fluorophore in the buffer. The inflection points of the curves, representing the thermal melting temperature of the protein in the respective conditions, were compared.

2.2.3.12 ATP binding studies

Transient kinetic measurements allow the detection of reaction intermediates and the determination of rate constants of individual steps on the reaction pathway. Therefore, in pre-steady state kinetics, the processes generally observed are reactions of changes in the enzyme, substrate or products from one species to another, when the reaction undergoes a single turnover event. The recorded time courses show an exponential dependency. Stopped flow is a suitable technique to study pre-steady state kinetics of reactions occurring in solution in the millisecond to second time range. These reactions can be followed directly and in real time using spectroscopic changes as observables.

In stopped-flow machines, two syringes, each filled with enzyme or substrate for example, are connected to a mixing chamber. These syringes are compressed to deliver about 50 to 200 µl each. The two reactant solutions are mixed rapidly in the mixing chamber and the reaction mixture is then transferred to an observation cuvette. The flow of the mixture into the cuvette is stopped with the help of a stop syringe and a photomultiplier mounted at 90° to the excitation beam records the progress of the reaction occurring after 1 ms. The spectroscopic changes are measured as a function of time.

Several observables can be used when working with a stopped-flow apparatus. Fluorescence changes due to binding or conformational rearrangements of a fluorescence-labeled component can be monitored. Additionally, two fluorescence-labeled components can be used which serve as a fluorescence donor and acceptor in a fluorescence resonance energy transfer (FRET) measurement.

The mechanism of FRET involves a donor fluorophore in an excited electronic state which can transfer its excitation energy to a nearby acceptor chromophore in a non-radiative fashion through long-range dipole-dipole interactions. The phenomenon of FRET is not mediated by photon emission and does not require the acceptor chromophore to be fluorescent. However, in most cases, both donor and acceptor are fluorescent and the occurrence of energy transfer manifests itself through quenching of donor fluorescence accompanied by an increase in acceptor fluorescence emission. Several criteria must be satisfied in order for FRET to occur. The FRET pair should be chosen in such a way that the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor. Furthermore, the two involved fluorophores must be positioned within a range of 1 to 10 nanometers of each other. As described by the Förster equation, the efficiency of the energy transfer process varies in proportion to the inverse sixth power of the distance separating the donor and acceptor molecules. An additional requirement is that the fluorescence lifetime of

the donor molecule must be long enough to permit the event to occur. According to Förster's theory, the rate of energy transfer is given by the equation:

$$K_T = \frac{1}{T_D} \cdot \left(\frac{R_0}{r} \right)^6,$$

where R_0 is the Förster critical distance, T_D is the donor life time in the absence of the acceptor and r is the distance separating the donor and acceptor chromophores. When the radius r equals the R_0, the transfer efficiency equals to 50%.

All requirements fulfilled, FRET is observed when the sample is excited with a light of wavelength corresponding to the absorption maximum of the donor fluorophore and detecting light emitted at wavelengths centered near the emission maximum of the acceptor. The FRET method is particularly suitable to follow the kinetics of ligand binding, although subsequent rearrangements are often observed as well.

2.2.3.12.1 Practical approach to rapid kinetic measurements

The kinetics of interaction of hBrr2 and its mutants/fragments with adenine nucleotides was studied by fluorescence stopped-flow, performed in a stopped-flow apparatus (SX-20MV). The fluorescence of *mant*-labeled nucleotides was excited at 290 nm by FRET from tryptophan residues in the proximity of the ATP biding pockets of hBrr2 and measured at 90° after passing a cut-off filter (KV 408). FRET was observed only when both donor and acceptor were present since negligible fluorescence change of tryptophan was observed when adenine nucleotides bound to hBrr2. Experiments were performed by rapidly mixing equal volumes (60 µl each) of the reactants in 20 mM HEPES pH 8.0, 150 mM NaCl and 1.5 mM $MgCl_2$ and monitoring fluorescence change over time. In all cases, 1000 data points were acquired in logarithmic sampling mode. This mode yields reliable data since it collects an appropriate number of points both in the initial part of the curve (where signal changes occur rapidly) and at the end of the curve (where the signal changes slowly or becomes constant). The data was visualized using the software Pro-Data Viewer. The final curves were obtained by averaging 7–10 individual traces. Data were evaluated by fitting to a single exponential function with a characteristic time constant (k_{app}), amplitude (F_1) and another variable for the final signal (F_∞) according to the equation:

$$F = F_\infty + F_1 \exp\left(-k_{app} t\right),$$

where F is the fluorescence at time t. Where necessary, two exponential terms were used with two characteristic time constants (k_{app1}, k_{app2}), amplitudes of the signal change (F_1, F_2) and another variable for the final signal (F_∞) according to the equation:

$$F = F_{\infty} + F_1 \exp\left(-k_{app1}t\right) + F_2 \exp\left(-k_{app2}t\right).$$

Calculations and statistical analysis were performed using Prism (GraphPad software).

2.2.3.13 Electrophoretic mobility shift assay (EMSA)

[^{32}P]-5'-end labeled RNA oligonucleotides were mixed with varying concentrations of protein in an appropriate binding buffer. *S. cerevisiae* tRNA was added, when stated, to a final concentration of 0.5 µg/µl, as a non-specific competitor. Reactions were incubated at 20°C for 30 min. Samples were mixed with native gel loading buffer and fractionated on a 6% polyacrylamide gel (acrylamide/N,N'-methylene-*bis*-acrylamide ratio 29:1) in 1×TBE at 4°C. The gel was dried and visualized using a phosphorimager (Typhoon 8600).

2.2.3.14 Mass spectrometry

Mass spectrometry was used to analyze the protein fragments produced by limited proteolysis and the contaminants in protein preparations. The proteins/protein fragments were separated by SDS-PAGE and stained with Coomassie staining solution and further supplied to Monika Raabe, who performed the mass spectrometric analyses of the protein fragments (Group of Dr. Henning Urlaub; Bioanalytical Mass Spectroscopy, Max Planck Institute of Biophysical Chemistry, Göttingen).

Bands of interest were excised and proteins were digested in-gel with trypsin and extracted as previously described by Shevchenko *et al.*, 1996. Tryptic fragments were analyzed in a LC-offline MALDI-ToF/ToF-MS (4800 MALDI ToF/ToF) or in a LC-coupled ESI Q-ToF (QToF Ultima) and/or in an orbitrap (LTQ-Orbitrap XL) mass spectrometer under standard conditions. Proteins were identified by searching fragment spectra against the NCBI non-redundant database using Mascot as a search engine.

2.2.4 Crystallographic methods

2.2.4.1 Principles of X-ray crystallography

The first protein structure models based on X-ray diffraction data were derived in the 1950s, beginning with the structure of haemoglobin by Max Perutz and Sir John Cowdery Kendrew, whose pioneer work was rewarded with the Nobel Prize in Chemistry in 1962. Since this successful example, the crystal structure of thousands of biological molecules have been determined.

The main goal in X-ray crystallography is to derive the three dimensional structure of a given protein, other biomacromolecules or complex based on a set of X-ray scattered

intensities measured at different directions in space. This process can be divided in three basic steps and will be briefly explained as follows.

The first step is to obtain a protein crystal. The scattering of X-rays by a single protein molecule is extremely weak and still not possible to detect routinely. However the periodic arrangement of proteins inside a crystal creates interference effects, which greatly enhance the intensity of the scattered X-rays in particular directions allowing them to be measured. For the purpose of structure determination, there are specific concerns about the quality of the crystals of proteins. A crystal useful for X-ray diffraction should be relatively large, in the range of 50 μm to about 0.5 mm, with a single internal structure and no significant imperfections such as cracks and twinning.

In the second step, the crystal is exposed to an intense monochromatic beam of X-rays generating a distribution of scattered radiation in different directions in spaces, known as the diffraction pattern. A detector records the snapshots of the scattered X-rays as the crystal is rotated. The complete pattern is then retrieved based on the measured snapshots and symmetry considerations.

The third step consists of the determination of a structural model from the computational analysis of the diffraction pattern. Firstly, an electron density map is derived from the diffraction pattern. Then, the initially derived map is fitted with a structural model, which describes the position of every atom inside the protein. New phases are calculated and the structural model is refined by repeating this process iteratively until some statistical quantities related to the goodness of fitness achieve a certain desired value.

The most crucial part in step 3 is the determination of the initial electron density map from the diffraction pattern. It relies on the knowledge of both the amplitude and phase of the scattered X-rays, however only amplitudes can be derived from the experimentally measured intensities. This lack of experimental information about the phases constitutes a fundamental limitation, commonly known as the phase problem of X-ray crystallography. Many methods exist to retrieve the phase information including experimental substructure phasing, density modification, direct methods and molecular replacement, which will be further explained.

In the kinematic theory of diffraction, the amplitude and phase of the scattered X-rays in a particular direction in space are expressed by a complex function known as the structure factor **F**(**h**), given by

$$\mathbf{F}(\mathbf{h}) = \int_{cell} \rho(\mathbf{x}) e^{2\pi i \mathbf{h}\cdot\mathbf{x}} dV ,$$

in which $\rho(\mathbf{x})$ is the electron density at the position determined by the vector $\mathbf{x}$ = (x, y, z), $\mathbf{h}$ is a vector in the reciprocal space with components (h, k, l) that represent the Miller indices and the integration is performed over the entire volume of the unit cell, which generates the crystal lattice.

This equation expresses the structure factor as the superposition of the X-rays scattered by each infinitesimal volume element *dV* within the crystal unit cell with an amplitude that depends on the number of electrons and a phase angle that depends on its position inside the crystal and the diffraction geometry.

Within this formalism, the structure factor is related to the electron density by a Fourier transform. Hence, if the structure factor is known, the electron density can be retrieved by an inverse Fourier transform:

$$\rho(\mathbf{x}) = \frac{1}{\mathrm{V}} \sum_{\mathrm{h,k,l}} \mathbf{F}(\mathbf{h}) e^{-2\pi i \mathbf{h}\cdot\mathbf{x}} .$$

In general, an inverse Fourier transform would involve an integral like the forward Fourier transform, but if the object is periodic (like a crystal), it can be substituted with a summation.

The measured intensities provide only the real part of the complex number **F**(**h**), i.e. its amplitude which is proportional to the square root of the intensity. Since the phase, that is the complex part of **F**(**h**), is unknown, the Fourier transform equation cannot be directly used to determine the electron density.

Experimental phasing methods do not depend on any prior structural information about the protein molecule. The basis of experimental phasing lies in recording differences in intensity data resulting from electronic differences caused by heavy or anomalous marker atoms. In general, X-rays do not change their phase when scattered by atoms, only their amplitudes vary proportionally to the atom's electron density. However, for X-rays with energies in the vicinities of an absorption edge of an atom, quantum mechanical effects induce an additional phase shift in the scattered X-rays and modify their amplitude. This process, known as resonant or anomalous scattering, is mathematically represented by adding a complex energy/wavelength dependent term to the normal atomic scattering factor (F_0):

$$F_{an} = F_0 + f'(\lambda) + if''(\lambda) .$$

The real term, *f'*, is the modification in the scattering amplitude while the phase shift is given by the complex term, *f''*.

According to Friedel's law, in the absence of anomalous scattering, centrosymmetrically opposed pairs of structure factors **F**(**h**) and **F**(**-h**) have exactly the same magnitude

(|**F**(**h**)|=|**F**(**-h**)|) and conjugate phase angle (φ(**h**)=-φ(**-h**)). Therefore, the diffraction intensities I(**h**) and I(**-h**) are equal. The complex contribution due to anomalous scattering change both phase and amplitude of the structure factor, leading to differences between the paired intensities. As a consequence, Friedel's law no longer holds under these conditions.

The great majority of biological molecules do not contain atoms with absorption edges at energies typically used for X-ray diffraction experiments. Hence, to exploit anomalous effects in solving the phasing problem, foreign atoms (typically heavy metals) need to be introduced in the protein crystal.

Marker atom positions can be determined using Patterson analyses. Paterson maps can be reconstructed directly from intensities, unlike electron density maps, which require both the phases and the structure factor amplitudes as Fourier coefficients. The Patterson function is the Fourier transform of the reflection intensities and a contoured 2- or 3-dimensional representation of this function provides the Patterson map. A difference Patterson map constructed from anomalous intensity differences reveals the interatomic distances between anomalous marker atoms and is used for marker atom substructure solution. The difference data provide only the marker atom substructure and this information is enough to solve the phasing equations for the protein phases. Because of the centrosymmetry of the Patterson space, the handedness of the substructure solution is not determined. The solution can either be the correct substructure enantiomorph or have the wrong handedness.

Various methods to determine the protein phases exist and are often combined. Single isomorphous replacement (SIR), generic single wavelength anomalous dispersion (SAD) and native sulfur based single wavelength anomalous dispersion (S-SAD) require the support of direct methods or density modification to break the inherent twofold phase ambiguity of the phasing equations. Direct methods exploit the deviations from random-atom structure factor distributions as a result of the non-random distribution of atoms in a true molecular structure. The non-randomness leads to implicit relations between structure factor amplitudes and phase relations. Density modification techniques are powerful phase improvement methods, do not require an atomic model and can be considered an extension of experimental phasing.

The phase ambiguity can be also broken by adding orthogonal anomalous signal from the same crystal (SIRAS) or providing phases from at least a second derivative in multiple isomorphous replacement (MIR) or by combining both possibilities (MIRAS). Multi-wavelength anomalous dispersion (MAD) provides orthogonal dispersive and anomalous data from the same crystal and the phase ambiguity can be resolved directly from the phasing equations.

Another phasing method is molecular replacement (MR). The term replacement should, however, be interpreted in the sense of relocating and not substitution. In practice, a known structure model, presumed to be similar to the unknown structure with a sequence identity of usually 30% or higher, is rotated and translated in the unit cell or asymmetric unit the solution with the best fit between the calculated diffraction data from the replaced model and the observed data from the unknown structure is obtained. The 3-dimensional rotational and translational searches can be done based on Patterson search methods, 6-dimensional stochastic searches or modern maximum likelihood-based search functions. While the map reconstruction by Fourier synthesis is robust against noisy and poorly measured structure factor amplitudes, the phase errors strongly affect the electron density reconstruction. Therefore, model phases can seriously bias the resulting map towards the model when no bias minimization method is applied.

2.2.4.2 General crystallography setup

To identify initial crystallization conditions, different commercially available crystallization reagents were screened in 96-well MRC plates by sitting drop vapor diffusion technique. Drops of 200 nl (100 nl protein solution + 100 nl reservoir) were dispensed using a Cartesian liquid dispensing robot with 4 or 8 channels.

Initial hits were usually refined by manual setups in a 24-well format. Commercial additive screens were routinely tested to improve crystallization conditions.

2.2.4.3 Crystallization and diffraction data collection

Crystallization of $Brr2^{HR}$ and $hBrr2^{HR,S1087L}$ was carried out at 20ºC using the sitting drop vapor diffusion method. Crystals were obtained by mixing 1 µl of protein solution at 10 mg/ml with 1 µl of reservoir solution (0.1 M sodium acetate or sodium citrate, pH 4.6 or pH 5.0, 1.2 M sodium malonate) and optimized by microseeding and addition of a cocktail of additives (Silver Bullets condition 12; Hampton Research). The crystals were cryo-protected by transfer into a solution containing 0.1 M sodium acetate or sodium citrate, pH 4.6 or pH 5.0, 3.0 M sodium malonate and 0.1 M sodium chloride and then flash-cooled in liquid nitrogen.

Co-crystallization with nucleotides failed due to the high salt concentration required for $hBrr2^{HR}$ crystallization. To soak nucleotides into the crystals, $hBrr2^{HR}$ crystals were stabilized by cross-linking (Lusty, 1999) and transferred for 30 min at 20ºC into a fresh 2 µl drop containing a low salt soaking buffer (0.1 M sodium acetate or sodium citrate, pH 4.6 or pH 5.0, 0.1 M sodium malonate, 10 mM $MgCl_2$, 25 mM nucleotide). The soaked crystals were

cryo-protected by transfer into soaking buffer plus 30% (v/v) glycerol or ethylene glycol and then flash-cooled in liquid nitrogen. Diffraction data were collected at beamline 14.2 of BESSY II (HZB, Berlin, Germany), beamline PXII of SLS (Paul Scherrer Institute, Villigen, Switzerland) and beamline P14 of PETRA III (DESY, Hamburg, Germany) and processed with XDS (Kabsch, 2010) and HKL2000 (Minor *et al.*, 2006).

2.2.4.4 Structure solution, model building and refinement

The structure of hBrr2^{HR,S1087L} was solved by multiple isomorphous replacement with anomalous scattering (MIRAS). Samarium and tantalum derivatives were prepared by soaking crystals in mother liquor containing 0.3 mM samarium chloride (Hampton Research) or 5 mM tantalum bromide (Jena Bioscicence) for 12 h at 20ºC. Derivatized crystals were cryo-protected by transfer into a solution containing 0.1 M sodium acetate or sodium citrate, pH 4.6 or pH 5.0, 3.0 M sodium malonate, 0.1 M sodium chloride, and flash-cooled in liquid nitrogen. A bromide derivative was prepared by soaking a crystal for 1 min in cryo-buffer supplemented with 1 M sodium bromide and flash-cooling in liquid nitrogen.

Samarium sites were located and initial phases were calculated using the SHELX program suite (Sheldrick, 2008). Initial phases were used to locate tantalum and bromide sites by difference Fourier analyses. MIRAS phases were calculated and refined using SHARP (Bricogne *et al.*, 2003) and improved by solvent flattening with DM (Cowtan, 1994). Model building was done using COOT (Emsley and Cowtan, 2004; Emsley *et al.*, 2010) and refinement using REFMAC5 (Vagin *et al.*, 2004). To verify the chain tracing, a highly redundant data set was collected from a native crystal at 2.071 Å X-ray wavelength and used with combined model and experimental phases to calculate an anomalous difference Fourier map, which revealed the position of sulfur atoms in cysteine and methionine side chains. The structure of hBrr2^{HR} and nucleotide-bound structures were solved with Molrep (Vagin and Teplyakov, 1997) using the coordinates of the hBrr2^{HR,S1087L} structure as a search model and refined using REFMAC5, including TLS refinement (Winn *et al.*, 2003), with manual model building in COOT.

3. Results

3.1 Expression of Prp8, Snu114 and Brr2

Prp8, Brr2 and Snu114 form a salt-stable complex that can be detached from the U5 snRNP (Achsel *et al.*, 1998), suggesting that these proteins are organized as a functional unit, which is essential for spliceosome catalytic activation and disassembly. Prp8 is the largest and most highly conserved protein in the spliceosome (Fig 3.1A). Sequence and structural similarities between Prp8 and other protein domains indicate the presence of a putative bromodomain (Br) in the N-terminal region of Prp8 which is preceded by a proline rich region (P) and a nuclear localization signal (NLS) (Dlakić and Mushegian, 2011). Additionally, the N-terminus of Prp8 contains a Snu114-Cwc21-interacting domain (SCwid) (Grainger *et al.*, 2009). The central conserved domain of Prp8 is related to the catalytic domain of reverse transcriptases (RT) (Dlakić and Mushegian, 2011). The RT domain is followed by a functionally and structurally uncharacterized sequence region, which may be functionally equivalent to maturase-specific X/thumb (Th/X) domain of retroelements (Dlakić and Mushegian, 2011). Finally, the C-terminal domain displays RNase H-like and Jab1/MPN folds. The latter domains are presently the only experimentally characterized folds of Prp8 (Pena *et al.*, 2007, 2008; Ritchie *et al.*, 2008; Zhang *et al.*, 2007).

Snu114 is a large G-protein that bears close resemblance to the eukaryotic ribosomal elongation factor 2 (eEF2) (Bartels *et al.*, 2002). Aside from an N-terminal acidic domain, which is predicted to be intrinsically unstructured, Snu114 exhibits the same domain structure as eEF2, including the G-domain (Fig. 3.1B). It seems that the GTP/GDP state of Snu114 regulates the unwindase activity of Brr2 (Small *et al.*, 2006). Based on this finding, it has been suggested that Snu114 functions as a classical signalling G protein transducing signals to Brr2 to control spliceosome dynamics. Presently, it is not known by which mechanisms this regulation occurs and whether it requires direct involvement of Prp8. However, considering the similarity with the ribosomal translocase, it is also expect that Snu114 may additionally provoke translocation-like events during splicing, e.g. during restructuring of the active site between the two transesterification steps (Wahl *et al.*, 2009).

The Brr2 helicase is a U5 snRNP protein, which is stably associated with the spliceosome during catalytic activation and splicing catalysis. Brr2 seems to be required twice during each splicing event; first, during catalytic activation (Raghunathan and Guthrie, 1998; Laggerbauer

et al., 1998) and second, during spliceosome disassembly (Small *et al.*, 2006). Sequence analyses indicate that Brr2 is composed of a mostly unstructured N-terminal domain followed by a tandem repeat of helicase cassettes. Each cassette is predicted to be comprised of dual RecA-like domains linked *via* a WH domain to a Sec63 homology unit (Pena *et al.*, 2009). Due to the structural similarities shared between the Sec63 domains of Brr2 and the DNA helicase Hel308, it has been proposed that Brr2 consists of two expanded Hel308-like helicase cassettes (Zhang *et al.*, 2009; Pena *et al.*, 2009) (Fig. 3.1C). Only the N-terminal cassette of Brr2 appears to be catalytically active (Kim and Rossi, 1999) while the C-terminal cassette may comprise a versatile protein-protein interaction device (Liu *et al.*, 2006; van Nues and Beggs, 2001).

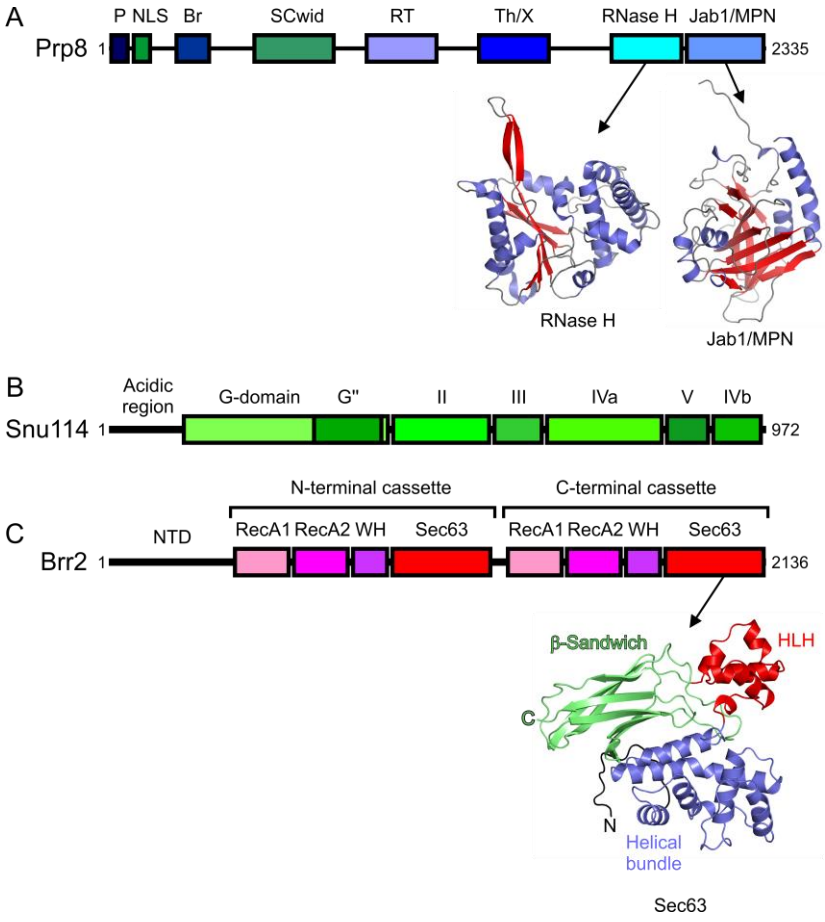


Fig 3.1: Domain arrangement of Prp8 (A), Snu114 (B) and Brr2 (C). Numbers indicate start and end amino acids of the human proteins. Domains are indicated as coloured boxes and labeled above the schemes. Prp8 – blue colors (P – proline rich region; NLS – nuclear localization signal; Br – bromodomain; SCwid – Snu114/Cwc21-interacting domain; RT – reverse transcriptase domain; Th/X – maturase-specific X/thumb domain; RNase H – RNase H-like domain; Jab1/MPN – Jab1/MPN-like domain). Snu114 – green colors (G" – Snu114/eEF-2-specific insertion in the G-domain; II-V – domains II-V; domain IV is divided in two parts, IVa and IVb). Brr2 – red colors (NTD – N-terminal domain; RecA – RecA-like domains; WH – winged helix domain; Sec63 – Sec63 homology region). The two tandem cassettes of Brr2 are indicated above the scheme. The proteins are not drawn to scale relative to each other. Below the Prp8 and Brr2 schemes, experimental structures of domains are depicted (RNase H and Jab1/MPN in Prp8; C-terminal Sec63 in Brr2).

Multiple interactions have been reported among distinct domains of Prp8, Brr2 and Snu114 (Liu *et al.*, 2006). The C-terminus of Prp8 was shown to interact with Brr2. Consistent with this finding, a Prp8 fragment encompassing the RNase H and Jab1/MPN domains was shown to couple the ATPase activity of Brr2 to its RNA unwinding activity and to stimulate Brr2 helicase activity by increasing the association rate of Brr2 with its RNA substrate (Pena *et al.*, 2009; Maeder *et al.*, 2009). The N- and C-terminal regions of Prp8 also seem to interact with Snu114. Supporting previous Y2H results, an intramolecular fold, denoted as a "Snu114-Cwc21 interacting domain" (SCwid), at the N-terminus of Prp8 has been identified (Grainger *et al.*, 2009). The documented interaction network among these factors is consistent with the strong interaction reported among Prp8, Snu114 and Brr2 in native complexes (Achsel *et al.*, 1998).

The production of significant amounts of pure samples suitable for crystallization is one of the most time consuming steps during high resolution structure-based studies. Since all three proteins of interest are large nuclear factors, they are notoriously difficult to express in heterologous systems. A complex containing these three proteins can be natively purified from HeLa and yeast nuclei. However, the material is non-homogeneous and not amenable to crystallization. Therefore, for further structural and functional investigations, we set out to design strategies for co-expression of these three proteins.

3.1.2 Expression of the human Prp8-Snu114-Brr2 complex

Considering that Prp8, Snu114 and Brr2 are among the largest proteins in the spliceosome, which additionally may require post-translational modifications, we had to devise a strategy to express these U5 snRNP proteins in a eukaryotic system. Therefore, we established a baculovirus-based expression facility using insect cells in order to produce this ternary complex as well as its individual components. Since we mainly pursued the structural investigation of these three factors individually or as complexes by means of X-ray crystallography, we worked with expression-optimized synthetic genes and looked in parallel at orthologs from several organisms in order to optimize chances for success.

In order to co-express the human Prp8-Snu114-Brr2 complex, we have cloned hPrp8 in a Donor plasmid (pIDS), hBrr2 with a C-terminal His_6-tag in an Acceptor plasmid (pFL) and hSnu114 with a C-terminal strepII-tag in another Donor plasmid (pIDK). The plasmids containing the proper insert were fused by CreLox recombination. Subsequent to the Cre-mediated fusion of Acceptor and Donor plasmids, we performed *in vitro* Tn7 transposition into EMBacY. The recombinant bacmids were isolated and used to transfect Sf9 cells. The

virus amplification was performed in Sf21 cells. Monitoring of the cell expression showed that hBrr2 and hSnu114 were well produced while hPrp8 was produced in only minute amounts (Fig. 3.2).

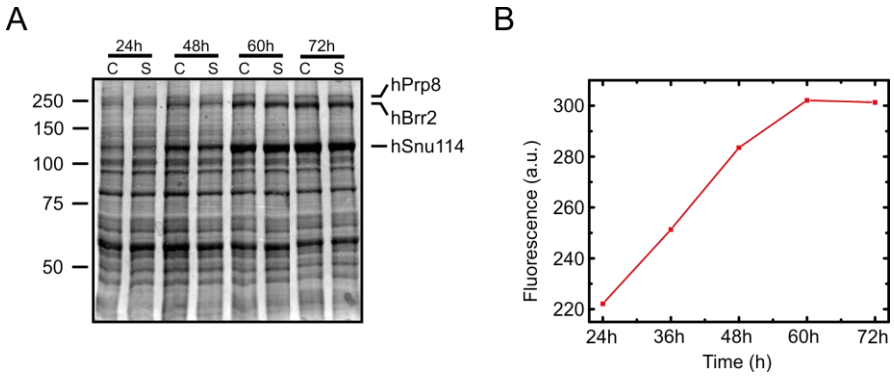


Fig. 3.2: Production of the human ternary complex in insect cells. (A) Coomassie-stained SDS gel showing a time course (from 24h to 72h) of the expression of human full-length Prp8-Snu114-Brr2 complex using the MultiBac system. The sizes in kDa of the molecular weight marker are given on the left. The abbreviations "C" stands for whole cell extract and "S" – for supernatant. (B) By following the YFP fluorescence signal, which depends on YFP expression, we could follow heterologous protein production levels. When YFP expression reaches a plateau, the expression of heterologous proteins under the same promoter as YFP also typically reaches their peak production.

We next tried to purify the three proteins as a complex through the His_6-tag present at the C-terminus of hBrr2. However, only hBrr2 was efficiently bound to the Ni^{2+}-NTA beads and could be eluted in the presence of 250 mM imidazole. We further performed size exclusion chromatography where hBrr2 eluted as a single, symmetric peak and migrated as a monomer (Fig. 3.3B). We additionally tried to purify the complex through a Strep-Tactin column taking advantage of the StrepII-tag present at the C-terminus of hSnu114. Unfortunately, the complex could not be isolated using this strategy since hSnu114 did not bind efficiently to the column.

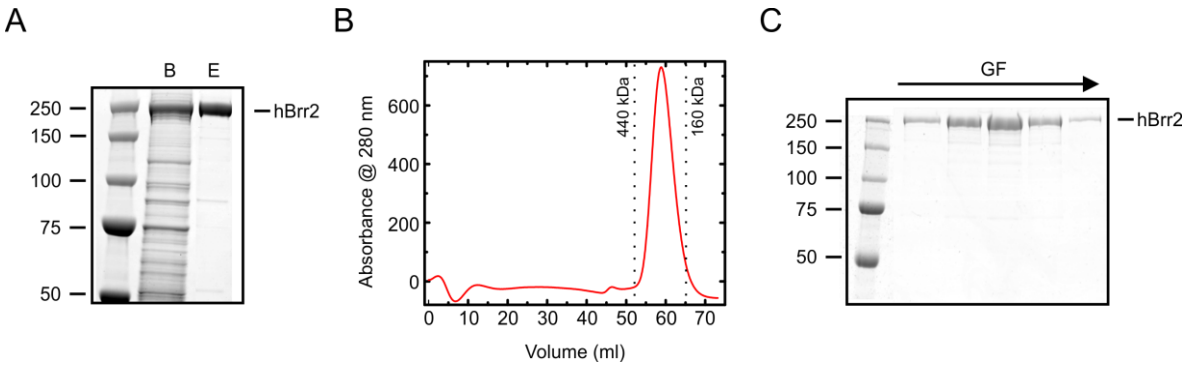


Fig. 3.3: Human ternary complex purification trial. (A) Coomassie-stained SDS gels showing the purification of full-length hBrr2, which was the only protein of the human ternary complex that could be isolated. The molecular weight marker is displayed on the left (sizes are in kDa); B – protein bound to Ni^{2+}-NTA beads and E – elution from a Ni^{2+}-NTA beads. (B) Chromatogram showing the migration profile of hBrr2 on a Superdex 200 16/60. The retention volume of commercial molecular weight standards is indicated (C) Coomassie-stained SDS

gel of a gel filtration run of hBrr2 on a Superdex 200 column.

3.1.3 Expression of the yeast Prp8-Snu114 complex

Our previous results indicated that hBrr2 was not stably bound to hPrp8 and hSnu114 when co-produced in insect cells. We presently do not know if the C-terminal tags on hBrr2 and hSnu114 might have affected the interaction among the three factors. Considering that the interaction between hPrp8 and hSnu114 is particularly strong and this binary complex remains completely intact even in the presence of high concentrations of chaotropic agents (0.4 M sodium thiocyanate) (Achsel *et al.*, 1998), we have decided to test the co-production of yPrp8 with ySnu114 in insect cells. yPrp8 was cloned in an Acceptor plasmid (pFL) and ySnu114 in a Donor plasmid (pIDS) and both plasmids were again fused by Cre-lox recombination. We made use of a FLAG-tag in the C-terminus of Prp8 in order to purify a stoichiometric complex since previous results obtained with the human protein suggested that yPrp8 may be produced only in minor amounts. yPrp8-ySnu114 binary complex may subsequently be complemented with individually produced and purified yBrr2. The virus amplification was performed in Sf21 cells as before. The protein expression, monitored by SDS-PAGE at various times after infection, showed that the yPrp8-ySnu114 complex was expressed, although in small amounts. A large scale expression trial in High Five cells resulted in a clear production of the complex, although ySnu114 was produced in higher amounts than yPrp8 (Fig. 3.4).

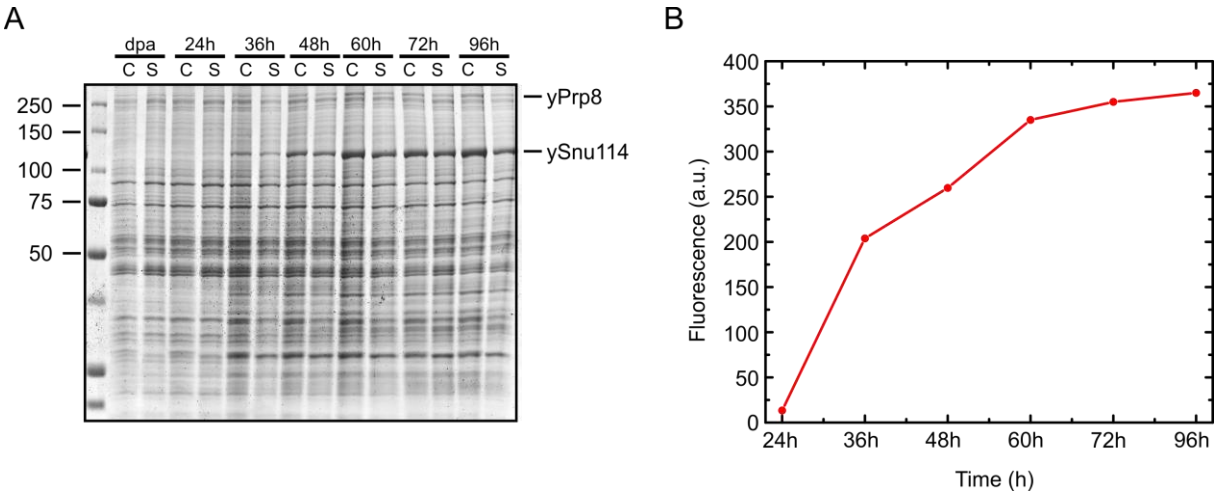


Fig. 3.4: Production of the yPrp8-ySnu114 complex in insect cells. (A) Coomassie-stained SDS gel showing a time course (from dpa to 96h; dpa – day of proliferation arrest) of the production of yeast full-length Prp8-Snu114 complex using the MultiBac system. The molecular weight marker is displayed on the left (sizes are in kDa). C – whole cell extract, S – soluble supernatant. (B) YFP fluorescence signal measurements were used to follow heterologous protein production levels.

The complex between yPrp8 and ySnu114 could be purified using Anti-FLAG beads (Fig. 3.5A and B), albeit in very small amounts. Most of the protein complex could not be eluted from the beads using competing FLAG peptides or by a change in pH. Nevertheless, with the recovered material, we were able to test the migration of the complex on a Superdex 200 size-exclusion column. We observed two peaks in the chromatogram, one at the void volume of the column containing the protein complex (Fig. 3.5C and D) and the other one at full column volume, containing the peptide used for elution from the antibody resin, confirmed by mass spectrometry analysis (data no shown). No peaks corresponding to the individual proteins were observed.

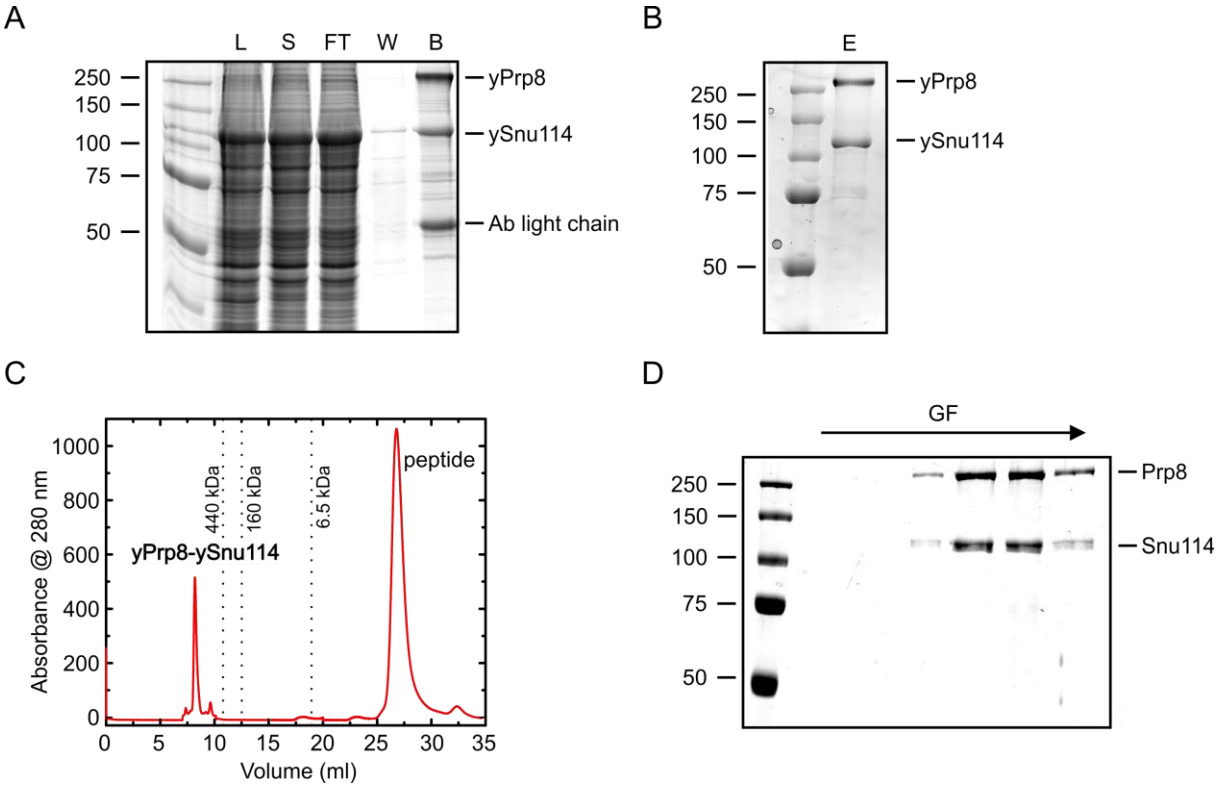


Fig 3.5: Purification trial of the yPrp8-ySnu114 complex. (A, B) Coomassie-stained SDS gels showing the purification of full-length yPrp8-ySnu114 complex. The molecular weight marker is displayed on the left (sizes are in kDa). Cell lysate, L, was clarified by centrifugation and the resulting soluble fraction, S, was loaded on anti-FLAG antibody beads. The unbound sample was collected, FT, and the beads were washed (W) and further tested for bound proteins, B. The proteins were eluted in a single step, E. (C) Chromatogram showing the migration profile of yPrp8-ySnu114 complex on a Superdex 200 10/300. (D) The protein complex was purified on a Superdex 200 10/300 column. The peak fraction at the void volume of the column (~ 7.5 ml) corresponds to the protein complex which was analyzed by SDS-PAGE.

We, therefore, concluded that these two proteins form a stable complex which can be purified for further investigation. However, major improvements in this system are necessary in order to achieve stoichiometric expression of the components which could increase the yield of the purified complex. Additionally, it is necessary to devise an optimized strategy to

purify a homogenous complex suitable for crystallization by selecting proper tags and avoiding major loss of protein on the beads.

3.2 Structural analysis of the Ski2-like helicase Brr2

Our initial strategy of co-expressing Prp8, Snu114 and Brr2 as a complex failed, since we were unable either to isolate the human complex or purify homogeneous material in amounts suitable for crystallization in the case of the yeast complex. Nevertheless, we have successfully isolated human Brr2 during the purification trials of the human ternary complex (Fig. 3.3).

The full length yeast Brr2 (ca. 250 kDa) has been previously expressed in yeast cell culture, but in insufficient amounts for X-ray crystallography studies. Additionally, only a small fragment of the yeast enzyme (the C-terminal Sec63 unit) has been expressed in *E. coli* and purified in large scale for structural studies (Zhang *et al.*, 2009; Pena *et al.*, 2009). Therefore, relying on our previous achievement in isolating hBrr2, we aimed at producing highly purified and homogeneous human and yeast Brr2 for further structural and functional investigations.

3.2.1 Expression and purification of human and yeast Brr2

The full-length human and yeast Brr2 were initially cloned with a C-terminal His_6-tag and N-terminal His_{10}-tag, respectively, in an Acceptor plasmid, pFL, and produced in large amounts in insect cell culture using the MultiBac system (Fig. 3.6A and B). Both proteins could be purified to near homogeneity using Ni^{2+}-NTA affinity column followed by an anion exchange step. yBrr2 and hBrr2 migrated in size exclusion chromatography as monomers (Fig. 3.6C, D, E). The U4/U6 unwinding activities of the human and yeast enzymes were assessed in collaboration with Sina Mozaffari-Jovin (MPI-BPC). Human and yeast Brr2 were active in ATP-dependent U4/U6 duplex unwinding (Fig. 3.6F) but failed to crystallize.

In order to remove putatively flexible regions that may hinder crystallization, we treated hBrr2 and yBrr2 with proteases, several of which gave rise to a stable ca. 200 kDa fragment (Fig. 3.7). Mass spectrometric fingerprinting and N-terminal micro-sequencing of hBrr2 samples showed that chymotrypsin yielded a fragment whose N-terminus coincided with the predicted start of the first RecA domain (residue 458), while subtilisin left about 60 additional N-terminal residues (start residue 395). For yBrr2, elastase yielded a large fragment encompassing the PWI domain while the chymotrypsin fragment contained 130 residues before the predicted start of the first RecA domain (Fig. 3.7). The cleavage pattern observed

in the proteolysis experiments indicate that the N-terminal extensions of human and yeast Brr2 may show slight structural differences. The regions encompassing the two helicase cassettes remained intact even upon prolonged protease treatment for both Brr2 proteins.

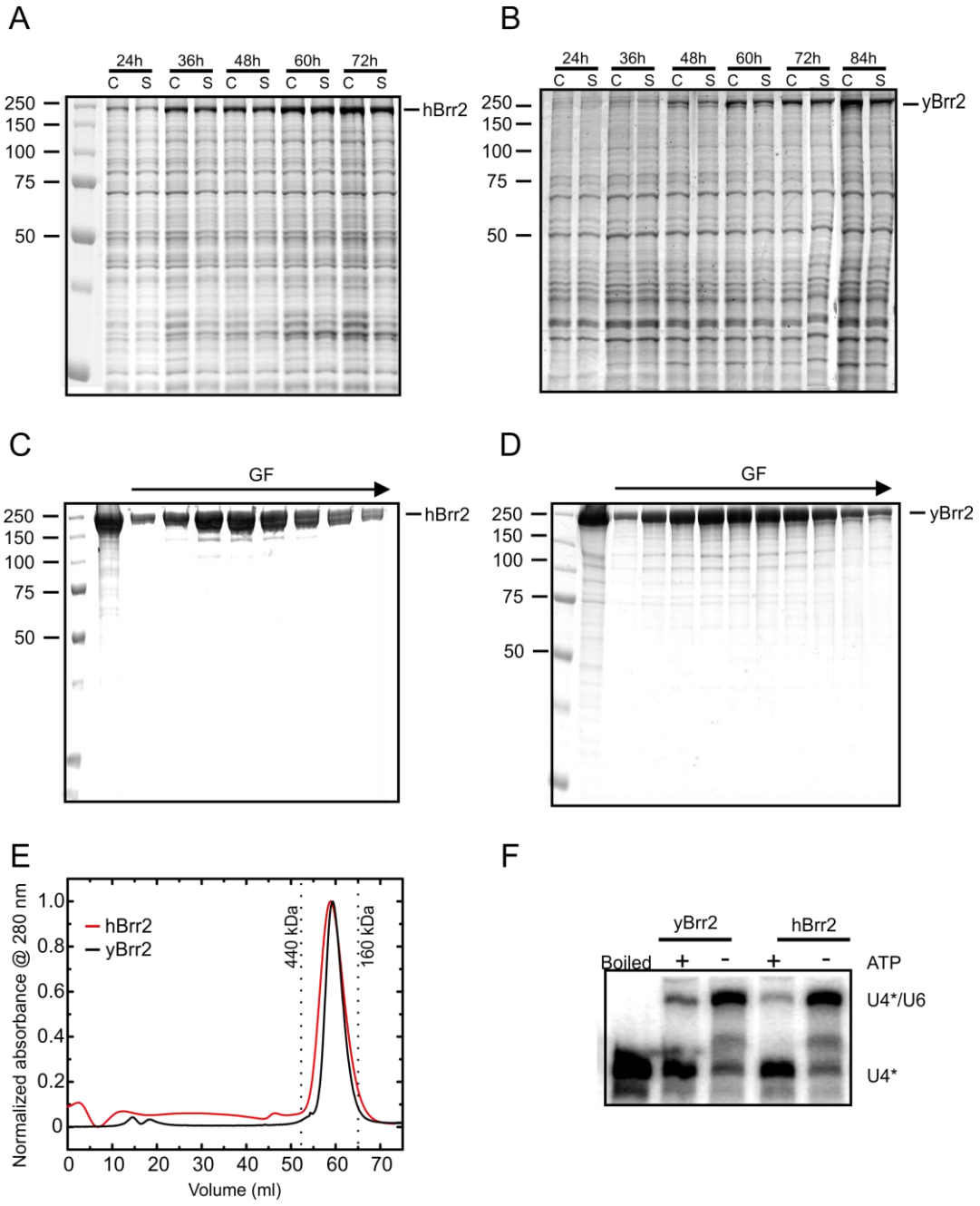


Fig. 3.6: Production and purification of h and yBrr2. Coomassie-stained SDS gels showing expression (A and B) and purification (C and D) of full-length human and yeast Brr2. The molecular weight marker is shown on the left (sizes in kDa). Both yeast and human Brr2 were solubly expressed in insect cells and purified using Ni^{2+}-NTA affinity followed by an anion exchange and size exclusion chromatography. C – whole cell extract, S – supernatant. (E) As a final purification step, yeast and human proteins were loaded on a Superdex 200 16/60 size exclusion column. We concluded that both proteins migrated as monomers by comparison with the retention volume of commercial molecular weight standards. (F) Native gel analysis of U4/U6 unwinding. 0.25 nM of *in vitro* transcribed, annealed and gel-purified U4/U6 di-snRNA (U4* – [^{32}P]-labeled U4 snRNA) were treated with 25 nM recombinant protein (indicated) at 40°C for 15 minutes in the presence or absence of Mg^{2+}-ATP.

We designed and screened for production a large number of fragments based on the proteolysis experiments. Based of their efficient production, we subsequently mainly focused on producing hBrr2 fragments. Furthermore, we gave priority to the human enzyme considering its medical relevance. We expressed and purified six truncated proteins, differently tagged at the N- or C-terminus, corresponding to the protease-resistant portions of hBrr2 (residues 395-2136; 395-2129; 458-2136 and 458-2129) with and without a short C-terminal peptide that lacked electron density in the crystal structure of the hBrr2 Sec63 unit (PDB ID 2Q0Z) (Fig. 3.8).

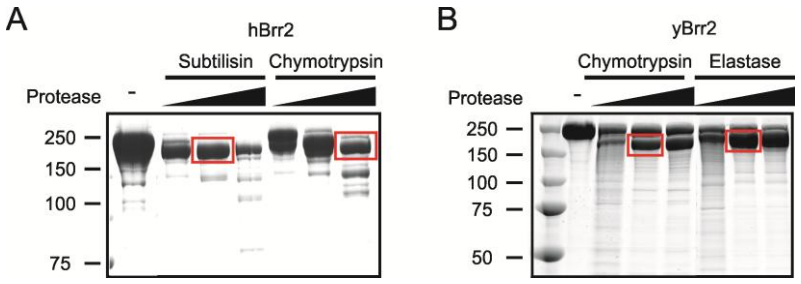


Fig. 3.7: Experimental definition of yeast and human Brr2 stable fragments. (A) Limited proteolysis of full-length hBrr2 and (B) yBrr2. Stable, ca. 200 kDa fragments obtained with subtilisin and chymotrypsin for hBrr2 and chymotrypsin and elastase for yBrr2 are boxed.

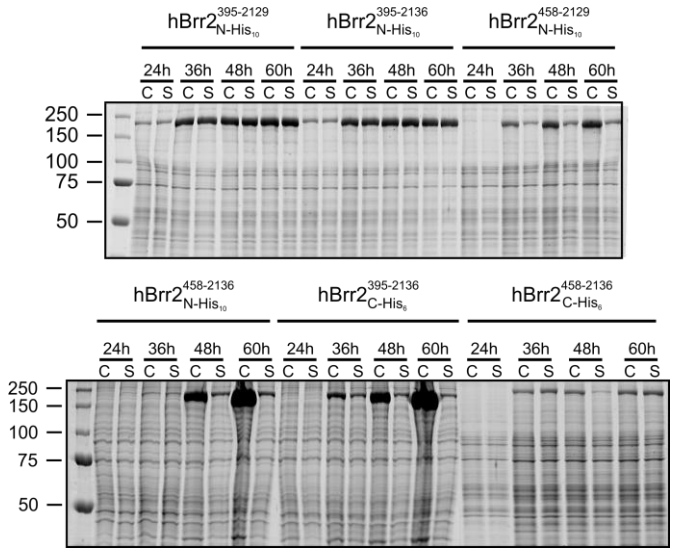


Fig. 3.8: Coomassie-stained SDS gels showing an expression time course of the various hBrr2 constructs in insect cells. The molecular weight marker is shown on the left (sizes in kDa). Although all constructs were well produced, only two constructs, hBrr2$^{395\text{-}2129}$ and hBrr2$^{395\text{-}2136}$ with N-terminal His_{10} tags, were completely soluble and yielded high amounts of purified protein. C – whole cell extract, S – supernatant.

While removal of the last seven residues had no effect on the helicase activity, deletion of the ca. 60 residues preceding the first RecA domain led to a severe drop in duplex unwinding (Fig. 3.9).

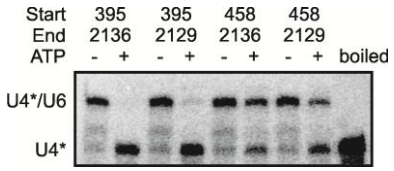


Fig. 3.9: Helicase activity of hBrr2 fragments. Borders of the fragments analyzed are shown above the gel. Running positions of the U4/U6 duplex (U4* – U4 labeled) and U4 snRNA are indicated.

We thus focused in further work on the hBrr$^{395\text{-}2129}$ fragment, henceforth referred to as hBrr2 "helicase region" (hBrr2^{HR}). hBrr2^{HR} was among the two most solubly produced constructs (Fig 3.8) and could be produced in large amounts in High Five cells (Fig. 3.10A). The MultiBac system indeed allowed us to routinely produce 10–15 mg of Brr2^{HR} per liter of insect cell culture. hBrr2^{HR} was purified to near homogeneity through Ni^{2+}-NTA chromatography, followed by an anion exchange and a size exclusion step (Fig. 3.10B, C, D).

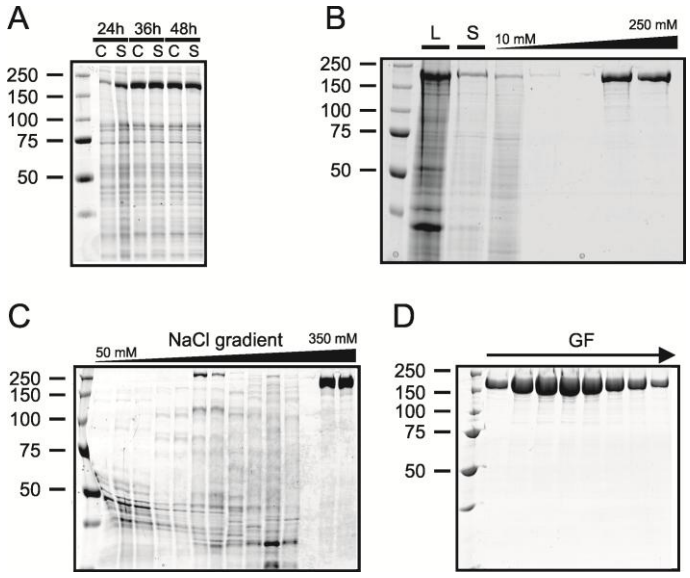


Fig. 3.10: Production and purification of Brr2^{HR}. (A) Coomassie-stained SDS gel showing an expression time course of hBrr2^{HR} in insect cells. The molecular weight marker is shown on the left (sizes in kDa). C – whole cell extract, S – supernatant. (B) Cell lysate, L, was clarified by centrifugation and the resulting soluble fraction, S, was loaded on a HisTrap FF column. The protein was eluted with a linear gradient from 10 to 250 mM imidazole. (C) The protein was transferred to low salt buffer and loaded on a MonoQ 10/100 GL column. hBrr2^{HR} was eluted with a linear 50 to 600 mM sodium chloride gradient and further purified by (D) gel filtration on a 26/60 Superdex 200 column. The peak fractions corresponding to the protein were analyzed by SDS-PAGE.

3.2.2 Structural analysis of an active, protease-resistant portion of hBrr2

We have obtained crystals of hBrr2^{HR} as well as of a Brr2 point mutant linked to the RP disease. As crystals for the mutant protein were first available and diffracted to 2.65 Å resolution (Fig. 3.11A), we have solved its structure by multiple isomorphous replacement with anomalous scattering (MIRAS). $R_{work}/(R_{free})$ values were refined to 22.9%/(26.8%) with

good stereochemistry (Table 3.1). The structures of hBrr2^{HR} in the absence and presence of nucleotides were solved by molecular replacement. Residues 403–2125 of hBrr2^{HR} could be fully traced with only a few exposed loop regions exhibiting weaker than average electron density. Chain tracing was verified using the anomalous scattering of sulfur atoms, which revealed the positions of the vast majority of cysteine and methionine side chains (Fig. 3.11 B, C).

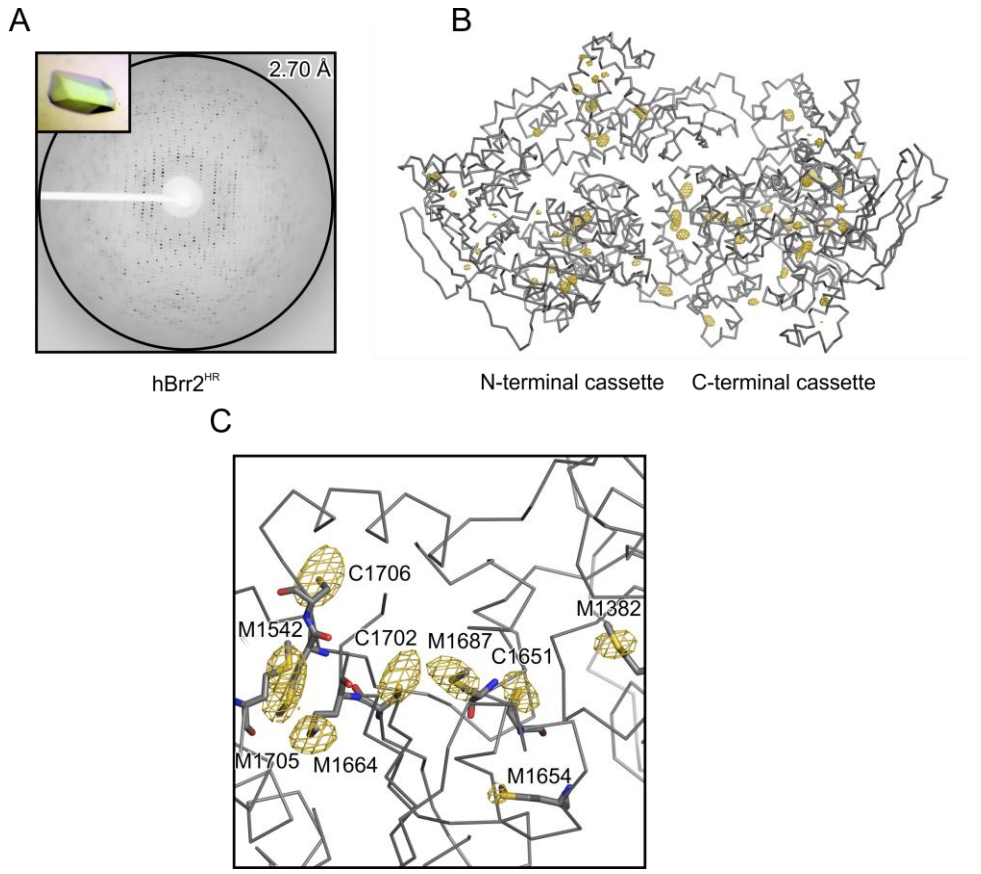


Fig. 3.11: Structural model. (A) Crystal and a diffraction image of hBrr2^{HR}. (B) Anomalous difference Fourier map contoured at the 3σ level (golden mesh) superimposed on a Cα-trace of hBrr2^{HR,S1087L} (gray). The map was calculated using the anomalous differences collected on a native crystal at an X-ray wavelength of 2.071 Å and phases obtained from the final model. (C) Close-up view.

Table 3.1. Crystallographic data.

Data collection								
Dataset[a]	**S1087L-1**	**S1087L-2**	**S1087L-3**[b]	**NaBr**[c]	**$SmCl_3$**[c]	**Ta_6Br_{14}**[b]	**wt**	**Mg-ATP**	**Mn-ATP**[c]
Wavelength (Å)	0.9999	0.9184	2.071	0.9199	1.8457	1.2546	1.1270	0.9184	1.8536
Space Group	C2	C2	C2	C2	C2	C2	C2	C2	C2
Unit Cell (Å, °)									
a	145.6	146.2	145.8	146.1	146.2	146.7	142.0	147.1	146.8
b	149.2	149.5	148.8	148.7	147.2	151.2	150.5	154.6	154.8
c	141.0	141.3	140.9	141.2	140.3	142.0	143.7	143.3	143.2
ß	120.1	120.3	120.9	120.4	120.4	120.1	118.4	120.6	120.7
Resolution (Å)[d]	50.0-3.1 (3.20-3.10)	50.0-2.65 (2.70-2.65)	50.0-4.0 (4.1-4.0)	50.0-2.96 (3.07-2.96)	50.0-4.0 (4.07-4.00)	50.0-3.3 (3.40-3.30)	50.0-2.69 (2.84-2.69)	50.0-2.92 (3.00-2.92)	50.0-3.5 (3.56-3.50)
Reflections									
Unique	46827 (4192)	74403 (3706)	42927 (3168)	53710 (5305)	21905 (2169)	71292 (11031)	72077 (9976)	57617 (2884)	34425 (1759)
Completeness (%)	98.7 (97.5)	99.7 (99.7)	98.8 (99.1)	99.8 (99.4)	100 (100)	98.0 (96.0)	98.7 (93.9)	99.5 (100)	99.7 (100)
Redundancy	7.1 (6.8)	2.3 (2.3)	11.5 (11.2)	7.3 (6.8)	7.7 (7.6)	3.5 (3.4)	3.6 (3.1)	3.9 (3.8)	3.7 (3.7)
R_{sym}[e]	0.07 (0.69)	0.042 (0.549)	0.081 (0.310)	0.13 (0.86)	0.14 (0.79)	0.06 (0.87)	0.08 (0.76)	0.035 (0.524)	0.045 (0.682)
I/σ	17.5 (3.2)	14.8 (1.1)	26.7 (7.8)	14.0 (2.1)	19.8 (2.4)	11.7 (1.4)	13.8 (1.6)	24.5 (1.5)	16.0 (1.4)

Phasing				
Dataset	**S1087L-1**	**NaBr**	**$SmCl_3$**	**Ta_6Br_{14}**
No. Sites		12	12	1
Phasing Power[f]				
iso (acentric)		0.52	0.46	1.15
iso (centric)		0.40	0.35	0.81
ano (acentric)		0.41	0.54	0.89
R_{Cullis}[g]				
iso (acentric)		0.81	0.93	0.58
iso (centric)		0.80	0.92	0.59
ano (acentric)		0.98	0.97	0.89
FOM[h]				
Before DM	0.31			
After DM	0.72			

Refinement			
Dataset	**S1087L-2**	**Wt**	**Mg-ATP**
Resolution (Å)	50.0-2.66 (2.73-2.66)	50.0-2.7 (2.77-2.70)	50.0-2.92 (3.00-2.92)
Reflections			
Number	70636 (4814)	68192 (4564)	55588 (3932)
Completeness (%)	99.2 (93.5)	98.7 (89.0)	99.7 (100)
Test set (%)	5	5	5
R-factors[j]			
R_{work}	22.8 (34.2)	24.2 (36.0)	23.0 (33.2)
R_{free}	26.7 (36.1)	27.8 (39.6)	27.0 (35.2)
Ramachandran Plot			
Favored	99.2	97.9	98.6
Outlier	0.8	2.1	1.4
Rmsd[k] **geometry**			
Bonds (Å)	0.009	0.010	0.007
Angles (°)	1.25	1.30	1.15
PDB ID	4F92	4F91	4F93

[a] Datasets: **S1087L-1** – $hBrr2^{HR,S1087L}$ native for phasing; **S1087L-2** – $hBrr2^{HR,S1087L}$ native for refinement; **S1087L-3** – $hBrr2^{HR,S1087L}$ native for sulfur anomalous signal; **NaBr** – $hBrr2^{HR,S1087L}$ NaBr soak; **$SmCl_3$** – $hBrr2^{HR,S1087L}$ $SmCl_3$ soak; **Ta_6Br_{14}** – $hBrr2^{HR,S1087L}$ Ta_6Br_{14} soak; **wt** – $hBrr2^{HR,wt}$ for refinement; **Mg-ATP** – $hBrr2^{HR,S1087L}$ ATP/$MgCl_2$ soak for refinement; **Mn-ATP** – $hBrr2^{HR,S1087L}$ ATP/$MnCl_2$ soak for manganese anomalous signal.

[b] Processed with XDS, anomalous pairs counted as different reflections.

[c] Processed with HKL2000, anomalous pairs counted as one reflection.

[d] Values for the highest resolution shell in parentheses.

[e] $R_{sym}(I) = (\Sigma_{hkl}\Sigma_i[|I_i(hkl) - \langle I(hkl)\rangle|] / \Sigma_{hkl}\Sigma_i I_i(hkl)$, in which $I_i(hkl)$ – intensity of the i^{th} measurement of reflection hkl; $\langle I(hkl)\rangle$ – average value of the intensity of reflection hkl for all i measurements.

[f] Phasing power = $P = \Sigma_n|F_{H,calc}| / \Sigma_n|E|$, in which $|E| = |F_{PH,obs}| - |F_{PH,calc}|$ = mean lack of closure error; n – number of observed scattering amplitudes for the derivative; $F_{PH,obs}$, $F_{PH,calc}$ – observed and calculated structure factor amplitudes of the derivative; $F_{H,calc}$ – calculated structure factor amplitudes of the heavy atom substructure.

[g] $R_{Cullis} = \Sigma_{hkl}||F_{PH} \pm F_P| - F_{H,calc}| / \Sigma_{hkl}|F_{PH} \pm F_P|$; F_{PH}, F_P – observed structure factor amplitudes of the derivative, native; $F_{H,calc}$ – calculated structure factor amplitudes of the heavy atom substructure; "+" if signs of F_{PH} and F_P are equal, "-" if opposite.

[h] FOM = Figure of Merit = $m = |F(hkl)_{best}| / |F(hkl)|$, in which $\boldsymbol{F}(hkl)_{best} = \Sigma_\alpha [P(\alpha)\, \boldsymbol{F}_{hkl}(\alpha)] / \Sigma_\alpha P(\alpha)$; P – phasing power; α – phase angle.

[i] DM – Density modification

[j] $R_{work} = \Sigma_{hkl}[||F_{obs}| - k|F_{calc}||] / \Sigma_{hkl}|F_{obs}|$; $R_{free} = \Sigma_{hkl\in T}[||F_{obs}| - k|F_{calc}||] / \Sigma_{hkl\in T}|F_{obs}|$; $hkl\in T$ – test set; F_{obs}, F_{calc} – observed and calculated (from model) structure factor amplitudes.

[k] Rmsd - root-mean-square deviation.

The structure of hBrr2^{HR} is very compact and comprises two globular helicase cassettes (N-terminal: residues 463–1288; C-terminal: residues 1310–2125), which are closely associated with each other and bury ca. 1200 Å^2 of combined surface area at their interface (Fig. 3.12). Both cassettes exhibit similar overall architectures (Cα root-mean-square deviation [rmsd] of 2.5 Å; Fig. 3.12), in spite of only 27% sequence identity (Fig. 3.13). Relative to the N-terminal cassette, the C-terminal cassette is rotated by 115° and translated by 63 Å within hBrr2^{HR} (Fig. 3.12). Both cassettes comprise two prototypical RecA-like ATPase domains followed by a winged helix (WH), a seven helix bundle (HB), a helix-loop-helix (HLH) and an immunoglobulin-like (IG) domain (Fig. 3.12A, B). The latter three domains constitute the Sec63 homology region and resemble the structure of the isolated C-terminal Sec63 units from yeast (Zhang *et al.*, 2009; Pena *et al.*, 2009) and human Brr2 (PDB ID 2Q0Z). In each cassette, two RecA domains and a HB domain form the floor and the roof, respectively, of a central tunnel (Fig 3.14). The second RecA domain is connected through an extended peptide (residues 886-896/1721-1731 in the N-/C-terminal cassette) to a WH domain, which constitutes one side of the tunnel and fastens the first RecA domain to the HB domain. On the other side, the second RecA and the HB domain approach (N-terminal cassette) or contact (C-terminal cassette) each other. A prominent loop of the RecA-2 domain extends across the tunnel entry towards a long scaffolding helix of the HB domain (Fig. 3.14). Equivalents of these elements have been suggested to constitute a strand separation device and a ratchet, respectively, in other SF2 proteins (Büttner *et al.*, 2007). Although direct evidence for a ratcheting function of the HB scaffolding helix is missing, it has been shown to contact the substrate nucleic acid in related enzymes (Büttner *et al.*, 2007; Weir *et al.*, 2010). The HLH and IG domains lie on one flank of the HB domain, forming a similar triangular arrangement in both cassettes (Cα rmsd of 3.1 Å) as previously seen for the isolated C-terminal Sec63 units from yeast (Zhang *et al.*, 2009; Pena *et al.*, 2009) and human Brr2 (PDB ID 2Q0Z). In both Sec63 units, the IG domain is wedged between the HB and HLH domains (Fig. 3.12), which in turn do not directly contact each other except immediately around their linking peptide. Our structural findings suggest that the individual Brr2 cassettes resemble the complete SF2 DNA helicase Hel308 (Büttner *et al.*, 2007) expanded by an IG domain. As in Hel308, the circular domain arrangement in both hBrr2^{HR} cassettes leads to the clustering of conserved ATPase and helicase motifs, known to bind and hydrolyze nucleotide tri-phosphates and to bind nucleic acids (Fig. 3.15).

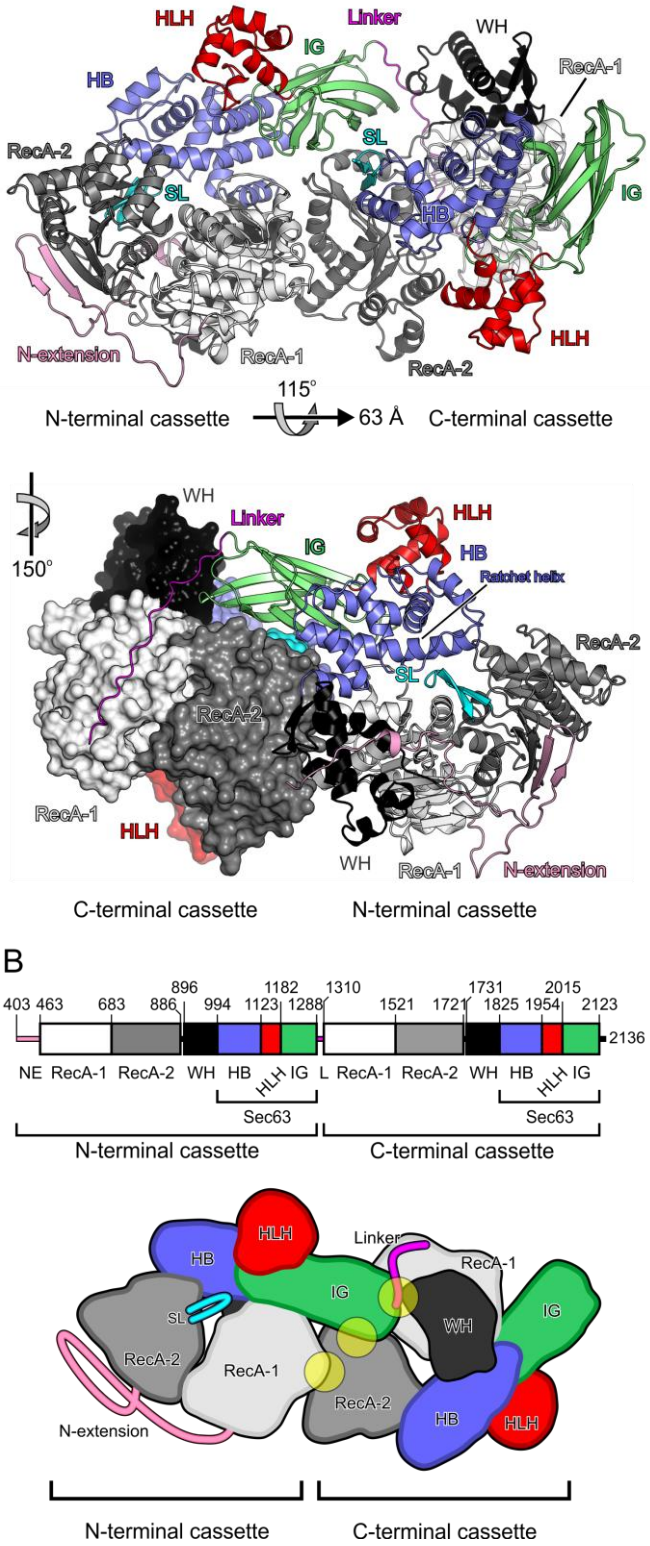
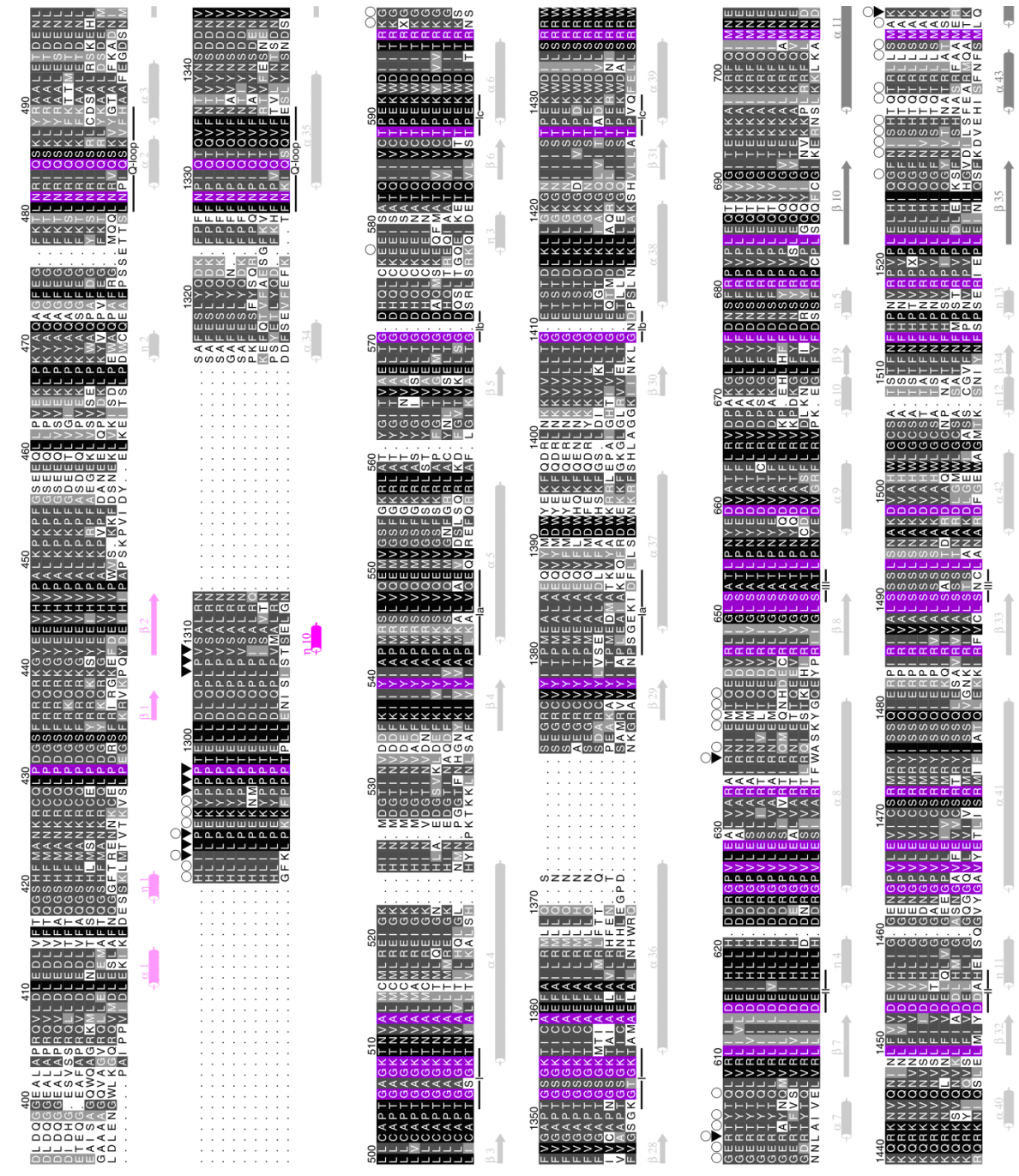
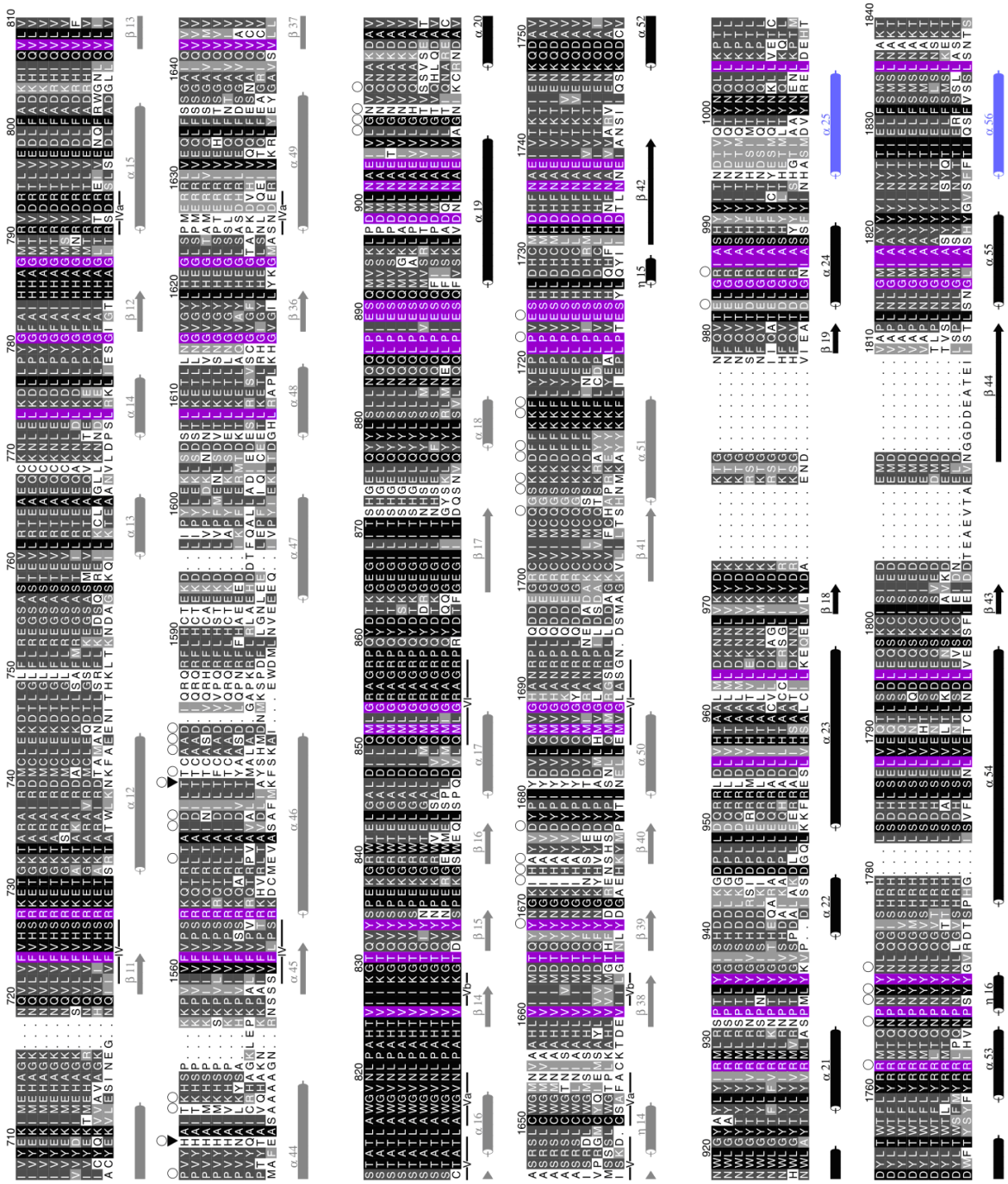


Fig. 3.12: Overall structure of hBrr2^{HR}. (*A*) Top: Ribbon plot of hBrr2^{HR}. Coloring by domains and functional elements: N-terminal extension – pink; RecA-1 – light gray; RecA-2 – dark gray; WH – black; HB – blue; HLH – red; IG – green; linker – magenta; separator loop (SL) – cyan. Rotation/translation symbols below the panel indicate the relationship between the cassettes within hBrr2^{HR}. Bottom: combined ribbon (N-terminal cassette) and surface plot (C-terminal cassette) showing the inter-cassette linker. The linker can be divided into three regions. The N-terminal part (region 1; residues 1289–1295) lies along a cleft between the N-terminal IG domain and the C-terminal WH domain and contacts the tip of the C-terminal RecA-1 domain. A central triple-Pro motif (region 2; residues 1296–1298) lacks interactions with any domain, while the C-terminal part (region 3; residues 1299–1309) runs snugly around the C-terminal RecA-1 domain. Rotated 150° counter-clockwise as indicated. (*B*) Schematic representations of Brr2^{HR}. Top: Domain borders. Bottom: 2D scheme of Brr2^{HR}. Inter-cassette contacts are indicated by semitransparent yellow circles.


N-terminal cassette
C-terminal cassette
N-terminal cassette
C-terminal cassette
N-terminal cassette
C-terminal cassette
H. sapiens
M. musculus
M. gallopavo
X. tropicalis
D. rerio
D. melanogaster
C. elegans
A. thaliana
S. cerevisiae
Helicase motif
Secondary structure


N-terminal cassette
C-terminal cassette
N-terminal cassette
C-terminal cassette
N-terminal cassette
C-terminal cassette
H. sapiens
M. musculus
M. gallopavo
X. tropicalis
D. rerio
D. melanogaster
C. elegans
A. thaliana
S. cerevisiae
Helicase motif
Secondary structure

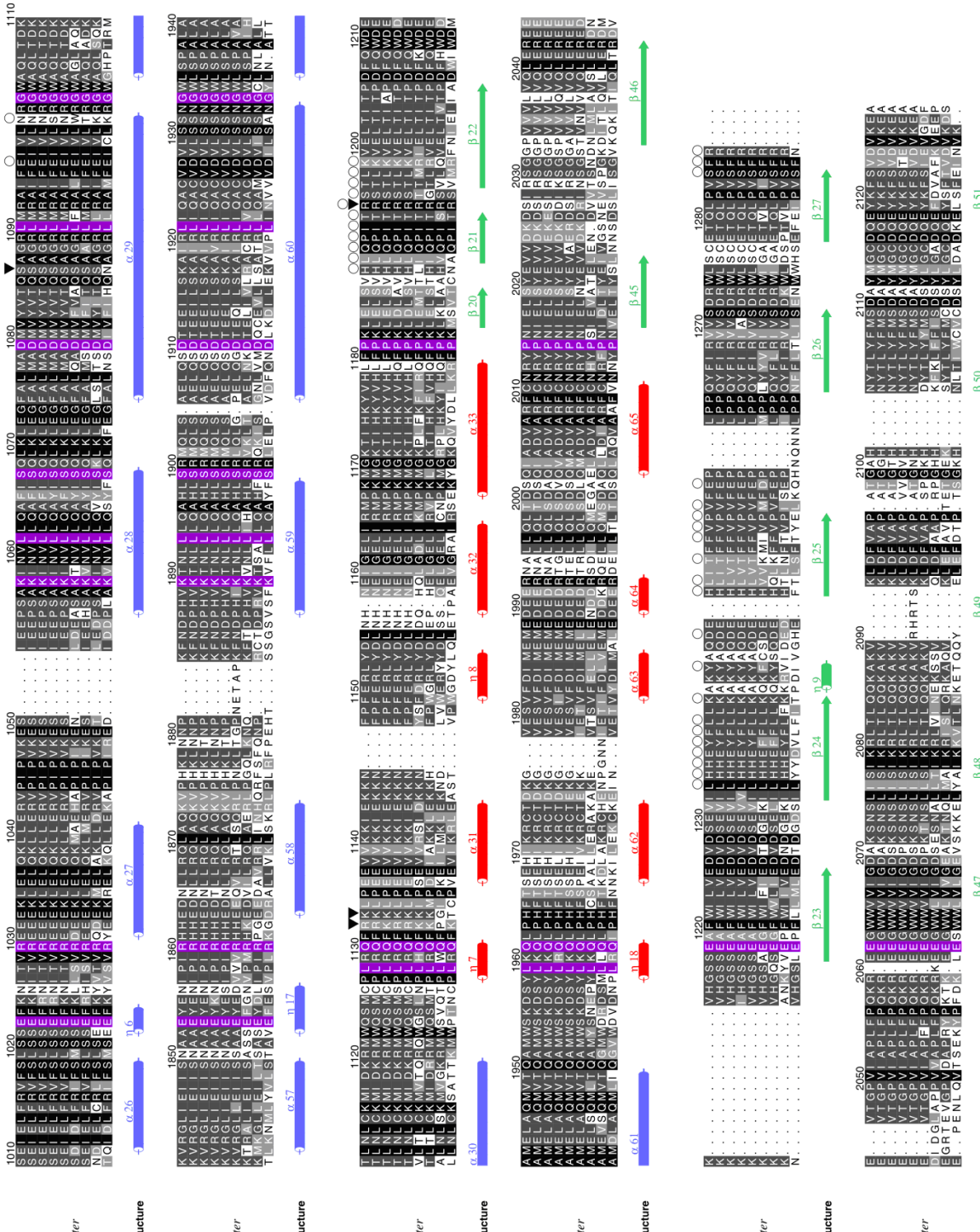

N-terminal cassette
C-terminal cassette
N-terminal cassette
C-terminal cassette
N-terminal cassette
C-terminal cassette
H. sapiens
M. musculus
M. gallopavo
X. tropicalis
D. rerio
D. melanogaster
C. elegans
A. thaliana
S. cerevisiae
Helicase motif
Secondary structure

Figure 3.13: Multiple sequence alignment of Brr2 orthologs. First block of sequences – N-terminal cassette, second block of sequences – C-terminal cassette. Residue numbers refer to the human Brr2 sequence. The cassettes have been aligned within each block and with respect to each other. The background coloring of the residues is according to the conservation within each cassette, darker background corresponding to a higher degree of conservation. Residues that are invariant across both cassettes are shown with a purple background. Secondary structure elements are indicated by icons and colored according to their domains (N-terminal extension – pink; RecA-1 – light gray, RecA-2 – dark gray, WH – black, HB – blue, HLH – red, IG – green, inter-cassette linker – magenta). ATPase/helicase motifs (Q and Roman numerals) are indicated below each block by a black line (Fairman-Williams *et al.*, 2010). Open circles denote residues involved in inter-cassette contacts. Filled triangles denote point mutations investigated herein. Organisms: *Homo sapiens*, *Mus musculus*, *Meleagris gallopavo*, *Xenopus tropicalis*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*.

In our hBrr2^{HR} structure, the N-terminal cassette is preceded by 60 conserved residues (Fig. 3.12A top) that adopt an irregular but well defined structure along conserved surfaces (2219 Å^2 buried), surrounding both RecA domains and the WH domain (Fig. 3.12A top). The reduction in helicase activity upon deletion of this N-terminal expansion (Fig. 3.9) shows that it may support a productive domain orientation or contact substrate RNA.

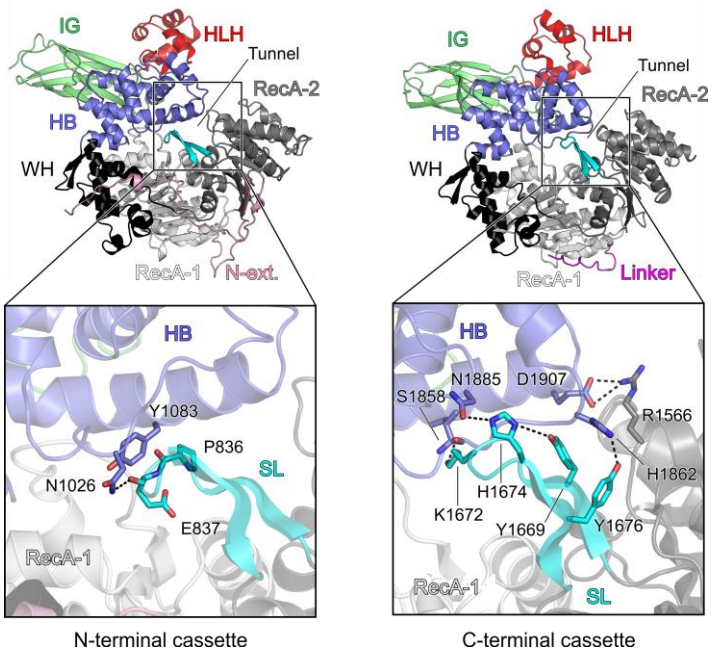


Fig. 3.14: Organization of the individual cassettes. Ribbon plots of N-terminal (left) and C-terminal (right) cassettes with expanded views on the domain closure and contacts between the separator loops and the ratchet helices. Domains and elements are colored as in Fig. 3.12. Interacting residues are shown as sticks and colored by atom type (carbon – as the respective structural element; nitrogen – blue; oxygen – red). Dashed lines – hydrogen bonds or salt bridges.

A conserved 20 residue linker (residues 1289–1309; Fig. 3.12B) that runs along an extended, highly conserved surface (996 Å2 buried; Fig. 3.12A bottom) covalently connects

the two cassettes. The linker can be divided into three regions. Its N-terminal part (region 1; residues 1289–1295) lies along a cleft between the N-terminal IG domain and the C-terminal WH domain and contacts the tip of the C-terminal RecA-1 domain. A central triple-Pro motif (region 2; residues 1296–1298) lacks interactions with any domain, while its C-terminal part (region 3; residues 1299–1309) runs snuggly around the C-terminal RecA-1 domain.

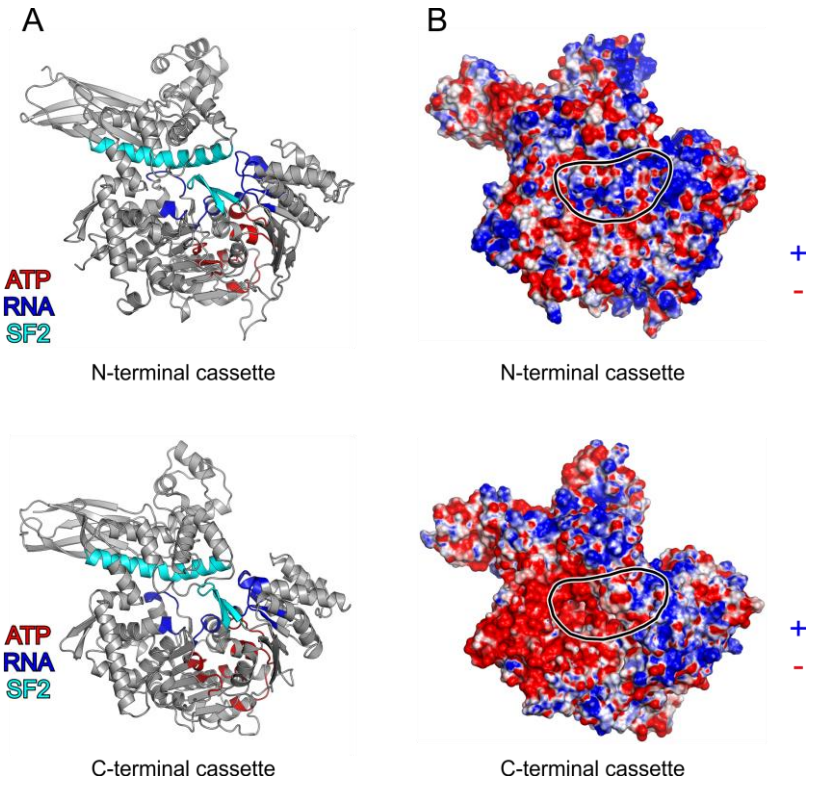


Fig. 3.15: Helicase motifs and surface electrostatics. (A) Ribbon plots of both cassettes showing the location of the helicase motifs: Red – motifs involved in nucleotide binding and hydrolysis; blue – motifs involved in RNA binding; cyan – SF2-specific motifs (ratchet helix and separator loop). The view of the N-terminal cassette (top) is the same as in Fig. 3.12, bottom. The C-terminal cassette is shown in an identical orientation. (B) Electrostatic surface potential mapped at the surface of the cassettes. Blue – positive charge; red – negative charge. Black outlines mark the central tunnels.

The cassette interface can be divided into two areas. The first involves the N-terminal IG domain fitted squarely between the second RecA domain and the WH domain of the C-terminal cassette (Fig. 3.12, 3.16). The second involves contacts between the RecA-1 and WH domains of the N-terminal cassette and the RecA-2 domain of the C-terminal cassette (Fig. 3.12, 3.16). The extensive contacts observed between the two cassettes suggest that they form a functional unit.

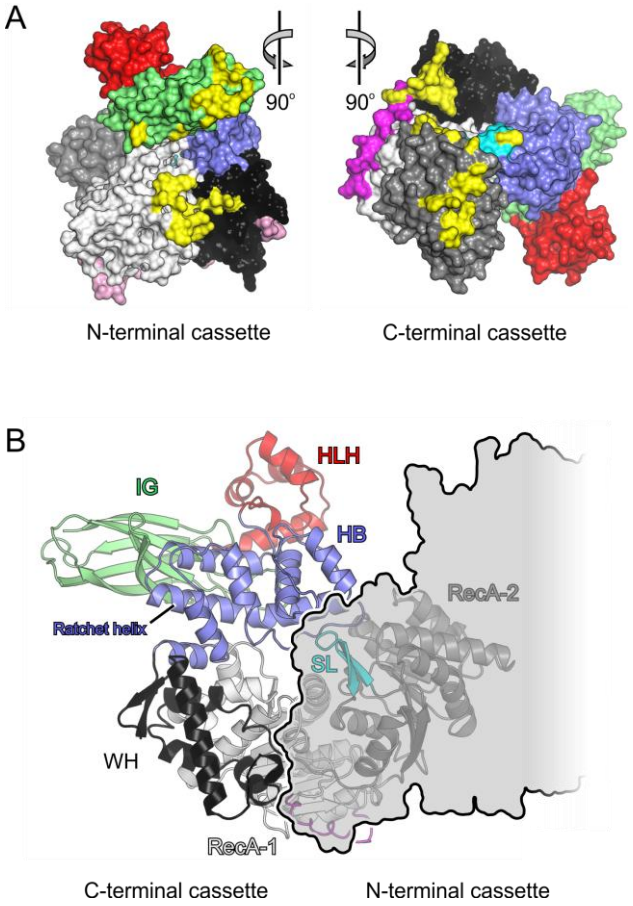


Fig. 3.16: Inter-cassette interactions. (A) Book view onto the interacting surfaces of the N-terminal (left) and C-terminal (right) cassettes. Domains are shown in surface representation and colored as in Fig. 3.12. Inter-cassette contact residues are colored in yellow. Interaction region I involves the N-terminal IG domain contacting the C-terminal RecA-2, including the C-terminal equivalent of the separator loop, and WH domains. Interaction region II ensues between the N-terminal RecA-1 and WH domains and the C-terminal RecA-2 domain. Indicated views are relative to Fig. 3.12A, top. (B) View along the central tunnel of the C-terminal cassette with the N-terminal cassette as a semi-transparent outline covering part of the tunnel. The orientation of the C-terminal cassette is the same as the orientation of the N-terminal cassette in Fig. 3.12A, bottom.

3.2.3 Functional analysis of the hBrr2 cassettes

Previous genetic analyses have shown that the ATPase and helicase activities of the N-terminal cassette of Brr2 are required for splicing, while putatively inactivating mutations were tolerated at the C-terminal cassette (Kim and Rossi, 1999), suggesting that the C-terminal cassette may not be an active ATPase or helicase. Furthermore, from the crystal structure of hBrr2^{HR}, we have observed that both cassettes interact intimately.

To address the ATPase and helicase activities of the individual cassettes as well as whether the inter-cassette interactions translate into a functional cooperation, we systematically screened for soluble fragments encompassing the N- or C-terminal cassette alone. Among 26 different constructs tested (13 encompassing the N-terminal cassette and 13 encompassing the C-terminal cassette), only fragments 395–1324 (hBrr2^{NC}, comprising the

N-terminal extension, N-terminal cassette and the inter-cassette linker) and 1282–2136 (hBrr2^{CC}, comprising the inter-cassette linker and the C-terminal cassette) could be solubly expressed and purified. All fragments that did not contain the inter-cassette linker were either not expressed or insoluble. Melting analyses (with proteins based on the wt sequence) showed that both fragments underwent cooperative unfolding with similar melting temperatures (44.7°C and 42.8°C, respectively) (Fig. 3.20), suggesting that both cassettes are independent folding units. In gel filtration analysis, no stable complex was formed between the separately produced and mixed cassette constructs, possibly due to the overlap in the linker element (data not shown).

While hBrr2^{NC} retained ATPase and U4/U6 di-snRNA unwinding activities, hBrr2^{CC} was entirely inactive as an ATPase or helicase (Fig. 3.17B, C), which is consistent with some of its ATPase/helicase motifs deviating from the consensus (Noble and Guthrie, 1996; Oyama *et al.*, 2009) and with the lack of severe phenotypes upon mutating these motifs in yBrr2 (Kim and Rossi, 1999). Surprisingly, hBrr2^{NC} exhibited diminished intrinsic (~0.01 ATP/hBrr2^{NC}/s) and U4/U6-stimulated (0.30±0.05 ATP/hBrr2^{NC}/s) ATPase activities compared to hBrr2^{HR} (Fig. 3.17C). In addition, U4/U6 di-snRNA unwinding by hBrr2^{NC} was markedly reduced compared to the dual cassette construct, hBrr2^{HR} (Fig. 3.17C and D). hBrr2^{HR} also more efficiently unwound a simple model duplex with a single-stranded 3'-overhang (Fig. 3.17D) compared to hBrr2^{NC}, showing that the stimulatory effects are not simply due to special sequences or structures of U4/U6 di-snRNA, but rather reflect an intrinsic organizational feature of hBrr2^{HR}. These results show that the N-terminal cassette harbors the helicase activity of Brr2, while the C-terminal cassette acts as an intramolecular cofactor for this helicase.

To investigate why the C-terminal cassette is inactive as an ATPase and whether it may nevertheless still bind ATP, we attempted to determine structures of hBrr2^{HR} in complex with Mg^{2+}-ATP, a non-hydrolysable analog (Mg^{2+}-AMPPNP) or transition state analogs (ADP-AlF_x, ADP-BeF_x). Co-crystallization attempts failed due to the high salt concentrations required by hBrr2^{HR}. To soak nucleotides into the crystals at lower ionic strength, hBrr2^{HR} crystals were stabilized by cross-linking. Soaking of cross-linked crystals with ATP or non-hydrolysable ATP analogs yielded the same results; nucleotides bound at both cassettes without significant conformational changes and in a manner incompatible with ATP hydrolysis (Fig. 3.18A, B, C).

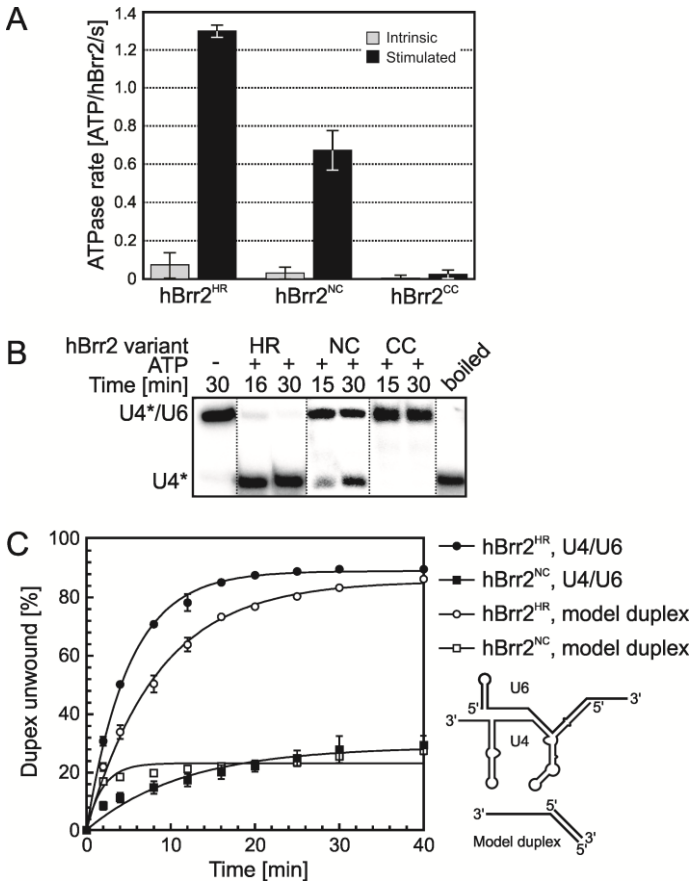


Fig. 3.17: Activities of the individual cassettes. (A) Intrinsic (gray bars) and RNA-stimulated (black bars) ATPase activities of $Brr2^{HR}$ compared to the individual cassettes. Error bars represent standard errors of the mean for three independent measurements. (B) Unwinding of U4/U6 di-snRNA by $Brr2^{HR}$ and the isolated N-terminal cassette and lack of helicase activity in the isolated C-terminal cassette. Lanes were compiled from three identically processed gels. (C) Unwinding time courses comparing the activities of $hBrr2^{HR}$ and $hBrr2^{NC}$ towards U4/U6 and a model duplex with a single-stranded 3'-overhang. Error bars represent standard errors of the mean for two independent measurements. Apparent unwinding rate constants (k_u) and amplitudes (A): $hBrr2^{HR}$/U4/U6 – k_u = 0.200±0.006 min^{-1}; A = 88.9±0.6%; $hBrr2^{HR}$/model duplex – k_u = 0.118±0.006 min^{-1}; A = 85.5±1.2%; $hBrr2^{NC}$/U4/U6 – k_u = 0.087±0.016 min^{-1}; A = 28.6±1.9%; $hBrr2^{NC}$/model duplex – k_u = 0.521±0.13 min^{-1}; A = 22.9±0.9%). The activity assays with $hBrr2^{HR}$ as well as $hBrr2^{NC}$ and $hBrr2^{CC}$ were carried out in collaboration with Sina Mozaffari Jovin (MPI-BPC) using proteins provided by Karine Santos.

Irrespective of the nucleotide employed, an ADP moiety, presumably originating from contamination in the nucleotide preparations, was accommodated at the N-terminal cassette (Fig. 3.18A). The nucleotide was bound almost exclusively by motifs from the first RecA domain and lacked interactions with RecA-2 required for hydrolysis (Fig. 3.18A). Unlike in spliceosomal DEAH/RHA helicases (Walbott *et al.*, 2010; He *et al.*, 2010), Q485 (Q loop) interacts with the N6 and N7 positions of the adenine, explaining the ATP/CTP specificity of hBrr2 (Laggerbauer *et al.*, 1998). While crystal packing may have restricted conformational changes upon nucleotide binding, very similar non-hydrolytic binding modes were recently seen in Mtr4 (Weir *et al.*, 2010) and Hjm (Oyama *et al.*, 2009) proteins, suggesting that RNA binding is additionally required to elicit an active ATPase conformation in Ski2-like helicases and explaining the low intrinsic ATPase activity of $hBrr2^{HR}$.

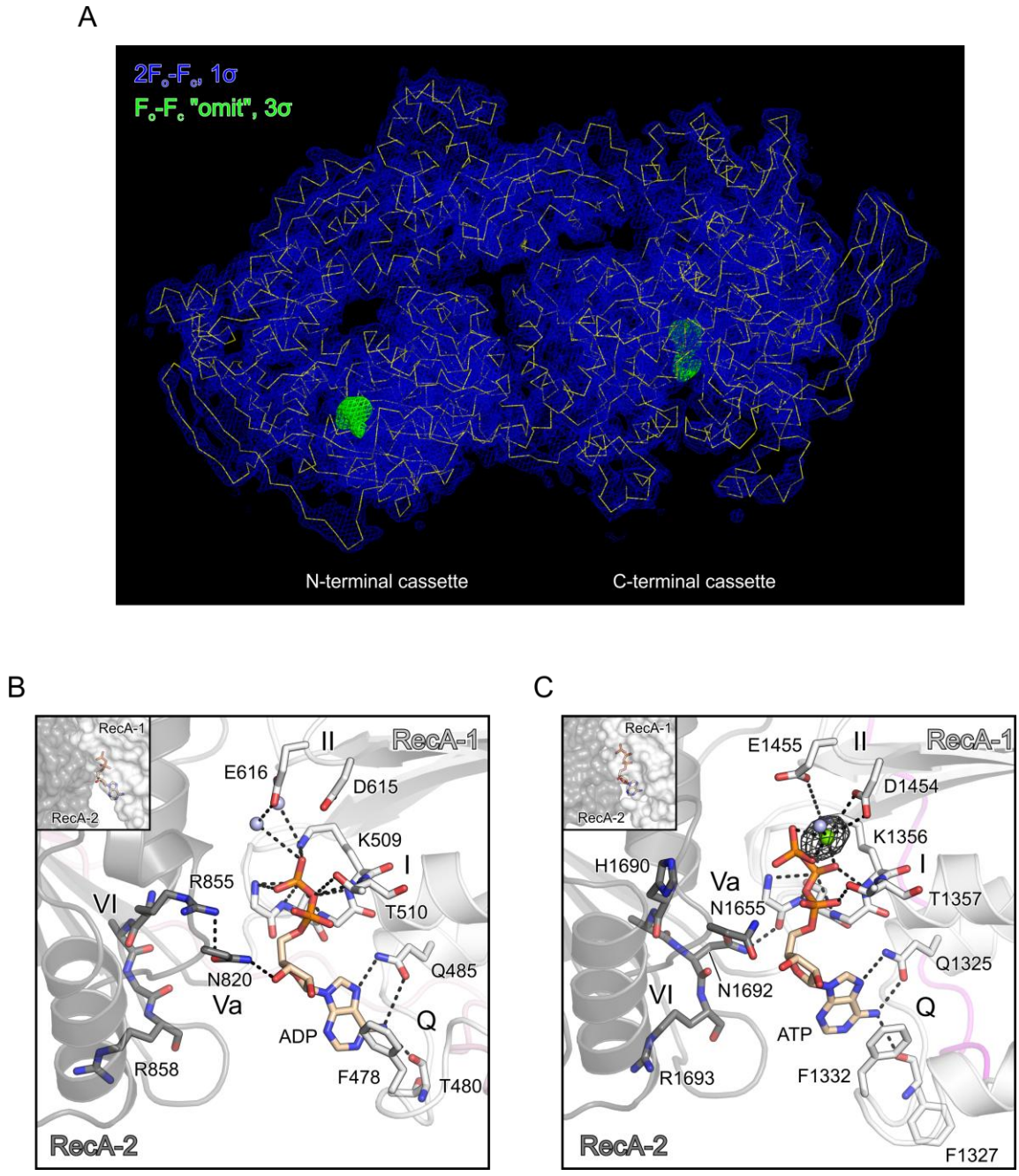


Fig. 3.18: Nucleotide binding. (A) $2F_o$-F_c electron density covering the entire hBrr2^{HR,S1087L} molecule (blue, contoured at the 1σ level) with a Cα trace in yellow. Green patches show F_o-F_c "omit" electron density (contoured at the 3σ level) at both cassettes upon omission of the nucleotides from the refinement. The view is the same as in Fig. 3.12A, top. (B and C) Close-up stereo views of the ATP pockets of the N-terminal (B) and C-terminal (C) cassettes. Domains and elements are colored as in Fig. 3.12. Nucleotides and selected residues are shown as sticks and colored by atom type (carbon – as the respective structural element; nitrogen – blue; oxygen – red; ATP carbon – beige; phosphorus – orange). Water molecules and a Mn^{2+} ion are shown as blue and green spheres, respectively. Dashed lines – hydrogen bonds or salt bridges. Q and Roman numerals – helicase motifs. Gray mesh in (C) – anomalous difference electron density contoured at the 4σ level, indicating the position of the Mn^{2+} ion.

In contrast, a Mg^{2+}-ATP (or analog) complex, again selected through a Q loop (Q1332), was bound at the C-terminal cassette (Fig. 3.18B). Presence of a single divalent metal ion coordinated by the ß and γ-phosphates of the nucleotide, D1454 (motif II) of Brr2^{HR} and a

water molecule was verified by anomalous difference density in a long-wavelength data set collected on a crystal soaked with Mn^{2+}-ATP (Fig. 3.18B). As in the N-terminal cassette, contacts of the RecA-2 domain to the bound nucleotide were completely lacking. However, at the C-terminal cassette the non-hydrolytic state is specifically stabilized via non-canonical residues (Fig. 3.18C). The H1690 (motif VI), which is an arginine in canonical SF2 helicases, may be too short to contact the ATP phosphates. The following residue, N1692, is a glycine or alanine in active helicases (e.g. G857 in the N-terminal cassette). N1692 engages in a hydrogen bond with the backbone carbonyl of G1353 (motif I) from the first RecA domain and thereby locks both H1690 and R1693 (motif VI) in orientations pointing away from the ATP phosphates, hindering the C-terminal cassette to adopt a conformation conducive to hydrolysis. As a further consequence, N1655 (motif V) in the second RecA domain is pushed away from the nucleotide sugar, which it contacts in active SF2 helicases (Fig. 3.18B).

3.2.4 RNA binding to hBrr2

3.2.4.1 Modelling of the RNA path through $Brr2^{HR}$

We next investigated whether the C-terminal cassette stimulation of the N-terminal helicase activity depends on direct interactions of the C-terminal cassette with RNA during unwinding. So far, we have failed to crystallize $hBrr2^{HR}$ in complex with RNA. To investigate whether and how the C-terminal domain may contribute to RNA substrate binding, we modeled RNA binding at the active N-terminal cassette in analogy to nucleic acid binding by the related SF2 DNA helicase Hel308 (Büttner *et al.*, 2007) and the SF2 RNA helicase Mtr4 (Weir *et al.*, 2010). To this end, we superimposed the Hel308-DNA structure (PDB ID 2P6R) onto the N-terminal cassette of $Brr2^{HR}$, converted DNA to RNA and manually adjusted the nucleic acid to reduce clashes with the protein. In the model, one RNA strand is threaded through the central tunnel of the N-terminal cassette, with 3'-to-5' directionality, running across the conserved RNA-binding motifs of the RecA domains, alongside the separator loop and beneath the ratchet helix of the HB domain (Fig. 3.19A). The model suggested that upon emergence from the N-terminal tunnel, the RNA strand may either turn away from the C-terminal cassette and exit via a positively charged surface on the N-terminal HLH domain (Fig. 3.19B, path 1) as seen in Hel308 (Büttner *et al.*, 2007) or continue in the direction of the

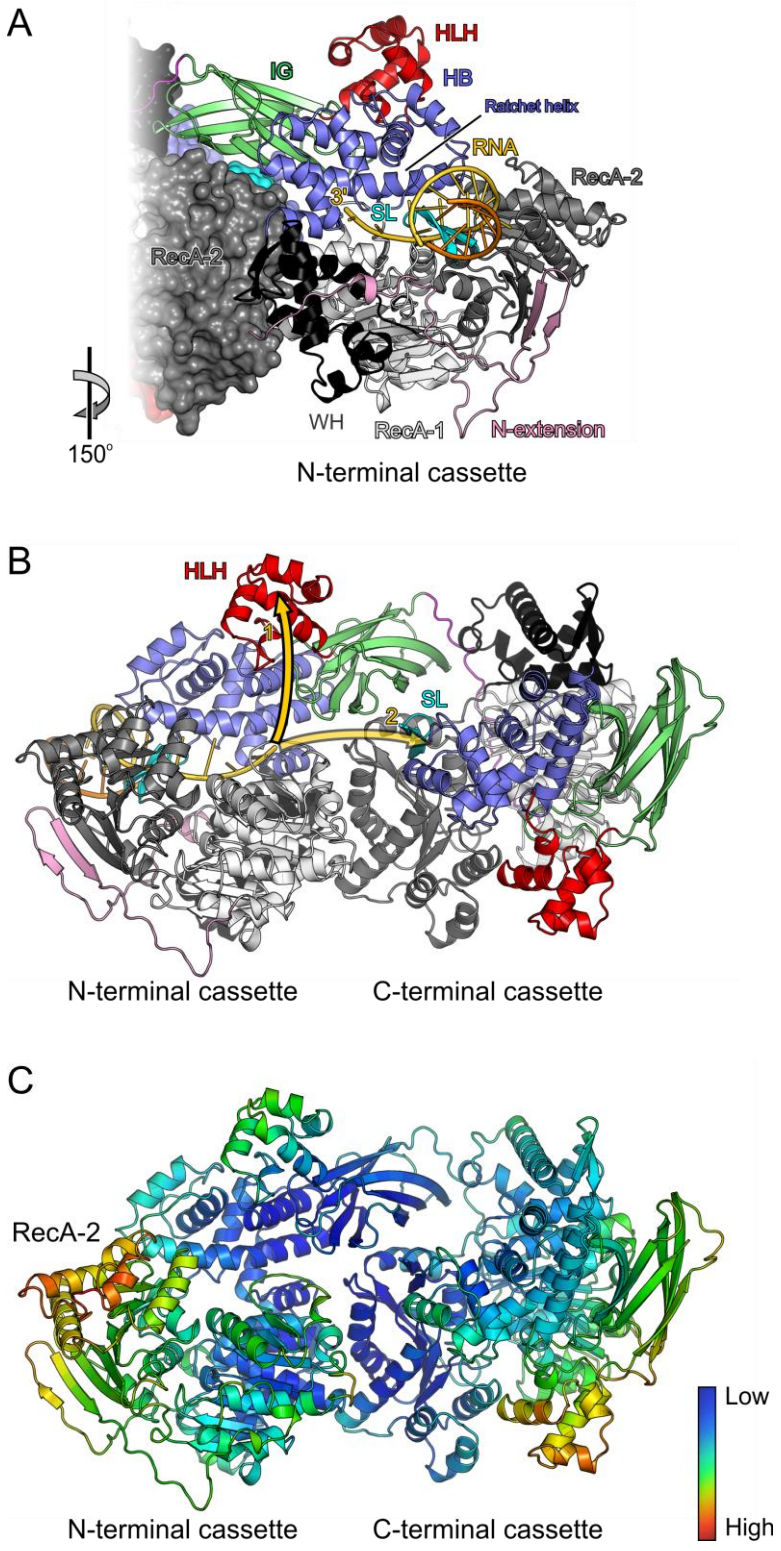


Fig 3.19: Model for RNA binding and loading. (A) View along the central tunnel of the N-terminal cassette with a modeled RNA ligand (threaded strand – gold; complementary strand – orange). Domains and elements are colored as in Fig. 3.12A. Rotated 150° counter-clockwise as indicated compared to Fig. 3.12A, top. (B) Ribbon plot of the same model viewed as in Fig. 3.12A, top. The golden arrows indicate two possible paths of the RNA strand exiting the N-terminal cassette. (C) Ribbon plot of hBrr2^{HR,S1087L} colored according to the crystallographic temperature factors of the Cα atoms. Red – high mobility/flexibility; blue – low mobility/flexibility. Same orientation as in Fig. 3.12A, top.

putative separator loop of the C-terminal cassette (Fig. 3.19B, path 2). We resorted to a mutational strategy to distinguish between these alternatives.

3.2.4.2 Design and characterization of mutant proteins

The following mutational analyses were based on the RP33-linked S1087L variant of $Brr2^{HR}$ (GenBank ID: gi|45861371) (see section 3.2.6). The effects of this mutation on $Brr2^{HR}$ RNA binding, ATPase and helicase activities are rather mild (Fig. 3.23). Furthermore, no significant conformational changes were observed compared to the $hBrr2^{HR}$ structure (Cα rmsd of 0.4 Å). Additional mutations are expected to show the same trend in $hBrr2^{HR}$ as in $hBrr2^{HR,S1087L}$. We reasoned that the primary S1087L mutation would "sensitize" the protein, rendering the phenotypes of other mutations more easily experimentally accessible.

All mutants used in this study were efficiently expressed, purified and migrated as a monomer in size exclusion chromatography (data not shown). In thermofluor-based thermal melting analyses, all variants exhibited cooperative transitions with comparable melting temperatures (Fig. 3.20). Furthermore, equilibrium CD spectra were indicative of a high content of regular secondary structure in all hBrr2 variants (data not shown). These data indicate that all hBrr2 variants tested herein were well folded and that mutant phenotypes were not simply a result of a loss of stable tertiary structure.

3.2.4.3 Structure-based mutational analyses

We mutated positively charged residues on the surface of the N-terminal HLH domain (RK1133-4EE) that do not directly interact with other residues from the HLH or neighboring domains (Fig. 3.21A). Compared to the parent enzyme, the RK1133-4EE double mutant exhibited enhanced intrinsic and RNA-stimulated ATPase activities while its helicase activity, estimated by single-point unwinding assays, was dramatically reduced (Fig. 3.21E, lane 2; Fig. 3.21F, lane 4). Furthermore, binding to an RNA model duplex was almost abrogated in the RK1133-4EE double mutant (Fig. 3.21H), supporting the idea that its dysfunction roots in impaired RNA binding.

A tunnel with equivalents of a ratchet helix and a separator loop is also seen in the C-terminal cassette (Fig. 3.14, left). However, part of the rim and the inner walls of the C-terminal tunnel are negatively charged (Fig. 3.15B, right). In addition, the C-terminal cassette exhibits more extensive contacts between RecA-2 and the HB domain than the N-terminal cassette which could counteract RNA loading into the tunnel (Fig. 3.14, bottom right).

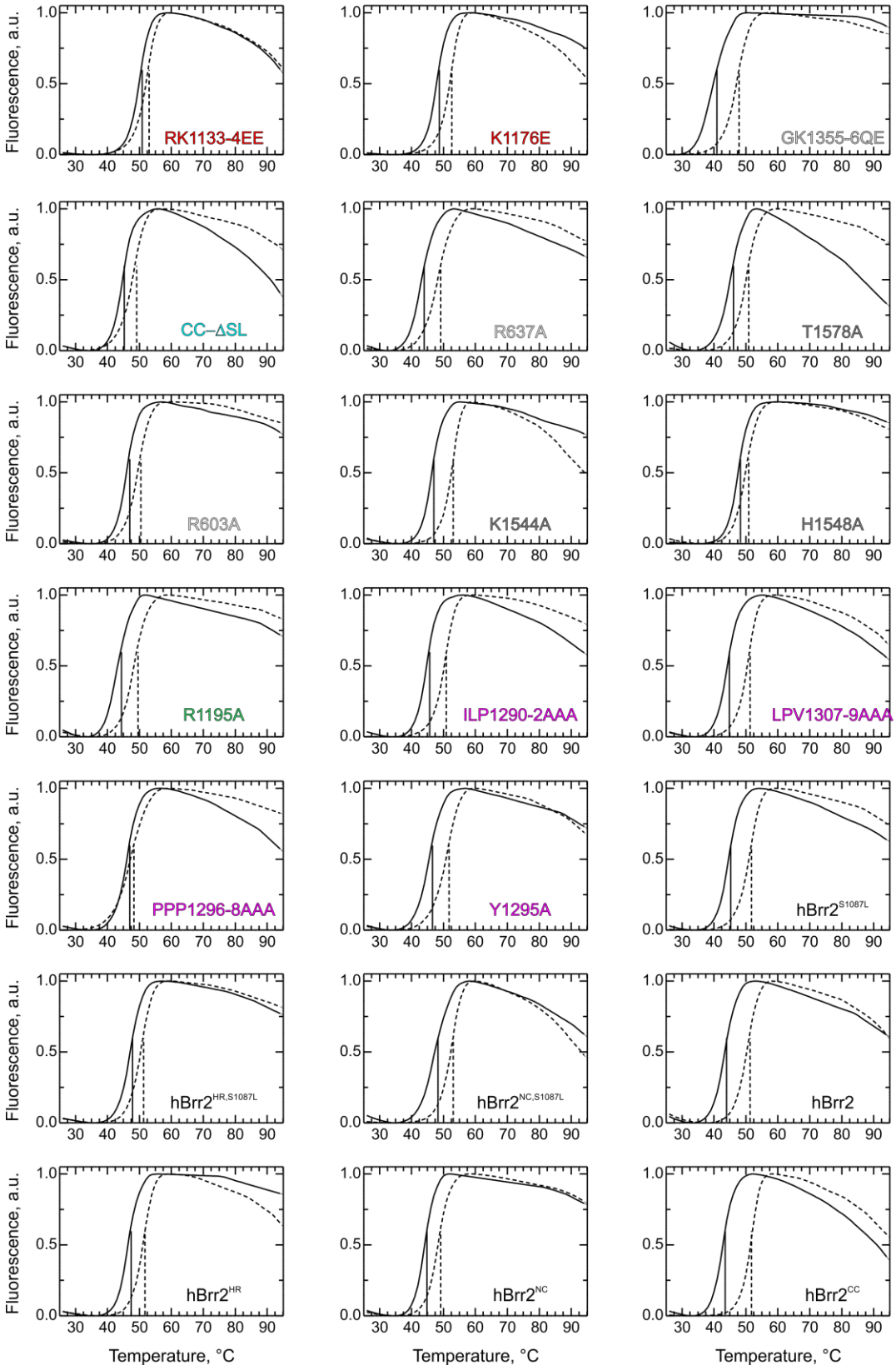


Figure 3.20: DSF analyses of the hBrr2 variants used in this work. Black lines indicate the absence of Mg^{2+}-ATP while dashed lines indicate the presence of Mg^{2+}-ATP in the buffer. All variants exhibited cooperative transitions with similar melting temperatures.

Furthermore, the N-terminal cassette contacts the C-terminal RecA-2, including the element that is the equivalent of the N-terminal cassette separator loop, as well as the neighboring RecA-1 and WH domains, thereby sterically blocking the entrance to the C-terminal cassette tunnel (Fig. 3.16B). We, therefore, predicted that the equivalent of the separator loop in the C-terminal cassette should be dispensable for RNA duplex unwinding. Indeed, replacement of this loop (residues Y1668-V1677) by a single serine had essentially no effect on ATPase activities (Fig. 3.21E, lane 3) or RNA duplex unwinding (Fig. 3.21F, lane 5; Fig. 3.21G) and did not significantly affect the RNA affinity (K_d hBrr2$^{HR,CC\text{-}\Delta SL}$ – 31.0±6.3 nM; Fig. 3.21H), in stark contrast to the essential function of this element in the N-terminal cassette of yBrr2 (Pena *et al.*, 2009; Zhang *et al.*, 2009). These findings support the idea that an unwound RNA strand traverses the N-terminal HLH domain, as seen for DNA in Hel308 (Büttner *et al.*, 2007), and is guided away from the C-terminal cassette.

Taken together, these results suggest that RNA contacts essential for duplex unwinding are apparently fostered only by the N-terminal cassette, which is consistent with a lack of RNA binding to an isolated yBrr2 C-terminal cassette (Zhang *et al.*, 2009). Since the stimulatory effect of the C-terminal cassette, therefore, seems to originate from its interaction with the N-terminal cassette rather than from binding of the RNA substrate, we investigated these inter-cassette contacts in more detail (see section 3.2.5).

3.2.4.4 RNA loading

As a member of the Ski2-like family of SF2 helicases, Brr2 is thought to translocate in a 3'-to-5' direction on one of the substrate strands. However, in the U4/U6 di-snRNP, the 3'-ends of U4 and U6 snRNA are occluded by secondary structures and/or bound proteins (Leung *et al.*, 2011; Achsel *et al.*, 1999) and are thus unavailable for Brr2 binding. Psoralen-crosslinking of the RNA network in the minor spliceosome indicated that U4atac/U6atac stem 1 (equivalent to U4/U6 stem I in the major spliceosome) is unwound before stem II during catalytic activation, implying that Brr2 translocates on U4 (U4atac) snRNA in 3'-to-5' direction (Frilander and Steitz, 2001). We suggest that Brr2 circumvents the sequestered 3'-end of U4 (U4atac) snRNA by intermittent opening of its N-terminal RecA-2 and HB domains and loading onto the internal single-stranded U4 (U4atac) snRNA region immediately downstream of stem I. N-terminal cassette opening appears feasible considering the limited interactions between the RecA-2 and HB domains (Fig. 3.14) and in light of the crystallographic B-factor distribution, showing that the tip of the N-terminal RecA-2 domain is one of the most flexible portions of the hBrr2^{HR} crystal structure (Fig. 3.19C).

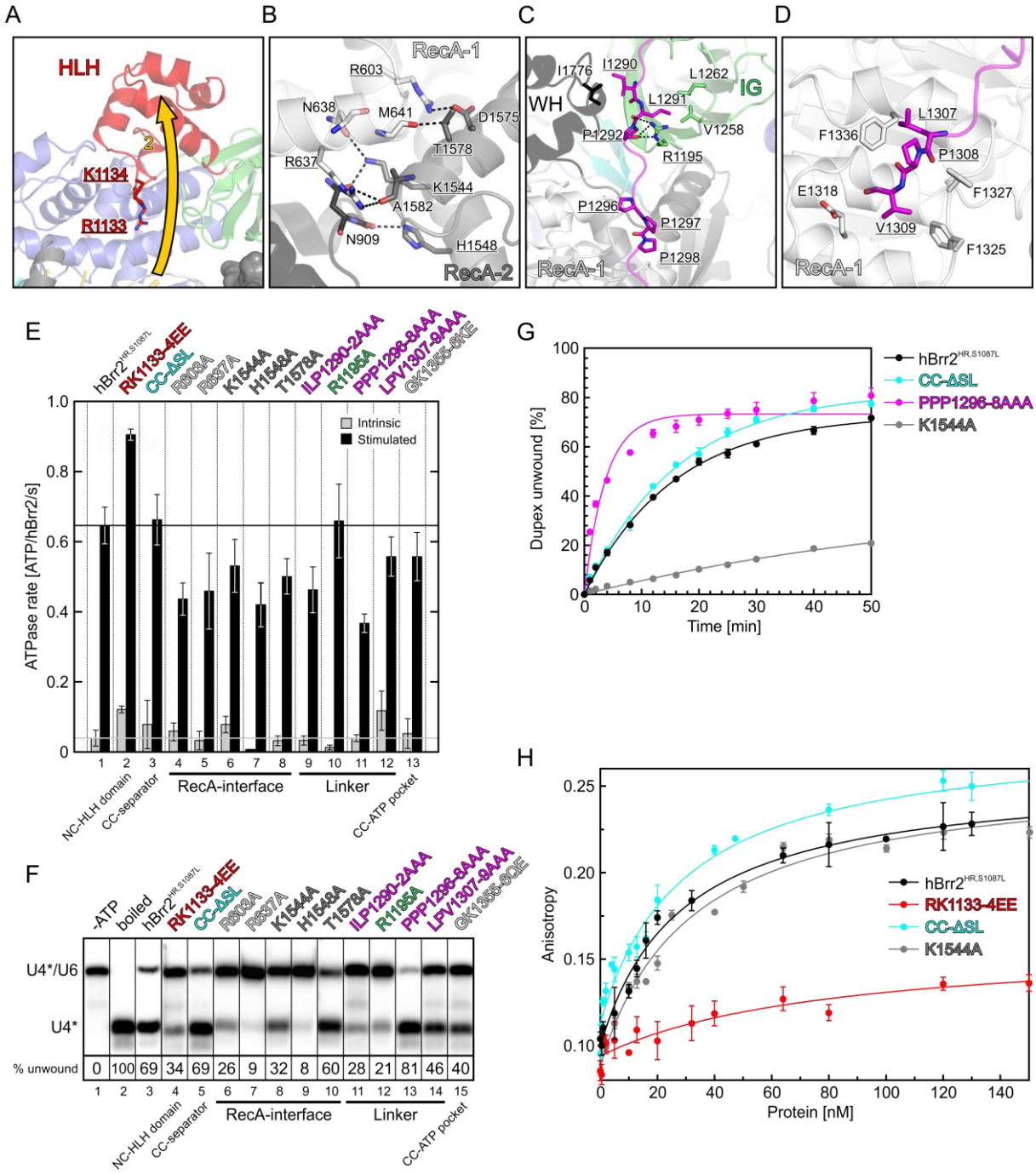


Fig 3.21: Mutational analysis of hBrr2^{HR}. (A) Close-up view on the N-terminal HLH domain. R1133 and K1134 comprise *bona fide* RNA contact sites. Gold arrow, putative path of the RNA. Image is in the same orientation as in Fig. 3.12A, top. (B) Contacts between the N-terminal RecA-1 and the C-terminal RecA-2 domain. Image is rotated 90° about the horizontal axis (top to back) compared with Fig. 3.12A, top. (C) Upper portion of the linker. Image is rotated 30° about the horizontal axis (top to front) compared with Fig. 3.12A bottom. (D) Lower portion of the linker. Image is rotated 30° about the horizontal axis (top to back) compared with Fig. 3.12A bottom. Mutated residues in A–D are underlined. (E) Intrinsic (gray bars) and U4/U6-stimulated (black bars) rates of ATP hydrolysis of the hBrr2^{HR,S1087L} variants indicated. NC, N-terminal cassette; CC, C-terminal cassette; CC-ΔSL, replacement of the C-terminal separator loop by a single serine. Error bars represent SEMs for three independent measurements. (F) Single point unwinding assays comparing the hBrr2^{HR,S1087L} variants indicated above the gel. Quantification (percent unwound after 50 min) is shown below the image. Lanes were compiled from two identically processed gels. (G) Unwinding time courses of selected hBrr2^{HR,S1087L} variants. Apparent unwinding rate constants (ku) and amplitudes (A): hBrr2^{HR,S1087L}, k_u = 0.064±0.003 min^{-1}, A =

73.0±1.2%; hBrr2$^{HR,CC-\Delta SL}$, k_u = 0.062±0.003 min^{-1}, A = 82.4±1.8%; hBrr2$^{HR,PPP1296-8AAA}$, k_u = 0.27±0.04 min^{-1}, A = 73.3±2.2%; hBrr2^{HR,K1544A} , k_u = 0.015±0.003 min^{-1}, A = 39.3±6.5%. Error bars represent SEMs for two independent measurements. (H) RNA binding by the indicated hBrr2^{HR,S1087L} variants measured by fluorescence polarization. Error bars represent SEMs for three independent measurements. K_d hBrr2^{HR,S1087L}, 28.5±3.8 nM; K_d hBrr2$^{HR,RK133-4EE}$, not determined; K_d hBrr2$^{HR,CC-\Delta SL}$, 31.0±6.3 nM; K_d hBrr2^{HR,K1544A}, 35.0±7.2 nM. Mutant labels and curves in E–H are colored according to their domains or elements (hBrr2^{HR,S1087L} reference, black). Data from panels E-H were acquired by Sina Mozaffari-Jovin using protein provided by Karine Santos.

3.2.5 Inter-cassette communication

In all nucleic acid helicases, RecA domains couple nucleotide binding, hydrolysis and release to nucleic acid deformation or translocation. Direct contacts of the Brr2 C-terminal cassette to the N-terminal RecA domains are thus obvious candidate sites for inter-cassette communication. To test the importance of such contacts, we introduced single alanine substitutions expected to weaken interactions between the N-terminal RecA-1 or WH and the C-terminal RecA-2 domains (R603A, R637A, K1544A, H1548A and T1578A; Fig. 3.21B).

Residues 603 and R637 (N-terminal RecA-1) form a salt bridge with D1575 and engage in a bifurcated hydrogen bond with the backbone carbonyl of A1582 (C-terminal RecA-2), respectively (Fig. 3.21B). While mutation of either residue had little effect on the intrinsic ATPase, stimulated ATPase was moderately reduced (Fig. 3.21E, lanes 4 & 5) and helicase activity was virtually abrogated (Fig. 3.21F, lanes 6 & 7). K1544 and T1578 (C-terminal RecA-2) form hydrogen bonds with the side chain of N638 and with the backbone carbonyl of M641 (N-terminal RecA-1), respectively (Fig. 3.21B). Upon mutation to alanine, ATPase activities were only mildly affected (Fig. 3.21E, lanes 6 & 8), while helicase activity in K1544A was strongly and in T1578A slightly reduced (Fig. 3.21F, lanes 8 & 10). As an example, we quantified the helicase effect of the K1544A variant by recording full unwinding time courses. The apparent rate constant of unwinding in this mutant was reduced more than fourfold and the protein unwound only about half the amount of U4/U6 within 50 min compared to the parent enzyme (Fig. 3.21G). Finally, we also tested H1548 (C-terminal RecA-2) that forms a hydrogen bond with the backbone carbonyl of G908 (N-terminal WH) (Fig. 3.21B). The H1548A mutation showed reduced intrinsic and moderately reduced stimulated ATPase activities (Fig. 3.21E, lane 7), while helicase activity was again almost entirely lost (Fig. 3.21F, lane 9). None of the mutated residues belongs to the canonical ATPase/helicase motifs of either cassette, suggesting that all phenotypes are due to disturbed cassette interactions. Consistent with this idea, RNA binding by the K1544A mutant was essentially unchanged (Fig. 3.21H).

Since the linker establishes interaction networks with both cassettes, it may serve as an additional tool for inter-cassette communication. To test this idea, we mutated all of its structural regions. Residues I1290, L1291 and P1292 (region 1) run between the N-terminal IG and C-terminal WH domains (Fig. 3.21C). Concomitant mutation to alanines reduced intrinsic as well as RNA-stimulated ATPase activities (Fig. 3.21E, lane 9) and essentially wiped out helicase activity (Fig. 3.21F, lane 11). We also tested a reciprocal contact of R1195 on the N-terminal IG domain to the backbone carbonyl of P1292 in region 1 of the linker (Fig. 3.21C). The R1195A mutant showed reduced intrinsic ATPase, with its stimulated ATPase almost unchanged (Fig. 3.21E, lane 10), and was severely defective in duplex unwinding (Fig. 3.21F, lane 12). Residues L1307, P1308 and V1309 (region 3) tie the lower part of the linker to the C-terminal RecA-1 domain (Fig. 3.21D). Similar to the alanine substitution effects in region 1 and R1195, exchange of residues in region 3 for alanines increased the intrinsic ATPase but reduced the RNA-stimulated ATPase and helicase activities of hBrr2^{HR} (Fig. 3.21E, lane 12; Fig. 3.21F, lane 14). Furthermore, mutation of the highly conserved proline motif, P1296, P1297 and P1298 (region 2), to alanines is expected to increase the flexibility in this part of the linker (Fig. 3.21C). In contrast to residues in regions 1 and 3, residues in region 2 do not directly contact the bulk of hBrr2^{HR}. This mutant showed reduced intrinsic and RNA-stimulated ATPase activities (Fig. 3.21E, lane 11). Strikingly, its apparent unwinding rate constant was more than quadrupled (Fig. 3.21F, lane 13; Fig. 3.21G), showing that certain modulations of inter-cassette contacts may also lead to up-regulation of Brr2 helicase activity.

To investigate whether nucleotide binding at the C-terminal cassette influences the N-terminal helicase, we introduced changes in the C-terminal ATP pocket of hBrr2^{HR,S1087L} designed to interfere with nucleotide accommodation (GK1355-6QE; Fig. 3.18C). While ATPase activity was only mildly affected (Fig. 3.21E, lane 13), helicase activity was strongly reduced in this mutant (Fig. 3.21F, lane 15).

3.2.6 Structural basis for dysfunctional hBrr2 variants

The hBrr2^{HR} structure offers explanations for the malfunction of several previously investigated Brr2 alleles (Fig. 3.22). Of particular interest, a number of mutations in Brr2 have been linked to the RP33 form of retinitis pigmentosa (Benaglio *et al.*, 2011; Zhao *et al.*, 2006, 2009). One set of affected residues lies in the connection between the N-terminal RecA domain (R681C, R681H, V683L) and the first ß-strand of the RecA-2 domain (Y689C), where they establish inter-domain contacts and stabilize domain folds (Fig. 3.22A, B).

Another set of mutations maps to the ratchet helix of the N-terminal HB domain (S1087L, R1090L; Fig. 3.22C, D, E) and corresponding changes in yBrr2 were detrimental to U4/U6 unwinding and splicing (Zhao *et al.*, 2009; Zhang *et al.*, 2009; Pena *et al.*, 2009). R1090 (conserved as R1107 in yBrr2) extends from the underside of the ratchet helix (Fig. 3.22C) and is thus likely important for interaction with RNA (Pena *et al.*, 2009). Considering that S1087 is an asparagine in yBrr2 (N1104), we directly tested the consequences of a leucine at this position in the disease-relevant human protein. hBrr2^{HR,S1087L} exhibited decreased RNA binding and reduced ATPase and helicase activities as compared to the wild type (wt) variant, Brr2^{HR} (Fig. 3.23).

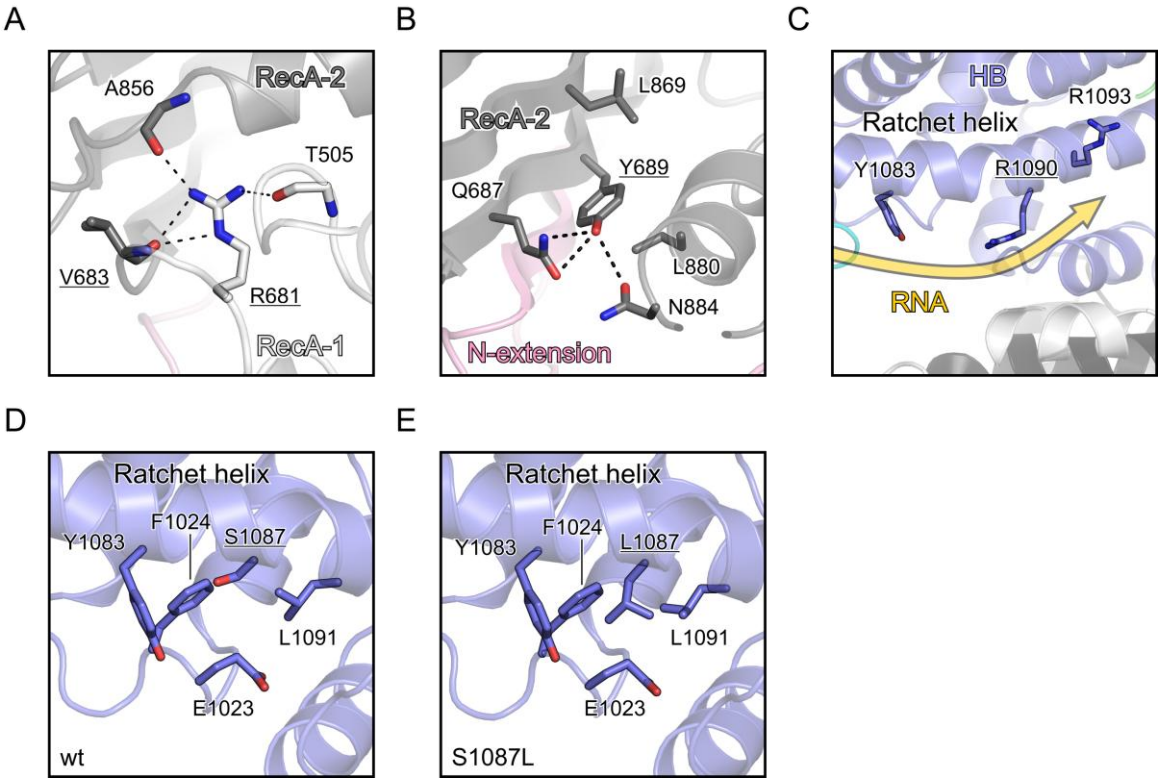


Fig. 3.22: RP33-linked hBrr2 mutations. (A-E) Ribbon plots showing details of the environments of the RP33-linked hBrr2 residues with coloring as in Fig. 3.12. Dashed lines – hydrogen bonds or salt bridges. Affected residues are underlined. The semi-transparent golden arrow in (C) indicates the modeled RNA path. In (D) and (E), the residue 1087 environment in hBrr2^{HR} and hBrr2^{HR,S1087L} is shown, respectively. Views relative to Fig. 3.12, top: (A) 60° clockwise about the vertical axis; (B) 90° clockwise about the vertical axis; (C) unchanged; (D and E) 90° about the horizontal axis (top to back).

The crystal structure of hBrr2^{HR,S1087L} at 2.65 Å resolution revealed no significant conformational changes compared to hBrr2^{HR} (Cα rmsd 0.4 Å) except that a leucine at position 1087 interacts more intimately with a neighboring hydrophobic/aromatic cluster on the HB domain (Fig. 3.22D, E). Additionally, only a serine, instead of a leucine, could foster hydrogen bonding interactions to bound RNA. Therefore, S1087L abrogates functionally important S1087-RNA contacts and/or counteracts conformational changes in the ratchet

helix of the HB domain, which have been postulated in related DEAH/RHA helicases as a means to drag RNA across the separator loop (Walbott *et al.*, 2010).

In addition to the above disease-related mutations, our hBrr2^{HR} structure also offers explanations for the malfunction of several other Brr2 mutants that have been investigated in the past. Previously reported yBrr2 variants with mutations in helicase motifs I (G526D/K527N) and II (D634G) in the N-terminal cassette did not support yeast viability, U4/U6 di-snRNA unwinding or yBrr2 ATPase activity (Kim and Rossi, 1999). The equivalent residues in other helicases are known to be involved in nucleotide binding and hydrolysis. As expected, the corresponding residues in hBrr2 (G508, K509 and D615) line the ATP pocket of the N-terminal cassette and contact the nucleotide phosphates (G508, K509) or are expected to coordinate a metal ion (D615) upon productive accommodation of ATP (Fig. 3.18B).

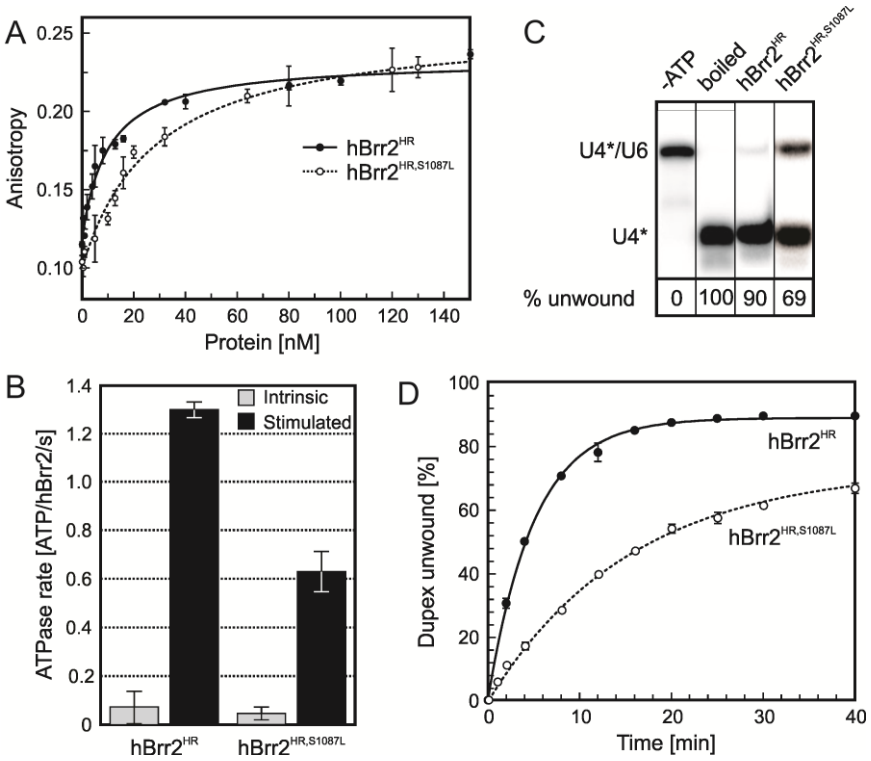


Fig 3.23: Effects of the RP33-linked S1087L mutation. (A) RNA binding by hBrr2^{HR} and hBrr2^{HR,S1087L} measured by fluorescence polarization. Error bars represent standard errors of the mean for three independent measurements. K_d hBrr2^{HR} – 12.2±2.0 nM; K_d hBrr2^{HR,S1087L} – 28.5±3.8 nM. (B) Intrinsic (gray bars) and RNA-stimulated (black bars) rate of ATP hydrolysis by hBrr2^{HR,wt} and hBrr2^{HR,S1087L}. Error bars represent standard errors of the mean for three independent measurements. (C) Single point unwinding assays of hBrr2^{HR} and hBrr2^{HR,S1087L}. Quantification (% unwound after 50 min) at the bottom. Lanes were compiled from two identically processed gels. (D) Unwinding time courses of hBrr2^{HR} and hBrr2^{HR,S1087L}. Apparent unwinding rate constants (k_u) and amplitudes (A): hBrr2^{HR} – k_u = 0.200±0.006 min^{-1}; A = 89.0±0.6%; hBrr2^{HR,S1087L} – k_u = 0.064±0.003 min^{-1}; A = 73.0±1.2%. Data acquired by Sina Mozaffari-Jovin using proteins produced by Karine Santos.

The yeast *brr2-1* allele directs the exchange of E610 in yBrr2 for a glycine leading to impeded U4/U6 di-snRNA unwinding (Raghunathan and Guthrie, 1998) and failure to release

the excised intron and to dissociate snRNAs during spliceosome disassembly (Small *et al.*, 2006). The equivalent E591 of hBrr2 lies at the center of motif Ic (Fig. 3.24A), which has been seen to interact with nucleic acids in other SF2 helicases (Büttner *et al.*, 2007; Sengoku *et al.*, 2006; Weir *et al.*, 2010). Consistently, in the present structure, E591 is exposed on the inner surface of the presumed RNA-binding tunnel across from the ratchet helix (Fig. 3.24A). It also interacts with the neighboring R624, thus positioning this residue for RNA binding and contributing to the stability of the first RecA domain (Fig. 3.24A).

The E909K exchange in yBrr2 led to a block of pre-mRNA splicing before the first catalytic step (Xu *et al.*, 1996). The affected glutamate (E890 in hBrr2) is positioned in a peptide linking the RecA-2 and WH domains of the N-terminal cassette and stabilizes the domain arrangement by interacting concomitantly with the side chains of N657 (second RecA domain), R928 and Y936 (WH domain; Fig. 3.24B).

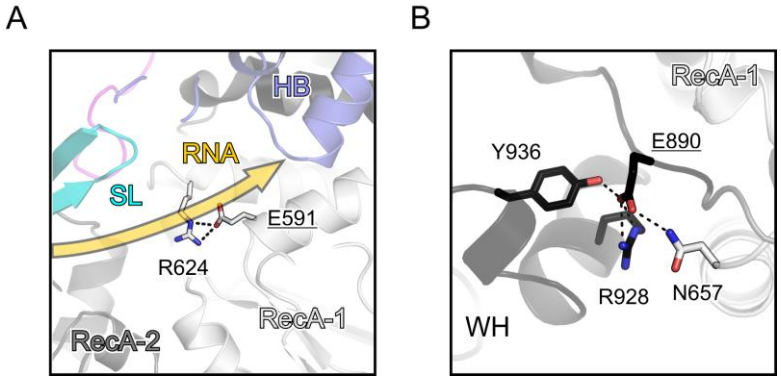


Fig. 3.24: Brr2 variants. (A and B) Structural contexts of Brr2 residues in the N-terminal cassette, which, upon mutation, give rise to a dysfunctional enzyme. The bulk of hBrr2^{HR} is shown as a semi-transparent ribbon with domains and elements colored as in Fig. 3.12. In (A), the modeled RNA path is indicated by a semi-transparent, golden arrow. Selected residues are shown as sticks and colored according to atom type (carbon – as the respective structural element; nitrogen – blue; oxygen – red). Dashed lines indicate hydrogen bonds or salt bridges. Mutated residues are underlined. Views relative to Fig. 3.12, top: (A) 90° about the horizontal axis (top to front); (B) 90° counter-clockwise about the vertical axis.

3.3 Interaction of hBrr2 with nucleotides in solution

To obtain an estimate for the affinity of hBrr2 to nucleotides and, specifically, to investigate if N- and C-terminal cassettes bind nucleotides with different affinities, we performed interaction studies with ADP and with a non-hydrolysable ATP analog, ATPγS. Although the intrinsic ATPase activity of hBrr2^{HR} is low, by using ATPγS we intended to avoid any degree of conversion of ATP to ADP during the experiments. All data was acquired in collaboration with Dr. Pohl Milon, Dept. of Physical Biochemistry, MPI-BPC, Göttingen.

Interactions of nucleotides with hBrr2 were studied using FRET from tryptophan residues located in the vicinity of the nucleotide binding pockets of hBrr2 to the *mant* group of *mant*-ATPγS or *mant*-ADP. It has been shown for many GTPases that nucleotide binding and nucleotide exchange reactions are not perturbed by the *mant* group (Jameson and Eccleston, 1997; Pittman *et al.*, 2006).

Control experiments showed that FRET signals were observed only when donor (tryptophan residues) and acceptor (*mant* group) were simultaneously present (Fig. 3.25). Therefore, fluorescence increase was observed only upon mixing labeled nucleotide with hBrr2^{HR}, hBrr2^{NC} or hBrr2^{CC} (Fig. 3.25). Signal changes were absent when hBrr2^{HR}, hBrr2^{NC} or hBrr2^{CC} were mixed with unlabeled nucleotides or buffer lacking nucleotides (Fig. 3.25).

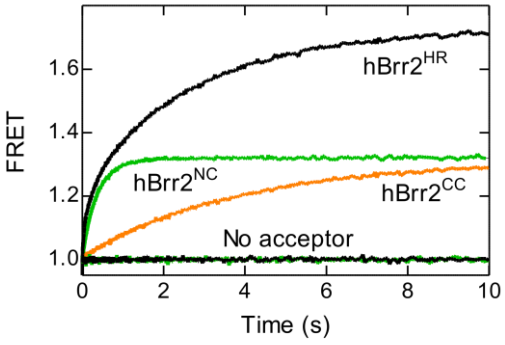


Fig. 3.25: Binding of *mant*-ATPγS to hBrr2^{HR}, hBrr2^{NC} and hBrr2^{CC}. Upon mixing in a stopped flow apparatus nucleotide-free hBrr2^{HR} or fragments thereof (0.2 µM) with *mant*-ATPγS or unlabeled ATPγS (5 µM), the emitted fluorescence of *mant* group (acceptor) is recorded as a function of time. In the absence of the fluorescence acceptor, no signal change was observed.

To determine association rate constants, a fixed concentration of nucleotide-free hBrr2 or fragments thereof was mixed with varying concentrations of *mant*-nucleotides. The data were analyzed by exponential fitting to determine the apparent rate constant (k_{app}) value for each titration point (Fig. 3.26).

hBrr2^{NC} titrations curves were fit with a double-exponential equation (Fig. 3.27), resulting in apparent rate constants k_{appNCa}, corresponding to a very rapid phase, and k_{appNCb}, associated with a second, rapid phase. Nevertheless, due to the fact that this very fast phase (represented by k_{appNCa}) results in a very small amplitude change, we only considered the apparent rate constant of the rapid phase (k_{appNCb}), which contributes most of the fluorescence signal change. The time courses obtained with hBrr2^{CC} were fit using a single-exponential equation (Fig. 3.26), with apparent rate constant k_{appCC}.

The hBrr2, hBrr2^{HR} and hBrr2^{HR,S1087L} time courses were best described by double-exponential fitting (Fig. 3.26), resulting in apparent rate constants k_{appNC} and k_{appCC}, once they

were similar to the apparent rate constants obtained for the individual cassettes. It is important to mention that titrations performed with hBrr2 and its fragments containing both cassettes (hBrr2^{HR}, hBrr2^{HR,S1087L}) showed a third, very rapid phase, which we could not fit due to the very small fluorescence change (amplitude) associated with this phase. Henceforth, we refer to k_{appNC} as a fast apparent rate constant corresponding to the ATP/ADP binding velocity to the N-terminal cassette, while k_{appCC} is the constant that describes the apparent velocity of nucleotide binding to the C-terminal cassette.

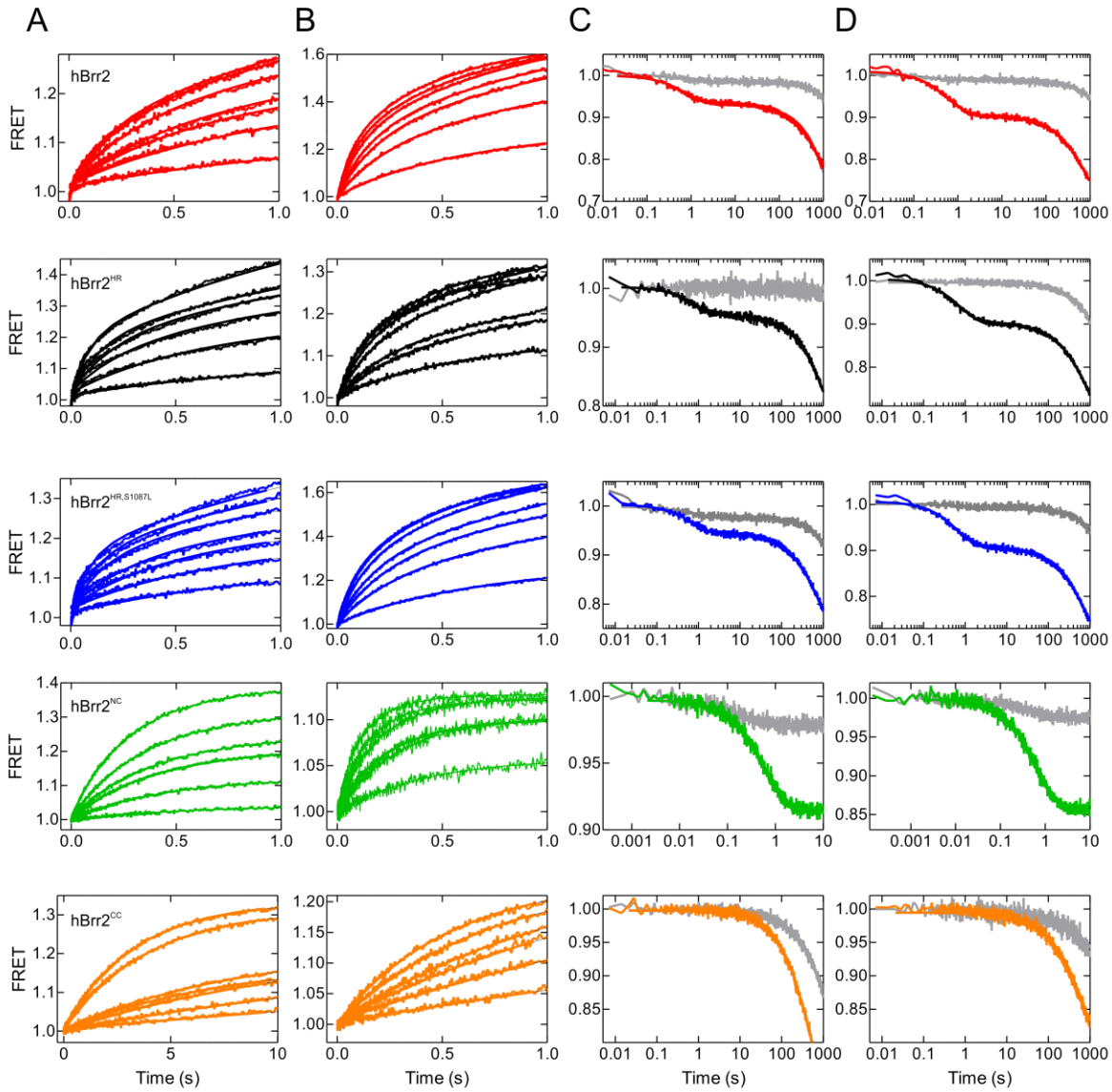


Fig. 3.26: Time courses of nucleotide binding at increasing concentrations of *mant*-ATPγS (A) and *mant*-ADP (B) to hBrr2, hBrr2^{HR}, hBrr2^{HR,S1087L}, hBrr2^{NC} and hBrr2^{CC} (top to bottom). Dissociation of hBrr2·*mant*-ATPγS (C) and hBrr2·*mant*-ADP (D) complexes upon mixing with unlabeled nucleotides (colored traces). Gray traces represent the decrease in FRET related to the dissociation of protein·*mant*-nucleotide by dilution in buffer. For each row of graphs, a different protein fragment (indicated) was used to perform nucleotide binding and dissociation experiments. It is not clear at the moment if the observed decrease in fluorescence upon dilution of hBrr2/hBrr2^{HR}/hBrr2^{HR,S1087L}/hBrr2^{CC}·*mant*-nucleotide with buffer either reflects the dissociation of the *mant*-nucleotide or it is a consequence of photobleaching due to extensive light exposure.

The apparent rate constants obtained from the single- and double- exponential fittings were plotted as a function of the nucleotide concentration (Fig. 3.27).

In the case of hBrr2^{NC}, the concentration dependence of k_{appNCb} deviated from a linear behavior. Instead, the k_{appNCb} values tended to saturation at high nucleotide concentration, showing a hyperbolic dependency. Such concentration dependency could indicate that the fluorescence change observed is a monomolecular rearrangement, rather than a bimolecular binding reaction, and would be consistent with a two-step binding mechanism (Fig. 3.27A):

$$\text{hBrr2}^{NC} + \text{ATP}\gamma\text{S} \rightleftharpoons \text{hBrr2}^{NC}\cdot\text{ATP}\gamma\text{S} \rightleftharpoons \text{hBrr2}^{NC}\cdot\text{ATP}\gamma\text{S}^*.$$

In this case, the first bimolecular reaction:

$$\text{hBrr2}^{NC} + \text{ATP}\gamma\text{S} \rightleftharpoons \text{hBrr2}^{NC}\cdot\text{ATP}\gamma\text{S},$$

is too fast to be measured or does not result in an appreciable fluorescence change, whereas the first-order reaction (the rearrangement corresponding to the second step):

$$\text{hBrr2}^{NC}\cdot\text{ATP}\gamma\text{S} \rightleftharpoons \text{hBrr2}^{NC}\cdot\text{ATP}\gamma\text{S}^*,$$

results in an increase of fluorescence intensity with velocity corresponding to k_{appNCb}. Such accommodation step observed for hBrr2^{NC} is not observed for hBrr2^{CC} although nucleotides bound to the N and C-terminal cassettes in Brr2^{HR} structure assume similar conformations, i.e. nucleotide binding is mainly fostered by motifs from the first RecA domain lacking interactions with RecA-2 (Fig. 3.18B and C).

According to k_{appNCa}, k_{appNCb} and to the proposed model, we assume that binding of nucleotide to hBrr2 is significantly faster than the following rearrangement. In this case, we imply that the first step is rapid compared to the second step and, therefore, the equilibrium dissociation constant for the first step, K_{SNC}, can be calculated from the nucleotide concentration dependencies of the k_{appNCb} values according to the equation:

$$k_{appNCb} = k_{-NCb} + k_{NCb}\frac{[ATP\gamma S]}{[ATP\gamma S] + K_{SNC}}$$ (Fig. 3.27).

From the same equation, we can also obtain the forward and reverse rates of nucleotide accommodation, k_{NCb} and k_{-NCb}. The equilibrium dissociation constant of the second step represented by nucleotide accommodation, $K_2'_{NC}$, was calculated as a ratio between forward and reverse rates, k_{-NCb}/k_{NCb}. The overall dissociation constant K_d for ADP and ATP could not be derived from the amplitude dependencies on nucleotide concentration. Therefore, the global K_d for the N-cassette interaction with nucleotides was then calculated using the following equation: $K_d = K_{SNC} \times K_2'_{NC}$. The equilibrium dissociation constants for the first step

(K_{SNC}) were 6.17±1.83 and 5.74±2.0 µM and for the second step ($K_2'_{NC}$) were 0.50±0.05 and 0.22±0.1 for ATPγS and ADP respectively. Furthermore, these data indicate that the second step greatly contributes to the overall affinity for the nucleotides. Altogether, our structural and biochemical data suggest that, during the second step of nucleotide binding, cooperative conformational changes are triggered resulting in kinetic stabilization of the nucleotide bound form of hBrr2^{NC}.

Differently from the isolated cassette construct, hBrr2^{NC}, k_{appNC} of nucleotide binding to hBrr2 variants containing the C-terminal cassette were linearly dependent on the nucleotide

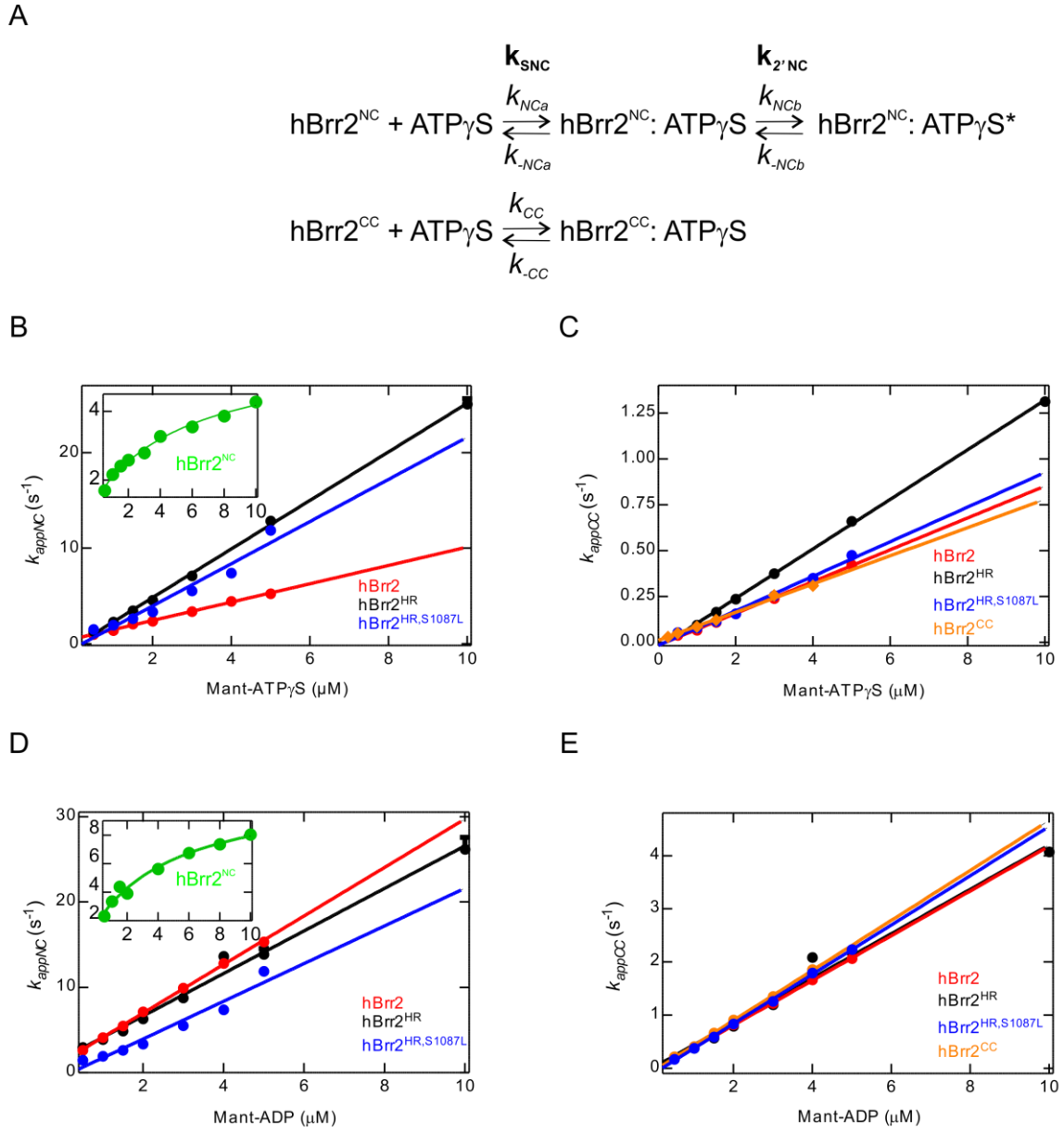


Fig. 3.27: (A) A two-step model illustrating ATP binding to hBrr2^{NC} (first step, k_{NCa}) and the subsequent accommodation (second step, k_{NCb}) (top) and a model depicting ATP binding to hBrr2^{CC} (bottom). Concentration dependency of the apparent rate constants for *mant*-ATPγS (B) and *mant*-ADP (D) binding to the N-terminal cassette of hBrr2 (red), hBrr2^{HR} (black), hBrr2^{HR,S1087L} (blue) and hBrr2^{NC} (green). Concentration dependency of the apparent rate constants for *mant*-ATPγS (C) and *mant*-ADP (E) binding to the C-terminal cassette of hBrr2

(red), hBrr2^{HR} (black), hBrr2^{HR,S1087L} (blue) and hBrr2^{CC} (orange). Error bars represent standard errors of the mean.

concentration, consistent with a single bimolecular event and indicating an influence of the C-terminal cassette on nucleotide binding by the N-terminal cassette. Furthermore, the measured apparent rate constants corresponding to k_{appCC} of hBrr2^{CC} together with hBrr2, hBrr2^{HR} and hBrr2^{HR,S1087L} also depended linearly as a function of the nucleotide concentration.

The association rate constants, k_{NC} and k_{CC}, were determined from the slope of the linear dependence of k_{app} on the concentration of *mant*-nucleotides. The intercept with the y-axis provided k_{-NC} and k_{-CC}. When the intercept with the y-axis was close to zero, the value of k_{-NC} and k_{-CC} could not be determined with precision therefore dissociation experiments were conducted.

The slope of the linear fitting of k_{appNC}, corresponding to the binding of ATPγS to the N-terminal cassette in the context of hBrr2, hBrr2^{HR}, hBrr2^{HR,S1087L}, indicated association rate constants k_{NC} of 0.96±0.02, 2.54±0.03 and 2.21±0.23 μM^{-1} s^{-1}, respectively. These rates, with the possible exception of k_{NC} for hBrr2, are similar to the ones obtained for ADP binding to the N-terminal cassette of different variants of hBrr2, which was about 2.5±0.11 μM^{-1} s^{-1}.

The hBrr2, hBrr2^{HR}, hBrr2^{HR,S1087L} rate constants for ATPγS binding to the C-terminal cassette, k_{CC}, were in the order of 0.1±0.01 μM^{-1} s^{-1}. We could observe that ADP binding to the C-terminal cassette of hBrr2, hBrr2^{HR} and hBrr2^{HR,S1087L} was faster compared to ATPγS, with association rates k_{CC} of about 0.45±0.01 μM^{-1} s^{-1}. This observation is in perfect agreement with the k_{CC} obtained for ATPγS and ADP binding to the isolated hBrr2^{CC}, 0.080±0.002 and 0.47±0.01 μM^{-1} s^{-1}, respectively. The linear concentration dependency of the k_{apps} is consistent with a bimolecular binding reaction, that, in the case of hBrr2^{CC}, could be represented as follows:

$$\text{hBrr2}^{CC} + \text{ATP}\gamma\text{S} \rightleftharpoons \text{hBrr2}^{CC}\cdot\text{ATP}\gamma\text{S}.$$

Unfortunately, we still cannot propose a full mechanism for nucleotide turnover for hBrr2 or hBrr2^{HR}. Apart from the difficulties involved in resolving the very rapid phase observed in the reactions with hBrr2 or hBrr2^{HR}, we presently do not know if nucleotide pre-bound to the C-terminal cassette could influence the velocity of nucleotide exchange or modulate nucleotide preference of the N-terminal cassette.

Nucleotide dissociation rate constants, k_{-NC} and $k_{-CC,}$ were determined upon mixing hBrr2·*mant*-ATPγS or hBrr2·*mant*-ADP with an excess of the respective unlabeled nucleotide. The release of the labeled nucleotide from hBrr2 resulted in fluorescence decrease. In this context, the rate by which the fluorescence decreases reflects the dissociation

rate constant of *mant*-ATPγS or *mant*-ADP from hBrr2, since, in the presence of a large excess of unlabeled nucleotide, rebinding of labeled nucleotide is negligible. Upon dilution of hBrr2^{NC}·*mant*-nucleotide with buffer, some *mant*-nucleotide dissociated. However, this dissociation impelled by the dilution did not occur to the same extent as in the presence of the dark nucleotide competitor (Fig. 3.26). Although we lack a complete description of the nucleotide binding mechanism to the N-terminal cassette, the observation of a single dissociation step for hBrr2^{NC} may suggest that this step is rate limiting in nucleotide turnover. However, we cannot exclude other reactions occurring upstream of the pathway to be slower, i.e. phosphate release.

The hBrr2^{NC} dissociation rates (k_{-NC}) for ATPγS and ADP were 1.62±0.016 and 1.89±0.1 s^{-1}. ATPγS and ADP dissociation rates for hBrr2^{CC} (k_{-CC}) were about 0.0020±0.0002 s^{-1}, extremely slow compared to k_{-NC}. Two phases could be distinguished during nucleotide dissociation from hBrr2, hBrr2^{HR}, hBrr2^{HR,S1087L}. The first phase is represented by a fast dissociation rate constant comparable to that for the N-terminal cassette, k_{-NC}. The nucleotide dissociation rates related to the N-terminal cassette for hBrr2, hBrr2^{HR} and hBrr2^{HR,S1087L} were very similar, 1.40±0.04 s^{-1} for ATPγS and 1.37±0.02 s^{-1} for ADP, respectively. The second phase had a dissociation rate constant similar to the one associated with the very slow release of nucleotide from hBrr2^{CC}, k_{-CC}. Nucleotide release from the C-terminal cassette in the context of hBrr2, hBrr2^{HR} and hBrr2^{HR,S1087L} was very slow, about 0.0020±0.0002 s^{-1}.

The K_d values for the interaction of ATPγS and ADP to hBrr2, hBrr2^{HR}, hBrr2^{HRS1087L}, hBrr2^{CC} were calculated from the association ($k_{NC,}$ *kcc*) and dissociation rate constants (k_{-NC} and k_{-CC}), using either k_{-NC}/k_{NC} *or* k_{-CC}/k_{CC}. The K_d's for the Brr2^{NC}·nucleotide complexes were 3.08±0.55 µM and 1.26±0.18 µM for ATPγS and ADP, respectively, while the K_d's for N-terminal cassette·nucleotide complexes in the context of hBrr2, hBrr2^{HR}, hBrr2^{HRS1087L} were about 0.50±0.07 µM for both nucleotides. This indicates that, in the absence of the C-terminal cassette, the N-terminal cassette exhibits lower affinities for nucleotides. The K_d's of the C-terminal cassette·nucleotide complexes are similar for all Brr2 variants measured, about 0.01±0.001 µM.

The results described above show that binding of ATPγS and ADP to the N-terminal cassette is faster compared to the velocity of nucleotide association with the C-terminal cassette. The same trend is observed for nucleotide dissociation; ATPγS and ADP release from the N-terminal cassette is much faster for the N-terminal cassette compared to the C-terminal cassette. On the other hand, the differences in the mechanism of nucleotide binding to N-terminal cassette in the absence or presence of the C-terminal cassette may indicate

conformational constrains in the active domain of Brr2 that are released by covalent and surface contacts with the C-terminal cassette. In the absence of the C-terminal cassette, the N-terminal cassette seems to rearrange before achieving a kinetically stable complex with adenosine nucleotides. In the presence of the C-terminal cassette, it is possible that the N-terminal cassette may be already conformationally oriented to accept the nucleotide in a stable manner. These solution studies also verified our structural finding that the inactive C-terminal cassette binds nucleotides, with a K_d that suggests it will be saturated at cellular concentrations.

The extremely slow nucleotide dissociation from the C-terminal cassette reflects its higher affinity for ATPγS or ADP. Considering the slow nucleotide release from the C-terminal cassette and the high concentration of ATP in the nucleus compared to ADP, we suggest that ATP is stably bound to the C-terminal cassette throughout a whole splicing cycle. This scenario may be tuned by other proteins, interacting with Brr2 in the spliceosome, functioning as adenine nucleotide exchange factors for Brr2.

Table 3.2: Rate constants and equilibrium dissociation constants of interactions between hBrr2 or fragments thereof and nucleotides.

See next page.

	k_{NC} ($\mu M^{-1} s^{-1}$)		k_{-NC} (s^{-1})		K_{dNC} (μM)		k_{CC} ($\mu M^{-1} s^{-1}$)		k_{-CC} (s^{-1})		K_{dCC} (μM)	
	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP
hBrr2	0.96±0.02	2.84±0.03	1.53±0.04	1.53±0.02	1.60±0.05	0.54±0.05	0.09±0.01	0.42±0.01	10^{-3}±10^{-4}	10^{-3}±10^{-4}	0.01±10^{-3}	2×10^{-3}±10^{-4}
hBrr2^{HR}	2.54±0.03	2.49±0.11	1.19±0.04	1.30±0.02	0.47±0.07	0.52±0.14	0.14±0.01	0.42±0.02	9×10^{-4}±10^{-5}	10^{-3}±10^{-4}	0.006±10^{-4}	2×10^{-3}±10^{-4}
hBrr2$^{HR\text{-}S1087L}$	2.21±0.23	2.73±0.08	1.40±0.04	1.49±0.03	0.63±0.06	0.54±0.09	0.09±0.01	0.46±0.01	10^{-3}±2.10^{-4}	10^{-3}±10^{-4}	0.01±10^{-3}	2×10^{-3}±10^{-4}
hBrr2^{CC}	-	-	-	-	-	-	0.080 ± 10^{-3}	0.47±0.01	2×10^{-3}±10^{-5}	2×10^{-3}±10^{-4}	0.025±10^{-3}	4×10^{-3}±10^{-4}
	$K_{S\,NC}$ (μM)		$K_{2\,NC}$		k_{NCb} (s^{-1})		k_{-NCb} (s^{-1})		k_{-NC} (s^{-1})		K_{dNC} (μM)	
	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP	ATPγS	ADP
hBrr2^{NC}	6.17±1.83	5.74±2.0	0.50±0.05	0.22±0.01	4.38±0.43	9.68±1.42	1.47±0.14	1.79±0.50	1.62±0.02	1.89±0.1	3.08±0.55	1.26±0.18

4. Discussion

We have established the expression of yeast and human Prp8, Snu114 and Brr2 using synthetic, expression-optimized genes and the MultiBac system. We could co-express and co-purify a complex of full-length yPrp8 and ySnu114. Additionally, we were able to reproducibly express in large amounts and purify full length Brr2 from both organisms as well as truncations thereof. Notably, using limited proteolysis, we have defined a stable 200 kDa portion of hBrr2 comprising both helicase cassettes. This fragment yielded crystals that diffracted to beyond 2.7 Å resolution.

We have solved the crystal structure of the entire human Brr2 helicase region, clarifying how two helicase cassettes interact and are arranged with respect to each other in a tandem SF2 enzyme. We also provide a structure of the RP33-linked S1087L variant of $Brr2^{HR}$ which, along with the wild type structure, offers possible explanations for the dysfunction of this Brr2 mutant as well as other Brr2 variants that have been investigated in the past. Furthermore, our work revealed the mode of binding of adenine nucleotides as well as nucleotide preferences of both cassettes.

Guided by these structures, we have interrogated which structural elements and organizational aspects are fundamental for the enzyme mechanism using mutational analyses. The results allowed speculation about novel regulatory principles underlying Brr2 function. We have also performed pre-steady state kinetic studies to understand how ATP binds to the two different nucleotide binding pockets of hBrr2. While our work still does not clarify all the details of the Brr2 helicase mechanism and of the precise roles of this enzyme during pre-mRNA splicing, we have provided a solid framework on which to interpret mechanistic studies.

4.1 Brr2 structure as a tandem repeat of two expanded Hel308 modules

Our structural studies show that Brr2 consists of two consecutive Hel308-like modules expanded by IG domains. The structures of other Ski2-type helicases, Hjm (a DNA helicase that binds fork-related Y-structured DNAs unwinding their double stranded regions) and Mtr4 (a RNA helicase working as an auxiliary factor of the exosome), as well as of Prp43 (a DEAH helicase involved in splicing and ribosome biogenesis) also revealed a similar domain

architecture as Hel308 (Walbott *et al.*, 2010; Weir *et al.*, 2010; Oyama *et al.*, 2009). Interestingly, the sequence similarities to Hel308 beyond the RecA domains of Hjm, Mtr4, Prp43 and Brr2 were not previously recognized. It seems that the Hel308 domain organization is a common architecture used by both Ski2-like and DEAH helicases and raises the possibility that these helicases may share structural and potentially mechanistic similarities in regions beyond their helicase domains.

The tandem RecA domains of nucleic acid helicases bear conserved sequence motifs to bind and hydrolyze nucleotide triphosphates (Q, I and II in the first RecA domain; VI in the second RecA domain), to bind nucleic acids (Ia, Ib and Ic in the first RecA domain; IV, IVa, V and Vb in the second RecA domain) and to coordinate nucleic acid and nucleotide binding (III in the first RecA domain; Va in the second RecA domain) (Jankowsky, 2011). Corresponding motifs could also be located in both cassettes of hBrr2. Due to the circular arrangement of the RecA, WH and HB domains in both $hBrr2^{HR}$ cassettes, the helicase motifs of the RecA domains are proximal to each other and form *bona fide* substrate binding sites (Fig. 3.15) unlike in many other helicase apo structures (Jankowsky and Fairman, 2007).

Apart from the conserved RecA-like catalytic modules, nucleic acid unwindases frequently harbor accessory domains that influence catalytic activities or mediate interaction with other factors. Upon sequence and structural analysis, it was shown that Brr2 encompasses helicase-associated Sec63 units. Previous mutagenesis studies suggested that the N-terminal RecA-like domains and the HB domain of the N-terminal Sec63 unit form a functional entity and that this architecture serves to integrate a ratcheting device on the Sec63 unit (HB) with the catalytic RecA-like domains (Pena *et al.*, 2009; Zhang *et al.*, 2009), thereby coupling the ATPase activity to nucleic acid translocation. The overall similarity to Hel308, including the presence of a ratchet in the HB and the WH as a connecting element between RecA domains and HB, may suggest that Brr2 can processively plow through extended RNA duplexes. This potential processivity is attractive, considering that Brr2 needs to unwind U4/U6, which contains long stem regions and is highly stable in yeast (Brow and Guthrie, 1988). However, there is presently no direct evidence showing that Brr2 is a processive helicase.

Brr2 additionally contains 60 conserved residues preceding the first RecA domain that tightly encircle the N-terminal cassette (Fig. 3.12). The reduction in helicase activity observed upon deletion of this N-terminal expansion suggests that it supports a productive domain organization or may be involved in substrate binding. Differently structured and positioned N-terminal expansions have also been seen in the SF2 helicases Mtr4 (Weir *et al.*, 2010) and

Prp43 (Walbott *et al.*, 2010). In Mtr4, a 60-residue extension N-terminal to RecA-1 folds into a long β-hairpin loop. This N-terminal β-hairpin packs with its β-strands against the first RecA and latches onto the second RecA. Thus, the active conformation of the two RecA domains in Mtr4 seems to be stabilized by intramolecular interactions *via* the N-terminal β-hairpin on one side and by contacts between the HB and the RecA domains on the opposite side. The N-terminal extension of Prp43 consists of 94 residues encompassing three α-helices. Helix 1 packs between the WH and HB domains whereas helices 2 and 3 interact with RecA-1. These helices are separated by an ordered 20 residue linker that braces the helicase ring structure, apparently once more supporting a productive domain conformation. Thus, although these N-terminal extensions are poorly conserved and may vary considerably in length between yeast and human proteins, they seem to serve similar functions. Topologically, the N-terminal extension at the N-terminal cassette of Brr2 is equivalent to the inter-cassette linker at the C-terminal cassette. The extensive interaction of the linker with the C-terminal cassette and its lack of exposed loops explain why it was not cut during limited proteolysis.

Both cassettes of Brr2 interact through an extensive interface (Fig. 3.16). The N-terminal IG domain is fitted squarely between the C-terminal RecA-2 and WH domains. An additional inter-cassette contact area ensues between the N-terminal RecA-1 and WH domains and the C-terminal RecA-2 domain. This intimate cassette interaction is likely a hallmark of all Brr2 enzymes and functionally important, considering that the majority of interacting side chains are conserved among Brr2 orthologs (Fig 3.13). Additionally, it explains why deletion of the C-terminal cassette or RecA and WH domains is detrimental to the overall stability of yBrr2 *in vivo* (Zhang *et al.*, 2009).

4.2 RNA binding by Brr2

In SF2 helicases, nucleic acid substrates preferentially bind with a 3'-to-5' directionality across the first and second RecA like domains, which could explain the 3'-to-5' unwinding directionality commonly observed with these enzymes (Hopfner and Michaelis, 2007). Particularly, a 3'-to-5' translocation has been experimentally confirmed for Hel308 (Guy and Bolt, 2005), Mer3/HFM1 (Nakagawa and Kolodner, 2002) and Mtr4 (Bernstein *et al.*, 2008), which are all members of the Ski2-like subfamily to which Brr2 belongs (Bleichert and Baserga, 2007; Jankowsky, 2011). Therefore, Brr2 is also expected to translocate in the 3'-to-5' direction on its nucleic acid substrates. This mode of engaging the RNA fits our current model of how Brr2 accommodates its substrate (Fig. 3.19), which is further supported by several pieces of evidence. First, Hel308-like contacts to a single-stranded RNA were also

seen in the related Ski2-like helicase Mtr4 (Weir *et al.*, 2010). Second, the entrance and walls of the N-terminal tunnel are positively charged, suitable for interaction with the negatively charged RNA backbone (Fig. 3.15). Third, defective Hel308-like RNA binding at the N-terminal cassette can explain the loss of activities in several characterized Brr2 variants (Fig. 3.12. Fourth, a Hel308-like mode of RNA binding at the N-terminal cassette could also explain the decreased activity we observed for the S1087L mutant compared to hBrr2^{HR} (Fig. 3.23).

In the context of the spliceosome, Brr2 has been suggested to translocate on U4 during catalytic activation (Frilander and Steitz, 2001). However, the 3'-end of U4 is occluded by secondary structures and/or bound proteins in the U4/U6 di-snRNA (Achsel *et al.*, 1999; Leung *et al.*, 2011). In agreement with the general trend for SF2 helicases, the hBrr2^{HR} structure suggests that the enzyme could possibly load onto an internal single-stranded region of U4 downstream of stem I by opening the N-terminal cassette between the RecA-2 and HB domains.

Consistent with our modeling exercise and mutational analyses, recent data demonstrated that only the N-terminal cassette of Brr2 form RNA interactions and that Brr2 initiates U4/U6 disruption by translocating on U4 and unwinding U4/U6 stem I (Hahn *et al.*, 2012). Additionally, Brr2 was shown to specifically bind to the single-stranded region of U4 preceding U4/U6 stem I supporting our structural findings. This same region on U4 is recognized by the RNase H domain of Prp8, indicating that the RNase H domain interferes with U4/U6 unwinding by blocking Brr2's interaction with U4 snRNA (Mozaffari-Jovin *et al.*, 2012).

4.3 The C-terminal cassette as an intramolecular cofactor

Brr2 is one of the few known helicases (together with Slh1 (Martegani *et al.*, 1997) and ASCC3 (Dango *et al.*, 2011)) that contain two helicase-like modules. The function of the second helicase-like module has long been elusive and intriguing. The C-terminal cassette displays several substantial deviations in the canonical helicase motifs. The glutamate of the typical motif II, DExD/H, is replaced with an aspartate (DDAH) in the C-terminal cassette yBrr2. This glutamate has been postulated to be the key catalytic residue that activates a water molecule to hydrolyze NTP in DExD/H-box proteins and other helicases (Cordin *et al.*, 2006; Sengoku *et al.*, 2006; Caruthers and McKay, 2002). Likewise, the Ser-Ala-Thr (SAT) residues in motif III are replaced with SNC or SSS in the case of yeast and human sequences, respectively. The SAT residues are thought to participate in interdomain interactions between

the two RecA domains upon ATP and RNA binding and to play a role during unwinding by DExD/H-box proteins (Sengoku *et al.*, 2006; Zhang *et al.*, 2009). The C-terminal cassette also lacks obvious motifs IVa and V.

Most enzyme families contain inactive members thought to have arisen by gene duplication events followed by subsequent accumulation of inactivating mutations. Evolutionary conservation suggests that such pseudoenzymes are functionally important. However, in most cases, their functions are unknown (Adrain and Freeman, 2012). During this work, we have shown that the C-terminal cassette of Brr2 is a pseudo-helicase that has been converted into an intramolecular regulator of a neighboring, similarly structured active helicase. This non-enzymatic role of the C-terminal cassette is in agreement with its non-canonical ATPase and helicase motifs (Noble and Guthrie, 1996; Lauber *et al.*, 1996) and with previous genetics analyses showing that mutations in putative active site residues of the C-terminal cassette still support splicing (Kim and Rossi, 1999). However, our results additionally show that the C-terminal cassette can still bind ATP but specifically lost its ability to hydrolyze the nucleotide. Furthermore, we have been able to show that, apart from the substantial deviations in the canonical helicase motifs, the C-terminal cassette exhibits an increased barrier to adopt a hydrolytic conformation which seems to further consolidate the inactivity of this cassette.

Our modeling studies corroborated by mutational analyses suggest that the C-terminal cassette most likely also does not bind RNA during duplex unwinding. Our findings are consistent with recent results from (Hahn *et al.*, 2012). In agreement with this finding, the C-terminal cassette does not seem to rely on specific sequences or structures of its RNA substrate to modulate the N-terminal cassette activity since such stimulatory effect due to the presence of the C-terminal cassette was not only observed with U4/U6 di-snRNA as a substrate but also with a simple model duplex (Fig. 3.17).

Furthermore, our results identify a number of features that are required for the cofactor function of the C-terminal cassette and suggest mechanisms by which it may act. Our mutational studies show that direct inter-cassette contacts are essential for cassette communication. Due to their large contact area, the cassettes most likely mutually stabilize the specific conformational states they adopt in the apo form of $hBrr2^{HR}$. We suggest that crystal crosslinking has cemented these conformations considering that we do not observe any significant conformational changes in the nucleotide-bound states. Our nucleotide preparations obviously contained both ATP (or analog) and ADP. The outcome of our soaking experiments, therefore, indicates the nucleotide preference of the cassette

conformations in the apo form (ADP at the N-terminal cassette, Mg^{2+}-ATP at the C-terminal cassette). Bound ADP suggests that we observed the post-ATP hydrolysis and post-phosphate release state at the N-terminal cassette in agreement with the enhanced N-terminal ATPase activity in the presence of the C-terminal cassette. Thus, the C-terminal cassette may drive ATP hydrolysis and/or phosphate release at the N-terminal cassette, thereby facilitating associated changes in nucleic acid binding and unwinding.

Our structural analyses show that the C-terminal cassette preferentially binds Mg^{2+}-ATP in the presence of the N-terminal cassette, but we did not indicate whether it cycles between nucleotide-bound and free states during RNA unwinding. However, our preliminary pre-steady state kinetic studies on nucleotide binding to hBrr2 indicated that nucleotide release from the C-terminal cassette binding pocket is very slow suggesting that ATP is stably bound to the C-terminal cassette when Brr2 is not in the presence of RNA. Since the adoption of the hydrolytic conformation is hindered at the C-terminal cassette, it seems to be conformationally more restricted than the N-terminal cassette and may remain stably associated with the nucleotide. The function of the nucleotide at the C-terminal cassette may thus be to rigidify its structure and allow it to act as a scaffold on which the N-terminal cassette could efficiently undergo conformational changes required for duplex unwinding.

The C-terminal cassette may also exploit inter-cassette contacts to directly influence the positioning of active site domains in the N-terminal cassette. Interactions between the HLH and HB domains are important for duplex unwinding in the related Hel308 (Richards *et al.*, 2008; Woodman *et al.*, 2007). In Brr2, the N-terminal IG domain intervenes between the HLH and HB domains and is connected to the upper part of the inter-cassette linker. Mutations in the linker affected $Brr2^{HR}$ activity both negatively and positively (Fig. 3.21). It is conceivable that different functional states (such as ATP bound, ATP hydrolyzed, ADP bound) are associated with different relative orientations of the cassettes and that such conformational changes may be transmitted *via* the linker and the IG domain to the N-terminal HLH and HB domains, thereby modulating the activity of the N-terminal helicase.

It has been suggested that Brr2 may acquire processivity by oligomerization of its ATPase units in a fashion analogous to ring-forming helicases (Staley and Guthrie, 1998). While our data indeed indicate that Brr2's full helicase activity only unfolds given a direct interaction of the N- and C-terminal RecA domains, the linear disposition of these four domains in Brr2 is decisively different from the ring-like arrangement of the corresponding domains in hexameric helicases. Furthermore, our results also suggest that the C-terminal RecA domains do not engage RNA during unwinding. Rather, the dependence of the stimulation on inter-

cassette RecA contacts may be explained by stabilization of the N-terminal RecA and WH domains and by their positioning relative to other active site elements.

4.4 Remote regulation of the N-cassette activity

Mutations that interfere with ATP binding at the C-terminal cassette, which is remote from the N-terminal active site and from the cassette interface, also exhibit strong effects on the N-terminal helicase. This observation demonstrates that, in principle, ligand binding at the C-terminal cassette can be sensed by the N-terminal helicase, which is consistent with our preliminary data showing that the N-terminal cassette binds nucleotides differently in the presence of the C-terminal cassette. While we presently cannot trace this long range communication on the atomic level, it is likely also conducted through the direct inter-cassette RecA or linker contacts discussed above.

Mutational analyses furthermore support that Brr2 is not simply optimized for helicase activity but for versatile regulation on various levels. The inter-cassette communication uncovered here suggests novel regulatory functions for at least some of the many spliceosomal proteins that interact at the C-terminal cassette of Brr2, which include factors essential for different steps of splicing catalysis (Van Nues and Beggs, 2001; Liu *et al.*, 2006). The ability of the C-terminal cassette to transmit signals to the N-terminal cassette suggests that these proteins may not merely use the C-terminal cassette as a passive landing pad but also to influence the N-terminal cassette from a distance, by modulating its nucleotide binding, hydrolysis, RNA binding or coupling of ATPase and helicase activities. The observation that, while many mutations reduced Brr2 helicase activity, one variant (PPP1296–8AAA in the linker) exhibited significantly enhanced unwinding activity indicates that interacting factors may either down or up-regulate Brr2. Sequentially binding proteins may thus switch the enzyme on or off as required during particular phases of the splicing process.

The above principles also offer one possible solution to the intriguing problem of how a large number of factors can influence alternative splicing. Several of these proteins may directly or indirectly target the expanded surface provided by the C-terminal cassette to modulate Brr2 activity, which would affect splicing kinetics and consequently the choice or proofreading of alternative splice sites. Analogous kinetic switches, influencing splice site choice *via* the modulation of U1 snRNP interaction with 5' SS, have recently been uncovered (Yu *et al.*, 2008). This study showed that the presence of silencer sequences altered the U1 snRNP/5' SS complex rendering it less able to engage U2 snRNP. These silencer sequences affect the fast kinetic step representing the joining of the 5' and 3' SS in a complex committed

to splicing. These findings demonstrate that minor perturbations of spliceosomal protein-protein or protein-RNA interactions can alter delicate kinetic balance thereby influencing splice site choices.

We also note that similar regulatory principles as in Brr2 may also be at work in other proteins from diverse cellular contexts, which are composed of active and inactive nucleotide binding domains, such as the membrane associated guanylate kinase CASK (Mukherjee *et al.*, 2011) or the cystic fibrosis transmembrane conductance regulator (CFTR) (Aleksandrov *et al.*, 2002; Basso *et al.*, 2003). CASK is composed of an N-terminal CaM-kinase domain and a C-terminal set of domains typical of membrane-associated guanylate kinases, including a guanylate kinase domain that is catalytically inactive and supposed to function as a protein interaction platform. Interestingly, the phosphorylation rate of the active CASK CaM-kinase domain is dramatically increased in the context of the full-length protein compared to the CaM-kinase domain alone. CFTR contains two nucleotide binding domains that share low sequence similarity and exhibit functional asymmetry. The N-terminal domain displays unorthodox residues in the nucleotide binding site and is catalytically inactive although it binds ATP. In contrast, the C-terminal nucleotide binding domain binds and hydrolyses ATP. It has been suggested that cooperative interaction between N- and C-terminal domains results in functional coupling of nucleotide binding at the two domains and channel gating.

4.5 Molecular basis for the RP33 form of retinitis pigmentosa

The present structure rationalizes the dysfunction of many previously investigated Brr2 variants. In particular, effects of mutations that have been recently linked to RP33 (Zhang *et al.*, 2009; Benaglio *et al.*, 2011) are seen to interfere with hBrr2 function.

Mapping RP33-linked Brr2 mutations on the structure suggested that these mutations interfere in diverse ways with Brr2 activity as shown directly for the S1087L mutation. Leucine 1087 is located in the ratchet helix of the N-terminal HB domain and thus may exert its effect through dysfunctional RNA binding or translocation at the N-terminal cassette. Since the S1087L mutation has no discernable effect on the folding of Brr2, it is likely that at least some of the RP33-linked Brr2 variants are incorporated into spliceosomes *in vivo*. The presence of such RP33-linked variants may result in a slow-down of spliceosome catalytic activation due to impairment of hBrr2 activity. This inefficiency of the splicing reaction is likely tolerated by the majority of human tissues. However, RP33-linked mutations may be

critical for cell survival in tissues that require particularly elevated splicing activity, as observed in the retina (Tanackovic *et al.*, 2011). Alternatively, a retina-specific splicing event may no longer be supported by the reduced levels of the splicing machinery (Yin *et al.*, 2011; Linder *et al.*, 2011).

5. Outlook

Already before incorporation into the spliceosome, Brr2 is associated with its U4/U6 substrate in the context of the U4/U6.U5 tri-snRNP. Brr2 unwinds the U4/U6 di-snRNA *in vitro* (Laggerbauer *et al.*, 1998; Raghunathan and Guthrie, 1998) and it is required for spliceosome activation *in vivo* (Kim and Rossi, 1999). However, U4/U6 unwinding has to be delayed until splice sites have been reliably located. Brr2 also acts a second time during spliceosome disassembly (Small *et al.*, 2006). Therefore, Brr2 apparently has to be turned on and off several times during a splicing event. Our crystal structure of $hBrr2^{HR}$ provided insights into the mechanism of action whereby the two ring-like helicase domains of Brr2 cooperate to unwind the U4/U6 snRNA duplex. The C-terminal helicase unit, while inactive in ATP hydrolysis and RNA duplex unwinding, strongly stimulates the N-terminal helicase. Using structure-guided mutagenesis, we delineated communication lines between the cassettes required for this modulation. Using pre-steady state kinetics, we also probed the nucleotide binding preferences and worked out possible nucleotide binding mechanisms of either cassette. Our results revealed the structural and functional interplay between two helicase cassettes in a tandem SF2 enzyme and suggested how Brr2 interactors may exploit the C-terminal cassette as a "remote control" to regulate the N-terminal helicase of the enzyme (Fig 5.1A and B).

However, several mechanistic aspects of Brr2 helicase action remain unclear, including the molecular mechanisms whereby the helicase may be switched on and off during splicing. Brr2 tightly interacts with Prp8 and Snu114 (Achsel *et al.*, 1998) and both of these U5 proteins have been shown to modulate Brr2 activity (Kuhn and Brow, 2000; Kuhn *et al.*, 2002; Small *et al.*, 2006). A C-terminal fragment of Prp8, which contains its RNase H-like and Jab1/MPN-like domains, has been shown to stimulate Brr2-mediated U4/U6 unwinding activity *in vitro*, but the mechanism whereby this is achieved is presently unknown (Maeder *et al.*, 2009; Pena *et al.*, 2009).

During this thesis, we could also demonstrate, using analytical gel filtration analysis, that both $yPrp8^{RNaseH\text{-}Jab1/MPN}$ as well as the Jab1/MPN domain of yPrp8 alone ($yPrp8^{Jab1/MPN}$) form stable complexes with recombinant yBrr2 (Fig. 5.2, Weber *et al.*, 2011). These initial findings paved the way for large scale production of human and yeast Brr2-$Prp8^{Jab1/MPN}$ complexes for structural analyses and further elucidation of the role of the $Prp8^{Jab1/MPN}$ domain in the regulation of Brr2 activity (Mozaffari-Jovin *et al.*, 2013).

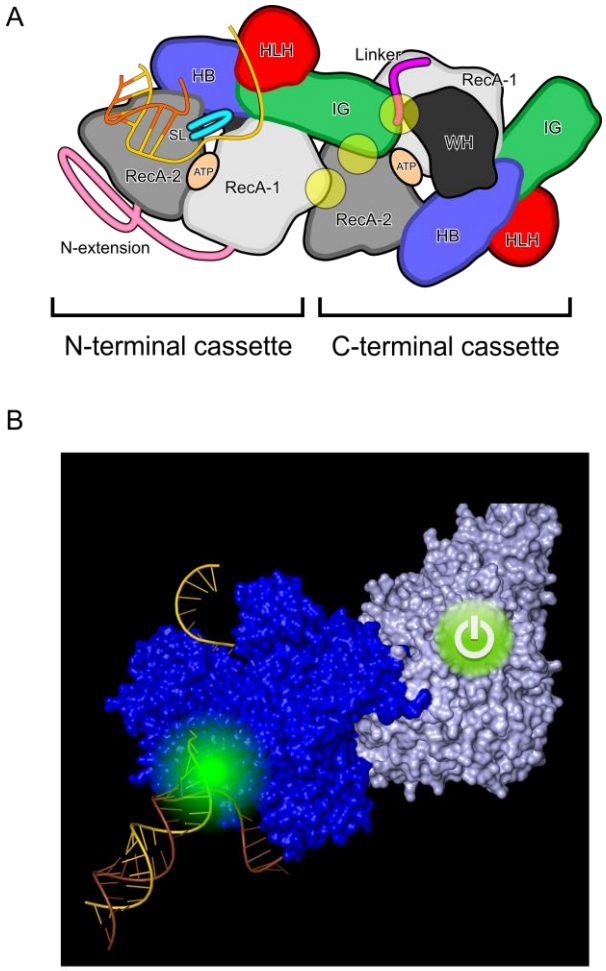


Fig 5.1: (A, B) Schematic representations of Brr2^{HR}. 2D scheme of Brr2^{HR} summarizing the main findings of this thesis: N- and C-terminal cassettes intimately interact and changes in the interaction surface modulate Brr2 activity (inter-cassette contacts are indicated by semitransparent yellow circles). Both cassettes bind nucleotides, although with different affinities. RNA interactions are mainly fostered by the N-terminal cassette. Domains and elements are colored as in Fig. 3.12 (B) The C-terminal cassette as an intramolecular power switch in the Brr2 helicase. RNA duplex unwinding by the active N-terminal cassette (blue) is stimulated by the C-terminal cassette (light blue). The RNA duplex is represented in gold and orange colors.

Finally, we have delineated a crystallizable fragment of hBrr2, hBrr2^{HR}, that can be further used as a platform for crystallization trials with RNA and other splicing factors. Indeed, steps in this direction have been successful, since our laboratory has recently obtained the structure of hPrp8$^{Jab1/MPN}$-Brr2^{HR} complex at 3.4 Å resolution (Mozaffari-Jovin *et al.*, 2013). We have also further optimized yBrr2 fragments based on hBrr2^{HR} that are now successfully used to obtain crystals of yBrr2 complexes. By means of rapid kinetic measurements, we have demonstrated that N- and C-terminal cassettes bind nucleotides with different affinities. These findings can be further complemented by studies investigating the influence of RNA, protein interactors (e.g. Prp8$^{Jab1/MPN}$) and mutations (e.g. in the cassette interface and ATP binding pocket) on nucleotide binding by Brr2.

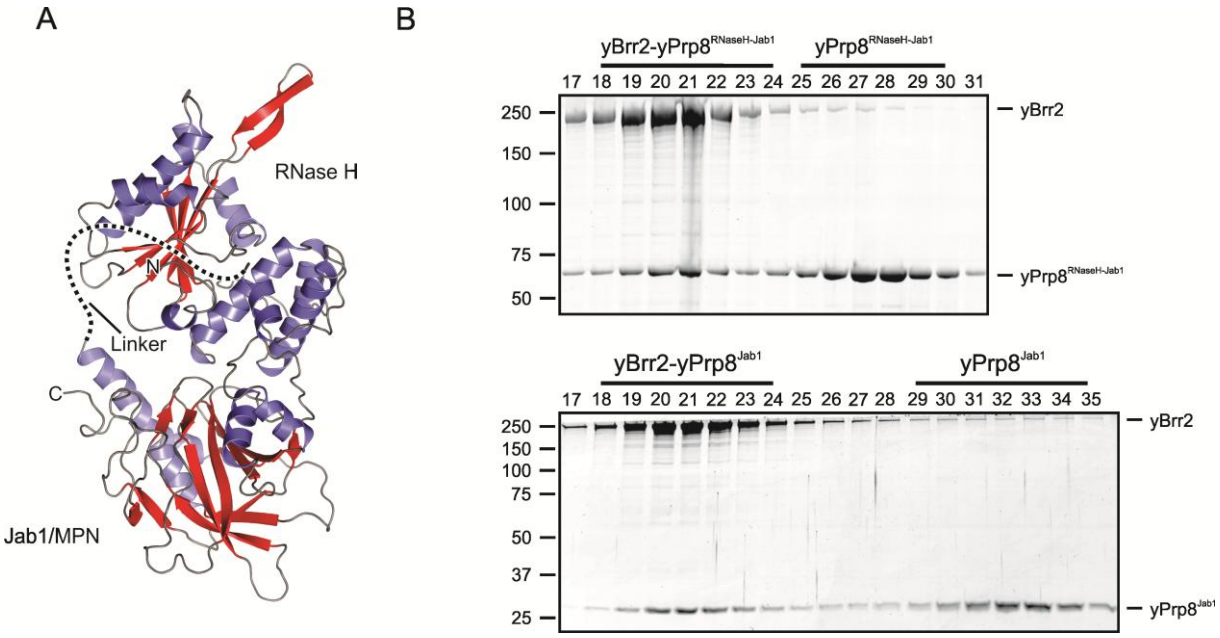


Fig. 5.2: (A) Ribbon plot of the crystal structure of yPrp8$^{RNaseH\text{-}Jab1/MPN}$. Helices – blue; strands – red; loops – gray. A disordered linker is indicated by a dashed lined. The domains and protein termini are labeled. (B) Coomassie-stained SDS gels of analytical gel filtration runs (Superdex 200) of a yBrr2/yPrp8$^{RNaseH\text{-}Jab1/MPN}$ mixture (top) and of a yBrr2/yPrp8$^{Jab1/MPN}$ mixture (bottom). Fractions eluting from the column are indicated above the gels; molecular weight markers are shown on the left. Both yPrp8$^{RNaseH\text{-}Jab1/MPN}$ and yPrp8$^{Jab1/MPN}$ elute together with yBrr2 (peak fractions 18-24). Excess of yPrp8$^{RNaseH\text{-}Jab1/MPN}$ and yPrp8$^{Jab1/MPN}$ elute as the individual proteins (peak fractions 25-30 on top panel and 29-35 on bottom panel). These results indicate that yBrr2 forms stable and presumably stoichiometric complexes with yPrp8$^{RNaseH\text{-}Jab1/MPN}$ and yPrp8$^{Jab1/MPN}$.

6. References

Abelson, J., Trotta, C. R., and Li, H. (1998). tRNA splicing. The Journal of Biological Chemistry *273*, 12685–12688.

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, R. (1998). The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. Molecular and Cellular Biology *18*, 6756–6766.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. The EMBO Journal *18*, 5789–5802.

Adrain, C., and Freeman, M. (2012). New lives for old: evolution of pseudoenzyme function illustrated by iRhoms. Nature Reviews. Molecular Cell Biology *13*, 489–498.

Alcid, E. A., and Jurica, M. S. (2008). A protein-based EM label for RNA identifies the location of exons in spliceosomes. Nature Structural & Molecular Biology *15*, 213–215.

Aleksandrov, L., Aleksandrov, A. A., Chang, X.-B., and Riordan, J. R. (2002). The First Nucleotide Binding Domain of Cystic Fibrosis Transmembrane Conductance Regulator Is a Site of Stable Nucleotide Interaction, whereas the Second Is a Site of Rapid Turnover. The Journal of Biological Chemistry *277*, 15419–15425.

Anantharaman, V., Koonin, E. V, and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Research *30*, 1427–1464.

Arenas, J. E., and Abelson, J. N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proceedings of the National Academy of Sciences of the United States of America *94*, 11798–11802.

Aronova, A., Bacíková, D., Crotti, L. B., Horowitz, D. S., and Schwer, B. (2007). Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. RNA *13*, 1437–1444.

Ast, G. (2004). How did alternative splicing evolve? Nature Reviews. Genetics *5*, 773–782.

Bach, M., Winkelmann, G., and Lührmann, R. (1989). 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. Proceedings of the National Academy of Sciences of the United States of America *86*, 6038–6042.

Barrass, J. D., and Beggs, J. D. (2003). Splicing goes global. TRENDS in Genetics *19*, 295–298.

Bartels, C., Klatt, C., Lührmann, R., and Fabrizio, P. (2002). The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome. EMBO Reports *3*, 875–880.

Basso, C., Vergani, P., Nairn, A. C., and Gadsby, D. C. (2003). Prolonged nonhydrolytic interaction of nucleotide with CFTR's NH2-terminal nucleotide binding domain and its role in channel gating. The Journal of General Physiology *122*, 333–348.

Behrens, S., and Lührmann, R. (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. Genes & Development *5*, 1439–1452.

Behzadnia, N., Golas, M. M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C. L., Urlaub, H., Stark, H., *et al.* (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. The EMBO Journal *26*, 1737–1748.

Bellare, P., Kutach, A. K., Rines, A. M. Y. K., Guthrie, C., and Sontheimer, E. J. (2006). Ubiquitin binding by a variant Jab1 / MPN domain in the essential pre-mRNA splicing factor Prp8p. RNA *12*, 292–302.

Bellare, P., Small, E. C., Huang, X., Wohlschlegel, J. A., Staley, J. P., and Sontheimer, E. J. (2008). A role for ubiquitin in the spliceosome assembly pathway. Nature Structural & Molecular Biology *15*, 444–451.

Benaglio, P., McGee, T. L., Capelli, L. P., Harper, S., Berson, E. L., and Rivolta, C. (2011). Next generation sequencing of pooled samples reveals new SNRNP200 mutations associated with retinitis pigmentosa. Human Mutation *32*, E2246–58.

Berglund, J. A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes & Development *12*, 858–867.

Bernstein, J., Patterson, D. N., Wilson, G. M., and Toth, E. A. (2008). Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3'->5' helicase partner of the nuclear exosome. The Journal of Biological Chemistry *283*, 4930–4942.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M. M., Sander, B., Will, C. L., Urlaub, H., Stark, H., and Lührmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA *16*, 2384–2403.

Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846–850.

Bieniossek, C., Richmond, T. J., and Berger, I. (2008). MultiBac: multigene baculovirus-based eukaryotic protein complex production. Current Protocols in Protein Science *51*.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annual Review of Biochemistry *72*, 291–336.

Bleichert, F., and Baserga, S. J. (2007). The long unwinding road of RNA helicases. Molecular Cell *27*, 339–352.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713–725.

Bono, F., and Gehring, N. H. (2011). Assembly, disassembly and recycling: The dynamics of exon junction complexes. RNA Biology *8*, 24–29.

Boon, K.-L., Grainger, R. J., Ehsani, P., Barrass, J. D., Auchynnikava, T., Inglehearn, C. F., and Beggs, J. D. (2007). prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. Nature Structural & Molecular Biology *14*, 1077–1083.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry *72*, 248–254.

Braunagel, S. C., Parr, R., Belyavskyi, M., and Summers, M. D. (1998). Autographa californica nucleopolyhedrovirus infection results in Sf9 cell cycle arrest at G2/M phase. Virology *244*, 195–211.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallographica Section D Biological Crystallography *59*, 2023–2030.

Bringmann, P., Rinke, J., and Appel, B. (1983). of snRNPs U1, U2, U4, U5 and U6 with 2, 2, 7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1) RNP. The EMBO Journal *2*, 1129–1135.

Brow, D. A. (2002). Allosteric cascade of spliceosome activation. Annual Review of Genetics *36*, 333–360.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. Acta Crystallographica Section D Biological Crystallography *54*, 905–921.

Burgess, S., Couto, J. R., and Guthrie, C. (1990). A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. Cell *60*, 705–717.

Burgess, S. M., and Guthrie, C. (1993). A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell *73*, 1377–1391.

Büttner, K., Nehring, S., and Hopfner, K.-P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. Nature Structural & Molecular Biology *14*, 647–652.

Carthew, R. ., and Sontheimer, E. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642–655.

Caruthers, J. M., and McKay, D. B. (2002). Helicase structure and mechanism. Current Opinion in Structural Biology *12*, 123–133.

Chakarova, C. F., Hims, M. M., Bolz, H., Abu-Safieh, L., Patel, R. J., Papaioannou, M. G., Inglehearn, C. F., Keen, T. J., Willis, C., Moore, A. T., *et al.* (2002). Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. Human Molecular Genetics *11*, 87–92.

Chan, R. T., Robart, A. R., Rajashankar, K. R., Pyle, A. M., and Toor, N. (2012). Crystal structure of a group II intron in the pre-catalytic state. Nature Structural & Molecular Biology *19*, 555–557.

Chan, S.-P., and Cheng, S.-C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. The Journal of Biological Chemistry *280*, 31190–31199.

Chan, S.-P., Kao, D.-I., Tsai, W.-Y., and Cheng, S.-C. (2003). The Prp19p-associated complex in spliceosome activation. Science *302*, 279–282.

Chapman, K., and Boeke, J. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. Cell *65*, 483–492.

Chen, J. H., and Lin, R. J. (1990). The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors. Nucleic Acids Research *18*, 6447.

Chen, J. Y., Stands, L., Staley, J. P., Jackups, R. R., Latus, L. J., and Chang, T. H. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Molecular Cell *7*, 227–232.

Cheng, S. C., and Abelson, J. (1987). Spliceosome assembly in yeast. Genes & Development *1*, 1014–1027.

Colgan, D. F., and Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes & Development *11*, 2755–2766.

Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature *349*, 487–493.

Cordin, O., Banroques, J., Tanner, N. K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. Gene *367*, 17–37.

Cordin, O., Hahn, D., and Beggs, J. D. (2012). Structure, function and regulation of spliceosomal RNA helicases. Current Opinion in Cell Biology *24*, 431–438.

Cowtan, K. D. (1994). “dm”: An Automated Procedure for Phase Improvement by Density Modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography *31*, 34–38.

Dango, S., Mosammaparast, N., Sowa, M. E., Xiong, L.-J., Wu, F., Park, K., Rubin, M., Gygi, S., Harper, J. W., and Shi, Y. (2011). DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation. Molecular Cell *44*, 373–384.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Molecular and Cellular Biology *26*, 5528–5543.

Deutsch, M., and Long, M. (1999). Intron-exon structures of eukaryotic model organisms. Nucleic Acids Research *27*, 3219–3228.

Dlakić, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. RNA *17*, 799–808.

Doma, M. K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature *440*, 561–564.

Dong, A., Xu, X., Edwards, A., and Chang, C. (2007). In situ proteolysis for protein crystallization and structure determination. Nature Methods *4*, 1019–1021.

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

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallographica Section D Biological crystallography *60*, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica Section D Biological crystallography *66*, 486–501.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Molecular Cell *36*, 593–608.

Fairbanks, G., Steck, T., and Wallach, D. (1971). Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry *10*, 2606–2617.

Fairman, M. E., Maroney, P. A., Wang, W., Bowers, H. A., Gollnick, P., Nilsen, T. W., and Jankowsky, E. (2004). Protein displacement by DExH/D “RNA helicases” without duplex unwinding. Science *304*, 730–734.

Fairman-Williams, M. E., Guenther, U.-P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Current Opinion in Structural Biology *20*, 313–324.

Fitzgerald, D., Berger, P., and Schaffitzel, C. (2006). Protein complex expression by using multigene baculoviral vectors. Nature Methods *3*, 1021–1032.

Fleckner, J., Zhang, M., Valcarcel, J., and Green, M. R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes & Development *11*, 1864–1872.

Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. (1986). Correlation between sites of limited proteolysis and segmental mobility in thermolysin. Biochemistry *25*, 1847–1851.

Fontana, A., and Laureto, P. P. de (1997). Probing the partly folded states of proteins by limited proteolysis. Folding and Design *2*, 17–26.

Fontana, A., De Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004). Probing protein structure by limited proteolysis. Acta Biochimica Polonica *51*, 299–321.

Fox-Walsh, K. L., Dou, Y., Lam, B. J., Hung, S.-P., Baldi, P. F., and Hertel, K. J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proceedings of the National Academy of Sciences of the United States of America *102*, 16176–16181.

Friesen, P. D., and Miller, L. K. (1987). Divergent Transcription of Early 35- and 94-Kilodalton Protein Genes Encoded by the HindIII K Genome Fragment of the Baculovirus Autographa californica Nuclear Polyhedrosis Virus. Journal of Virology *61*, 2264–2272.

Frilander, M. J., and Steitz, J. A. (2001). Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. Molecular Cell *7*, 217–226.

Frischmeyer, P. A., Van Hoof, A., O'Donnell, K., Guerrerio, A. L., Parker, R., and Dietz, H. C. (2002). An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science *295*, 2258–2261.

Girard, C., Verheggen, C., Neel, H., Cammas, A., Vagner, S., Soret, J., Bertrand, E., and Bordonné, R. (2008). Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. The Journal of Biological Chemistry *283*, 2060–2069.

Golas, M. M., Sander, B., Will, C. L., Lührmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. Molecular Cell *17*, 869–883.

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

Gottschalk, A., Kastner, B., Luhrmann, R., Fabrizio, P., Gottschalk, A., Kastner, B., Lührmann, R., and Fabrizio, P. (2001). The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. RNA *7*, 1554–1565.

Gozani, O., Potashkin, J., Reed, R., and Gozani, O. R. (1998). Potential Role for U2AF-SAP 155 Interactions in Recruiting U2 snRNP to the Branch Site. Molecular and Cellular Biology *18*, 4752–4760.

GRAINGER, R., and BEGGS, J. (2005). Prp8 protein: at the heart of the spliceosome. RNA *11*, 533–557.

Grainger, R. J., Barrass, J. D., Jacquier, A., Rain, J.-C., and Beggs, J. D. (2009). Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. RNA *15*, 2161–2173.

Graveley, B. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197–1211.

Guthrie, C., and Patterson, B. (1988). Spliceosomal snRNAs. Annual Review of Genetics *22*, 387–419.

Guy, C. P., and Bolt, E. L. (2005). Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. Nucleic Acids Research *33*, 3678–3690.

Häcker, I., Sander, B., Golas, M. M., Wolf, E., Karagöz, E., Kastner, B., Stark, H., Fabrizio, P., and Lührmann, R. (2008). Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. Nature Structural & Molecular Biology *15*, 1206–1212.

Hahn, D., and Beggs, J. D. (2010). Brr2p RNA helicase with a split personality: insights into structure and function. Biochemical Society Transactions *38*, 1105–1109.

Hahn, D., Kudla, G., Tollervey, D., and Beggs, J. D. (2012). Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. Genes & Development *26*, 2408–2421.

Halls, C., Mohr, S., Campo, M. Del, Yang, Q., and Lambowitz, A. M. (2007). Involvement of DEAD-box proteins in groups I and II intron splicing. Biochemical characterization of Mss116p, ATP-hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. Journal of Molecular Biology *365*, 835–855.

Haugen, P., Dawn, M. S., and Bhattacharya, D. (2005). The natural history of group I introns. TRENDS in Genetics *21*, 111–119.

He, Y., Andersen, G. R., and Nielsen, K. H. (2010). Structural basis for the function of DEAH helicases. EMBO Reports *11*, 180–186.

Heinrichs, V., Bach, M., Winkelmann, G., and Lührmann, R. (1990). U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. Science *247*, 69–72.

Herschlag, D. (1988). The Role of Induced Fit and Conformational Changes of Enzymes in Specificity and Catalysis. Bioorganic Chemistry *96*, 62–96.

Le Hir, H., Izaurralde, E., Maquat, L. E., and Moore, M. J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. The EMBO Journal *19*, 6860–6869.

Hoffman, B. E., and Grabowski, P. J. (1992). U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. Genes & Development *6*, 2554–2568.

Van Hoof, A., Frischmeyer, P. A., Dietz, H. C., and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science *295*, 2262–2264.

Hopfner, K.-P., and Michaelis, J. (2007). Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. Current Opinion in Structural Biology *17*, 87–95.

Inada, T., and Aiba, H. (2005). Translation of aberrant mRNAs lacking a termination codon or with a shortened 3'-UTR is repressed after initiation in yeast. The EMBO Journal *24*, 1584–1595.

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of Escherichia coli with plasmids. Gene *96*, 23–28.

Isken, O., and Maquat, L. E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes & Development *21*, 1833–1856.

Jackson, I. (1991). A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Research *19*, 3795–3798.

Jády, B. E., and Kiss, T. (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. The EMBO Journal *20*, 541–551.

Jameson, D., and Eccleston, J. (1997). Fluorescent Nucleotide Analogs: Synthesis and Applications. In Methods in Enzymology, pp. 363–390.

Jankowsky, A., Guenther, U.-P., and Jankowsky, E. (2011). The RNA helicase database. Nucleic Acids Research *39*, D338–41.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. TRENDS in Biochemical Sciences *36*, 19–29.

Jankowsky, E., and Fairman, M. E. (2007). RNA helicases--one fold for many functions. Current Opinion in Structural Biology *17*, 316–324.

Jankowsky, E., Fairman, M. E., and Yang, Q. (2005). RNA Helicases: Versatile ATP-Driven Nanomotors. Journal of Nanoscience and Nanotechnology *5*, 1983–1989.

Jurica, M. S., and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Molecular Cell *12*, 5–14.

Jurica, M. S., Sousa, D., Moore, M. J., and Grigorieff, N. (2004). Three-dimensional structure of C complex spliceosomes by electron microscopy. Nature Structural & Molecular Biology *11*, 265–269.

Kabsch, W. (2010). XDS. Acta Crystallographica Section D Biological crystallography *66*, 125–132.

Kambach, C., Walke, S., Young, R., Avis, J. M., De la Fortelle, E., Raker, V. a, Lührmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell *96*, 375–387.

Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell *106*, 595–605.

Kim, D. H., and Rossi, J. J. (1999). The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. RNA *5*, 959–971.

Kim, S. H., and Lin, R. J. (1993). Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. Proceedings of the National Academy of Sciences of the United States of America *90*, 888–892.

Kim, S., and Lin, R. (1996). Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of Pre-mRNA Splicing. Molecular and Cellular Biology *16*, 6810–6819.

Kim, S., Smith, J., Claude, A., and Lin, R. (1992). The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. The EMBO Journal *1*, 2319–2326.

Kiss, T. (2004). Biogenesis of small nuclear RNPs. Journal of Cell Science *117*, 5949–5951.

Kistler, A., and Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. Genes & Development *15*, 42–49.

Kong, J., and Liebhaber, S. A. (2007). A cell type-restricted mRNA surveillance pathway triggered by ribosome extension into the 3' untranslated region. Nature Structural & Molecular Biology *14*, 670–676.

Koodathingal, P., Novak, T., Piccirilli, J. A., and Staley, J. P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. Molecular Cell *39*, 385–395.

Korneta, I., Magnus, M., and Bujnicki, J. M. (2012). Structural bioinformatics of the human spliceosomal proteome. Nucleic Acids Research *40*, 7046–7065.

Kudlinzki, D., Nagel, C., and Ficner, R. (2009). Crystallization and preliminary X-ray diffraction analysis of the C-terminal domain of the human spliceosomal DExD/H-box protein hPrp22. Acta Crystallographica Section F Structural Biology and Crystallization Communications *65*, 956–958.

Kuhn, A. N., and Brow, D. A. (2000). Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. Genetics *155*, 1667–1682.

Kuhn, A. N., Reichl, E. M., and Brow, D. A. (2002). Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. Proceedings of the National Academy of Sciences of the United States of America *99*, 9145–9149.

Kuhn, A., Santen, M. van, and Schwienhorst, A. (2009). Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. RNA *15*, 153–175.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature *227*, 680–685.

Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proceedings of the National Academy of Sciences of the United States of America *95*, 4188–4192.

Laggerbauer, B., Liu, S., Makarov, E., Vornlocher, H.-P., Makarova, O., Ingelfinger, D., Achsel, T., and Lührmann, R. (2005). The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation. RNA *11*, 598–608.

Last, R., Maddock, J., and Woolford, J. (1987). Evidence for related functions of the RNA genes of Saccharomyces cerevisiae. Genetics *117*, 619–631.

Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W. S., Hartmann, E., and Luhrmann, R. (1996). The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases. The EMBO Journal *15*, 4001–4015.

Lauber, J., Plessel, G., Prehn, S., Will, C. L., Fabrizio, P., Gröning, K., Lane, W. S., and Lührmann, R. (1997). The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. RNA *3*, 926–941.

Leung, A., Nagai, K., and Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. Nature *473*, 536–539.

Li, N., Mei, H., MacDonald, I. M., Jiao, X., and Hejtmancik, J. F. (2010). Mutations in ASCC3L1 on 2q11.2 are associated with autosomal dominant retinitis pigmentosa in a Chinese family. Investigative Ophthalmology & Visual Science *51*, 1036–1043.

Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S. L., and Quackenbush, J. (2000). Gene index analysis of the human genome estimates approximately 120,000 genes. Nature Genetics *25*, 239–240.

Linder, B., Dill, H., Hirmer, A., Brocher, J., Lee, G. P., Mathavan, S., Bolz, H. J., Winkler, C., Laggerbauer, B., and Fischer, U. (2011). Systemic splicing factor deficiency causes tissue-specific defects: a zebrafish model for retinitis pigmentosa. Human Molecular Genetics *20*, 368–377.

Linder, P. (2006). Dead-box proteins: a family affair--active and passive players in RNP-remodeling. Nucleic Acids Research *34*, 4168–4180.

Linder, P., and Slonimski, P. P. (1989). An essential yeast protein, encoded by duplicated genes TIF1 and TIF2 and homologous to the mammalian translation initiation factor eIF-4A, can suppress a mitochondrial missense mutation. Proceedings of the National Academy of Sciences of the United States of America *86*, 2286–2290.

Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Lührmann, R., Carlomagno, T., and Wahl, M. C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. Science *316*, 115–120.

Liu, S., Rauhut, R., Vornlocher, H.-P., and Lührmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. RNA *12*, 1418–1430.

Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. Science *294*, 1098–1102.

Long, J. C., and Caceres, J. F. (2009). The SR protein family of splicing factors: master regulators of gene expression. The Biochemical Journal *417*, 15–27.

Lopez, P., and Séraphin, B. (1999). Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. RNA *5*, 1135–1137.

Lu, A., and Miller, L. K. (1997). Regulation of baculovirus late and very late gene expression. In The Baculoviruses, L. K. Miller, ed. (N.Y: Plenum Press), pp. 193–216.

Lu, P., Lu, G., Yan, C., Wang, L., Li, W., and Yin, P. (2012). Structure of the mRNA splicing complex component Cwc2: insights into RNA recognition. The Biochemical Journal *441*, 591–597.

Luckow, V. A., and Summers, M. D. (1988). Trends in the Development of Baculovirus Expression Vectors. Nature Biotechnology *6*, 47–55.

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell *144*, 16–26.

Lusty, C. J. (1999). A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. Journal of Applied Crystallography *32*, 106–112.

Madhani, H. D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell *71*, 803–817.

Maeder, C., Kutach, A. K., and Guthrie, C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. Nature Structural & Molecular Biology *16*, 42–48.

Makarov, E. M., Makarova, O. V, Achsel, T., and Lührmann, R. (2000). The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is

stably associated with the U5 snRNP via protein-protein interactions. Journal of Molecular Biology *298*, 567–575.

Makarov, E. M., Makarova, O. V, Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., and Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science *298*, 2205–2208.

Makarova, O. V, Makarov, E. M., Liu, S., Vornlocher, H.-P., and Lührmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing. The EMBO Journal *21*, 1148–1157.

Makarova, O. V, Makarov, E. M., Urlaub, H., Will, C. L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. The EMBO Journal *23*, 2381–2391.

Maquat, L. E., Carmichael, G. G., and York, N. (2001). Quality Control of mRNA Function. Cell *104*, 173–176.

Martegani, E., Vanoni, M., Mauri, I., Rudoni, S., Saliola, M., and Alberghina, L. (1997). Identification of Gene encoding a Putative RNA-Helicase , Homologous to SKI2 , in Chromosome VII of Saccharomyces cerevisiae. Yeast *13*, 391–397.

Martin, A., Schneider, S., and Schwer, B. (2002). Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. The Journal of Biological Chemistry *277*, 17743–17750.

Mathew, R., Hartmuth, K., Möhlmann, S., Urlaub, H., Ficner, R., and Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nature Structural & Molecular Biology *15*, 435–443.

Matlin, A. J., Clark, F., and Smith, C. W. J. (2005). Understanding alternative splicing: towards a cellular code. Nature Reviews. Molecular Cell Biology *6*, 386–398.

Mayas, R. M., Maita, H., Semlow, D. R., and Staley, J. P. (2010). Spliceosome discards intermediates via the DEAH box ATPase Prp43p. Proceedings of the National Academy of Sciences of the United States of America *107*, 10020–10025.

Mayas, R. M., Maita, H., and Staley, J. P. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nature Structural & Molecular Biology *13*, 482–490.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D. L. (1997). McCracken_Nature_1997.pdf. Nature *385*, 357–361.

McKie, a B., McHale, J. C., Keen, T. J., Tarttelin, E. E., Goliath, R., Van Lith-Verhoeven, J. J., Greenberg, J., Ramesar, R. S., Hoyng, C. B., Cremers, F. P., *et al.* (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). Human Molecular Genetics *10*, 1555–1562.

McManus, C. J., Schwartz, M. L., Butcher, S. E., and Brow, D. A. (2007). A dynamic bulge in the U6 RNA internal stem-loop functions in spliceosome assembly and activation. RNA *13*, 2252–2265.

Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallographica Section D Biological crystallography *62*, 859–866.

Montpetit, B., Thomsen, N. D., Helmke, K. J., Seeliger, M. A., Berger, J. M., and Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. Nature *472*, 238–242.

Mozaffari-Jovin, S., Santos, K. F., Hsiao, H.-H., Will, C. L., Urlaub, H., Wahl, M. C., and Lührmann, R. (2012). The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. Genes & Development *26*, 2422–2434.

Mozaffari-Jovin, S., Wandersleben, T., Santos, K. F., Will, C. L., Lührmann, R., and Wahl, M. C. (2013). Inhibition of RNA Helicase Brr2 by the C-Terminal Tail of the Spliceosomal Protein Prp8. Science.

Mukherjee, N., Corcoran, D. L., Nusbaum, J. D., Reid, D. W., Georgiev, S., Hafner, M., Ascano, M., Tuschl, T., Ohler, U., and Keene, J. D. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. Molecular Cell *43*, 327–339.

Nakagawa, T., and Kolodner, R. D. (2002). The MER3 DNA helicase catalyzes the unwinding of holliday junctions. The Journal of Biological Chemistry *277*, 28019–28024.

Newman, a J., Teigelkamp, S., and Beggs, J. D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. RNA *1*, 968–980.

Nielsen, T. K., Liu, S., Lührmann, R., and Ficner, R. (2007). Structural basis for the bifunctionality of the U5 snRNP 52K protein (CD2BP2). Journal of Molecular Biology *369*, 902–908.

Niesen, F. H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature Protocols *2*, 2212–2221.

Nilsen, T. W. (1994). RNA-RNA Interactions in the Spliceosome : Unraveling the Ties That Bind Minireview. Cell *78*, 1–4.

Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463.

Noble, S., and Guthrie, C. (1996). Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. Genetics *143*, 67–80.

Van Nues, R. W., and Beggs, J. D. (2001). Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae. Genetics *157*, 1451–1467.

Oberer, M., Marintchev, A., and Wagner, G. (2005). Structural basis for the enhancement of eIF4A helicase activity by eIF4G. Genes & Development *19*, 2212–2223.

Oubridge, C., Ito, N., Evans, P., Teo, C., and Nagai, K. (1994). Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexes with and RNA hairpin. Nature *372*, 432–438.

Oyama, T., Oka, H., Mayanagi, K., Shirai, T., Matoba, K., Fujikane, R., Ishino, Y., and Morikawa, K. (2009). Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm. BMC Structural Biology *9*, 2.

Pandit, S., Lynn, B., and Rymond, B. C. (2006). Inhibition of a spliceosome turnover pathway suppresses splicing defects. Proceedings of the National Academy of Sciences of the United States of America *103*, 13700–13705.

Parker, R., Siliciano, P. G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell *49*, 229–239.

Passarelli, A. L., and Guarino, L. A. (2007). Baculovirus late and very late gene regulation. Current Drug Targets *8*, 1103–1115.

Patel, A. A., and Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. Nature Reviews. Molecular Cell Biology *4*, 960–970.

Patel, S. B., and Bellini, M. (2008). The assembly of a spliceosomal small nuclear ribonucleoprotein particle. Nucleic Acids Research *36*, 6482–6493.

Pena, V., Jovin, S. M., Fabrizio, P., Orlowski, J., Bujnicki, J. M., Lührmann, R., and Wahl, M. C. (2009). Common design principles in the spliceosomal RNA helicase Brr2 and in the Hel308 DNA helicase. Molecular Cell *35*, 454–466.

Pena, V., Liu, S., Bujnicki, J. M., Lührmann, R., and Wahl, M. C. (2007). Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. Molecular Cell *25*, 615–624.

Pena, V., Rozov, A., Fabrizio, P., Lührmann, R., and Wahl, M. C. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. The EMBO Journal *27*, 2929–2940.

Pennock, G. D., Shoemaker, C., and Miller, L. K. (1984). Strong and Regulated Expression of Escherichia coli beta-Galactosidase in Insect Cells with a Baculovirus Vector. Molecular and Cellular Biology *4*, 399–406.

Perriman, R. J., and Ares, M. (2007). Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. Genes & Development *21*, 811–820.

Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature *458*, 475–480.

Ponting, C. P. (2000). Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. Biochemistry Journal *535*, 527–535.

Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP 4 program suite. Acta Crystallographica Section D Biological Crystallography *59*, 1131–1137.

Price, S. R., Evans, P. R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645–650.

Pyle, A. M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. Annual Review of Biophysics *37*, 317–336.

Query, C. C., and Konarska, M. M. (2004). Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. Molecular Cell *14*, 343–354.

Query, C. C., Moore, M. J., and Sharp, P. A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes & Development *8*, 587–597.

Raghunathan, P. L., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Current Biology *8*, 847–855.

Rajkowitsch, L., and Schroeder, R. (2007). Dissecting RNA chaperone activity. RNA *13*, 2053–2060.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Current Opinion in Cell Biology *12*, 340–345.

Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. Genes & Development *3*, 2113–2123.

Richards, J. D., Johnson, K. a, Liu, H., McRobbie, A.-M., McMahon, S., Oke, M., Carter, L., Naismith, J. H., and White, M. F. (2008). Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. The Journal of Biological Chemistry *283*, 5118–5126.

Rinke, J., Appel, B., Digweed, M., and Lührmann, R. (1985). Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. Journal of Molecular Biology *185*, 721–731.

Ritchie, D. B., Schellenberg, M. J., Gesner, E. M., Raithatha, S. A., Stuart, D. T., and Macmillan, A. M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. Nature Structural & Molecular Biology *15*, 1199–1205.

Ritchie, D. B., Schellenberg, M. J., and MacMillan, A. M. (2009). Spliceosome structure: piece by piece. Biochimica et Biophysica Acta *1789*, 624–633.

Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. Nature Reviews. Molecular Cell Biology *5*, 232–241.

Rowen, L., Young, J., Birditt, B., Kaur, A., Madan, A., Philipps, D. L., Qin, S., Minx, P., Wilson, R. K., Hood, L., *et al.* (2002). Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. Genomics *79*, 587–597.

Roy, J., Kim, K., Maddock, J. R., Anthony, J. G., and Woolford, J. L. (1995). The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. RNA *1*, 375–390.

Ruby, S., and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science *242*, 4–11.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning : a laboratory manual.

Sawa, H., and Abelson, J. (1992). Evidence for a base-pairing interaction between U6 small nuclear RNA and the 5' splice site during the splicing reaction in yeast. Proceedings of the National Academy of Sciences of the United States of America *89*, 11269–11273.

Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., and Lührmann, R. (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. The EMBO Journal *23*, 3000–3009.

Schellenberg, M. J., Edwards, R. A., Ritchie, D. B., Kent, O. a, Golas, M. M., Stark, H., Lührmann, R., Glover, J. N. M., and MacMillan, A. M. (2006). Crystal structure of a core spliceosomal protein interface. Proceedings of the National Academy of Sciences of the United States of America *103*, 1266–1271.

Schmitzová, J., Rasche, N., Dybkov, O., Kramer, K., Fabrizio, P., Urlaub, H., Lührmann, R., and Pena, V. (2012). Crystal structure of Cwc2 reveals a novel architecture of a multipartite RNA-binding protein. The EMBO Journal *31*, 2222–2234.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Molecular Cell *30*, 743–754.

Schwer, B., and Gross, C. H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. The EMBO Journal *17*, 2086–2094.

Schwer, B., and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. The EMBO Journal *11*, 5033–5039.

Seetharaman, M., Eldho, N. V, Padgett, R. A., and Dayie, K. T. (2006). Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. RNA *12*, 235–247.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Molecular Cell *11*, 965–976.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. Cell *125*, 287–300.

Séraphin, B. (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. The EMBO Journal *14*, 2089–2098.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell *59*, 349–358.

Sharma, S., Kohlstaedt, L. a, Damianov, A., Rio, D. C., and Black, D. L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nature Structural & Molecular Biology *15*, 183–191.

Shatkin, A. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645–653.

Shatkin, A. J., and Manley, J. L. (2000). The ends of the affair: capping and polyadenylation. Nature Structural Biology *7*, 838–842.

Sheldrick, G. M. (2008). A short history of SHELX. Acta Crystallographica Section A Foundations of Crystallography *64*, 112–122.

Shevchenko, A., Wilm, M., Vorm, O., Jensen, O. N., Podtelejnikov, A. V, Neubauer, G., Mortensen, P., and Mann, M. (1996). A strategy for identifying gel-separated proteins in sequence databases by MS alone. Biochemical Society Transactions *24*, 893–896.

Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R., and Kielkopf, C. L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. Molecular Cell *23*, 49–59.

Siliciano, P. G., Kivens, W. J., and Guthrie, C. (1991). More than half of yeast U1 snRNA is dispensable for growth. Nucleic Acids Research *19*, 6367–6372.

Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annual Review of Biochemistry *76*, 23–50.

Small, E. C., Leggett, S. R., Winans, A. A., and Staley, J. P. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Molecular Cell *23*, 389–399.

Smith, D. J., Query, C. C., and Konarska, M. M. (2008). "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. Molecular Cell *30*, 657–666.

Smith, G. E., Summers, M. D., and Fraser, M. J. (1983). Production of Human Beta Interferon in Insect Cells Infected with a Baculovirus Expression Vector. Molecular and Cellular Biology *3*, 2156–2165.

Sontheimer, E. J. (2001). The spliceosome shows its metal. Nature Structural Biology *8*, 11–13.

Soulard, M., Della Valle, V., Siomi, M. C., Piñol-Roma, S., Codogno, P., Bauvy, C., Bellini, M., Lacroix, J. C., Monod, G., and Dreyfuss, G. (1993). hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. Nucleic Acids Research *21*, 4210–4217.

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

Spingola, M., Grate, L., Haussler, D., and Ares, M. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA *5*, 221–234.

Staley, J. P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Molecular Cell *3*, 55–64.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315–326.

Stanek, D., and Neugebauer, K. M. (2006). The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze. Chromosoma *115*, 343–354.

Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature *409*, 539–542.

Stark, H., and Lührmann, R. (2006). Cryo-electron microscopy of spliceosomal components. Annual Review of Biophysics and Biomolecular Structure *35*, 435–457.

Stephens, R. M., and Schneider, T. D. (1992). Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. Journal of Molecular Biology *228*, 1124–1136.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification *41*, 207–234.

Tanackovic, G., Ransijn, A., Thibault, P., Abou Elela, S., Klinck, R., Berson, E. L., Chabot, B., and Rivolta, C. (2011). PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. Human Molecular Genetics *20*, 2116–2130.

Tanaka, N., Aronova, A., and Schwer, B. (2007). Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. Genes & Development *21*, 2312–2325.

Tanner, N. K., Cordin, O., Banroques, J., Doère, M., and Linder, P. (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. Molecular Cell *11*, 127–138.

Tanner, N. K., and Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. Molecular Cell *8*, 251–262.

Toor, N., Keating, K. S., Taylor, S. D., and Pyle, A. M. (2008a). Crystal structure of a self-spliced group II intron. Science *320*, 77–82.

Toor, N., Rajashankar, K., Keating, K. S., and Pyle, A. M. (2008b). Structural basis for exon recognition by a group II intron. Nature Structural & Molecular Biology *15*, 1221–1222.

Tritschler, F., Braun, J. E., Eulalio, A., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009). Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. Molecular Cell *33*, 661–668.

Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., and Berger, I. (2010). New baculovirus expression tools for recombinant protein complex production. Journal of Structural Biology *172*, 45–54.

Tsai, R.-T., Tseng, C.-K., Lee, P.-J., Chen, H.-C., Fu, R.-H., Chang, K., Yeh, F.-L., and Cheng, S.-C. (2007). Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. Molecular and Cellular Biology *27*, 8027–8037.

Umen, J. G., and Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. RNA *1*, 869–885.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallographica Section D Biological crystallography *60*, 2184–2195.

Vagin, A., and Teplyakov, A. (1997). MOLREP : an Automated Program for Molecular Replacement. Journal of Applied Crystallography *30*, 1022–1025.

Valadkhan, S. (2010). Role of the snRNAs in spliceosomal active site. RNA Biology *7*, 345–353.

Valadkhan, S., Mohammadi, A., Jaladat, Y., and Geisler, S. (2009). 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.

Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J. L. (2007). Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. RNA *13*, 2300–2311.

Vaughn, J. L., Goodwin, R. H., Tompkins, G. J., and McCawley, P. (1977). The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). In Vitro *13*, 213–217.

Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Molecular Cell *6*, 1331–1342.

Vithana, E. N., Abu-Safieh, L., Allen, M. J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N. D., Willis, C., Moore, A. T., *et al.* (2001). A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). Molecular Cell *8*, 375–381.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell *136*, 669–687.

Wagner, J. D., Jankowsky, E., Company, M., Pyle, A. M., and Abelson, J. N. (1998). The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. The EMBO Journal *17*, 2926–2937.

Wahl, M. C., Will, C. L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., Van Tilbeurgh, H., Henry, Y., and Leulliot, N. (2010). Prp43p contains a processive helicase structural architecture with a specific regulatory domain. The EMBO Journal *29*, 2194–2204.

Walker, J. M. (2005). The Proteomics Protocols Handbook (Humana Press).

Walker, J., Saraste, M., Runswick, M., and Gay, N. (1982). Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding. The EMBO Journal *I*, 945–951.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., and Burge, C. B. (2008). Alternative Isoform Regulation in Human Tissue Transcriptomes. Nature *456*, 470–476.

Wang, Y., and Guthrie, C. (1998). PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. RNA *4*, 1216–1229.

Warkocki, Z., Odenwälder, P., Schmitzová, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Lührmann, R. (2009). Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. Nature Structural & Molecular Biology *16*, 1237–1243.

Wassarman, D. A., and Steitz, J. A. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science *257*, 1918–1925.

Weber, G., Cristão, V. F., De L Alves, F., Santos, K. F., Holton, N., Rappsilber, J., Beggs, J. D., and Wahl, M. C. (2011). Mechanism for Aar2p function as a U5 snRNP assembly factor. Genes & Development *25*, 1601–1612.

Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R., and Wahl, M. C. (2010). Functional organization of the Sm core in the crystal structure of human U1 snRNP. The EMBO Journal *29*, 4172–4184.

Weber, K., Pringle, J. ., and Osborn, M. (1972). Measurement of molecular weights by electrophoresis on SDS-acrylamide gel. In Methods in Enzymology (Elsevier), pp. 3–27.

Weir, J. R., Bonneau, F., Hentschel, J., and Conti, E. (2010). Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. Proceedings of the National Academy of Sciences of the United States of America *107*, 12139–12144.

Wickham, T. J., Davis, T., Granados, R. R., Shuler, M. L., and Wood, H. A. (1992). Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. Biotechnology Progress *8*, 391–396.

Will, C. L., and Lührmann, R. (2006). Spliceosome Structure and Function. In The RNA World, pp. 1–29.

Will, C. L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harbor Perspectives in Biology *3*, 1–23.

Will, C. L., and Lührmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. Biological Chemistry *386*, 713–724.

Will, C. L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. The EMBO Journal *21*, 4978–4988.

Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. In Methods in Enzymology, pp. 300–321.

Wolff, T., and Bindereif, A. (1993). Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of slicing. Genes & Development *7*, 1377–1389.

Woodman, I. L., Briggs, G. S., and Bolt, E. L. (2007). Archaeal Hel308 domain V couples DNA binding to ATP hydrolysis and positions DNA for unwinding over the helicase ratchet. Journal of Molecular Biology *374*, 1139–1144.

Wu, J., and Manley, J. L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes & Development *3*, 1553–1561.

Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061–1070.

Wyatt, J. R., Sontheimer, E. J., and Steitz, J. A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5’ splice site before the first step of pre-mRNA splicing. Genes & Development *6*, 2542–2553.

Xu, D., Nouraini, S., Field, D., Tang, S.-J., and Firesen, J. (1996). An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. Nature *381*, 709–713.

Xu, Y.-Z., and Query, C. C. (2007). Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. Molecular Cell *28*, 838–849.

Yang, K., Zhang, L., Xu, T., Heroux, A., and Zhao, R. (2008). Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. Proceedings of the National Academy of Sciences of the United States of America *105*, 13817–13822.

Yang, Q., and Jankowsky, E. (2005). ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. Biochemistry *44*, 13591–13601.

Yean, S. L., Wuenschell, G., Termini, J., and Lin, R. J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature *408*, 881–884.

Yin, J., Brocher, J., Fischer, U., and Winkler, C. (2011). Mutant Prpf31 causes pre-mRNA splicing defects and rod photoreceptor cell degeneration in a zebrafish model for Retinitis pigmentosa. Molecular Neurodegeneration *6*, 56.

Yu, Y., Maroney, P. A., Denker, J. A., Zhang, X. H.-F., Dybkov, O., Lührmann, R., Jankowsky, E., Chasin, L. A., and Nilsen, T. W. (2008). Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. Cell *135*, 1224–1236.

Yura, K., Shionyu, M., Hagino, K., Hijikata, A., Hirashima, Y., Nakahara, T., Eguchi, T., Shinoda, K., Yamaguchi, A., Takahashi, K.-I., *et al.* (2006). Alternative splicing in human transcriptome: functional and structural influence on proteins. Gene *380*, 63–71.

Zhang, L., Shen, J., and Guarnieri, M. (2007). Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. Protein Science *16*, 1024–1031.

Zhang, L., Xu, T., Maeder, C., Bud, L.-O., Shanks, J., Nix, J., Guthrie, C., Pleiss, J. a, and Zhao, R. (2009). Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. Nature Structural & Molecular Biology *16*, 731–739.

Zhang, M. (2001). Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. Genes & Development *15*, 30–35.

Zhang, M. Q. (1998). Statistical features of human exons and their flanking regions. Human Molecular Genetics *7*, 919–932.

Zhao, C., Bellur, D. L., Lu, S., Zhao, F., Grassi, M. A., Bowne, S. J., Sullivan, L. S., Daiger, S. P., Chen, L. J., Pang, C. P., *et al.* (2009). Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs. American Journal of Human Genetics *85*, 617–627.

Zhao, C., Lu, S., Zhou, X., Zhang, X., Zhao, K., and Larsson, C. (2006). A novel locus (RP33) for autosomal dominant retinitis pigmentosa mapping to chromosomal region 2cen-q12.1. Human Genetics *119*, 617–623.

Zhuang, Y., and Weiner, A. M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell *46*, 827–835.

List of figures

List of tables

List of abbreviations

Abbreviation	Expanded
AcMNPV	*Autographa californica* multiple nuclear polyhedrosis virus
APS	ammonium peroxodisulfate
ATP	adenosine triphosphate
BAC	bacterial artificial chromosome
BEVS	baculovirus expression vector system
BP	branch point
BPS	branch point sequence
BSA	bovine serum albumin
CD	circular dichroism
ddH_2O	double-distilled water
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
DSF	differential scanning fluorimetry
DTT	1,4-Dithiothreitol
e.g.	*exempli gratia*
EDTA	ethylenediaminetetraacetic acid
EJC	exon junction complex
EM	electron microscopy
ESE	exonic splicing enhancer
ESI	exonic splicing inhibitor
FRET	fluorescence resonance energy transfer
HB domain	helix bundle domain
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLH	helix-loop-helix domain
i.e.	*id est*
IG domain	immunoglobulin-like domain
IPTG	isopropyl-β-D-1-thiogalactopyranoside
ISE	intronic splicing enhancer
ISI	intronic splicing inhibitor
ISL	intramolecular stem-loop

Abbreviation	Expanded
kb	kilobase (unit of NA molecule length)
kDa	kilodalton (unit of molecular weight)
LSm	Sm-like protein
$m_3^{2,2,7}G$	2,2,7-trimethylguanosine
m_7G	7-methylguanosine
MAD	multi-wavelength anomalous dispersion
mant-NTP	2'/3'-O-(N-methylantranyloil)-nucleotide
MIR	multiple isomorphous replacement
MIRAS	multiple isomorphous replacement with anomalous scattering
MOI	multiplicity of infection
MR	molecular replacement
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein complex
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
nt	nucleotides
NTC	NineTeen Complex
NTP	nucleotide triphosphate
NTR complex	NineTeen related complex
OD	optical density
ORF	open reading frame
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PIPES	piperazine-N,N′-bis-(2-ethanesulfonic acid)
PMSF	phenylmethylsulfonyl fluoride
PNK	polynucleotide kinase
polh	polyhedrin
PPIase	peptidyl-prolyl isomerase
PPT	polypyrimidine tract
pre-mRNA	precursor mRNA
RES complex	pre-mRNA retention and splicing complex
rmsd or r.m.s.d.	root mean square deviation
RNA	ribonucleic acid
RP	retinitis pigmentosa

List of Abbreviations

Abbreviation	Expanded
RRM	RNA recognition motif
RS domain	arginine/serine rich domain
SAD	single wavelength anomalous dispersion
SDS	sodium dodecylsulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SF	superfamily
SIR	single isomorphous replacement
SIRAS	single isomorphous replacement with anomalous scattering
SL	stem loop
snRNP	small nuclear ribonucleoprotein complex
SR protein	serine-arginine protein
SS	splice site
S-SAD	sulfur-based single wavelength anomalous dispersion
TBE buffer	TRIS-borate-EDTA buffer
TBS	TRIS-buffered saline
TEMED	N,N,N',N'-tetraethylenediamide
TEV	tobacco etch virus
T_m	melting temperature
TRIS	Tristris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
U snRNA	uridine-rich small nuclear RNA
U2AF	U2 auxiliary factor
UTR	untranslated region
UV	ultra-violet
WH domain	winged helix domain
β-ME	2-Mercaptoethanol

Acknowledgements

I am truly thankful to my supervisor, Prof. Markus Wahl, for all his support and patience during my 4 years of PhD. I am especially grateful for his great ideas usually accompanied by fruitful discussions and unconditional enthusiasm that motivated me throughout my thesis. I appreciate all his encouragement and effort in guiding me and providing a good working atmosphere in the laboratory.

I am further grateful to the following people and organizations that helped me on my way towards this dissertation:

Prof. Reinhard Lührmann for his continuous support and valuable ideas.

Juliane Moses and Karin Hesse for solving a multitude of everyday problems.

The members of my thesis committee, Prof. Detlef Doenecke and Prof. Dirk Fasshauer for their helpful suggestions.

Claudia Alings, Bernhard Loll, Clemens Langer and Carsten Jacob for all the welcoming support when I joined our new group of Structural Biochemistry in the Free University in Berlin. Special thanks to Claudia for being always extremely helpful in daily lab issues.

Kirsten Poehlker, Christina Bach and Christin Fischer from the coordination office of the GGNB Program for their impressive organizational efforts, for continuous administrative support and for providing me with a travel grant that enabled me to attend the EMBO Practical Course: Computational structural biology – from data to structure to function in Cambridge, 2011.

Dr. Henning Urlaub, Dr. He-Hsuan Hsiao, Katharina Kramer, Monika Raabe and Uwe Plessmann for all mass spectrometric analyses.

Dr. Uwe Müller, Dr. Manfred Weiss, Dr. Sandra Pühringer and all MX-group at BESSY for the technical support. Special thanks to Dr. Uwe Müller and Dr. Manfred Weiss for helpful crystallography advices.

All the former and present members of the Department of Cellular Biochemistry for the friendly working atmosphere. While working in this department, my time in the lab was made particularly enjoyable due to my friends Tales, Inessa, Alexey, Homa and Anzhalika. I am very grateful for the good time we had together and for all interesting scientific and non-scientific discussions. Special thanks to Alexey for critical reading and great support during my thesis writing.

All my former and present lab mates in the Wahl group for sharing their experiences with me and helping me with all lab issues. I am heartily thankful for all the happy moments we had together. Special thanks to Traudy, Christian, Steffi, Eva and Dr. Gert Weber for critical reading of parts of this thesis and Dr. Bernhard Loll and Dr. Gert Weber for helping with crystallographic problems.

To Dr. Pohl Milon, Dr. Dmitry Burakovskiy, Prof. Marina Rodnina and all members of the Physical Biochemistry department for impressive assistance with rapid kinetics experiments, scientific support and patience during the last six months.

I am forever indebted to my parents, my sister and Tulio for their unconditional moral support and endless love. I express my sincerest thanks to my family and appreciate their encouragement and faithful support during these years.