Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structural and Biochemical Studies of the Human

DEAD-box Helicase Dbp5 and Nucleoporin Nup214

Involved in mRNA Export

Holger von Moeller

aus Leverkusen

2009

Meinen Eltern

Abstract

The hallmark of eukaryotic evolution was the development of the nucleus in cells. This compartmentalization requires the nucleocytoplasmic transport of thousands of molecules. The gate into and out of the nucleus is the nuclear pore complex (NPC). One of the molecules that needs to be exported from the nucleus is messenger RNA (mRNA). mRNA associates with proteins in the nucleus forming a messenger ribonucleoprotein particle (mRNP). mRNPs bind to dedicated transport factors that facilitate movement through the NPC. One protein that associates to mRNPs is the helicase Dbp5, which belongs to the DEAD-box family of RNA helicases. Dbp5 is essential for mRNA export in both yeast and humans. It binds RNA and is concentrated and locally activated at the cytoplasmic side of the nuclear pore complex, where it interacts with the cytoplasmic nucleoporin Nup214. In my PhD work, I have determined the crystal structures of human Dbp5 bound to RNA and AMPPNP, and bound to Nup214. I designed and performed *in vitro* assays, which show that binding of Dbp5 to nucleic acid and to Nup214 is mutually exclusive. The interactions are mediated by conserved residues, implying a conserved recognition mechanism. These results suggest a framework for the consecutive steps leading to the release of mRNA at the final stages of nuclear export. More generally, they provide a paradigm for how binding of regulators can specifically inhibit DEAD-box proteins.

Zusammenfassung

Die Entstehung des Zellkerns war eines der wichtigsten Ereignisse in der Entwicklung der eukaryotischen Zelle. Die daraus resultierende Unterteilung der Zelle in Zellkern und Zytoplasma erfordert jedoch den Transport von tausenden von Molekülen in und aus dem Zellkern. Als Schleuse hierfür dient der sogenannte Kernporenkomplex. Eines der Moleküle, welches aus dem Zellkern exportiert werden muss, ist die Boten-RNA (mRNA). mRNAs assozieren mit einer Vielzahl verschiedener Proteine im Zellkern und bilden sog. Boten-Ribonukleoprotein-Partikel (mRNP). mRNPs interagieren mit bestimmten Transportproteinen, welche die Bewegung durch den Kernporenkomplex und dadurch den Export aus dem Zellkern ermöglichen. Eines der Proteine welches mit mRNPs assoziert ist die DEAD-box Helikase Dbp5, welche in Mensch und Hefe essentiell am Kernexport von mRNAs beteiligt ist. Dbp5 bindet RNA und liegt konzentriert auf der cytoplasmatischen Seite des Kernporenkomplexes vor. Hier bindet es an das cytoplasmatischen Nukleoporin Nup214 und wird durch ein weiteres Protein lokal aktiviert. In meiner Doktorarbeit habe ich die Kristallstrukturen des menschlichen Dbp5 im Komplex mit RNA und im Komplex mit Nup214 bestimmt. Zusammen mit *in vitro* Experimenten konnte ich zeigen, dass RNA und Nup214 die gleiche Bindungstelle benutzen und sich deren gleichzeitige Bindung somit ausschliesst. Die für diese Interaktionen entscheidenden Aminosäuren sind konserviert und implizieren einen ähnlichen Erkennungs-mechanismus in Mensch und Hefe. Diese Ergebnisse stellen die Grundlage für ein Modell, indem die letzten Schritte des Kernexportes, welche in der Freigabe der mRNA enden, zusammengefasst sind. Desweiteren liefern diese Ergebnisse wichtige Einblicke in den Regulierungsmechanismen von DEAD-box Helikasen.

Publications and Presentations

Parts of this PhD thesis have been published:

Nat Struct Mol Biol. March 2009 (Feb 15. Epub ahead of print)

The mRNA export protein Dbp5 binds RNA and the cytoplasmic nucleoporin Nup214 in a mutually exclusive manner.

von Moeller H, Basquin C, Conti E.

I presented parts of this thesis at an international conference:

Thirteenth Annual Meeting of the RNA Society, Berlin 2008

"Structural insights into the function of the Dbp5 helicase in mRNA export"

Acknowledgements

I. Contents

Contents

1   Introduction

One highlight of modern biology was the enunciation of the central dogma of molecular biology half a century ago by Francis Crick (Crick 1958; Crick 1970). This dogma is the framework of the flow of information from DNA to protein. DNA is the carrier of genetic information that is transcribed into messenger RNA (mRNA), which is then translated into protein. This flow of information applies to all kingdoms of life.

The hallmark in the evolution of eukaryotic life was the development of the nucleus. The nucleus is separated from the cytoplasm by a semi permeable lipid bilayer, known as the nuclear envelope (NE). The NE consists of two membranes (inner and outer) that impede free diffusion of molecules in and out the nucleus. This compartmentalization required the (co-)development of nucleocytoplasmic transport. Since the genetic information is exclusively stored in the nucleus (except mitochondrial DNA) and proteins are only synthesized in the cytoplasm, mRNA that is transcribed in the nucleus has to be exported to the cytoplasm. The only exit from the nucleus (at interphase) for macromolecules like mRNA is the nuclear pore complex (NPC).

1.1 The nuclear pore complex

The nuclear pore complex is one of the biggest protein assemblies in cells and has been studied for several decades (recently reviewed in Suntharalingam and Wente 2003; Hetzer 2005; Antonin 2008; D'Angelo and Hetzer 2008; Lim 2008; Lim 2008; Meier and Brkljacic 2009). The NPC was discovered in the 1950s and the overall architecture was described more than a decade later (Gall 1967). The development of new techniques, such as cryo-electron tomography (ET), field-emission in-lens scanning electron microscopy (FEISEM), two photon 4pi fluorescent microscopy and atomic force microscopy, in the last few years has provided more detailed structural information (Stoffler 2003; Beck 2004; Maco 2006; Beck 2007; Huve 2008; Lim 2008). The NPC is a dynamic multiprotein complex made up of proteins called nucleoporins (Nups) and can be divided into three major structures: the NE embedded central core, the nuclear basket and the cytoplasmic fibrils (Figure 1a and 1b). The central core is symmetric and spans both membranes (outer and inner) of the NE with a central channel of ~ 40 nm in diameter. A frequently observed central plug occupying this channel was the subject of a long debate and has been

recently demonstrated to be cargo in transit in a cryo-electron microscopy study (Beck 2007).
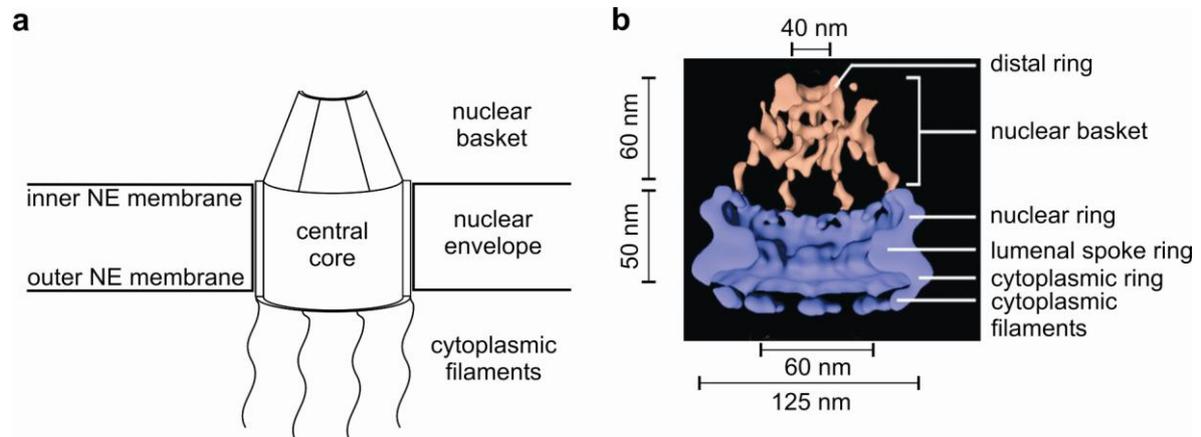
Figure 1: General architecture of the nuclear pore complex (NPC).
(a) cut away schematic illustration of the NPC. The NPC can be divided in three major structures: the nuclear envelope (NE) embedded core, the nuclear basket and the cytoplasmic filaments. (b) cryo-electron tomography structure of the *Dictyostelium* NPC adopted from (Beck 2004) in a similar cut away view as in panel a. The central cannel and the nuclear basket are further divided into substructures as indicated. The NPC has a maximum diameter of ~ 125 nm with a minimal luminal diameter of ~ 40 nm.

The cytoplasmic filaments and the nuclear basket consist of asymmetrically distributed nucleoporins (with respect to the NE plane) and are thought to play functional roles (reviewed in Marelli 2001; Rout and Aitchison 2001). From EM studies the overall size was estimated to be ~ 55 - 72 MDa for the yeast NPC (Rout and Blobel 1993; Yang 1998) and 125 MDa for the Xenopus NPC (Hinshaw 1992; Akey and Radermacher 1993). More recent calculations from proteomic approaches revealed a smaller total mass of ~ 44 MDa for the yeast (Rout 2000) and ~ 60 MDa for the mammalian NPC (Cronshaw 2002). This difference could be explained by the fact that the first estimations based on electron microscopy (EM) studies also included transiently bound proteins. In the metazoan case it is so far not clear why the Xenopus and the mammalian NPCs have such different masses (i.e. whether it reflects a larger/smaller assembly or experimental reasons).

Despite their enormous size, NPCs consist of a surprisingly small number of ~ 30 - 40 different Nups (Reichelt 1990; Yang 1998; Rout 2000; Cronshaw 2002). With a molecular weight (MW) between 100 and 358 kDa nucleoporins are above the average molecular weight of proteins in the cell. EM studies revealed an octagonal rotational symmetry and approximations of the abundances of the different Nups based on SDS PAGE quantification indicated that some Nups are represented by

multiples of eight copies - up to 48 copies of some Nups might be present per NPC (Rout 2000; Cronshaw 2002) resulting in a total number of 500 - 1000 Nups per pore (D'Angelo and Hetzer 2008). Depending on the cell cycle, a haploid yeast nucleus has 65 - 182 (Winey 1997) and mammalian nuclei have ~ 2000 - 3000 NPCs (Gerace and Burke 1988).

The NPC can be separated into subcomplexes that are thought to be defined units that are important for NPC assembly and disassembly during mitosis. The details are not fully understood (reviewed in Antonin 2008). Many interactions are not yet clear and under debate (e.g. Brohawn 2008 vs. Debler 2008). High-resolution structural data for some subcomplexes (Jeudy and Schwartz 2007; Melcak 2007; Boehmer 2008; Brohawn 2008; Debler 2008) and several domains (for e.g. Berke 2004; Weirich 2004; Napetschnig 2007, see Figure 3) are available from crystallographic studies. Together with computational predictions, it has been proposed that only a few common folds are present in the architecture of the NPC: beta-propeller, alpha-helical and coiled-coil domains are repetitively found/predicted in nucleoporins (reviewed in Schwartz 2005; Alber 2007). Only three Nups (Pom121, Gp210 and Ndc1, reviewed in Peters 2009) contain transmembrane domains that are believed to anchor the NPC in the NE (Mansfeld 2006; Stavru 2006).

Other common motifs found in Nups are the FG-repeats. These are peptide stretches of 4 to 48 repetitive phenylalanine-glycine motifs (GLFG, FxFG, PxFG or SxFG, where x is any residue) connected by linkers (5 - 50 residues). The motifs belong to the class of intrinsically disordered protein domains (Denning 2003). They are found in more than one third of the Nups, resulting in ~ 190 FG domains and ~ 2700 FG motifs per NPC (reviewed in Peters 2009). FG repeats play important roles in the selective permeability of NPCs (Macara 2001; Ribbeck and Gorlich 2001; Rout 2003; Peters 2005; Frey 2006). Ions, small metabolites, hydrophobic molecules and some molecules up to 30 - 40 kDa can diffuse through NPCs (Paine 1975). Large molecules up to a diameter of 39 nm and smaller molecules such as tRNAs or histones can be transported through the pore (Pante and Kann 2002), when associated with transport proteins that are able to interact with the FG repeats and escort the cargo through the NPC. Estimation of the traffic that occurs at a single NPC indicate that 1000 translocations per second per NPC and a total mass of one kg per day per eukaryotic cell are transported (Ribbeck and Gorlich 2001). It is

remarkable that the NPC is able to cope with these high numbers of transport events while maintaining selectivity.

Several models that account for the selective barrier have been proposed. In the 'virtual gate model' (Rout 2003) an energetic and not a physical barrier defines the permeability of the NPC. The FG Nups of the central channel generate an entropic barrier that prevents the passage of inert molecules. The entropy barrier increases with the size of the molecule and therefore the probability for large molecules to cross the NPC is low. In a second model, the 'oily spaghetti' model (Macara 2001), the FG repeats form a physical barrier by obstructing the central channel. Translocation in this model is achieved by transient interactions of transport factors with the FG repeats. In this case a 'binding and release mechanism' would allow a transport factor (+ cargo) to move through the pore. In the 'selective phase model (Ribbeck and Gorlich 2001), the FG Nups in the channel form a network of weak hydrophobic interactions. *In vitro,* FGs of the yeast nucleoporin Nsp1 can form a 3D-hydrogel (Frey 2006; Frey and Gorlich 2007). *In vivo,* such a gel-like sieve could hinder diffusion of molecules that are larger than its pore size. Transport factors would transiently disrupt the mesh by interacting with the FGs. The 'reduction of dimensionality model' (Peters 2005) predicts that the FG repeats act like a filter. In this model, the FG repeats form a hydrophobic layer coating the wall of the pore and therefore forming a ~ 10 nm nanopore in the center. Only molecules that are small enough to move through this nanopore or molecules that interact with the FGs (directly or via transport receptors) can slide through this filter. Although the details vary, all four models are based on the properties of the FG repeats. It is possible that a combination of the different mechanisms explains the selectivity of the NPC.

1.2 Nucleocytoplasmic transport

The FG repeats of the NPC do not solely define permeability but are also required for active transport (Bayliss 2002; Bednenko 2003). Independent work from Wente and Weis has shown that the FG repeats do not provide directionality (Strawn 2004; Zeitler and Weis 2004). Deletion and exchange of FG repeats of asymmetrically localized Nups have no effect on nucleocytoplasmic transport processes. How is the directional transport achieved? The transport between the two compartments is mediated by soluble transport factors that recognize specific cargos and are able to shuttle through NPCs. Most cargos associate with the largest family

of transport factors, the β-karyopherins, also known as importin β-like proteins (named after the first protein (importin β) identified). β-karyopherins belong to the family of HEAT (Huntington, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast PI3-kinase TOR1) repeat proteins. β-karyopherins involved in nuclear import are also called importins. Exportins mediate nuclear export in the reciprocal manner. In general transport factors can interact with their cargo either directly or via adaptor proteins. Binding to cargo and its release is regulated by the small GTPase Ran (RAs-related nuclear protein). The asymmetric distribution of RanGTP and RanGDP confers directionality in β-karyopherin dependent transport events. Ran is regulated by RanGTP-activating proteins (RanGAPs) in the cytoplasm. RanGDP is imported to the nucleus by the non-karyopherin transport factor NTF2 (nuclear transport factor 2). In the nucleus RanGDP is converted to RanGTP by the Ran guanine nucleotide exchange factor (RanGEF or RCC1). The interaction with the karyopherins is facilitated by another protein known as RanBP (Ran binding protein). Exportins bind RanGTP and the cargo in the nucleus and hydrolysis of GTP to GDP in the cytoplasm triggers release of both. Importins bind their cargo in the cytoplasm and are regulated in the reciprocal manner.

1.2.1 Nuclear export of messenger RNA

One of the cargos that need to be exported to the cytoplasm is messenger RNA (mRNA). mRNAs are transported as mRNPs (messenger ribonucleoprotein particles) and their export is independent of β-karyopherins and Ran (Clouse 2001). The mRNA export factor is a heterodimer consisting of TAP (NXF1 in metazoans and Mex67 in yeast) and p15 (NXT1 in metazoans and Mtr2 in yeast) recently reviewed in Cook 2007. The proteins p15 and Mtr2 are not conserved in sequence but are functionally and structurally homologous (Fribourg and Conti 2003). TAP is a multidomain protein with a N-terminal cargo (RNA binding domain (RBD) and leucine rich repeat (LRR) domain) and a C-terminal nucleoporin binding site (Braun 1999; Bachi 2000; Liker 2000). It has two binding sites for FG repeats (Fribourg 2001; Grant 2003) that are both required for efficient mRNA export (Braun 2001). The adaptor protein Aly/REF (Yra1 in yeast) binds the export factor and mRNA (Bruhn 1997; Strasser and Hurt 2000; Stutz 2000; Strasser and Hurt 2001) and is probably recruited by the DEAD-box helicase UAP56 (Sub2 in yeast) (Luo 2001; Strasser and Hurt 2001). UAP56 is essential for poly(A)$^+$-RNA in *S. cerevisiae, D. melanogaster*

and C. elegans (Gatfield 2001; Strasser and Hurt 2001). Another DEAD-box helicase that is involved in mRNA export is Dbp5 and is described in paragraph 1.3.1.1.

1.3 Helicase family of proteins

Helicases are ATPases that use the free energy of hydrolysis of adenosine triphosphate (ATP) to catalyze the separation of nucleic acid duplexes into single strands. These duplexes can consist of DNA-DNA, RNA-RNA, or DNA-RNA hybrids. Predictions for random sequence RNAs have revealed thermodynamic stable secondary structures with 50 % base pairing (Gralla and DeLisi 1974). This propensity to form stable structures is enhanced by the ability of RNA to form non Watson-Crick interactions. While some duplexes have defined conformations that fulfill cellular functions (e.g. tRNA, DNA double helix etc.), others are thermodynamically favored conformations with no functional relevance. Both types of duplexes need to be unwound at certain stages by helicases. In addition, some proteins that bind nucleic acids need to be removed (e.g. from mRNPs). Depending on their amino acid sequence, their substrates and unwinding polarity (5' - 3' or 3' - 5' or both), helicases are classified into five superfamilies (SF1 - SF5). Since the discovery of the first helicase more than 30 years ago, dozens of helicases have been identified in all kingdoms of life. Both DNA and RNA helicases belong to the same five superfamilies. DNA helicases are involved in processes such as DNA repair, recombination, replication and transcription. RNA helicases are involved in all aspects of RNA metabolism and many of them are essential. The superfamilies are further divided into subfamilies. SF1 and SF2 are the two largest superfamilies and are closely related. In yeast, 39 RNA helicases have been identified that belong to the SF2 family.

1.3.1 DEAD-box helicases

The largest SF2 subfamily is the so called DEAD-box family of RNA helicases. In yeast, 25 out of the 39 SF2 helicases belong to this subfamily. Most of them are involved in ribosome biogenesis or pre-mRNA splicing. Others play roles in translation, RNA export, RNA turnover and RNA quality control. The family of DEAD-box helicases was first described two decades ago (Linder 1989). At this time, sequence alignments identified eight conserved motifs. One of these, motif II, composed of the amino acids aspartate, glutamate, alanine and aspartate (DEAD)

gave the family its name. An additional motif (Q motif) consisting of a highly conserved glutamine residue (Tanner 2003) and a conserved aromatic residue upstream of the Q-motif was identified later (Cordin 2004). All DEAD-box proteins consist of a helicase core that is responsible for both ATP-binding/hydrolysis and RNA binding, which is achieved in a sequence independent manner. This core spans 350 - 400 residues and consists of two globular RecA-like domains that are connected by a linker (Caruthers and McKay 2002). Most DEAD-box helicases have N- and C-terminal flanking regions that are not conserved but are thought to play a role in specific interactions with binding partners and the localization within the cell (reviewed in Cordin 2006). The two RecA-like domains contain the 9 conserved motifs and are shown in Figure 2. Motif I (Walker A) and motif II (a variation of the Walker B motif) have been first described by John Walker (Walker 1982). They are not specific for RNA helicases and are found in other non helicase proteins (e.g. ABC transporters). Both motifs are responsible for NTP/ATP binding and hydrolysis. Motifs Ia, Ib, and IV, the partially conserved GG and the QxxR motif are involved in RNA binding. The Q motif is thought to be specific for the DEAD-box family of helicases and is involved in ATP binding. It has been suggested that the Q motif senses the nucleotide state of the protein and therefore regulates ATP binding and hydrolysis (Tanner 2003; Cordin 2004). Motifs V and VI are involved in the coordination and hydrolysis of ATP. The role of the motif III is somewhat ambiguous: it is thought to couple ATP hydrolysis with RNA unwinding. Some mutations in this motif have been shown to be ATPase deficient while other mutations have no effect on ATP hydrolysis but show loss of unwinding activity (Pause and Sonenberg 1992; Rocak 2005).
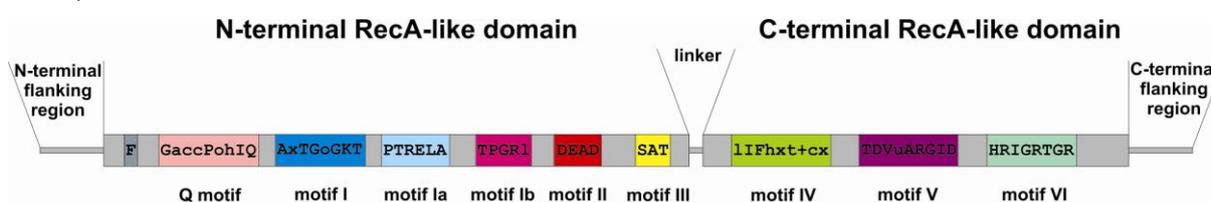
Figure 2: Schematic representation of the domain architecture of DEAD-box helicases. DEAD-box helicases consist of two RecA-like domains that form the minimal helicase core. The two domains are connected by a linker and include nine conserved motifs, shown here as different colored blocks with their consensus sequences. Most helicases have additional N- and C-terminal flanking regions of variable length or additional domains that are not conserved within the DEAD-box family.

From structural data it has become clear that despite their different functions and high specificity DEAD-box helicases bind RNA and ATP essentially in the same way (Andersen 2006; Bono 2006; Sengoku 2006). ATP is bound in between the two RecA-like domains. RNA is bound in a sequence independent manner by both domains. Although these proteins have been named for their activity (unwinding RNA helices/duplexes) actual unwinding activity has only been shown for a few helicases *in vitro* (e.g. Yang and Jankowsky 2006*)*. Unlike some SF1 and most DNA helicases DEAD-box proteins are not processive (reviewed in Jankowsky and Fairman 2007). Compared to DNA helicases, the intrinsic ATPase activity of these proteins is rather low (reviewed in Cordin 2006). Most DEAD-box helicases show an RNA dependent stimulation of the helicase activity that is independent of the RNA sequence. Some helicases have additional domains that confer specificity and activity. One example is DbpA, a helicase that binds specifically to 23S rRNA (Fuller-Pace 1993) via an additional domain. Activators have been described for only a few helicases (e.g. Alcazar-Roman 2006; Weirich 2006). Unwinding of blunt-end duplexes has been demonstrated for a few helicases (Rogers 1999; Liu 2008) but single stranded regions increase the unwinding efficiency, possibly by facilitating loading onto duplexes. Several helicases have been implicated in multiple cellular functions, as in the case of DEAD-box helicase 5 (Dbp5).

1.3.1.1 DEAD-box helicase 5

Dbp5 (DDX19B in *H. sapiens* and also called r̲ibonucleic a̲cid-t̲rafficking protein 8̲ (rat8) in yeast) belongs to the DEAD-box family of SF2 RNA helicases and is conserved across species. The 54 kDa human protein consists of an N-terminal flanking region (spanning 75 residues), the helicase core and a short C-terminal region (Figure 2 and Figure 6a). In yeast, the N-terminal flanking region (1 - 79) is dispensable (Snay-Hodge 1998) and its function is as yet unknown. Dbp5 has been shown to be essential for mRNA export in yeast (Snay-Hodge 1998; Tseng 1998) and human (Schmitt 1999). In yeast a temperature sensitive Dbp5 mutant results in the fast accumulation of poly(A)$^{+}$-RNA in the nucleus at permissive temperature (Snay-Hodge 1998). Dbp5 localizes at the cytoplasmic side of the NPC but is predominantly cytoplasmic (Snay-Hodge 1998; Tseng 1998). The helicase can shuttle between the nucleus and the cytoplasm. It docks at the NPC via the interaction with Nup214 (also called CAN) in humans (Kraemer 1994; Schmitt 1999) or Nup159 in yeast (Schmitt 1999). Nup214 is a cytoplasmic nucleoporin with a 45

kDa N-terminal β-propeller domain, a coiled-coil domain and C-terminal FG-repeats (Figure 3).

Dbp5 is an RNA-dependent ATPase with low activity *in vitro* and it is activated at the cytoplasmic side of the NPC by Gle1 and its cofactor IP_6 (Inositol hexakisphosphate) (Alcazar-Roman 2006; Weirich 2006). Mutations within the ATP binding motifs inhibit mRNA export and are lethal *in vivo* (Tseng 1998; Schmitt 1999). Dbp5 has been proposed to act in the final steps of mRNA export. It has been shown that Dbp5 can remove proteins from mRNA: *in vivo* studies in yeast indicate the removal of the transport factor Mex67 by the helicase (Lund and Guthrie 2005) and *in vitro* yeast Dbp5 displaces the RNA binding protein Nab2 in presence of ADP (Tran 2007). It has been suggested that the remodeling activities of the helicase (e.g. removal of the transport factor at the cytoplasmic side of the NPC) would bring directionality into mRNA export. Dbp5 has also been reported to bind cotranscriptionally in the nucleus with Balbani ring pre-mRNA in *C. tentans* (Zhao 2002) and might also interact with transcription factors (Estruch and Cole 2003). Furthermore, it is thought to play roles in translation termination, where the helicase has been shown to be involved in the recruitment of release factors (Gross 2007). More recently Dbp5 has been connected with P body components (Scarcelli 2008) and cell division in *S. cerevisiae* (van den Bogaart 2009). Whether these interactions are direct or mediated by another protein is at present unknown.
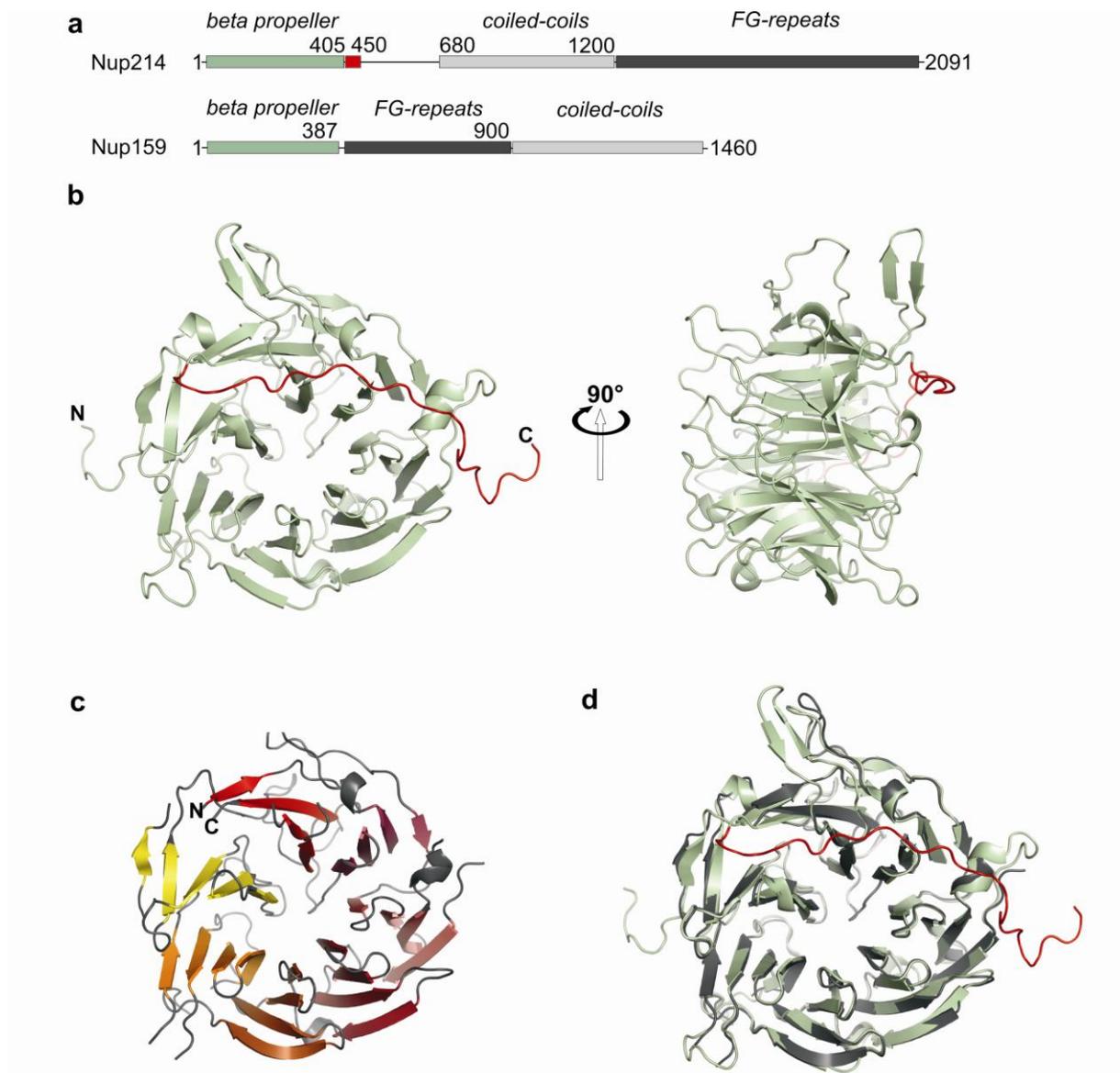
Figure 3: Comparison of *H. sapiens* Nup214 and *S. cerevisiae* Nup159.
(a) overall domain architecture of human Nup214 and yeast Nup159. Both proteins consist of a conserved N-terminal β-propeller domain (green), a coiled-coil domain (light grey) and FG-repeats (grey). Both in (b) human (Napetschnig 2007) and (c) yeast (Weirich 2004) the N-terminal domain mainly consists of anti parallel beta sheets connected with loops that form 7 beta propellers positioned around a central axis. (d) The two proteins superpose well (overall r.m.s.d of ~ 1.2 Å). The major difference lies in the C-terminal region. Here the crystal structure of the human protein features an extension (marked red in panel a and d) that binds across the β-propeller face in an unstructured fashion.

2   Aims and Scope of the Thesis

This PhD thesis examines the human DEAD-box helicase Dbp5 and its binding partner Nup214, both involved in nuclear export of messenger RNA. The work described in this thesis addresses the structural and biochemical details of the two proteins and tries to elucidate their role in nucleocytoplasmic transport.

Elucidation of the molecular mechanisms of complex recognition and formation requires detailed understanding of the three dimensional structures of the proteins. Even though the structures of the N-terminal domain of human Nup214 and the yeast orthologue Nup159 had been solved, it was not possible to draw conclusions on the functional role of the nucleoporin and its effect on the helicase. In order to understand how Dbp5 specifically docks at the nuclear pore complex via the interaction with Nup214 and to understand how the helicase binds to mRNA, it was necessary to solve the crystal structures of the two complexes.

The work focuses on two aspects: the Dbp5-AMPPNP-RNA complex and the Dbp5-Nup214 complex. In both parts the major aim was to determine the three-dimensional crystal structures of the two complexes. The insights and the conclusions drawn from the two structures were used to further characterize the DEAD-box helicase biochemically. Mutational studies, binding and ATPase assays were performed *in vitro* to support the structural conclusions.

The research of the presented thesis was carried out in the laboratory of Prof. Dr. Elena Conti at the European Molecular Biology Laboratories in Heidelberg from January 2006 to September 2007 and at the Max Planck Institute of Biochemistry in Martinsried from October 2007 until January 2009.

3 Materials and Methods

3.1 Materials

A detailed list of all suppliers is in appendix 6.6 and therefore only the names of the companies are specified in the text.

3.1.1 Chemicals and reagents

All common chemicals were purchased from Hampton, Fluka and Sigma-Aldrich unless otherwise stated.

3.1.2 Kits

Kit	Supplier
JBS Protein Methylation	Jena Bioscience
Qiaquick Gel Extraction	Qiagen
Qiaquick Spin Miniprep	Qiagen
QuikChange Site-directed Mutagenesis	Stratagene

3.1.3 Enzymes

All restriction enzymes including the appropriate buffers were purchased from New England Biolabs (NEB). For DNA amplification by PCR, Phusion polymerase from Finnzymes was used. Site directed mutagenesis was also performed with Phusion polymerase or with the polymerase supplied with the QuikChange site-directed mutagenesis kit (Stratagene). T4 DNA ligase, with its corresponding buffer, was purchased from Promega. T4 DNA Polymerase for LIC cloning was purchased from Novagen. TEV and PreScission (Rhinovirus 3C) protease were expressed and purified in the lab.

3.1.4 Vectors

	Vector	purchased from
E. coli	pETMCN	gift from C. Romier, IGBMC, Strasbourg
	LIC-system	designed & cloned by F. Martin and J. Basquin
	pGEX-4T	GE Healthcare
	pProEx-HTa	Invitrogen
	pETM-14	EMBL protein expression facility and G. Stier
	(& derivatives)	
Yeast	pYES 263	EUROSCARF
	pYES 260	
insect cells	pFASTBAC	Invitrogen

3.1.5 DNA oligos

DNA oligonucleotides were ordered from Sigma-Aldrich. PCR primers were purchased desalted and lyophilized and were diluted in H_2O. All sequences are given in 5´- 3´ direction.

Protein	Primer
yeast Dbp5	(82FWD)CTCTTAGAAATAAAAGTCCCGCGGCTGGGTTTAAGAGGTAACATGTCACG (83FWD)GCACTGTACAATGGAGAATTTGGGTCGGCGCCCTGAAAATAAAGATTCTC (482REV) GTCGGATCCCTACTAATCCTTTAACACTTTCTTAAC (457REV) TAAGGATCCCACGAGTCATTTCTACTACACCGAAG (293REV) GTCGGATCCCTATTCTAAAGTATTAGCATTTGGAACG
site directed mutagenesis	(E159AD160AFWD)GCTAACAAGAGTAAACCCAGCATCCGCATCTCCACAGGC (E159AD160AREV)GCCTGTGGAGATGCGGATGCTGGGTTTACTCTTGTTAGC (E202AK203AFWD)GTTCCCGATTCTTTTGCGGCAAACAAGCAAATTAATGCTC (E202AK203AREV)GAGCATTAATTTGCTTGTTTGCCGCAAAAGAATCGGGAAC (K205AQ206AFWD)CCCGATTCTTTTGAGAAAAACAAGCAAATTAATGCTCAAGTGATTG (K205AQ206AREV)CAATCACTTGAGCATTAATTTGCTTGTTTTTCTCAAAAGAATCGGG GCTATATGGTTTAATGACAATTGGATCTTCCATTATTTTTGTTGCA GCTATATGGTTTAATGACAATTGGATCTTCCATTATTTTTGTTGCA GCAGATAAGTTTGATGTTTTAACTGAGCTATATGGTTTAATGACAATTGGATCTTCC GCAGATAAGTTTGATGTTTTAACTGAGCTATATGGTTTAATGACAATTGGATCTTCC GACGACTTCAGAGAGGGTAGATCCAAAGTTTTGATTACTACTAATGTCCTGGCCCGT CCATCCAAAATTCAAGAAAGGGCGCTGCCATTATTATTAC (R224AR225AFWD)CGCCGGGCACTGTTCTTGACCTAATGGAAGAAAAATTGATGCAGC (R224ER225EREV)GCTGCATCAATTTTTCTTCCATTAGGTCAAGAACAGTGCCCGGCG (I255ER256EFWD)CAGCAGGGTCTAGGTGACCAGTGTGAAGAAGTTAAGAGATTTTTACCC (I255ER256EREV)GGGTAAAAATCTCTTAACTTCTTCACACTGGTCACCTAGACCCTGCTG
yeast Gle1	(1FWD)TTCTTCTTCCCATGGCAAGATTTGTGTTCGATGAGGTTTTCAATTCAG (118FWD)TTCTTCTTCCCATGGGCACGAAGAAATTGGTAAATGAGTCGG (135FWD)TTCTTCTTCCCATGGCGACAGCACCCTTACTCGAGGCAATAGAAG (239FWD)TTCTTCTTCCCATGGCACAAAACTTGGTGCTTGCCAATCAGAAAG (252FWD)TTCTTCTTCCCATGGGTAAAGCAGTTACAAACTTTGATAAAATCTCC (273FWD)TTCTTCTTCCCATGGCATTTTGGCATTATAAAGACAAAATTGCTC (238FWD)TTCTTCTTCCCATGGCTGATGTGAATGTAAGGAATTTGCTC (FLREV)AAAAAAAAAGGATCCCTAAGGAGACATTTCCGGAAAGGACTCC (REV2)AAAAAAAAAGGATCCGCTTTTTTTATAGGCAGAACTATGTCTTGC
human Dbp5	(1FWD)GGAATTCCATATGACATCTCCAAAATCGGATCTGGNTCGCG (68FWD)GGAATTCCATATGAGCAACCTTGTTGATAACACAAACC (55FWD)GGAATTCCATATGGACAGAGCTGCCCAGTCCTTACTC (58FWD)GGAATTCCATATGGCCCAGTCCTTACTCAACAAGCTG (72FWD)GGAATTCCATATGGATAACACAAACCAAGTGGAAGTCC (84FWD)GGAATTCCATATGCCAAACTCCCCTCTGTACTCGGTG (300REV)CCCGTCGGATCCCTACTCACGCTTCAGTTTGATAACGTTTGGG (302REV)CCCGTCGGATCCCTACTCTTCCTCACGCTTCAGTTTGATAACG (306REV)CCCGTCGGATCCCTAGGTGTCCAGGGTCTCTTCCTCACGC (479REV)CCCGTCGGATCCCTAGTTGGCGATTTTCTCAATCTCGTCC
site directed mutagenesis	(D223RFWD)GGCACCCCTGGGACTGTGCTGAGGTGGTGCTCCAAGCTCAAGTTCATTG (D223AFWD)GGCACCCCTGGGACTGTGCTGGCCTGGTGCTCCAAGCTCAAGTTCATTG (D223RREV)CAATGAACTTGAGCTTGGAGCACCACCTCAGCACAGTCCCAGGGGTGCC (D223AREV)CAATGAACTTGAGCTTGGAGCACCAGGCCAGCACAGTCCCAGGGGTGCC (I258AFWD)CCACCAAGATCAGAGCGCCCGCATCCAGAGGATGCTGCCC (I258DFWD)GGGCAGCATCCTCTGGATGCGGGCGCTCTGATCTTGGTGG (I258AREV)CCACCAAGATCAGAGCGACCGCATCCAGAGGATGCTGCCC (I258DREV)GGGCAGCATCCTCTGGATGCGGTCGCTCTGATCTTGGTGG (R259AFWD)CCACCAAGATCAGAGCATCGCCATCCAGAGGATGCTGCCC (R259DFWD)GGGCAGCATCCTCTGGATGGCGATGCTCTGATCTTGGTGG (R259AREV)CCACCAAGATCAGAGCATCGACATCCAGAGGATGCTGCCC (R259DREV)GGGCAGCATCCTCTGGATGTCGATGCTCTGATCTTGGTGG (R262AFWD)GATCAGAGCATCCGCATCCAGGACATGCTGCCCAGGAACTGCC (R262DFWD)GATCAGAGCATCCGCATCCAGGCGATGCTGCCCAGGAACTGCC (R262AREV)GGCAGTTCCTGGGCAGCATGTCCTGGATGCGGATGCTCTGATC (R262DREV)GGCAGTTCCTGGGCAGCATCGCCTGGATGCGGATGCTCTGATC

| human Nup214 | ```
(1FWD)GGAATTCCATATGGGAGACGAGATGGATGCCATGATTCCCG
(450REV)CCCGTCGGATCCCTAATTAATCATATAAAATGGACAAAGCACACCATCTGTTG
(405REV)CCCGTCGGATCCCTAAGAAGCATCCAGTTTCTGTGGGGCTTGAGAGGAG
``` |
|---|---|
| site directed mutagenesis | ```
(R359AFWD)GCCTGTGACAGACAAGAGTGATGCCTCCTTGCCCATGGGAGTTGTCG
(R359AREV)CGACAACTCCCATGGGCAAGGAGGCATCACTCTTGTCTGTCACAGGC
(R359DFWD)GCCTGTGACAGACAAGAGTGATAGGTCCTTGCCCATGGGAGTTGTCG
(R359DREV)CGACAACTCCCATGGGCAAGGACCTATCACTCTTGTCTGTCACAGGC
(R348AFWD)GGCTACTGGAGGATTCTAGTGCAGCTGAATTGCCTGTGACAGAC
(R348AREV)GTCTGTCACAGGCAATTCAGCTGCACTAGAATCCTCCAGTAGCC
(R348DFWD)GGCTACTGGAGGATTCTAGTGACGCTGAATTGCCTGTGACAGAC
(R348DREVGTCTGTCACAGGCAATTCAGCGTCACTAGAATCCTCCAGTAGCC
(V353AFWD)CTAGTCGAGCTGAATTGCCTGCGACAGACAAGAGTGATG
(V353AREV)CATCACTCTTGTCTGTCGCAGGCAATTCAGCTCGACTAG
(V353DFWD)CTAGTCGAGCTGAATTGCCTGACACAGACAAGAGTGATG
(V353DREV)CATCACTCTTGTCTGTGTCAGGCAATTCAGCTCGACTAG
``` |

3.1.6 RNA oligos

In all binding and crystallization experiments single stranded (ss) RNAs were used.

	length (bases)	purchased from
ss poly-U	5, 6, 7, 8, 10, 12, 15 & 20	Dharmacon
ss poly-A	10, 15, 20	Dharmacon
5´-biotinylated ss poly-U	20	Dharmacon
20γ polyuridylic acid potassium salt	~ 300 - 2000	Sigma-Aldrich

3.1.7 Media and buffers

Plates, buffers and media were prepared and autoclaved by EMBL's media kitchen or at MPIB.

Media	
Luria-Bertani (LB) (Miller 1972)	1 % (w/v) bacto trypton 0.5 % (w/v) bacto yeast extract 170 mM NaCl The medium was adjusted to pH 7.6 with NaOH
LB agarose plates	1.5 % (w/v) bacto agar in LB antibiotics (50 µg/ml Ampicillin or 25 µg/ml Kanamycin)
Terrific broth (TB)	1.2 % bacto trypton 2.4 % bacto yeast extract 0.4 % glycerol ddH$_2$O to 900 ml 0.017 M KH$_2$PO$_4$ 0.072 M K$_2$HPO$_4$

Auto inducing (Studier 2005)	929 ml ZY + 20 ml 5052 + 50 ml NPS + 1 mM $MgSO_4$
ZY	1 % N-Z-Amine (10 g /l) or Tryptone
	0.5 % Yeast-Extract (5 g/l)
5052 (50x stock)	25.0 % Glycerol
	2.5 % Glucose
	10 % α-lactose-monohydrate
NPS (20x stock)	0.5 M $(NH_4)_2SO_4$
	1.0 M KH_2PO_4
	1.0 M Na_2HPO_4
	pH 6.75
Purification buffers	
Lysis buffer	50 mM Hepes pH 7.5
	500 mM NaCl
	1 mM DTT
	complete protease inhibitor cocktail (Roche)
GSH elution buffer	20 mM HEPES pH 7.5
	100 mM NaCl
	1 mM DTT
	10 mM $MgCl_2$
	10 % Glycerol
	20 mM reduced GSH
	complete protease inhibitor cocktail (Roche)
Ni-NTA elution buffer	20 mM HEPES pH 7.5
	100 mM NaCl
	1 mM DTT
	10 mM $MgCl_2$
	500 mM Imidazole
	complete protease inhibitor cocktail (Roche)
Ionexchange buffer	20 mM HEPES pH 7.5
	1000 mM NaCl
	1 mM DTT
	complete protease inhibitor cocktail (Roche)
	(supplemented with 10 mM $MgCl_2$ and 10 % glycerol for the AMPPNP+RNA complex)
Gel filtration buffer	20 mM HEPES pH 7.5
	100 mM NaCl
	1 mM DTT
	(supplemented with10 mM $MgCl_2$ and 10 % glycerol for the AMPPNP-RNA complex)

Buffers for assays	
ATPase buffer	20 mM HEPES pH 6.5
	100 mM NaCl
	1 mM DTT
	10 mM $MgCl_2$
	10 % Glycerol
	0.1 % Nonidet P40
binding buffer	20 mM HEPES pH 7.5
	100 mM NaCl
	1 mM DTT
	10 mM $MgCl_2$
	10 % Glycerol
	0.1 % Nonidet P40
SAXS buffer	20 mM HEPES pH 7.5
	100 mM NaCl
	1 mM DTT
	10 mM $MgCl_2$
	5 % Glycerol

3.1.8 Cloning and expression strains

	Strain / Cell type	Genotypes
E. coli	XL1 blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacIqZ∆M15* Tn*10* (Tetr)]
	BL21 DE3	*E. coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal endA* Hte
	BL21 DE3 Gold	*E. coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal* λ(DE3) *endA* Hte
insect cells	SF9	wt
S. cerevisiae	BJ5459	*Matα, ura3-52, trp1, lys2-801, leu2∆1, his3∆200, pep4::HIS3, prb1∆1.6R, can1, GAL>*

3.1.9 Equipment

Instrument	Model	Supplier
Cell culture		
Shakers	KS-15	Edmund Bühler, Kühner
	Climo-Shaker ISF1X & W	
Centrifuges	Avanti J-20 XP	Beckman Coulter
	Micro centrifuge 5417C & 5810	Eppendorf
	Megafuge 1.0	
	Sorvall RC 5C	
		Sorvall

Chromatography		
FPLC	Aekta Explorer	GE Healthcare
	Aekta Purifier	
	Aekta Prime	
Columns	Hiload Superdex 200 16/60	
	Hiload Superdex 200 10/300	
	Superdex 200 5/150 gl	
	Hitrap Heparin, Hitrap Q	
Chromatography media	Heparin Sepharose, GSH-Sepharose 4B fast flow	
	His-select gel	Sigma Aldrich
	magnetic streptavidin beads	Dynal
Crystallization		
Plates	96well MRC96T	SwissCI
	24well vdx	Hampton Research
Pipetting robots	Phoenix	Art Robbins Instruments
Electrophoresis		
Gel casting	Made by workshops	EMBL & MPI
Precast gels	NuPAGE	Invitrogen
Gel chamber		
Power supply	PowerPac300	Biorad
	PowerPack515T	Biometra
	PowerEase	Invitrogen
	GPS200/400	GE Healthcare
	LKB-EPS 500/400	
	EPS301	
Gel imaging	Safe imager	Invitrogen
	Ebox3026	Peqlab
Electroporation		
Instrument	gene pulser	Biorad
	micro pulser	
Cuvettes	Gene pulser cuvette 0,2 cm electrode gap	
French-press	Emulsiflex	Avestin
Liquid scintillation counter	Tri-carb 2100TR LSA	Packard
PCR	Primus96	Peqlab
pH meter	Lab860	Schott
Pipettes		Eppendorf
Scales	LA1200S	Sartorius
	TE1502S	
Static light scattering	TDA302	Viscotech
Surface plasmon resonance		
Chips	CM5 chip	GE Healthcare
Instrument	BIAcore 3000	
Thermoblock/-shaker		Memert
	compact	Eppendorf
UV/visible spectrometer	BioPhotometer	Eppendorf
	Nanodrop ND-1000	Thermo Fisher
Water bath	Arius 611UF & 611DI	Sartorius
X-ray		
Sealed tube	PX scanner	Oxford Diffraction
Plate scanner	Xcalibur Nova	
Cryo loops	CryoLoop 20 micro	Hampton research
Caps and vials	Magnetic caps and vials	Molecular Dimensions

3.1.10 Synchrotron facilities and X-ray sources

All crystallographic data were collected at Swiss Light Source (SLS) at the Paul Scherrer Institute (PSI), Villigen, Switzerland at beam lines X06SA and X10SA. Crystals were tested in-house on PX-scanner and Excalibur before measuring at SLS. Small angle scattering data were collected at Advanced Light Source (ALS), Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA, USA at beam line SYBILS12.3.1.

3.1.11 Software

The following software was used to write this thesis, analyze data and generate figures and alignments.

Adobe Illustrator (www.adobe.com)
Adobe Photoshop (www.adobe.com)
BIAcore evaluation software (GE Healthcare)
Bioedit (Tom Hall, Ibis Biosciences)
CCP4 (CCP4 1994)
Coot (Emsley and Cowtan 2004)
Credo (Petoukhov 2002)
Crysol (Svergun 1995)
Damaver (Svergun 1999)
Dammin (Svergun 1999)
DSSP (Kabsch and Sander 1983)
Endnote (www.endnote.com)
Gnom (Svergun 1992)
Microsoft Office (www.microsoft.com)
Origin (www.originlab.com)
Primus (Konarev 2003)
Pymol (DeLano)
XDS (Kabsch 1993)

3.1.12 Servers

The following servers were used to analyze the refined structures and to generate input files for data refinement in Refmac (CCP4 1994).

3d-SS (http://cluster.physics.iisc.ernet.in/3dss/)
Aquaprot (http://bioinfo.weizmann.ac.il/aquaprot/)
Molprobity (http://molprobity.biochem.duke.edu/)
Prodrg (http://*davapc1.bioch.dundee.ac.uk/prodrg/*)
TLSMD (http://skuld.bmsc.washington.edu/~tlsmd/)

3.2 Methods

3.2.1 Cloning

3.2.1.1 PCR

PCR reactions were set up in a total volume of 50 µl:

> 0.5 µl Phusion Polymerase (1 U/µl)
>
> 10 µl 5x reaction buffer
>
> 1 µl DMSO (100 %)
>
> 1 µl dNTP mix (dATP, dTTP, dCTP, dGTP, 10 mM each)
>
> 1 µl forward primer (10 µM)
>
> 1 µl reverse primer (10 µM)
>
> 50 ng template DNA

The following cycling conditions were used:

> 30 s - 98°C initial denaturing step
>
> 10 s - 98°C denaturing
>
> 10 s - annealing 30 cycles*
>
>   x s - 72°C extension**
>
> 5 min - 72°C final extension

Annealing temperature was varied depending on the melting temperature (Tm) of the primers. In some cases touch-down PCR was applied with a decrease of the annealing temperature of 0.5°C per cycle.
**An extension time of 30 s per kb was used.

3.2.1.2 Agarose gel electrophoresis

DNA fragments generated by PCR or restriction digestion were separated by agarose gel electrophoresis, using a minigel system produced by the mechanical workshops, EMBL (Heidelberg; Germany). 1 % w/v agarose was melted in 1x TBE buffer (90 mM Tris-base, pH 8.3, 90 mM boric acid, 1 mM EDTA). After cooling to 65ºC, ethidium bromide or SYBR safe (Invitrogen) was added (0.5 µg/ml final concentration), and gels were casted. Before loading, the samples were supplemented with 0.5 volumes DNA dye (20 % Ficoll, 1 mM EDTA, 0.1 % SDS, 0.05 % Bromophenol Blue). DNA size marker (1 kb ladder, NEB) was loaded in a separate lane. Gel electrophoresis was performed at 7 V/cm in 1x TBE buffer. DNA fragments were visualized with UV light (ethidium bromide) or a blue light transilluminator (SYBR safe, Invitrogen) emitting blue light at a wavelength of ~ 470 nm.

3.2.1.3 Purification of DNA fragments

DNA fragments from restriction digestions were purified from an agarose gel. DNA bands of interest were cut from gel and purified with the gel extraction kit (Qiagen) according to the manufacturer's instructions. DNA fragments generated by PCR were purified directly from the reaction mixture using the PCR purification kit (Qiagen) and following the manufactures manual. All DNA fragments were eluted in buffer containing 10 mM Tris-Cl, pH 8.5

3.2.1.4 Restriction digest and ligation

Plasmids and DNA fragments generated by PCR were digested with appropriate enzymes (NEB). Optimal reaction buffers were used as recommended by the manufacturer. Reactions were typically carried out in a total volume of 50 µl, using 1 µl of each enzyme (10 U/µl), 5 µl of the appropriate 10x buffer (NEB), and if required 0.5 µl BSA (10 mg/ml). Reaction time and temperature varied depending on the enzymes used. Cloned plasmids were all tested by restriction digest prior to DNA sequencing. Plasmids and PCR fragments were digested with the appropriate restriction enzymes and were ligated using T4 DNA ligase (Promega). Generally, 6 µl of the gel-purified digested PCR product and 20 ng linearised plasmid were ligated. Ligations were performed in a 20 µl reaction containing the linearised plasmid and the PCR product, 1 µl T4 DNA ligase (400 U/ml, Promega), and 2 µl 10x T4 DNA ligase buffer (100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP, 300 mM Tris-Cl, pH 7.8). Ligation reactions were performed for at least 1 h at room temperature.

3.2.1.5 LIC cloning

3.2.1.5.1 Principle

Ligation independent cloning (LIC) is a fast and efficient way to integrate the desired ORF into a vector. Vectors and protocols were designed and produced by Jérôme Basquin and Florence Martin in our lab. The LIC system is using the 3' to 5' exonuclease activity of a T4 DNA polymerase to generate 12 - 15 bp overhangs on both the vector and the insert. The complementary overhangs of plasmid and insert anneal and are stable enough (without ligation) to be transformed into cells. The nicks in the plasmid get repaired by the *E. coli* replication machinery. Since this method is independent of restriction enzymes and their target sequences, this system provides a fast way to clone different constructs in different vectors. Here the principle and sequences are briefly described:

Linearization of vector with SacII:

```
5`actagtgaaaacctgtatttccagggagcagccgc      ggccggtgctttgcaggatcc`3
3`tgatcacttttggacataaaggtccctcgtcgg      cgccggccacgaaacgtcctagg`5
  ------ E  N  L  Y  F  Q  G  A ------------              ------
   SpeI        TEV Site  ^        SacII                   BamHI
```

Treatment with T4 DNA Pol + dTTP:

Digestion with T4 DNA polymerase with its 3' to 5' nuclease activity produces overhangs (underlined). The polymerase stops as soon as it reaches a thymine.

```
5`actagtgaaaacctgtattt                      ggccggtgctttgcaggatcc`3
3`tgatcacttttggacataaaggtccctcgtcgg                        tcctagg`5
```

<u>Insert processing:</u>

The LIC sequences are included in the primers.

```
5`ccagggagcagcctcgATG(desired ORF)TAAcgaggccggtgctttgc`3
3`ggtccctcgtcggagctac            attgctccggccacgaaacg`5
                M                   *
```

Digestion with T4 DNA polymerase (+ dATP) produces the overhangs (underlines).

```
5`ccagggagcagcctcgATG(desired ORF)TAAcga`3

3`          agctac            attgctccggccacgaaacg`5
```

After annealing of the complementary overhangs (underlined) the final plasmid (including two nicks) can be used for transformation:

```
5`tctaga(TAG)actagtgaaaacctgtatttccagggagcagcctcgATG(desiredORF)TAAcgaggccggtgctttgcaggatcc`3
3`agatct     tgatcacttttggacataaaggtccctcgtcggagctac            attgctccggccacgaaacgtcctagg`5
  ------     ------ E  N  L  Y  F  Q  G  A  A  S  M        *              ------
   XbaI       SpeI        TEV Site  ^                                      BamHI
```

The integration of the LIC system into different vectors allows different tags and protease sites.

3.2.1.5.2 Experimental procedure

<u>Insert processing:</u>

For processing of the PCR product, the following reaction was prepared:

gel purified PCR product	600 ng
T4 DNA Pol. buffer (10x)	2 µl
dATP (25 mM)	2 µl
DTT (100 mM)	1 µl
T4 DNA Pol. LIC qualified (Novagen)	0.4 µl
H_2O	add to 20 µl

The reaction mix was incubated for 30 min at room temperature. The enzyme was inactivated at 75°C for 20 min.

Vector processing:

The vector needs to be linearised first by digestion (or by PCR). Therefore 2 µg of vector were digested with 60 u SacII (or 20 u ZraI for 3C LIC vectors).

The cut vector was gel purified (0.8 % agarose gel) and processed as followed:

linearized vector	450 ng
T4 DNA Pol. buffer (10x)	3 µl
dTTP (25 mM)	3 µl
DTT (100 mM)	1.5 µl
T4 DNA Pol. LIC qualified (Novagen)	0.6 µl
H_2O	add to 30 µl

The reaction mix was incubated for 30 min at room temperature. The enzyme was inactivated at 75°C for 20 min.

Annealing reaction:

For the annealing, 2 µl of processed insert and 1 µl of processed vector were mixed and incubated for 10 min at room temperature. Then 1 µl of EDTA (25 mM) was added and the mix was incubated 10 min at room temperature. 2 µl of the reaction were used for transformation into *E. coli* cells.

3.2.1.6 Transformation

Transformation-competent bacteria were thawed on ice. 2 µl of the ligation reaction or 0.2 µg of plasmid were added to 50 µl of the competent bacteria, and transformed using the electroporation method. Here an electrical pulse of 1.8 kV (0.1 cm cuvettes) or 2.5 kV (0.2 cm cuvettes) was applied. After electroporation cells were slowly shaken on a thermomixer at 37°C for 30 min. The bacteria were then plated onto LB agar plates containing the appropriate antibiotic and were incubated over night at 37°C.

3.2.1.7 Site-directed mutagenesis

For the generation of point mutants PCR mutagenesis was performed, using the QuikChange Site-Directed Mutagenesis Kit (Stratagene) or Phusion polymerase (Finnzymes). The mutagenesis strategy involves the amplification of the target plasmid with sense and antisense oligonucleotide primers that contain the desired mutation. The same protocol as for the normal PCR was used. After amplification 1 µl of Dpn I enzyme (20 U/µl) was added to the amplification reaction and incubated for 5 h at 37°C in order to digest the template DNA.

3.2.1.8 DNA sequencing

All plasmids were validated by DNA sequencing performed by core facilities at EMBL Heidelberg or MPIB Martinsried.

3.2.2 Expression and purification of recombinant proteins

All proteins were expressed in baffled Erlenmeyer flasks with 10 % v/v media per flask. Traces of antifoam (Sigma-Aldrich) were used to prevent formation of foam. Cells were grown at an initial temperature of 37°C to different optical densities (OD_{600}) depending on the media used (auto inducing: OD_{600} ~ 2 - 3, TB: OD_{600} ~ 3, LB: OD_{600} ~ 0.8). For (overnight) expression the temperature was decreased to 18°C. Cells grown in TB and LB were induced with 0.1 - 0.5 mM IPTG. All purification steps were carried out at 4°C or on ice. Cells were lysed in lysis buffer by 2 - 3 french-press cycles.

3.2.2.1 Yeast Dbp5

All yeast Dbp5 constructs were expressed at 18°C in *E. coli* BL21 Gold cells using LB or auto inducing media. Cells were lysed in lysis buffer (paragraph 3.1.7) using an emulsiflex french-press. Lysed cells were centrifuged for 45 min at 48000 x g. GST-tagged constructs were batch purified using GSH-Sepharose. The GST-tag was cleaved over night at 4°C with TEV protease. Untagged proteins were eluted from resin with gel filtration buffer (paragraph 3.1.7), and proteins were further purified by gel filtration chromatography (S200).

3.2.2.2 Yeast Nup159

All yeast Nup159 constructs were expressed in *E. coli* BL21 Gold cells using LB or auto inducing media. In the case of LB 0.1 - 0.5 mM IPTG was used for induction at 18°C. Cells were lysed in lysis buffer (paragraph 3.1.7) using an emulsiflex french-press. Lysed cells were centrifuged for 45 min at 48000 x g. 6xHis-tagged constructs were batch purified using Ni-NTA resin. Protein was eluted in gel filtration buffer (paragraph 3.1.7) containing 500 mM Imidazole and tag was cleaved overnight using TEV protease at 4°C during dialysis against gel filtration buffer (paragraph 3.1.7). Protein was further purified by ion exchange chromatography (Hitrap Q) and a final gel filtration step (S200).

3.2.2.3 Yeast Dbp5-Nup159 complex

Yeast Dbp5 and Nup159 were both expressed at high levels when cotransformed into *E. coli* BL21 cells. Cells were lysed using french-press. The complex was captured in a GSH-Sepharose affinity batch purification step followed by tag cleavage on the resin with TEV protease overnight (4°C) on a rolling shaker. Cleaved proteins were eluted from the resin with gel filtration buffer and were further purified by gel filtration chromatography (S200). The complex was also reconstituted from separately purified proteins and further purified by size exclusion chromatography (S200).

3.2.2.4 Human Dbp5

Human Dbp5 constructs were expressed at 18°C overnight as GST-fusions in *E. coli* BL21 cells using auto inducing media. GST-Dbp5 constructs were captured on GSH-resin followed by extensive washing in gel filtration and ion exchange buffer. The 1 M NaCl wash (ion exchange buffer) was implemented to remove bound nucleic acids from the protein. GST-Dbp5 was either eluted with GSH elution buffer or the GST-tag was cleaved overnight with TEV/PreScission protease. Tagged and untagged Dbp5 constructs were further purified using a Heparin column. Bound proteins were eluted in gel filtration with a linear salt gradient of (100 to 1000 mM NaCl) and subjected to a size exclusion chromatography step (S200).

3.2.2.5 Human Nup214

Nup214 βΔC was expressed in *E. coli* BL21 Gold cells as N-terminal 6xHis- or GST-fusion proteins. Initial capturing using the appropriate resin (Ni-NTA or GSH-Sepharose) was followed by overnight tag cleavage using TEV protease. After tag cleavage untagged proteins were further purified by ion exchange chromatography (Hitrap Q) and size exclusion chromatography (S200).

3.2.2.6 Human Dbp5-Nup214 complex

For complex formation the two purified proteins were mixed at 4°C for one hour in gel filtration buffer. A slight excess of Dbp5 was typically used to saturate the nucleoporin. The formed complex was separated from an excess of unbound Dbp5 or Nup214 by size exclusion chromatography (S200).

3.2.3 Polyacrylamid gel electrophoresis

Proteins were analyzed on SDS PAGE using 15 % gels (according to Laemmli 1970). Separating gels were prepared according to the following recipe:

2.7 ml H_2O

2 ml 1.5 M Tris, pH 8.7

3.2 ml 30 % acrylamide/N,N'-methylene-bisacrylamide solution (37.5:1)

80 µl 10 % SDS

36 µl 10 % ammonium persulfate (APS)

12 µl 100 % N, N, N'N'-tetramethylethylene diamine (TEMED)

The stacking gel had the following composition:

2.63 ml H_2O

500 µl 1 M Tris-Cl, pH 6.8

800 µl 30 % acrylamide/N,N'-methylene-bisacrylamide solution (37.5:1)

40 µl 10 % SDS

24 µl 10 % APS

8 µl 100 % TEMED

3.2.4 Native gel electrophoresis

Native gels were used to analyze the homogeneity of protein samples and for electromobility shift assays (EMSA). Gels were cast and run in a glass-plate and a gel apparatus made by the EMBL workshop. 10 % polyacrylamide (acrylamide:bisacrylamide 29:1) in 50 mM Tris-acetate pH 8.0 and 10 mM magnesium-acetate was used. To avoid heat (denaturing), gels were run on ice or at 4°C with 80 V. 20 % glycerol was added to the samples as loading buffer. One lane was loaded with buffer supplemented with 20 % glycerol and traces of bromophenol blue as a control of migration. Gels were then stained for nucleic acids with toluidine (0.1 % toluidine in water).

3.2.5 ATPase assays

An initial reaction was performed to find the best pH condition for this assay, using different buffers with pH from 6.0 up to pH 8.0 in 0.5 pH steps.

Final ATPase reactions were performed in the buffer of maximal ATPase activity of Dbp5 containing 50 mM MES pH 6.5, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 10 % glycerol and 0.1 mg/ml BSA. 15 pmol of Dbp5, 2 mM ATP, 4 µg polyU RNA and traces of [γ-^{32}P] ATP (10 ci/mmol purchased from Perkin Elmer). Reactions were

incubated 0, 30, 60, 90 min at 30°C and stopped by adding 400 µl of ice-cold 10 % (w/v) acid-washed charcoal (Sigma-Aldrich) in 10mM EDTA as described. After 30 minutes of centrifugation at 16,000 g, supernatants containing [γ-^{32}P] P_i were counted by the Cerenkov method using a Packard liquid scintillation counter (Tri-carb 2100TR LSA).

3.2.6 Binding experiments

3.2.6.1 GST pull-downs

All GST-pull-down experiments were carried out at 4°C. About 2 µg of GST-tagged recombinant Dbp5 proteins were immobilized on 20 µl of glutathione agarose beads (GE Healthcare) per binding reaction. After 2 hours incubation, beads were washed three times with 1 ml of binding buffer (paragraph 3.1.7). Equimolar amounts of Nup214 proteins were added in a final volume of 100 µl binding buffer per reaction. A two-fold excess of Nup214 proteins was used for pull-downs with mutant proteins. After two-hour incubation, beads were washed three times with 1 ml binding buffer. Bound proteins were eluted with SDS sample buffer and analyzed by SDS PAGE.

3.2.6.2 RNA binding

All RNA-binding experiments were performed at 4°C using 5'-end-biotinylated U_{20} ssRNA (Dharmacon). 2 µg of each Dbp5 construct was incubated in binding buffer in the presence or absence of 1 mM AMPPNP and 170 nM U_{20} RNA. Samples were incubated overnight at 4°C. 50 µg magnetic streptavidin beads (Dynal) were added for 1h. Beads were washed three times with 0.5 ml binding buffer. Proteins mixtures before (input, 17 % of the total) and after co-precipitation (precipitate, eluted in SDS loading buffer) were separated on a 15 % (w/v) acrylamide SDS-PAGE (as described in paragraph 3.2.3) and visualized using Coomassie stain.

3.2.6.3 Electromobility shift assays (EMSA)

EMSA experiments were performed in order to analyze binding of Dbp5 to nucleic acids. EMSA were performed in buffer containing 50 mM HEPES pH 8.0, 100 mM NaCl, 10 mM MgCl$_2$, 1 mM DTT, 10 % glycerol supplemented or not with 1 mM ATP, AMPNP, ADP and AMP. Proteins were incubated with ssRNAs (6 - 20 nucleotides poly-U or poly-A, see paragraph 3.1.6) for 30 min on ice, before loaded on a native polyacrylamide gel. Gels were stained for nucleic acids using toluidine. After toluidine staining gels were destained in water and then stained for proteins using Coomassie.

3.2.6.4 ATP binding

Binding affinities of Dbp5 for ATP were measured using a spin column protocol. Here, the protein is incubated with the ligand. Bound and unbound ligands (in this case [γ-^{32}P] ATP) are separated using small spin gel filtration columns. Quantification of the amount of bound ligand in the presence of increasing concentrations of proteins allows the calculation of the binding constant (K_D). For K_Ds ≥ µM this method is limited by the amount (& solubility) of the proteins used. For the reaction the same buffer (pH 6.5) as for ATPase assay was used. For the gel filtration columns, a 5 % Sephadex G75 (GE Healthcare) suspension (incubated on rotating shaker for 2h at 4°C) in ATPase buffer was prepared. Minicolumns (Micro Bio-Spin, Bio-Rad) were loaded with 800 µl Sephadex G75 suspension and spun for 1 min at 750 x g (without drying the matrix). 5x10^{-5} - 5 nmols of protein were mixed with buffer in a total volume of 10 µl on ice and then pre-incubated for 10 min at 37°C. 10 µl of ATP mix containing traces (~ 2 fmol) of [γ-^{32}P] ATP (Perkin Elmer) buffer was added. Samples were incubated for 2 min at 37°C. The 20 µl samples were then added on the resin of the freshly prepared spin columns. Columns were spun for 2 min at 750 x g. The eluates were counted in a scintillation counter using the Cerenkov method (as described in 3.2.5). Data were analyzed with the following assumptions:

$E + L = EL$ (where E is the enzyme and L its ligand)

$K_D = \frac{[L]+[E]}{[EL]}$ and $[L]_0 = [L] + [EL]$ can be combined to

$$\frac{[EL]}{[L]_0} = \frac{[E]}{K_D + [E]} \qquad \text{(eq. 1)}$$

when $[E] \gg [L]$ it is assumed that $[E]\sim[E]_0$ (where $[E]_0$ is the range of protein concentration). Since negligible amounts of labeled ATP (L) are incubated with the protein (E) this assumption is valid. K_D can be estimated with the above equation with measured $[EL]$ and known concentration of the protein by plotting $\frac{[EL]}{[L]_0}$ against the concentration range $[E]_0$.

3.2.7 Surface plasmon resonance (SPR) spectroscopy

Binding constants for the Dbp5-Nup214 complex were determined using SPR spectroscopy. All experiments were designed together with Claire Basquin (MPIB) who performed the experiments.

3.2.7.1 Theory

Surface Plasmon Resonance (SPR) is based on surface plasmons. These are fluctuations of electron density at the boundaries of two materials (metal and dielectric) and can be described by different models (quantum mechanics, Drude model, Drude 1900). Surface plasmons can be excited by both electrons and photons and are then called polaritons. These propagate along the metal surface and decay. The excitation of plasmons by visible light is denoted as SPR. Incident polarized light interacts with the free oscillating electrons in the metal surface. Resonance occurs at a critical angle of incident, resulting in a minimum of intensity of reflected light (figure 3a and b). The angle depends on the refractive index (RI) at, or close to the metal surface. Therefore SPR instruments can be used to detect small changes of RI equivalent to a change in mass (for e.g. the immobilization of a ligand or a binding event). SPR chips are typically coated with gold, since it is a metal that supports surface plasmons.

Figure 4: Surface Plasmon Resonance spectroscopy. (a) schematics of an SPR experiment. Polarized light is used to exited surface plasmons. The reflected light is recorded by a detector. Molecules A (red) is immobilized on the functionalized gold surface (e.g. covalently bound antibodies) of the SPR chip. This results in a minimum of intensity of reflected light (see panel b, minimum A). The analyte (molecule B in yellow) binds to molecule A and results in a minimum that is shifted by angle φ (see panel b, minimum B). (b) Example of excitation of surface plasmons. At a certain angle φ excitation occurs, resulting in a minimum of intensity of reflected light. A change on the surface (e.g. immobilization of a protein) causes a change in refractive index resulting in an angular shift from A to B. c) example of a sensorgram, molecule A is immobilized on the gold surface and the analyte (molecule B) is injected. Binding of the two molecules results in a change of the resonance signal.

This technique is typically used to determine the binding affinity between two molecules. One of the molecules (ligand) is immobilized to a chemically modified matrix using well defined chemistry and the binding partner (analyte) is injected from the micro fluidic system. The binding signal observed is generally expressed in response units (RU) and reflects the association (on-rate) and dissociation (off-rate) of the reaction, from which the affinity constants can be calculated. Scatchard plots (SPR angle at equilibrium (Req) / concentration of analyte) can be used to evaluate data but can result in large errors especially in the low concentration range. Therefore a better approach is to use a non-linear regression using the Langmuir binding isotherm,

$$Req = (\frac{[A]}{[A] + K_D})Bmax \qquad \text{(eq. 2)}$$

where $Bmax$ is the maximum binding capacity, $[A]$ the concentration of the analyte and the dissociation constant K_D. Various binding models can be used to evaluate data (Karlsson and Falt 1997). In our case a simple bimolecular binding model could be used:

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB \quad \text{and} \quad K_D = \frac{k_{off}}{k_{on}}$$

(with k_{on} and k_{off} describing association and dissociation constants)

3.2.7.2 Experimental procedure and data analysis

SPR spectroscopy was performed at 25°C using a BIAcore 3000 (GE Healthcare). Anti-GST antibody (GE Healthcare) was immobilized on a CM5 chip using an amine coupling procedure (according to GE healthcare instructions) with a 30 µg/ml antibody at 5 µl/min in 10 mM Na-acetate at pH 5.0. Approximately 10000 response units (RU) were immobilized on the chip. The anti-GST chips were then derivatized with 100 to 150 RUs with GST-tagged Dbp5 constructs in running buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM DTT, 10 mM MgCl$_2$ supplemented or not with 1 mM AMPPNP or ADP). An anti-GST flow cell was used as the control. Untagged Nup214 constructs were injected across the chip at a flow rate of 30 µl/min. The BIAcore 3000 evaluation software was used for analysis of the experimental data. The saturation binding values were fitted according to a one-site Langmuir binding model using Origin software (http://www.originlab.com/).

3.2.8 Static light scattering

To determine the molecular weight (and the oligomerization state) of proteins and protein complexes and the monodispersity of samples static light scattering was applied.

3.2.8.1 Theory

A typical static light scattering experiment consists of a size exclusion chromatography column coupled to a static light scattering devise with a triple detector array. The detector array consists of a laser light scattering (LS), a UV absorbance (UV) and a refractive index (RI) detector. With this experimental setup the homogeneity, monodispersity, oligomerization state and the molecular weight of proteins can be accurately determined. The basic light scattering equation is

(eq. 3)

$$\frac{K^*c}{R(\theta)} = \frac{1}{M}\overbrace{\left[1 + \frac{16\pi^2}{3\lambda^2} <r_g^2> sin^2\left(\frac{\theta}{2}\right)\right]}^{P(\theta)} + 2A_2 c$$

Where $R(\theta)$ is the excess intensity of light scattered at angle θ (Rayleigh ratio) and $K^* = \frac{\left[4\pi^2 n^2 \left(\frac{dn}{dc}\right)^2\right]}{(\lambda_0^4 N_A)}$ the optical parameter (Debye constant) , c the concentration in mg/ml, λ_0 is the wavelength of the light in vacuum, M is the weight-average MW, λ the wavelength of the incident light, $<r_g^2>$ is the mean square of the radius of gyration, A_2 is the second virial coefficient, n is the refractive index, $\frac{dn}{dc}$ is the refractive increment and N_A is Avogadro's number. For low concentrations A_2 is negligible. For scattering molecules smaller than the wavelength of the incident light term $P(\theta)$ is ~1. This is a good approximation for proteins with $r_g < 15\ nm$, which corresponds to globular proteins with MW < 5x10^7 Da. Therefore this equation can be simplified to

$$\frac{K^* c}{R(\theta)} = \frac{1}{M} + 2A_2 c \qquad \text{(eq. 4)}$$

With K^* the intensity of the light-scattering signal is

$$(LS) = K_{LS} c M \left(\frac{dn}{dc}\right)^2 \quad (K_{LS} \text{ is an instrument calibration constant}) \qquad \text{(eq. 5)}$$

refractive index signal is described by

$$(RI) = K_{RI} c \left(\frac{dn}{dc}\right) \qquad (K_{RI} \text{ is an instrument calibration constant}) \qquad \text{(eq. 6)}$$

The UV absorbance is

$$(UV) = K_{UV} c \varepsilon \ (K_{UV} \text{ is an instrument calibration constant and } \varepsilon \text{ the extinction coefficient}) \qquad \text{(eq. 7)}$$

From this we can derive

$$M = \frac{K_{RI}^2}{K_{LS} K_{UV}} \frac{(LS)(UV)}{\varepsilon (RI)^2} \qquad \text{(eq. 8)}$$

Where M and ε are the molecular weight and the extinction coefficient of the protein. For most complexes ε is not known, but can be calculated from the amino acid composition. From this combination of three detectors accurate MWs can be measured and therefore stochiometries of protein-protein or protein-nucleic acid complexes can be determined. In our experimental setup the UV detection has only a monitoring function. The accurate determination of the concentration (RI detector) and the light scattering signal can be used to calculate the molecular weight.

3.2.8.2 Experimental procedures

All static light scattering experiments were performed at 25°C and in gel filtration buffer. A Superdex 200 column (5/150) coupled to a static light scattering devise (TDA302, Viscotech) with a triple detector array was used. Measurements were performed at a single angle (90°). 10 µl of protein sample at 2 mg/ml were injected.

3.2.9 Storage of purified proteins

For short term storage protein samples were stored on ice or at 4°C in gel filtration buffer. For long term storage purified protein samples (10 - 50 µl aliquots in thin wall PCR tubes at concentrations between 1 - 50 mg/ml) were flash frozen in liquid nitrogen and kept at - 80°C. Large protein quantities were directly (slowly) pipetted into liquid nitrogen (as described in Bergfors 1999), stored at - 80°C and for further use slowly thawed on ice, concentrated and subjected to size exclusion chromatography (S200).

3.2.10 Measurement of protein concentration

Protein concentration was determined using the Bradford assay (Bradford, 1976) or a spectrophotometer (Nanodrop). Dye reagent for Bradford assay was purchased from Bio-Rad. Bovine serum albumin (BSA) was used as a standard.

3.2.11 X-ray crystallography

Many textbooks (Blundell and Johnson 1976; Drenth 1999; Rhodes 2000; McPherson 2001) describe the theory and methods in detail and therefore I only describe the experimental procedures.

3.2.11.1   Crystallization

Initial screening was performed in 96well sitting drop plates using commercial and homemade screens. 100 nl protein was mixed with 100 nl of the mother liquor using a Phoenix pipetting robot. Two different temperatures (4°C and 18°C) were screened. Initial hits were then optimized and refined in 96well sitting drop or 24well hanging drop plates.

3.2.11.1.1   Human Dbp5-AMPPNP-RNA complex

Dbp5 ΔN was concentrated to 18 mg/ml and incubated with 1 mM AMPPNP and a 1.2 molar excess of U_{10} ssRNA for one hour at 4°C. Initial crystals grew in several conditions containing different PEGs. Optimized crystals were obtained using a reservoir solution consisting of 100 mM Tris pH 7.0, 20 % PEG 2000 mono-methyl-ether, 3 % PEG 3350 and 30 mM NaF. The crystals grew as needles and were cryoprotected by the addition of 15 % glycerol. Crystals were flash frozen in liquid nitrogen. The best data set extends to 2.2 Å resolution (statistics in Table1).

3.2.11.1.2   Human Dbp5-Nup214 Complex

Initial crystals of the Dbp5 ΔNΔC-Nup214 βΔC complex grew within 1 - 2 days in conditions with high concentrations of Na-Citrate, Na-Malonate, Na-Acetate or Na-Oxalate. Optimized crystals grew as plates by mixing equal volumes of protein complex at 50 mg/ml and a reservoir solution containing 50 mM MES pH 6.0 and 900 mM Na-Citrate. Crystals were cryoprotected in paraffin oil prior to data collection.

3.2.11.2   Data collection and processing

Crystals were mounted in cryo-loops (Hampton research, USA). All data were collected at 100°K at the Swiss Light Source (SLS) at beamlines X06SA and X10SA with an oscillation range of 1° and exposure times of 1 sec/frame. Diffraction images were recorded on mar225 CCD or mar225 mosaic CCD detector (mar research). Other experimental settings are listed in Table1. In order to collect complete datasets, collecting strategies using MOSFLM (Leslie 1992) and LABELIT (Sauter 2004) were applied. Raw data were indexed, integrated and processed with XDS (Kabsch 1993) and scaled/merged with XSCALE (Kabsch 1993).

3.2.11.3   Structure solution and model building

All structures were solved by molecular replacement using Phaser (McCoy 2004). Phases were improved using density modification with Pirate (CCP4 1994). Molecular replacement solutions were subjected to initial rigid body refinement with Refmac (CCP4 1994). Models were manually built using Coot (Emsley and Cowtan 2004) and refined with Refmac (CCP4 1994). Input files for TLS (Translation/Libration/Screw) refinement were generated using the TLSMD (TLS motion determination) server (Painter and Merritt 2006). TLS is a mathematical

model describing the local positional displacement from a mean position for atoms that are part of a rigid body. This deviation from a spherical electron density at the atomic position is similar to the individual atomic anisotropic thermal parameters and is caused by thermal vibration of the molecule/atoms and static disorder in the crystal lattice. In contrast to full anisotropic refinement, where the displacement for each atom is described by 6 parameters, TLS groups are described by more sophisticated models of displacement that add only 20 parameters of refinement per TLS group. Therefore TLS refinement, where typically only a few TLS groups are refined individually, does not require such high resolution (< 1.2 Å) data in order to satisfy the parameter/observation ratio. TLS refinement describes the model more accurately and typically results in improvement of R-factors by a few percent.

Figures were prepared with Pymol (DeLano). Structural superpositions were performed with the secondary structure matching function of Coot (Emsley and Cowtan 2004) or the 3D-SS server (Sumathi 2006). Secondary structures were assigned using DSSP (Kabsch and Sander 1983) and were implemented in the pdbs for correct representation in Pymol (DeLano) using the DSSP2PDB script (James Stroud).

3.2.11.3.1   Human Dbp5-AMPPNP-RNA complex

The Dbp5 ΔN-AMPPNP-U_{10}RNA structure was solved using human eIF4A3 from the EJC structure (PDB entry 2J0S; sequence identity 39 %) as a search model. The crystals contain two molecules per asymmetric unit. The model was refined using tight non crystallographic symmetry (NCS) restraints, which were released in the final stages of refinement to account for variation between the two molecules.

3.2.11.3.2   Human Dbp5-Nup214 complex

The Dbp5 ΔNΔC-Nup214 βΔC structure was solved by molecular replacement using the N-terminal domain of Dbp5 from the refined Dbp5 ΔN-AMPPNP-U_{10}RNA structure and the 1.65 Å human Nup214 structure (PDB entry 2OIT) as consecutive search models in Phaser (Emsley and Cowtan 2004). The statistics of the refined structures are in Table1.

3.2.12   Small angle X-ray scattering

To determine conformations of Dbp5 in presence of nucleotides and to investigate flexibility of some regions of Dbp5 and Nup214 (as described in paragraph 4.4) I collected solution small angle X-ray scattering (SAXS) data.

3.2.12.1   Theory

SAXS is a technique based on the elastic scattering of X-rays of macromolecules in solution. Scattering is recorded (scattering intensity as a function of the scattering angle) at low angles (0.1 - 10°) and can be used to derive structural information like global shape and size of macromolecules. In addition parameters such as molecular weight (MW), radius of gyration (R_g) and the maximum intramolecular distance (Dmax) can be determined. Rg and forward scattering intensity I(0) can be obtained from the Guinier formula ($I\,(q) \approx I(0)\,exp\left(-\frac{q^2 R_g^2}{3}\right)$) for small momentum transfer q ($q = \ 4\pi\frac{sin\,(\theta)}{\lambda}$), where 2θ is the total scattering angle and λ the X-ray wavelength) and in the Guinier plot (plotting ln I(q) versus q^2). Rg is a model free characterization of the molecular dimensions and I(0) can be related to the molecular weight with the relation $I(0) =$κc$(\varDelta\rho)^2(MW)^2$, where κ is a proportionality constant that can be determined from a measurement of a standard (protein of known molecular weight and concentration), c is the concentration of the macromolecule, $\varDelta\rho$ is the average electron density contrast of the molecule, and MW is the molecular weight. The scattering profile can be expressed as a distribution function $p(r)$ of intramolecular atomic distances ($Dmax$ being the maximum intramolecular distance). $p(r)$ can be obtained from an indirect Fourier transform of the scattering profile. In the ideal case (no aggregation, no heterogeneity of conformations etc. reviewed in Putnam 2007) it is possible to use the 1D-scattering profile for ab initio reconstructions of low-resolution 3D electron density. Typically a bead based approach is used for the reconstruction using program such as Gasbor (Svergun 2001).

3.2.12.2   Data collection, processing and analysis

All samples were subjected to an additional size exclusion chromatography step to remove aggregates. Samples were taken from the individual fractions of each gel filtration peak in order to get different concentrations of the sample. For higher concentrations the remaining fractions were pooled and concentrated to their final

concentration using spin concentrators. Samples were flash frozen in thin walled PCR tubes at concentrations from 1 - 10 mg/ml. Protein concentrations were verified by Bradford assay and a spectrophotometer (Nanodrop) before freezing. Samples were gently thawed on ice and spun in a table top centrifuge at maximum speed for 10 min prior to data collection.

SAXS data were collected at Advanced Light Source (ALS), Ernest Orlando Lawrence Berkeley National Laboratory at SIBYLS beamline BL12.3.1. Data were measured on a mar165 detector. 15 µl samples in a sample cell were illuminated at 10 keV (λ = 1.1159 Å) for data collection (q range = 0.010 A^{-1} - 0.250 A^{-1}) with a detector distance of 1.5 m. Scattering profiles in SAXS buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 5 % glycerol, and 1 mM DTT supplemented or not with 1 mM ADP) were collected from samples with concentrations between 1 and 10 mg/ml.

Guinier plot analysis was used to determine the particle radius of gyration (R_g) in Primus (Konarev 2003). Gnom (Svergun 1992) was used to evaluate the electron pair distribution function, $p(r)$ function and to calculate the maximum dimension of particles (Dmax). Theoretical scattering was computed with Crysol (Svergun 1995). *Ab initio* dummy residue models were calculated with Gasbor (Svergun 2001) and averaged with Damaver (Volkov and Svergun 2003). Missing loops and domains were added to the known crystal structures using Credo (Petoukhov 2002). Modeling of conformations of Dbp5 was performed by Friedrich Förster (MPIB). N- and C-terminal RecA-like domains of the Dbp5 ΔN-AMPPNP-RNA structure were used as a starting model. The two domains were treated as two rigid bodies connected with a flexible linker and were optimized against a scoring function that includes the experimental scattering profiles (as described in Forster 2008).

4 Results

During the first year of my PhD work I tried to crystallize the *S. cerevisiae* Dbp5-Nup159 complex. Despite good expression, solubility and homogeneity of the purified complex, crystallization experiments were not successful. I cloned, expressed and purified several constructs of both proteins and tested multiple combinations of those. In all cases a soluble and stable 1:1 complex was formed but never crystallized. Crystallization of the helicase alone and in presence of different ligands (ATP, AMPPNP, ADP, AMP and the combination with RNA) was also not successful. After testing several crystallization methods (sitting and hanging drop vapor diffusion and microbatch) and other methods that can promote crystallization, like reductive methylation of surface residues (Walter 2006; Kim 2008) and mutation of surface residues (Derewenda 2004), I started working on the human proteins. An overview of some of the attempts and results of the yeast proteins is shown in appendix 6.4.

4.1 Human Dbp5 in complex with single stranded RNA

4.1.1 Purification, crystallization and data collection

A vector encoding the full length (fl) human Dbp5 gene was available in our lab. Dbp5 had been cloned as an N-terminal glutathione S transferase (GST) and as an N-terminal 6xHis-fusion in pETMC vectors. Test-expressions in several *E. coli* expression strains revealed the highest protein yield in *E. coli* BL21 Gold cells. Even without a 6xHis-tag the protein had the tendency to stick to Ni-NTA (nickel-nitrilotriacetic acid) Sepharose and therefore I only used GST as a purification tag. The GST-fused protein was purified to homogeneity with an initial batch affinity purification step using glutathione Sepharose, followed by overnight cleavage of the tag by TEV protease as described in paragraph 3.2.2.4. For binding and pull-down experiments the tagged protein was eluted in buffer supplemented with 20 mM reduced glutathione without TEV cleavage. GST-tagged and untagged proteins were further purified by affinity chromatography using a heparin Sepharose matrix. Heparin is a highly sulphated glucosaminoglycan that mimics the polyanionic structure of nucleic acids and is therefore commonly used to purify DNA- and RNA-binding proteins. The fraction of the protein that bound to the heparin matrix was eluted with

a linear NaCl gradient (Figure 5a) and was then further purified to homogeneity by size exclusion chromatography (Figure 5b).

Figure 5: 15 % SDS gels of the 53 kDa human fl Dbp5.
(a) Elution from heparin sepharose and (b) gel filtration peak (S200)

In the first purifications I observed that a considerable amount of Dbp5 did not bind to the heparin sepharose. I collected this fraction and also purified it by size exclusion chromatography (S200), but despite apparently equal homogeneity on SDS gels this protein fraction never crystallized. With hindsight the heparin step is most likely selecting for properly folded and active proteins that are not bound to nucleic acids. To remove possible *E. coli* RNA contaminations, I introduced a high salt wash (1 M NaCl) in the first affinity step. This however reduced the amount of unbound Dbp5 in the heparin purification step.

The purified protein was first subjected to crystallization without any ligand. Several sparse matrix screens were tested but no crystals were obtained. Structural data for two other DEAD-box helicases in complex with RNA (Andersen 2006; Bono 2006; Sengoku 2006) suggested that the two RecA-like domains of Dbp5 might have a certain degree of flexibility without RNA bound and might be difficult to crystallize. Upon RNA binding the two RecA-like domains form a specific closed, more rigid conformation. I performed initial electro-mobility shift assays (EMSA, data not shown) and later pull-down experiments using biotinylated RNA (Figure 6b, lanes 2-5) to show that Dbp5 binds to single stranded (ss) RNA. In order to form the closed conformation and to promote crystallization a 20mer poly-A single stranded RNA (20A ssRNA) and the non-hydrolysable ATP analogue AMPPNP were added to the protein in a 1.2 molar excess. After incubation for 1h on ice the sample was used for crystallization. Initial crystals all shared a similar needle shaped habitus and grew within 5 days in a few conditions containing different sized poly-ethylenglycol (PEG) polymers (Figure 7a). Optimization of the PEG concentrations yielded marginally bigger crystals that were however still too small for data collection. Comparison of successful crystallization conditions of RNA helicases led to a more rational

optimization: several helicases (e.g. eIF4AIII) had been crystallized in conditions containing sodium cacodylate as buffer. Exchange of buffers further improved the crystal size (Figure 7b). Despite the small size of the needles (~ 150 x 10 x 10 µM), datasets up to 3.5 Å resolution were collected at the Swiss Light Source (SLS) synchrotron. Data were processed using XDS (Kabsch 1993) and the structure could be solved in spacegroup $P2_12_12$ using eIF4AIII from the exon junction complex (EJC) as a search model in molecular replacement (MR) with the program Phaser (McCoy 2004) as described in methods 3.2.11.3.

In order to get higher resolution data I continued optimization of the crystallization conditions. As the length of the RNA substrate is likely to influence crystallization (a too long RNA might interfere with crystal packing), I tested several oligoribonucleotides in order to find a minimal RNA. Preliminary electromobility shift assays (as described in 3.2.6.3) were performed (data not shown) with different ssRNA (20, 15, 10, 8, 7, 6 and 5mers of poly U or poly A). 5 - 8mer ssRNAs were not visible on the native gel and therefore not used for crystallization. The structures of DEAD-box helicases Vasa and eIF4AIII have visible electron density for six nucleotides, although longer ssRNAs were used for crystallization in both cases (Andersen 2006; Bono 2006; Sengoku 2006). This is consistent with RNAse protection assays of eIF4AIII that resulted in 6 - 7 protected nucleotides (Ballut 2005). Fl Dbp5 crystallized in the presence of 10, 15 or 20mer RNAs (+ AMPPNP), but surprisingly the 10mer poly-U RNA changed the crystal habitus from needles to slightly larger plate-like crystals (Figure 7c). Therefore a 10mer ssRNA was used in all crystallization experiments.

These crystals diffracted to 3.0 Å resolution at SLS and the structure was solved by molecular replacement using Phaser (CCP4 1994) in the same spacegroup ($P2_12_12$) using eIF4AIII as a search model. Manual model building was performed with the program Coot (Emsley and Cowtan 2004). Even without complete refinement (R_{free} = 31.5 %) it became apparent that a major part of the N-terminus (~ first 75 residues) had no visible electron density. Based on the partially refined model I cloned an N-terminally truncated construct (Dbp5 ΔN, residues 68 - 479, Figure 6a) since removing flexible parts of proteins is known to often improve the (diffraction) qualities of crystals. Dbp5 ΔN behaved like the fl protein during expression and purification. In RNA binding experiments (Figure 6b) I tested the minimal RNA-binding constructs and showed that the N-terminal truncated construct (Dbp5 ΔN) in

presence of AMPPNP binds to ssRNA (lane 4) to a similar extent as the GST-tagged (lane 3) and untagged fl Dbp5 (lane 2). The N-terminal RecA-like domain alone is not sufficient to bind RNA (lane 5). These pull-down experiments also confirmed that AMPPNP (compare lanes 2 and 11) and not ADP (lane 13) is needed for RNA binding. In addition I tested the effect of Nup214 on RNA binding that is discussed in paragraph 4.3.

Figure 6: The human Dbp5-AMPPNP-RNA complex. (a) Schematic representation of the domain organization of Dbp5 with the domain boundaries of the protein constructs used in this thesis. The RecA-like domains are colored blue. The N-terminal region is shown in orange. (b) Protein precipitations with biotinylated ssRNA identify the RNA-binding region of human Dbp5 for structural studies. Purified proteins were mixed with 5' end biotinylated 20-mer ssRNA and incubated with or without AMPPNP or ADP, as indicated. Protein mixtures before (input) and after co-precipitation (precipitate) were separated on a 15 % (w/v) acrylamide SDS-PAGE and visualized using Coomassie stain. The molecular weight standards are shown in lane 1. Dbp5 constructs are described in panel a. Nup214 βΔC corresponds to residues 1 - 405 (human sequence).

The N-terminal truncated protein (Dbp5 ΔN) crystallized in presence of AMPPNP and 10mer poly-U (U_{10}) ssRNA. The needle shaped crystals (Figure 7d) diffracted to 2.2 Å (see Figure 8 and Table 1). Data could be indexed and processed in the orthorhombic spacegroup $P2_12_12$, and a single MR solution (llg of 90, z-score of 7.5) was found in Phaser (McCoy 2004) using the partially refined fl model. One molecule per asymmetric unit was found corresponding to a solvent content of 37 %. Electron density maps were noisy and refinement showed no improvement of R-factors of ~ 40 %. I reprocessed the data that could also be indexed in the lower symmetry monoclinic spacegroup P2 and a beta angle of 90.022° (see Table 1). A clear MR solution (llg of 330, z-score of 17) was found in spacegroup $P2_1$, with 2 molecules per asymmetric unit corresponding to a solvent content of 41 %.

Superposition of the two molecules reveals that they relate to each other by rotation along a and c axes, and a small translational shift of 4 Å along the b axes. The model was manually built in Coot (Emsley and Cowtan 2004) and refined in Refmac (CCP4 1994) using individual isotropic B-factor refinement and tight NCS (non-crystallographic symmetry) restrains, which were released in the final rounds of refinement. The final model was refined to 2.2 Å with an R_{free} = 22.7 % and R_{work} = 17.1 % resolution using TLS refinement (Table1).

Figure 7: Crystals of human Dbp5 in complex with AMPPNP and RNA.
(a) initial crystals grew in a few PEG containing conditions (e.g. 100 mM HEPES pH 7.5, 10 % PEG 8000) using a 20mer poly-U RNA, (b) optimized crystals grew in 100 mM Na-Cacodylate pH 7.5, 26 % PEG 400 using a 20mer poly-U RNA, c) crystals grown in the same buffer as in panel b but in presence of a 10mer poly-U RNA d) final needle shaped crystals of Dbp5 ΔN-AMPPNP-RNA complex grew in 100 mM Tris pH 7.0, 20 % PEG 2000 mono-methyl-ether, 3 % PEG 3350 and 30 mM NaF in presence of a 10mer poly-U ssRNA.

Figure 8:
Diffraction image of a Dbp5 ΔN-AMPPNP-RNA complex crystal.
Data was recorded at beamline PXI (SLS, Villigen, Switzerland). Crystals diffracted to a maximum resolution of 2.2 Å.

4.1.2 Structure of human Dbp5 in complex with AMPPNP and RNA

There are two molecules in the asymmetric unit (molecule A and B) that are essentially identical (root mean square deviation (r.m.s.d.) of 0.2 Å over all amino-acid residues). Molecule A includes Dbp5 residues 75 - 466 and molecule B residues 75 - 467. Both molecules show no or very poor density for a disordered loop (residues 201 - 208). This loop is also disordered in other helicase structures (e.g. eIF4A, Schutz 2008). No density was observed for the C-terminal 12 - 13 residues. Each molecule includes six bases of poly-U RNA, one molecule AMPPNP and one magnesium ion (Figure 6c). The model of the Dbp5 ΔN-AMPPNP-RNA complex has good stereochemistry (Table 1 and Figure 9).

Figure 9: Ramachandran plot of Dbp5 ΔN in complex with AMPPNP and RNA. Plotted are the Φ - Ψ torsion angles of the protein backbone. The four basic Ramachandran plots are shown (a) general case (18 non-glycine non-proline amino acids) (b) glycine (c) proline and (d) pre-proline (referring to residues preceding a proline). 99.6 % (776/779) of all residues are in favored (98 %, light blue) regions, 100 % are in allowed (>99.8 %, blue) regions. There are no outliers.

Table 1: Data collection and refinement statistics (molecular replacement).

	Dbp5-AMPPNP-RNA		Dbp5-Nup214
Data collection			
Space group	*P2$_1$*	*P2$_1$2$_1$2*	*I222*
Cell dimensions			
a, b, c (Å)	41.60, 79.66, 124.54	91.3, 148.5, 42.1	117.41, 134.71, 152.85
α, β, γ(°)	90.0, 90.022, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	41.5 - 2.2 (2.3 - 2.2)	43.6 - 3.0 (3.2 - 3.0)	47.7 - 2.8 (3.0 - 2.8)
R$_{meas}$	10.6 (27.3)	12.7 (72.4)	13.8 (36.4)
I / σI	10.89 (4.54)	11.22 (2.27)	11.59 (3.31)
Completeness (%)	99.5 (99.3)	98.5 (99.3)	100 (100)
Redundancy	4.11 (4.11)	3.5 (3.5)	14.13 (13.6)
Refinement			
Resolution (Å)	41.5 - 2.2	43.6 - 3.0	47.7 - 2.8
No. reflections	39149	41938	30169
R$_{work}$ / R$_{free}$	16.8 / 22.5	26.0*/31.4*	21.8 / 26.0
Molecules/AU	2	1	1
Solvent content	41 %	54 %	70 %
No. Atoms			
Protein	6190		4824
Ligand/ion	326		
Water	299		85
B-factors			
Protein	18.7		40.5
Ligand/ion	14.3 (ATP), 33.6 (RNA)		-
Water	35.1		28.9
R.m.s. deviations			
Bond lengths (Å)	0.009		0.012
Bond angles (°)	1.4		1.4

Values in parentheses are for highest-resolution shell. R factors are as indicated:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}, \quad R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h - 1}} \sum_i^{n_h} |\hat{I}_h - I_h|}{\sum_h \sum_i^{n_b} I_{h,i}} \text{ with } |\hat{I}_h = \frac{1}{n_h} \sum_{n_h}^{1} I_{h,i}$$

4.1.2.1 Structural organization of Dbp5

Dbp5 can be divided into four parts (see Figure 6a and Figure 10): The N-terminal RecA-like domain (residues 75 - 296), the C-terminal RecA-like domain (residues 306 - 466), the connecting linker (residues 297 - 305) and a N-terminal flanking region (residues 75 - 92). Both RecA-like domains have a similar α-β topology with a central parallel β-sheet with loops on the N- and C-terminal ends of the strands. These loops contain the characteristic DEAD-box helicase motifs and converge to bind AMPPNP or RNA (see alignment Figure 17). The two RecA-like domains are in a closed conformation, binding AMPPNP and the magnesium ion in their cleft (Figure 10). The two β-sheets (7 strands in the N-terminal and 6 strands in the C-terminal RecA-like domain) are surrounded by 9 and 7 α-helices, respectively (see secondary structures in alignment, Figure 17). The short N-terminal flanking region preceding the N-terminal domain forms a small helix that is involved in AMPPNP binding. A more detailed description of this N-terminal region is given in paragraph 4.1.2.3.

Figure 10: Structure of Dbp5 in complex with RNA and AMPPNP.
Dbp5 ΔN (light blue) bound to AMPPNP (yellow) and single-stranded RNA (black). The magnesium ion at the ATP-binding site is shown in magenta. Six ordered nucleotides of a single-stranded poly-U RNA are present in the 2.2 Å resolution structure. The N- and C-terminal residues of Dbp5 visible in the electron density are indicated. An N-terminal helix (orange) preceding the N-terminal domain is involved in binding of AMPPNP.

Dbp5 has a similar structure (Figure 11) as that observed in the RNA-AMPPNP-bound complexes of Vasa and of eIF4AIII (Andersen 2006; Bono 2006; Sengoku 2006). According to the structural superposition of Dbp5 with eIF4AIII using 3d-SS (Sumathi 2006), about 90 % of the amino-acid residues superpose in their α-carbon positions with an r.m.s.d. of less than 1.2 Å. A major conformational variation is observed in the N-terminal region and is further described in Figure 34. Minor differences are in loops located on the surface of the molecules. Overall in the RNA-bound form the DEAD-box helicases Vasa, eIF4AIII and Dbp5 (and most likely other members of this protein family) assume a closed conformation.

Figure 11: Superposition of three DEAD-box helicases.
Stereo-view of Dbp5 in light blue superposed with the two DEAD-box helicases eIF4AIII in yellow (Bono 2006, pdb code 2J0S) and Vasa in magenta (Sengoku 2006, pdb code 2DB3) shown as ribbons. The three proteins superpose well except for conformational differences in surface loops and the N-terminal region preceding the N-terminal RecA-like domain.

4.1.2.2 Dbp5 binds RNA in a sequence independent manner

Dbp5 binds RNA by both RecA-like domains opposite the linker (Figure 6c, Figure 12a) region. The relatively high resolution of the structure allows to unravel the molecular details of the Dbp5-RNA interaction. Six nucleotides of single-stranded RNA bind with the 3' end at the N-terminal domain of Dbp5 and with the 5' end at the C-terminal domain (Figure 6c, Figure 12b). Both domains contribute to the binding site and although more contacts are from by the N-terminal domain (Figure 6c and Figure 12); this domain is not enough to bind RNA alone *in vitro* (Figure 6b, lane 5).

The interaction is mainly mediated via the ribose-phosphate backbone of the RNA and hence binding occurs in a sequence-independent manner. Only two contacts (involving Leu[226] and Phe[230]) are observed that contact the base of U6. A mesh of interactions involving four 2`-hydroxyl groups of the ribose explains the exclusive binding of Dbp5 to RNA and not DNA (Figure 12b). Normal base stacking of nucleotides 1 - 4 is disrupted by a sharp bend of the RNA between nucleotides 4 and 5 as also observed in the eIF4AIII and Vasa structure (Andersen 2006; Bono 2006; Sengoku 2006).

Compared to eIF4AIII, the overall properties of the RNA-binding pocket are conserved. More than 50 % of the residues are identical (Figure 12b). Binding of RNA is mediated by the conserved DEAD-box motifs: motifs Ia and Ib in the N-terminal domain and motifs IV and V in the C-terminal domain bind along the oligoribonucleotide chain (Figure 12b and Figure 17). Two other loops containing the so-called GG and QxxR motifs also contact RNA. This interaction is also observed in Vasa and eIF4AIII, although these motifs are only partially conserved in the Dbp5 sequence (Figure 17 and alignment in appendix). Interestingly, a different binding mode is observed in the 3.0 Å fl Dbp5-AMPPNP-RNA structure (Figure 13a). Here, nucleotides 1 - 4 show the same base stacking as in the Dbp5ΔN-AMPPNP-RNA structure. A sharp bend is also induced between nucleotides 4 and 5 but the gap is filled by the base of nucleotide 6, which turns ~ 180° and stacks in between the two bases (Figure 13a). Since there is limited space due to crystal contacts (Figure 13b), this unusual RNA conformation might be induced by crystal packing and might have no biological relevance *in vivo*.

Figure 12: Binding of RNA.
(a) Stereo-view of the RNA binding pocket. Electron density ($2F_O-F_C$) is contoured at 1 σ around the RNA. The N-terminal RecA-like domain is shown in blue and the C-terminal RecA-like domain in grey (b) Schematic representation of RNA binding. RNA is bound in a sequence independent manner with the 5'-end to the N-terminal domain (residues boxed blue) and the 3'-end to the C-terminal domain (residues boxed grey) of Dbp5. Interactions are mainly based on backbone contacts. The hydroxyl group of the ribose is involved in several contacts (in purple), explaining why Dbp5 binds RNA, but not DNA. Marked are the conserved motifs and for comparison the corresponding residues of eIF4AIII of the EJC in yellow.

Figure 13: RNA binding of fl Dbp5 in the crystal structure of space group P2$_1$2$_1$2 at 3.0 Å resolution. (a) Stereo-view of an unusual conformation of RNA bound to Dbp5. Electron density (2F$_O$-F$_C$) is contoured at 1.0 σ around the RNA. Base 6 turns 180° and stacks in the gap between bases 4 and 5. (b) Dbp5 (blue, labeled 1) is shown in the same orientation as in panel a, with bound RNA (black) and its symmetry related molecules (grey, numbered 2 - 4). In the crystal molecule 4 contacts molecule 1 close to its RNA binding side and thereby limits the space for the 3`end of the emerging RNA molecule.

4.1.2.3 AMPPNP is coordinated in a highly conserved binding pocket

AMPPNP is bound in between the two RecA-like domains. Conserved residues of the loops formed by motifs Q, I, II, and III in the N-terminal domain and motifs V and VI in the C-terminal domain converge to bind the nucleotide (AMPPNP, Figure 14, Figure 17). In addition to the two folded domains, a short N-terminal region preceding the first RecA-like domain (residues 75 - 92) stretches back towards the C-terminal domain and contributes to the ATP-binding site (Figure 6c). The adenine ring of the nucleotide is sandwiched between Phe^{112} of the N-terminal RecA-like domain and Phe^{433} of the C-terminal domain, which is surrounded by Val^{77} and Leu^{88} of this N-terminal region (Figure 15). A similar arrangement is observed in the structures of Vasa and eIF4AIII of the EJC (Andersen 2006; Bono 2006; Sengoku 2006), where the hydrophobic environment of the adenine ring is formed by apolar residues of the C-terminal domain and N-terminal region (Vasa Val^{583} and Phe^{225}), or by apolar residues of the C-terminal domain and an additional protein (eIF4AIII Tyr^{371} and Mago Ile^{146} in the EJC). Val^{77} and Arg^{114} are most likely the reason why Dbp5 exclusively binds ATP and not GTP. The apolar residue would interfere with the amino group of the purine ring and the ketogroup and its enol tautomere of GTP would be repulsed by the backbone carbonyl-group of Arg^{114}.

The majority of the residues binding AMPPNP are contributed by the N-terminal domain (Figure 14). The residues necessary for binding of the magnesium ion and hydrolysis of ATP are contributed by both RecA-like domains. Several DEAD-box helicases have been crystallized in complex with ADP. Several structures of only the N-terminal domain bound to AMPPNP or ADP have been deposited in the protein data bank (e.g. pdb codes: 2PIX, 3DKP, 3FE2, 3B7G). Comparison of the binding pocket reveals that in absence of the C-terminal domain ADP is coordinated by the same residues binding AMPPNP in the structure reported here.

Altogether an almost identical AMPPNP binding mode of Dbp5, Vasa and eIF4AIII is observed, although these proteins have very distinct functions in mRNA metabolism (Figure 14, Figure 15). This is further discussed in paragraph 5.3

Figure 14: Schematics of AMPPNP binding.
Interacting residues are boxed in blue (N-terminal RecA-like domain) and in grey (C-terminal RecA-like domain). Conserved DEAD-box helicase motifs are marked on top of the residues (see alignment Figure 17). Backbone contacts are represented by boxed residues without side-chains. Water molecules that contribute to the coordination and/or hydrolysis of ATP are shown in black and the magnesium ion in magenta.

So far the binding constants of DEAD-box helicases for ATP or ADP have not been reported in the literature. I measured binding data based on radiolabel ATP and spin column experiments (see methods 3.2.6.4) but due to experimental limitations (discussed in 3.2.6.4) I could not determine absolute binding constants. Nevertheless the results from these experiments suggest that the K_D for ATP is much higher than 5 µM and though extrapolations have a large error, K_Ds are likely in the high micromolar to millimolar range. Comparison with other helicases suggest that the N-terminal domain of Dbp5 (and other DEAD-box helicases) alone might bind ATP/ADP with low affinity. This would explain results from binding experiments discussed in 4.2.2.

Figure 15: Stereo view of AMPPNP binding.
AMPPNP is shown in yellow, Dbp5 residues in blue and the magnesium ion in magenta. Coordinated waters are not shown for clarity.

4.1.2.4 Hydrolysis of ATP

The non-hydrolysable ATP analogue AMPPNP is sandwiched between the two RecA-like domains as described in the preceding paragraphs. The helicase is trapped in an active conformation where the magnesium ion, several waters and the residues important for hydrolysis are in a productive state. The overall coordination of the nucleotide and the position of catalytic waters and the magnesium ion is similar to that observed for GTP in small GTPases such as H-Ras p21 (Pai 1990). Dbp5 (and other DEAD-box helicases) might use a similar mechanism for ATP hydrolysis (Figure 16). The reaction is most likely initiated by the abstraction of a proton from a catalytic water (probably the one coordinated by motif III, see Figure 14) by the γ-phosphate of ATP. The resulting nucleophilic hydroxide (probably stabilized by the surrounding His[425], Lys[144] and motif III) would then attack the protonated γ-

phosphate. A pentacovalent intermediate would be formed that would result in cleavage of the γ-phosphate group and the final products ADP and inorganic phosphate.

A water molecule responsible for the nucleophilic attack has been described in the structure of Vasa (Sengoku 2006). I observe two well ordered waters (b-factors of 9 and 16) that are both contacting the γ-phosphate of AMPPNP and are both in position for a nucleophilic attack (Figure 14). Both waters are also present in the structures of Vasa (2DB3) and eIF4AIII (2J0S). It is not obvious, which one acts as the catalytic water. A second water, that has not been described in the Vasa structure (Sengoku 2006) could be the candidate: it is less than 2.9 Å (compared to 3.2 Å for the other water) away from the γ-phosphorous atom and it is coordinated by residues (Glu[243] of motif II (that likely polarizes the nucleophile), the conserved His[425] and Ala[275] of motif III) that have been shown to be important for the ATPase activity of DEAD-box proteins. Mutations of the equivalents residues (Glu[243] and His[425]) of eIF4A reduced the ATPase activity and impaired unwinding (Pause and Sonenberg 1992; Pause 1993). Mutations in motif III (SAT to AAA) had no effect on the ATPase activity or RNA binding but showed defects in RNA unwinding (Pause and Sonenberg 1992). The contact of motif III to the putative catalytic water is mediated by the main-chain amide group of Ala[275] and therefore not affected in this mutation. Overall the role of motif III is somewhat ambiguous since this motif seems to couples ATP hydrolysis with RNA unwinding.

Another factor necessary for ATP hydrolysis is the bound magnesium ion (Figure 10, Figure 14 and Figure 15). The magnesium is coordinated in an octahedral geometry by four water molecules and one oxygen each of the β- and γ-phosphate of AMPPNP (Figure 14). The role of the metal ion in catalysis has been discussed for GTPases and ATPases (e.g. H-ras p21, Pai 1990). Several functions of the magnesium ion in nucleotide hydrolysis are possible: shielding the negative charge of the attacked γ-phosphate, increasing the acidic strength of the leaving group (β-phosphate) or stabilization of the transition state.

Figure 16: Possible ATP hydrolysis mechanism. A probable mechanism involves 4 steps: I) Hydrolysis of ATP is likely initiated by the abstraction of a proton from a water by the γ-phosphate of ATP. II) the resulting hydroxide ion attacks the γ-phosphate III) a pentavalent intermediate is formed that results in IV) cleavage of the γ-phosphate group.

Figure 17: Conservation of structural and functional features of Dbp5.
The structure-based sequence alignment includes orthologues from *H. sapiens* (Hs), *S. cerevisiae* (Sc), *S. pombe* (Sp) and *D. melanogaster* (Dm). The secondary structure elements are shown below the sequences (h for α-helices, b for β-strands and dots for extended or loop regions). Boxed in red are residues I mutated for in vitro studies. The Dbp5 sequence alignment includes a functionally different DEAD-box helicase (human eIF4AIII). Conserved sequence motifs shared by all DEAD-box helicases are boxed and defined below the sequences (motifs Q, I, Ia, GG, Ib, II, III, IV, QxxR, V and VI). Residues that are conserved specifically in Dbp5 orthologues are shaded in light blue. Highlighted above the sequences are the residues of human Dbp5 contacting AMPPNP (A, yellow) and RNA (R, white) in the Dbp5 ΔN-AMPPNP-RNA complex, or the nucleoporin (N, green) in the Dbp5 ΔNΔC-Nup214 βΔC complex. Residues 77 - 92 of human Dbp5 adopt different conformations in the two Dbp5 structures, forming a helix in the Dbp5 ΔN-AMPPNP-RNA complex (here indicated) and a β-strand in the Dbp5 ΔNΔC-Nup214 βΔC complex.

4.2 Human Dbp5 in complex with Nup214

Dbp5 not only binds to RNA but also interacts with the N-terminal region of the human nucleoporin Nup214 (Schmitt 1999; Napetschnig 2007) and in yeast with its orthologue Nup159 (Schmitt 1999; Weirich 2004). The structures of the N-terminal domains of Nup214 and Nup159 have been recently solved (Weirich 2004; Napetschnig 2007) and contain a conserved β-propeller domain (Figure 3 and Figure 20). In order to investigate how Dbp5 specifically docks to Nup214 (at the NPC) and how the nucleoporin influences the helicase, I solved the structure of the human Dbp5-Nup214 complex and performed several biochemical assays as discussed below

4.2.1 Purification

The complex of the N-terminal RecA-like domain of human Dbp5 (Dbp5 ΔNΔC, see Figure 6a) and the β-propeller domain of Nup214 (Nup214 βΔC) was reconstituted after purifying the two proteins separately. Dbp5 ΔNΔC was expressed as a GST-fusion and purified as the fl protein as described in paragraph 4.1.1. Untagged (for crystallization) and GST-tagged Dbp5 ΔNΔC (for surface plasmon resonance (SPR) and pull-down experiments) were purified to homogeneity as discussed before (paragraph 4.1.1). Tagged and untagged Dbp5 ΔNΔC eluted from the final size exclusion chromatography column (S200) in distinct peaks as shown for the untagged protein in Figure 18. In general Dbp5 ΔNΔC behaved like the fl protein during expression and purification with the exception that it precipitated slowly but continuously at 4°C. In contrast, the constructs consisting of both RecA-like domains and all Dbp5 constructs in presence of Nup214 were stable.

Figure 18: Purification of Dbp5 ΔNΔC.
(a) Size exclusion chromatogram (S200). Dbp5 ΔNΔC elutes as a distinct peak (b) 15 % SDS PAGE of selected fractions of the peak of the size exclusion chromatography shown in a.

Two protein constructs of Nup214, both consisting of the N-terminal β-propeller domain, Nup214 βΔC (residues 1 - 405) and Nup214 β (residues 1 - 450) were expressed in BL21 Gold cells as N-terminal 6xHis- or GST-fusions and purified using the appropriate resin (Ni-NTA or GSH-Sepharose), followed by overnight tag cleavage using TEV protease. Untagged proteins were further purified by ion-exchange (Hitrap Q) and size exclusion chromatography (S200) as described in methods 3.2.2.5.

For complex formation the two proteins were mixed for one hour on ice and further purified to homogeneity by size exclusion chromatography (Figure 19). The complex elutes in a distinct peak (peak II), well separated from the excess of Dbp5 ΔNΔC (peak III). Since homogeneity on SDS gels does not necessary equal monodispersity, which is prerequisite for crystallization, I determined the monodispersity and the molecular weight using static light scattering (sls). All sls experiments were performed by Jörg Tittor and Claire Basquin at MPIB. In these experiments (Figure 20), a single peak was observed the size exclusion chromatogram and was used to determine the concentration and the molecular weight as described in methods 3.2.7. The experimental MW is 71.124 KDa and corresponds to a 1:1 complex of Dbp5 ΔNΔC and Nup214 βΔC and is in good agreement with the calculated MW of 71.35 KDa. The sample is highly monodisperse with a MW/Mn ratio of 1.0008. We investigated all proteins before they were subjected to crystallization. We tested the Dbp5 ΔNΔC-Nup214βΔC complex (shown in Figure 20) before and after concentration of the sample up to 100 mg/ml and after 6, 24, and 48 h after the final size exclusion chromatography and observed no aggregation. Interestingly after 6 days (stored at 4°C) we determined a MW of 131 kDa with high monodispersity (MW/Mn = 1.008). This approximately corresponds to a dimer of the complex but was not further investigated.

Figure 19: Purification of Dbp5 ΔN-Nup214 βΔC-complex.
(a) Size exclusion chromatogram (S200) (b) 15 % SDS PAGE of selected fractions of gel filtration peaks shown in a. Peak I contains impurities of GST-Dbp5 ΔNΔC, peak II the 1:1 complex of Dbp5 ΔN-Nup214 βΔC and peak III the excess of Dbp5 ΔNΔC.

Figure 20: Chromatogram from static light scattering experiments.
(a) Chromatogram of the Dbp5 ΔNΔC-Nup214 βΔC complex. The three detectors (UV, refractive index and light scattering, as described in methods 3.2.7) are shown as indicated. The first peak contains the purified complex, the second peak does not contain any proteins (no UV detection) and most likely results from an excess of glycerol from the sample. (b) Close-up of the first peak shown in panel a. This peak was used to determine the concentration and the experimental MW (71.124 KDa).

4.2.2 The N-terminal RecA-like domain of Dbp5 is sufficient to bind to Nup214

There are seemingly conflicting reports in the literature on the requirements of Dbp5 for nucleoporin interaction. It has been shown that fl human Dbp5 binds the N-terminal region of Nup214 (residues 1 - 586) and that this interaction is destabilized in the presence of ATP nucleotides (Schmitt 1999). In addition it has been shown that a portion of yeast Dbp5 lacking the C-terminal RecA-like domain (Dbp5 ΔC) is

sufficient for Nup159 binding, and that this interaction is not affected by the presence of ATP nucleotides (Weirich 2004). To further analyze the minimal binding regions to generate a core complex suitable for crystallization I performed pull-down experiments with different constructs. GST-fused Dbp5 constructs were immobilized on GSH-resin and Nup214 constructs were then added. In these pull-down experiments with the human proteins, I find that even without the N-terminal region (Dbp5 residues 1 - 67), the N-terminal RecA-like domain alone is sufficient for Nup214 binding (Dbp5 ΔNΔC (residues 68 - 302), Figure 22a, lane 5).

Figure 21: Conservation of structural and functional features of N-terminal domain of Nup214. The structure-based sequence alignment includes orthologues from *H. sapiens* (Hs), *S. cerevisiae* (Sc), *S. pombe* (Sp) and *D. melanogaster* (Dm). The secondary structure elements are shown below the sequences (h for α-helices, b for β-strands and dots for extended or loop regions). Boxed in red are residues I mutated for in vitro studies. The β-strands are numbered with 1 to 7 (to indicate the blades of the β-propeller they belong to) and with A to D (to indicate the consecutive β-strands within each blade). Blades 6 and 7 contain an additional strand (E). Residues conserved in Nup214 orthologues are shaded in light green. Residues of human Nup214 contacting Dbp5 (D, blue) are highlighted above the sequence.

The structure of human Nup214 features a C-terminal extension (residues 405 - 450) that binds in an extended unstructured conformation across the β-propeller face (Napetschnig 2007). In pull-downs with full-length GST-Dbp5, I observed binding to the human Nup214 both in the presence and in the absence of this C-terminal extension (Figure 22a, Nup214 βΔC in lane 3 and Nup214 β in lane 4, respectively). Thus, the β-propeller core of human Nup214 is sufficient for Dbp5 binding. Moreover crystallization trials of the complex featuring this extension were not successful. This supports the idea that this part of the molecule is rather flexible and interferes with crystallization. Together with the fact that Napetschnig et al. observed no binding of this C-terminal tail in gel filtration and in binding experiments (Napetschnig 2007), it is likely that binding of this tail across the β-propeller face is a crystallization artifact.

Having the qualitative binding information from pull-down experiments I tested, with the help of Claire Basquin, the relative binding affinities of the different constructs quantitatively by surface plasmon resonance (BIAcore, GE Healthcare). Initially I designed the experiment using GST-Nup214 constructs in order to directly compare binding of different Dbp5 constructs to the same immobilized Nup214 molecule. However, under these experimental conditions Nup214 did not bind to Dbp5. Most likely the binding site is not accessible when immobilized on the surface. I therefore used GST-tagged Dbp5 constructs. Under the given conditions (methods 3.2.7.2) Dbp5 fl binds with a similar affinity of about 20 nM to either Nup214 β or Nup214 βΔC (Figure 22b, red and orange sensorgrams). These data indicated that the C-terminal extension of Nup214 does not influence Dbp5 binding and is not required. The dissociation constant is also similar when only the N-terminal RecA-like domain of Dbp5 is immobilized (Figure 22b, green sensorgrams), indicating that this domain alone is sufficient to bind to Nup214. The N-terminal domain includes all the binding determinants that are present in the fl protein. The C-terminal domain of Dbp5 does not contribute to binding of the nucleoporin. Data were fitted according to a bimolecular binding model using the non-linear Langmuir binding isotherm as shown in Figure 23.

To investigate the effect of nucleotides on the Dbp5-Nup214 complex we performed the same SPR experiments in presence of 500 µM AMPPNP or ADP. In the presence of these nucleotides, binding of Nup214 to fl Dbp5 is reduced by two orders of magnitude, to a K_D larger than 1 µM (AMPPNP, Figure 22b, blue sensorgrams) and larger than 0.6 µM (ADP. Figure 22b, purple sensorgrams). These

data show that the binding between Nup214 and fl Dbp5 is reduced when in the presence of nucleotides, even though the interaction with Nup214 is mediated by the N-terminal RecA-like domain alone**.** Interestingly we observed similar though slightly lower K_Ds for the Dbp5 ΔNΔC-Nup214 βΔC interaction in presence of ADP and AMPPNP (data not shown). This suggests that the conformation of the N-terminal domain (Dbp5 ΔNΔC) alone and the interaction with Nup214 is also altered in presence of nucleotides. This indicates that N-terminal Dbp5 domain is sufficient to bind to nucleotides to some extend and affects the Dbp5-Nup214 interaction.

Figure 22: The human Dbp5-Nup214 complex.
(a) Protein precipitations by GST pull-down assays qualitatively identify the binding regions of Nup214 and Dbp5 for structural studies. Glutathione agarose beads were pre-coated with GST or GST-Dbp5 proteins as indicated, and incubated with the purified β-propeller domain of Nup214 either with the C-terminal extension (residues 1 - 450, Nup214 β) or without it (residues 1 - 405, Nup214 βΔC). One fifth of the input and one eighth of the bound fractions were analyzed by SDS-page. Nup214 βΔC precipitates with full-length Dbp5 at least as efficiently as Nup214 β (lanes 3 and 4). A Dbp5 construct containing the first RecA-like domain and lacking both the N-terminal region and the second RecA-like domain (Dbp5 ΔNΔC, residues 68 - 302) binds Nup214 βΔC to a similar extent as full-length Dbp5 (compare lanes 5 and 3). (b) BIAcore analyses of the interaction between Dbp5 and Nup214. Sensorgrams were obtained from the injection of Nup214 over GST-Dbp5 immobilized on chip surfaces coated with anti-GST antibody. The horizontal axis of the sensorgram indicates the time (seconds) and the vertical axis indicates the response units (RU). In the experiments, the flow cells were derivatized with about 100 - 150 RU of the respective GST-Dbp5 proteins. Nup214 proteins were injected at different concentrations (indicated). Each dissociation constant (K_D) was obtained from at least two separate experiments and derived using a bimolecular interaction model.

Model: one site binding

Equation: $Req = (\frac{[A]}{[A]+K_D})Bmax$

$R^2 = 0.99833$

$Bmax = 36.00674 \pm 0.4044$

$K_D = 22.76216 \pm 1.10482$

Figure 23: Fit of SPR binding data.
This example shows the fit of binding data of GST-Dbp5 ΔNΔC (immobilized) with Nup214 βΔC (analyte) according to a non-linear regression using the Langmuir binding isotherm. Plotted is Req (SPR angle at equilibrium in RU) against the concentrations (in nM) of Nup214 βΔC used. With $Req = (\frac{[A]}{[A]+K_D})Bmax$ the K_D can be calculated. R^2 describes the quality of the fit (with R^2 = 1.0 describing a perfect fit).

4.2.3 Crystallization and data collection

The core complex between Dbp5 ΔNΔC and Nup214 βΔC was crystallized in a sitting drop vapor diffusion experiment. Plate-like crystals were optimized (Figure 24b), cryoprotected with parafine oil, flash frozen in liquid nitrogen and diffracted to 2.8 Å (Figure 25 and Table1) as described in 3.2.11.1.2.

Figure 24: Crystals of the Dbp5 ΔNΔC-Nup214 βΔC complex.
Initial crystals grew in 50 mM MES pH 6.0 and 1.1 M Na citrate and several other condition containing high concentrations of sodium salts of organic acids like citric, malonic, acetic and oxalic acid. (b) Crystals were optimized in 50 mM MES pH 6.0 and 900 mM Na citrate and cryoprotected in parafine oil.

Figure 25: Diffraction image of a Dbp5 ΔNΔC-Nup214 βΔC complex crystal.
Data were recorded at beamline PXII (SLS, Villigen, Switzerland). Crystals diffracted to a maximum resolution of 2.8 Å.

4.2.4 Structure of human Dbp5 in complex with Nup214

The structure was solved by molecular replacement as described in methods 3.2.11.3.2. The 1.65 Å crystal structure of Nup214 (Napetschnig 2007) and the N-terminal domain of Dbp5 (from the Dbp5-AMPPNP-RNA structure) were used as consecutive search models. The model was refined to 2.8 Å resolution to an R_{free} of 26.0 % and R_{work} of 21.8 % and good stereo chemistry (Table 1 and Figure 26). One molecule of the Dbp5-Nup214 complex is present in the structure, where the N-terminal RecA-like domain of Dbp5 (Dbp5 ΔNΔC) binds to the side of the β-propeller face of nucleoporin (Nup214 βΔC, Figure 27). The overall fold of the N-terminal RecA-like domain of Dbp5 is essentially the same as observed in the Dbp5-AMPPNP-RNA structure. The major difference lies in the N-terminal region of Dbp5 (Figure 6, residues 77 - 92, orange). While in the AMPPNP-RNA-bound structure this region forms a helix that contributes to ATP binding (Figure 6), in the Nup214-bound structure (where AMPPNP is not present) the same region wraps around the opposite side of Dbp5, forming an additional β-strand in the core domain. Other structural differences are smaller and localized at the regions containing motifs Ib and GG, and at the helix between motifs II and III. These regions contact Nup214 (Figure 17) and have different conformations compare to the RNA-AMPPNP bound structure.

The β-propeller domain of Nup214 adopts the typical ring-like structure assembled of seven blades arranged around a central axis (Figure 3 and Figure 27). Each blade is formed by β-strands connected by loops of variable length. Long loops extend on one side of the ring (top side), while the other side (bottom side) features shorter extensions (Figure 27, left side). Dbp5 binds to the top side and the contact is mainly mediated by these loops from Nup214 protruding from

blades 1, 6 and 7 and loops and one helix of Dbp5 (Figure 27). The β-propeller domain of Nup214 is in a very similar conformation as observed in the apo structure solved in the lab of Hoelz and Blobel (Napetschnig 2007). Structural differences are mainly in the region responsible for Dbp5 binding. Here loops have slightly different conformations as compared to the unbound structure, while the core of the molecule is essentially unchanged (r.m.s.d. of 0.7 Å between all α-carbon atoms).

Figure 26:
Ramachandran plot of Dbp5 ΔNΔC in complex with Nup214 βΔC.
Plotted are the Φ - Ψ torsion angles of the protein backbone. The four basic Ramachandran plots are shown (a) general case (18 non-glycine non-proline amino acids) (b) glycine (c) proline and (d) pre-proline (referring to residues preceding a proline). 98.5 % (588/597) of all residues are in favored (98 %, light blue) regions and 100 % are in allowed (>99.8 %, blue) regions. There are no outliers.

Figure 27: Structure of the human Dbp5-Nup214 complex.
View of the Dbp5 ΔNΔC-Nup214 βΔC complex in two orientations: On the left, the structure is shown with the N-terminal RecA like domain of Dbp5 in the same orientation as in the Dbp5 ΔN-AMPPNP-RNA structure in Figure 6. On the right, the complex is rotated by 90° around a vertical axis and with a 20° rotation around a horizontal axis to better view the β-propeller ring, where the blades are numbered 1 to 7. The N- and C-terminal residues of Dbp5 and Nup214 in the structure are indicated. The N-terminus of Dbp5 (residues 77 - 92) forms an additional β-strand (shown in orange). The N-terminus of Nup214 (residues 9 - 22, shown in magenta) folds back from blade 1 to complete the neighboring blades 7 and 6 of the β-propeller.

4.2.5 Evolutionary conservation of the Dbp5-Nup214 interaction

Analysis of the Dbp5-Nup214 binding interface by sequence alignments and the program AquaProt (Reichmann 2008) reveals an evolutionary conserved interaction (see alignments Figure 17 and Figure 21). The loop between blade 6 and 7 of Nup214 that is involved in Dbp5 binding is the most conserved part of the β-propeller domain. Other residues involved in the interaction with Dbp5 are also conserved (see residues marked with blue D in alignment, Figure 21). The character of the binding interface is predominantly hydrophilic. In particular, there are two direct salt bridges: Nup214^{Arg348} and Dbp5^{Asp223}, and Nup214^{Asp359} and Dbp5$^{Arg259 \text{ and } Arg262}$

(Figure 28a). These residues are conserved across species (corresponding yeast residues: Nup159^{Arg321} and yDbp5^{Asp221}; Nup159 ^{Asp333} and yDbp5$^{Arg256 \text{ and } Arg259}$).

I mutated those residues in human Dbp5 and Nup214 and performed pull-down experiments (as described in methods 3.2.6.1) in order to investigate their role in complex formation. The mutants were expressed and purified as the wild type (wt) proteins. During expression and purification all mutants showed a similar behavior as the wt proteins, with the exception of Nup214^{R348D} that was not expressed under the given conditions. In the pull-down experiments the single point mutation of Dbp5^{D223R}, of Dbp5^{R259D} or Dbp5^{R262A} is sufficient to abrogate the interaction with the nucleoporin (Figure 28b, lanes 3, 5, 6). The reverse charge mutation of Nup214^{D359R} also inhibits complex formation (Figure 28b, lane8).

Only three amino-acid residues of Dbp5 involved in Nup214 binding are hydrophobic. The main contact is Dbp5^{Ile258} that interacts with Nup214^{Val353} (Figure 28a). Previous work from Weis and coworkers (Weirich 2004) showed that a mutation of the corresponding residue of yeast nucleoporin, Nup159^{Ile326E} together with Nup159^{Val323E}, cause Dbp5 mislocalization and block mRNA export. Since human Nup214 already features a glutamic acid (Nup214^{Glu350}) at the corresponding position of Nup159^{Val323} (*Figure 21*, Figure 28), the drastic effect of the double mutant in yeast might be due solely to the impairment of the interaction between Nup159^{Ile326} and Dbp5^{Ile355}. This is supported by the fact that mutation of human Dbp5^{I258A} is sufficient to disrupt the interaction with Nup214 βΔC by *in vitro* pull-down experiments (Figure 28d, lane 4). Several other mutants of Dbp5 have been reported to affect mRNA export in *in vivo* experiments (Snay-Hodge 1998; Schmitt 1999). A summary of the mutants is shown in Appendix 6.3). Most mutations that showed severe effects in mRNA export carry mutations either in the conserved DEAD-box motifs or in the core of the protein and are therefore likely to impair ATP/RNA binding or to disturb the overall fold and stability of the protein.

I also tested the Dbp5 mutants for their ability to bind to RNA (Figure 28c and d). A detailed description is given in paragraph 4.3.

Figure 28: The Dbp5-Nup214 interaction is mediated by conserved residues.
(a) Close-up of the region of interaction between the N-terminal RecA-like domain of Dbp5 (in blue) and Nup214 (in green) in the Dbp5 ΔNΔC-Nup214 βΔC structure. The residues involved in hydrophobic and electrostatic interactions are highlighted. (b) Protein precipitations by GST-pull-downs with Dbp5 and Nup214 mutant proteins. The pull-down assays were carried out as described in Figure 22a. Mutations of Dbp5^{D223R}, Dbp5^{I258A}, Dbp5^{R259D} or Dbp5^{R262A} impair the interaction with Nup214. Nup214^{V353A} shows reduced binding to Dbp5, whereas Nup214^{D359R} impairs the interaction. (c) Close-up of the region of interaction between the N-terminal RecA-like domain of Dbp5 (in light blue) and RNA (in black) in the Dbp5 ΔN-AMPPNP-RNA structure (shown in a similar view as for the Nup214-interaction close-up in panel a). For clarity, the C-terminal domain and AMPPNP have been omitted in this figure. Binding of RNA at nucleotides 5 and 6 occurs via Dbp5 residues involved in binding Nup214. (d) Protein precipitations with biotinylated ssRNA and mutant Dbp5 proteins. The pull-down assays were carried out as described in Fig. 1b. Binding to RNA is impaired in the Dbp5^{D223R} mutant and reduced in the Dbp5^{R259D} mutant. In contrast, neither of the substitutions I258A or R262A impairs RNA binding, although slower binding compared to the WT protein was observed.

4.3   Dbp5 binds to either RNA/ATP or Nup214 in a mutually exclusive manner

Comparison and superposition of the two described structures reveal that the N-terminal domain of Dbp5 uses part of the binding site for Nup214 to form the binding site for RNA (Figure 28a and c). In particular, Asp^{223} in motif Ib contacts ribonucleotides 5 and 6 (Figure 28c and Figure 17). Mutation of Dbp5^{D223R} not only impairs Nup214 binding as discussed before (Figure 28b, lane 3), but also impairs RNA binding in a biotin pull-down experiment (Figure 28d, lane 3). The other half of the Nup214-binding site, which is formed by the helix between motifs II and III, contacts RNA only marginally. Consistently, Dbp5 mutants with an I258A or an R262A substitution are still able to bind RNA (Figure 28 d, lanes 4 and 6), while they are unable to bind Nup214 (Figure 28b, lanes 4 and 6). Although Dbp5^{I258A} and Dbp5^{R262A} still bind to RNA, I observed that compared to the wt protein they bind RNA much slower. After one hour (as described for the wt proteins in 3.2.6.2) barely any binding was detected for those mutants. In order to show that Dbp5^{D223R} is really not capable of binding RNA the incubation time with RNA for all mutants in the pull-downs (using biotinylated RNAs) was increased to 12 h.

The observation from the structural analysis that Nup214 and RNA occupy overlapping binding sites on Dbp5 and the observation from pull-down experiments with biotinylated RNA, where only Dbp5 and not a Dbp5-Nup214 complex was precipitated (Figure 6, lanes 6 - 10, 12 and 15), suggests that binding of RNA and Nup214 is mutually exclusive. This prediction can be recapitulated in competition experiments. While Dbp5 is efficiently precipitated by biotinylated RNA immobilized on streptavidin beads in the presence of AMPPNP, the precipitation decreases when adding increasing amounts of Nup214 (Figure 29a). Conversely, binding of Nup214 to GST-Dbp5 decreases when adding increasing amounts of RNA (in the presence of AMPPNP, Figure 29b). The same result is observed in EMSAs, where bound (shifted) RNAs are efficiently competed out by increasing amounts of Nup214 (Figure 30, lanes 4 - 6 compared to lane 1).

Figure 29: Mutually exclusive binding of Dbp5 to RNA - AMPPNP and to Nup214.
(a) Effect of Nup214 on the RNA-binding properties of Dbp5. The pull-down assay with biotinylated ssRNA shows that the efficiency of Dbp5 binding to RNA decreases progressively with increasing concentrations of Nup214. (b) Effect of RNA on the Nup214-binding properties of Dbp5. The pull-down assay with GST-tagged Dbp5 shows that the efficiency of Dbp5 binding to Nup214 decreases progressively with increasing concentrations of RNA. Dbp5 and Nup214 proteins where incubated at 1.7 µM, and then mixed or not with 1 mM AMPPNP or 1 mM ADP and increasing amounts of RNA. The RNA concentration in lane 10 equals that in lane 8.

Figure 30: Electromobility shift assay (EMSA) of Dbp5 and 20U ssRNA.
Effect of Nup214 on the RNA-binding properties of Dbp5. Dbp5 shifts RNA only in the presence of AMPPNP (lane 1) and not without (lane 3). Increasing amount of Nup214 (lanes 4 - 6) decrease the level of bound RNA. Nup214 alone does not shift RNA (lane 7). Nup214 (lane 7, green asterisk) and the Dbp5-Nup214 complex (lane 8, red asterisk) are faintly stained by toluidine. Dbp5 alone does not migrate into the gel (lane 3, blue asterisk). These faint proteins bands were approved by Coomassie staining.

These results are supported by ATPase experiments. I qualitatively tested the ATPase activity of Dbp5 in presence of Nup214 (as described in methods 3.2.5). I first tested the pH dependency for this reaction to find the condition with the highest activity (pH 6.5, Figure 31). This condition was used for further experiments, where increasing amounts of Nup214 were used. The amount of ATP hydrolyzed by Dbp5

constantly decreases with increasing amounts of Nup214. (Figure 32). This is further supported by ATPase assays using the previously described mutants of Dbp5 and Nup214 (Figure 33). Dbp5^{I258A}, Dbp5^{R259D} and Dbp5^{R262D} have a lower ATPase activity as compared to the wt Dbp5 (Figure 33a). This is most likely due to the previously described slower RNA binding properties of these mutants. Nevertheless the effect of Nup214 on the ATPase activity of Dbp5 mutants unable to bind to the nucleoporin is negligible (Figure 33b), their ATPase activities are not affected in presence of the nucleoporin.

Figure 31:
Relative pH dependency of the ATPase activity of Dbp5.
Plotted are the counts against the time. Fl Dbp5 has the highest ATPase activity at pH 6.5 (red line). No ATPase activity is observed when RNA (bottom blue) or Dbp5 is omitted (orange)

Figure 32:
Effect of Nup214 on the ATPase activity of Dbp5.
The ATPase assay shows that the ATPase activity of Dbp5 progressively decreases with increasing amounts of Nup214. Plotted is the relative amount of ATP hydrolyzed (%) vs. the time (min). The values are the mean of 3 independent reactions with a linear fit. Error bars represent the standard deviation.

Figure 33: Relative ATPase activity of Dbp5 WT and mutants.
a) Comparison of Dbp5 WT and mutants Dbp5^{I258A}, Dbp5^{R259A} and Dbp5^{R262A}. All mutants have lower ATPase activity than the WT, which can be explained by slower RNA binding properties of the mutants (observed in biotin pull-downs). (b) Comparison of Dbp5 WT and mutants in present of Nup214. As already shown in Figure 32 Dbp5 WT has a lower ATPase activity in presence of Nup214. Mutants unable to bind to Nup214 are not affected. For clarity only Dbp5^{I258A} (purple and light purple) is plotted, but the other Dbp5 mutants showed the same behavior.

A second observation from the structures is that in the closed conformation of Dbp5 bound to RNA and AMPPNP, the C-terminal domain of Dbp5 would clash sterically with an incoming Nup214 (Figure 34, compare panels a and b). Interestingly, in the presence of ADP alone (pdb code 3EWS) recently deposited by the Structural Genomics Consortium SGC, the conformation of Dbp5 is more open than the AMPPNP-RNA bound form. In the ADP-bound structure the N-terminal region of Dbp5 forms a helix that is sandwiched between the N- and C-terminal domains (Figure 34c, in orange) and would partially clash with an incoming Nup214. This is further described in the paragraph 5.2.

Figure 34: Comparison of Dbp5 structures.
(a) The Dbp5 ΔNΔC-Nup214 βΔC structure is shown in a similar orientation as in Figure 22c. Nup214 βΔC is in green. The N-terminal RecA-like domain of Dbp5 is in light blue, with the N-terminal region in orange. In red are regions of the Dbp5 N-terminal domain engaged in the interaction with the nucleoporin and with nucleic acid (see panel b). (b) The Dbp5 ΔN-AMPPNP-RNA structure is shown with the N-terminal domain in the same orientation and color scheme as in panel a. The C-terminal domain is shown in a lighter shade. (c) The structure of Dbp5-ADP from the Structural Genomics Consortium SGC (pdb code 3EWS) is shown with the N-terminal domain in the same orientations as in panels a and b. In the structure, there are 20 additional residues ordered at the N-terminus (ΔN'). The N-terminal region (residues 53 - 67) forms a helix sandwiched between the two RecA-like domains.

4.4 Small Angle X-ray Scattering

In order to investigate the relative conformations of the two RecA-like domains of Dbp5 in the presence of nucleotides and Nup214 I performed SAXS experiments. I tried to address the question if a single conformation is induced by ATP or ADP alone, since Dbp5 has been shown to remove Nab2 from RNA only when in presence of ADP (Tran 2007). I also wanted to see the conformation of the N-terminal flanking region of Dbp5 (1 - 68) in solution, which lacks electron density in the crystal structure. I therefore measured data for the fl and the N-terminally truncated helicase in presence and absence of nucleotides. No aggregation or intramolecular interactions was observed for any of the samples (experimental (red) scattering profiles in Figure 35 and Figure 36c), prerequisite for data analysis. Nevertheless I could not reconstruct 3D-electron density for the helicase in presence of ATP, ADP or Nup214. Resulting bead models were typically much bigger than expected. From SAXS and sls experiments I knew the molecular weight, which

excluded the possibility of a different oligomerization state. Altogether this suggests major flexibility or heterogeneity of conformations. Reconstructions based on simple bead models (used in e.g. Gasbor) typically fail when proteins have major flexible regions. I therefore tried to model conformations of the helicase treating the two RecA-like domains as rigid bodies. With the help of Friedrich Förster (MPIB), who performed the computational work, we tried to model different conformations of the helicase, based on the crystal structure. Dbp5 was treated as a molecule that consists of two rigid bodies (the two RecA-like domains) that are connected with a flexible linker. 1000 independent optimizations of single models against a scoring function including the experimental scattering profiles were carried out (as described in Forster 2008). Each optimization probed ~ 100 different conformations and therefore a total number of ~ 100000 different models were tested. Comparison of the experimental scattering curves with calculated ones (using Crysol, Svergun 1995) show that the experimental scattering profiles cannot be fitted to and explained by a single conformation (Figure 35). Neither a closed conformation (as observed in the Dbp5-AMPPNP-RNA structure), nor a (modeled) open conformation explain the conformation of Dbp5 in presence of ADP (compare Figure 35 a and b). Comparison of relative radii of gyration (RG) derived from experimental data using Guinier plots (data not shown) reveal that Dbp5 without nucleotides (apo form) has the largest R_G of 35 Å. In presence of ADP R_G is 32.5 Å (vs. 31 Å calculated from pdb 3EWS). The closed conformation (as observed in the Dbp5-AMPPNP-RNA structure) has the smallest R_G of 28 Å (calculated from 3FHT). The observation that the experimental scattering profiles miss essential features suggests that an ensemble of different conformations exists in solution. This would explain the 'smoothness' of the experimental profiles, since the experimental SAXS profile would be the average of multiple conformations.

One of the major disadvantages of SAXS is the fact that the information content of SAXS profiles is very limited. Our measured profiles contain approximately 10 independent data points according to the Wittaker-Shannon theorem (Svergun and Koch 2003, reviewed in Putnam 2007). Therefore simultaneous fitting of an ensemble of models (e.g. 10 different conformations) hits the limit of the data-parameter ratio. Finding rigid body coordinates and the

relative abundances of those different conformations is simply underdetermined, and therefore reliability of those conformations would be very poor.

Figure 35: Comparison of experimental and calculated small angle X-ray scattering profiles of Dbp5. Experimental scattering profiles are shown in red and calculated profiles (blue) from atomic structures using Crysol (Svergun 1995) as shown in blue as surface representation with (a) a closed conformation as observed in the AMPPNP and RNA bound structure and (b) a more open conformation (modeled as two rigid bodies connected by a flexible linker). The experimental scattering profiles cannot be fitted with the calculated ones. The experimental profiles miss essential features compared to the calculated ones and profiles are overall smoothened.

I also recorded SAXS data for the β-propeller domain of Nup214 alone. For this experiment a construct including the C-terminal tail was used in order to see if this part of the molecule is ordered (like in the crystal structure) or unstructured in solution. Reconstructions with the programs Credo (Petoukhov 2002) and Gasbor (Svergun 1992) both result in electron density that fits the shape of the β propeller. The reconstruction is shown in (Figure 36b). I observe additional density outside the core (Figure 36b) , which could be explained by a flexible C-terminal tail. However fitting of the crystal structure in the SAXS density is not conclusive since the β-propeller is a flat, ring like molecule and can therefore be positioned into the density in two different orientations and all rotations. A comparison of the calculated scattering profile using the crystal structure of Nup214 (Napetschnig 2007) with the experimental one (Figure 36c) shows that they are in good agreement. The differences observed could be explained by a certain degree of

flexibility of surface loops and the C-terminal tail. All in all the SAXS data for both proteins and the complex do not provide unambiguous and conclusive data. Reconstructions are either not possible due to flexibility or fitting of the data is impaired by the overall globular shape and the similar size of the proteins (domains).

Figure 36: SAXS reconstruction of Nup214 β-propeller domain.
a) Surface representation of Nup214 β crystal structure (Napetschnig 2007) , (b) bead model reconstruction of Nup214 β using the program Credo (Petoukhov 2002). Additional density is observed and marked with red circles. c) Comparison of calculated (blue) and experimental (red) scattering profiles of Nup214 β.

5   Conclusions and Outlook

5.1   Mutually exclusive binding - a paradigm of DEAD-box helicase regulation

Structural and biochemical analysis of the human DEAD-box helicase Dbp5 reported in this thesis show that binding of RNA-AMPPNP and Nup214 are mutually exclusive. Nup214 binds the N-terminal domain of Dbp5 and occupies the binding pockets for RNA nucleotides on the surface of this domain. The C-terminal domain of Dbp5 does not provide binding sites for Nup214. However, the C-terminal domain would sterically hinder the access of Nup214 to the N-terminal domain when the molecule is in the closed, nucleotide-bound conformation. In cells, Dbp5 is concentrated at the cytoplasmic side of the nuclear pore complex by interaction with Nup214 (Nup159 in yeast). Nup214 not only localizes Dbp5, but also traps it in an inactive/inhibited conformation. In this conformation both ATPase activity and RNA binding of Dbp5 are impaired. It is conceivable that the C-terminal RecA-like domain is a target of another yet unknown regulator, acting in a similar way like Nup214.

Notably, the binding sites for RNA and Nup214 on Dbp5 partially overlap. This prevents the concomitant binding of RNA and Nup214 but could allow Dbp5 to switch between RNA-bound and Nup-bound state. The incoming molecule (Nup214 or RNA) would be able to dock to Dbp5 at the sites left empty by the other and possibly accelerate its release. The overall hydrophilic character of the Dbp5-Nup214 interface probably facilitates this switch between states, since major conformational changes (e.g. to bury a hydrophobic surface) are not necessary for both proteins in their apo and bound form. This raises the question of whether Nup214 removes RNA and ATP from Dbp5 or whether RNA and ATP dissociate Dbp5 from Nup214 during mRNA export. Both possibilities are discussed in paragraph 5.4.

DEAD-box helicases are highly conserved in sequence and structure (further discussed in 5.3) and binding of RNA is therefore most likely similar for most DEAD-box helicases. Hence, the interaction of Dbp5 and Nup214 suggest a paradigm of regulation that is likely to apply to other helicases and their regulators. This concept, where a regulator can act by decreasing the activity towards RNA and ATP, presents an efficient and probably universal way to inhibit this class of helicases.

5.2 DEAD-box helicase 5 is modulated in a nucleotide dependent fashion

In its active state, upon concomitant RNA and ATP binding, Dbp5 assumes a closed conformation similar to that observed in the structures of Vasa (Sengoku 2006) and eIF4AIII in the EJC (Andersen 2006; Bono 2006). ATP is bound in the cleft of the two RecA-like domains and RNA is bound by both domains opposite to the linker region, thus forming a compact globular conformation.

Another crystal structure of the human Dbp5-AMPPPNP-RNA complex has been determined independently (Collins 2009, pdb code 3G0H at 2.7 Å resolution) and is essentially identical. The superposition of both RNA-bound structures reveals an r.m.s.d of 0.73 Å over all C_α-atoms. In the same paper, the authors describe the crystal structure of human Dbp5 in complex with ADP (Collins 2009, 3EWS at 2.7 Å resolution). Hydrolysis of ATP to ADP triggers the transition to a more open conformation (see Figure 34c). In the ADP-bound structure, an additional helix is bound in between the two RecA-like domains. The authors speculate that this is a functionally important feature that regulates the helicase activity. In yeast, the N-terminal region including this helix has been shown to be dispensable for cell viability and mRNA export (Snay-Hodge 1998). It is possible that despite high conservation yeast and human Dbp5 are different in this respect. The SAXS analysis of human Dbp5 in presence of 1 mM ADP does not support a single conformation, but the two RecA-like domains adopt a more closed conformation in solution as compared to the apo form. Collins et al. used high concentrations of ADP (20 mM) to crystallize the complex. Altogether the conformation observed in the crystal structure could be favored by crystal contacts and the protein is likely to be more dynamic in solution.

In the inhibited conformation, Nup214 is bound to the N-terminal RecA-like domain of the helicase. Dbp5 presumably adopts an open conformation since in the closed state the C-terminal domain would clash with the nucleoporin (compare Figure 34 panel a and b). Another crystal structure of the human Dbp5-Nup214 complex was published (Napetschnig 2009) and appears similar, although the coordinates are so far not accessible and therefore a superposition and direct comparison has not been possible. Napetschnig et al. crystallized the N-terminal β-propeller domain of Nup214 in complex with a Dbp5 construct consisting of both RecA-like domains. In this 3.2 Å resolution structure (pdb code 3FMO) the authors

did not observe electron density for the C-terminal domain of Dbp5, suggesting that it is flexible.

Yet another crystal structure of the *S. pombe* Dbp5 (3FH0, coordinates not accessible) together with NMR structures (2KBE and 2KBF, coordinates not accessible) of the two individual RecA-like domains of *S. cerevisiae* Dbp5 were published later (Fan 2009). Although the structural statistics of the NMR data are rather poor (~ 65 % in favored Ramachandran regions), the data suggest that the N-terminal flanking region is unstructured in solution and the N-terminal region (residues 71 - 121) is highly dynamic. The authors conclusions are in agreement with our data, where the fl human Dbp5 lacks ordered electron density for the first 74 residues and where the N-terminal region (75 - 92) of the molecule adopts different conformations, depending on the nucleotide state. Furthermore the authors have shown that the N-terminal RecA-like domain is sufficient for ADP and ATP binding.

Thus, it appears that Dbp5 transition from an open to a closed conformation occurs in an RNA and nucleotide dependent manner. The presence of RNA and ATP induce the closed conformation. Binding of ADP might induce a single conformation where the two RecA-like domains are separated by a helix, or form more open and dynamic conformations. In the nucleotide-free (apo) form the helicase adopts flexible conformations where the two RecA-like domains are best described as two rigid bodies connected by a flexible linker. The ADP bound state is compatible with Nup214 binding (Napetschnig et al. crystallized the Dbp5-Nup214 complex in presence of ADP), although the nucleotide state affects the affinity for the nucleoporin. Compared to the nucleotide-free state binding to Nup214 in presence of AMPPNP is reduced by a factor of 50 and in presence of ADP by a factor of 30. Since the nucleotide binding site is the most conserved part of DEAD-box helicases this nucleotide dependent modulation likely applies to other DEAD-box proteins.

5.3 DEAD-box helicases have a functionally and evolutionary conserved core

Comparison with other DEAD-box helicase structures shows that these proteins have conserved structural features: they consist of two globular RecA-like domains that bind ATP in their cleft with highly conserved residues. The two domains bind concomitantly to ssRNA in a sequence-independent manner. Including my work structural data for three DEAD-box helicases in complex with an ATP analogue and ssRNA are available (Vasa, Sengoku 2006; eIF4AIII, Andersen 2006; Bono 2006; Dbp5, Collins 2009) and show that ATP binding/hydrolysis is achieved by a similar mechanism involving equivalent amino acids, coordinated waters and metal ions. Most residues binding RNA are also conserved within this family (alignment in appendix 6.5).

The high conservation of DEAD-box proteins on a sequence, structural and mechanistical level raises the question how they are able to specifically act on different substrates at different stages? Most DEAD-box helicases bind RNA in a sequence independent manner and therefore specificity is unlikely to be conferred by the RNA substrate, although some helicases have other domains outside the helicase core that do provide specificity. One example is DbpA that binds specifically to a hairpin of the 23S rRNA via its additional C-terminal domain (Fuller-Pace 1993). So far N- and C-terminal flanking regions, that are found in many DEAD-box helicase are poorly understood but could play roles in specificity by localizing helicases to specific compartments within the cell. Specificity might arise from direct and indirect interactions with regulators that activate or inhibit, transport and localize helicases, but so far only few interactions have been shown *in vitro* (e.g. eIF4A-eIF4G complex, Schutz 2008). Some of those interactions might only be transient and possible in the context of certain multiprotein complexes (e.g. in the case of the EJC, Andersen 2006; Bono 2006). The reconstitution of native complexes such as mRNPs *in vitro* seems to be an important but difficult task for the future in order to understand how helicases act on substrates and how they are regulated.

Studies to date have focused on the activation of Dbp5 and of DEAD-box proteins in general. The activity of Dbp5 is stimulated by the protein Gle1 and its cofactor IP_6 (Alcazar-Roman 2006; Weirich 2006). Gle1 binds the nucleoporin CG1 in humans or Nup42 in yeast and is believed to be in close proximity to human Nup214 and yeast Nup159 in the architecture of the nuclear pore complex

(Miller 2004). The molecular details of how Gle1 acts on the ATP-bound form of Dbp5 are currently unknown. Structural studies of the DEAD-box proteins eIF4A and eIF4AIII have given some insights into the molecular mechanism for their specific activation by eIF4G and Barentsz, respectively. The next aim is to get insights into the activation of Dbp5 at a molecular level, which could present a general paradigm of how this class of proteins is regulated. In addition to this, most mechanistic details of how ATP hydrolysis is coupled to unwinding are not yet understood. The use of different ATP transition state analogues like beriliumfluoride-ADP and aluminiumfluoride-ADP could provide some mechanistic insights.

Comparison of crystal structures of various DEAD-box helicases reveal a common feature: independent of the ligand or regulator bound, the globular structures of the two individual RecA-like domains remain similar (Figure 37). DEAD-box helicases are highly conserved on a sequence level and probably evolved from the same ancestor. Together with the overall conservation of structural features this suggests that a stable core and an efficient mechanism for ATP binding/hydrolysis and RNA binding remained unchanged, while variations of the surface properties and modular addition of domains outside the core made different functions within the cell possible.

Figure 37: Overview of crystal structures of different DEAD-box helicases.
In this schematic, structures of different helicases have been superposed and are shown in the same orientation. The N-terminal RecA-like domains are shown in dark grey (labeled N), the C-terminal RecA-like domains (if present) in lighter grey. RNA is shown as a black line, regulators that are bound are shown in red, green or blue. (a) DEAD-box helicases bound to a non-hydrolysable ATP analogue and RNA assume a closed conformation where the ATP analogue is bound in the cleft of the two RecA-like domains (e.g. Vasa, Sengoku 2006) (b) DEAD-box helicase eIF4AIII in the exon junction complex (Andersen 2006; Bono 2006) is locked on the RNA by interactions with Barentsz (green) and Mago+Y14 (blue) (c) eIF4AIII assumes an open conformation when only bound to Barentsz (Bono 2006). Here, the C-terminal RecA-like domain rotates more than 180 degrees. The same conformation is observed in the absence of Barentsz (Andersen 2006) (d) only a few DEAD-box helicases have been crystallized in the apo form, where both domains are present (e.g. eIF4A, Caruthers 2000). The two domains have a very flexible conformation in respect to each other, preventing crystallization in most cases. The conformation of eIF4 shown here is most likely stabilized by crystal contacts and resembles one out of many possible conformations. (e) in the eIF4A+eIF4G complex (also known as eIF4F, Schutz 2008), the helicase is in an activated conformation. eIF4G binds opposite the RNA binding side and brings the two domains in close proximity. (f) several helicases have been crystallized with ADP (e.g. pdb code 2PL3), but in general only the N-terminal RecA-like domain was used or ordered. In the case of human Dbp5, both domains are present separated by a helix (shown in orange). The helicase assumes a more closed conformation compared to the apo conformation, which is probably stabilized due to crystal contacts. (g) In the Dbp5-Nup214 complex (pdb code 3FHC, 3FMO), the helicase is in an inhibited conformation where both ATP hydrolysis and RNA binding are impaired.

5.4 Nuclear export of messenger RNA - an updated model

Messenger RNA is transcribed in the nucleus and needs to be exported to the cytoplasm where translation occurs. Transport through the NPC is mediated by the nuclear transport factor TAP-p15 in metazoans and Mex67-Mtr2 in yeast. The N-terminal half of TAP binds the mRNP, while the C-terminal half binds the FG repeats of nucleoporins (reviewed in Cook 2007) . Export competent mRNPs include many proteins, some of which bridge the mRNA to the transport factor, others are removed prior to or during export and some that are not dissociated until translation. The current model for mRNA export envisages that Dbp5 could bind to mRNA in the nucleus (Cole and Scarcelli 2006; Stewart 2007; Iglesias and Stutz 2008), where Dbp5 is found in both *S. cerevisiae* (Estruch and Cole 2003) and *C. tentans* (Zhao 2002). Upon reaching the cytoplasmic side of the NPC, docking of mRNP-bound Dbp5 to Nup214 would position it close to its activator Gle1. The subsequent activation of Dbp5 would therefore occur specifically at the cytoplasmic side of the NPC and result in ATP hydrolysis, dissociation of Dbp5 and other mRNP-bound factors such as yeast Mex67 (Lund and Guthrie 2005) or Nab2 (Tran 2007), thus releasing the mRNP to the cytoplasm.

However, the data reported are not compatible with this model since Dbp5 binding to the mRNP and to Nup214 are mutually exclusive. The association of an mRNP-bound Dbp5 to Nup214 would result in the dissociation of RNA and ATP and the simultaneous activation of Dbp5 by Gle1 would therefore not be possible at this step. A way to reconcile this model with the results from my structural analysis would be if Gle1 binding/activation occurs prior to Nup214 binding (Figure 38), or in a second scenario, that binding to the emerging mRNP and binding to Gle1 and subsequent activation occurs after Nup214 binding (Figure 39): In the first case, the helicase would have to bind to the mRNP in the nucleus. As the mRNP-bound Dbp5 would move through the NPC, it would encounter Gle1. Gle1 would stimulate ATP hydrolysis, promoting the remodeling activity of Dbp5, resulting in displacement of the mRNP-bound proteins and thus releasing the (remodeled) mRNP. Dbp5 in its apo (open) form would then be captured by Nup214, possibly by docking initially to the ADP-bound form and accelerating ADP displacement.

Figure 38: mRNA export and release model 1.
The nuclear pore complex (NPC) is shown schematically with the nuclear basket, the central scaffold and the cytoplasmic fibrils. Dbp5 is shown in light blue, with the two RecA domains in a closed conformation (bound to ATP and RNA), in a slightly more open conformation (when bound to ADP) and in a completely open conformation (when unbound or bound to Nup214). The cytoplasmic nucleoporin Nup214 is shown in green, with the globular Dbp5-binding domain and the FG-repeat domain. The FG repeats of nucleoporins are shown as zigzagged lines in green (Nup214), grey (hCG1) and in black (other Nups). The cytoplasmic nucleoporin hCG1 is in grey and is shown bound to Dbp5-activating protein Gle1. The mRNP is shown with various proteins bound to it (adaptors, DEAD-box proteins etc.). In the nucleus, the export competent mRNP is shown bound to the transport factor TAP-p15 (in dark gray). TAP-p15 binds the FG-repeats of nucleoporins via the C-terminal domain, and adaptors to bridge to RNA via the N-terminal domain. When Dbp5 bound to the mRNP approaches Gle1 it is activated (red asterisk) and displaces the export factor by local unwinding and thereby prevents backsliding of the mRNP. The export factor shuttles back to the nucleus and free Dbp5 is captured by Nup214 to prevent removal of other proteins.

In the second scenario (Figure 39), association of Dbp5 with RNA in the nucleus might not be required for the mRNP to become export competent. In this model, the emerging mRNP would encounter Dbp5 attached to Nup214 in an open form. Dbp5 would bind to ATP and the mRNP, resulting in dissociation from Nup214. The juxtaposition of Gle1 (Gle1 binds to CG1 (Nup42 in yeast) that is thought be in close proximity to Nup214 (Nup159 in yeast) in the architecture of the NPC) would then activate Dbp5 locally and result in remodeling of the mRNP,

removal of the transport factor and release of the RNA. Unbound Dbp5 would then either diffuse to the cytoplasm or be sequestered by Nup214 (possibly by docking initially to the ADP-bound form and accelerating ADP displacement), which would then position the helicase for the next emerging mRNP. This model predicts that fractions of Dbp5 bound to the mRNP in the nucleus would not necessarily be dissociated from the mRNP upon nuclear export. Whether this is a mechanism by which Dbp5 might be retained from the nucleus to cytoplasmic mRNPs until it reaches the translation machinery (Gross 2007) or P body components (Scarcelli 2008) remains to be shown.

Both models raise the question of how Dbp5 dissociates from the nucleoporin. In *in vitro* assays we have determined the binding affinities for the Dbp5-Nup214 complex. The two proteins form a strong interaction with K_Ds in the nanomolar range. One way to disrupt the complex would be an unknown (protein-)factor that helps to dissociate this complex. A second way would be a nucleotide-dependent regulation, since different nucleotide states of the helicase influence the binding affinity for the nucleoporin. In the presence of AMPPNP or ADP, binding of the nucleoporin is reduced. Dbp5 bound to Nup214 would bind ATP, possibly by its N-terminal domain alone and thus reduce the affinity for the nucleoporin. Cooperative binding of RNA and ATP would then result in dissociation from Nup214. Recent NMR titration experiments (Fan 2009) determined binding affinities of the *S. cerevisiae* N-terminal domain of Dbp5 for ATP ($K_D \sim 1.5$ mM) and ADP ($K_D \sim 0.5$ mM). It is surprising that the helicase has a higher affinity for ADP than ATP. Repeating these experiments with the human helicase in presence of the nucleoporin could provide more details. In the presence of Nup214, Dbp5 could have a preference for ATP. In this case the nucleoporin would help to preload the helicase with ATP, while inhibiting its ATPase activity. This would result in reduced affinity of the Dbp5-Nup214 complex and concurrently prepare the helicase to dissociate.

Another open question in both models is how Dbp5 discriminates between proteins that must be dissociated and proteins that need to be retained on the cytoplasmic mRNP. As RNA binding by Dbp5 is independent of the nucleic acid sequence, and since DEAD-box helicases are not processive and therefore their activity does not propagate over long distances, discrimination is likely to arise from proximity effects. This would be the case if Dbp5 and the proteins that need

to be removed from the mRNP interact directly or via adaptors. However these interactions with possibly targets such as Mex67 and Nab2 have not been shown *in vitro* so far. The positioning of Dbp5-Gle1 in the vicinity of specific mRNP-bound proteins to be remodeled could confer selective mRNP remodeling without the necessity of direct interactions. Binding of the TAP-p15 transport factor to the nucleoporin FG repeats might position it near Dbp5 and Gle1, which are associated with the only two cytoplasmic nucleoporins Nup214 (yeast Nup159) and CG1 (yeast Nup42) that contain FG repeats (Terry and Wente 2007). The local activation of Dbp5 in the proximity of TAP-p15 would thus lead to the dissociation of the transport factor from the mRNP providing directionality in the export of mRNAs.

Many steps in the final mRNA export and release mechanism remain unclear. In the case of Dbp5 one unresolved question is whether the cytoplasmic and the nuclear fractions of the helicase are separate pools with independent functions or if the same molecule travels all the way from transcription to translation on the same mRNA. It is also currently unknown how the helicase is imported to the nucleus, since a bipartite nuclear localization sequence (NLS) is not present in the sequence. Ultimately, *in vivo* studies in cells are likely to uncover more details of the interactions of Dbp5 in the cell. Single molecule studies and the use of fluorescence correlation spectroscopy could show, which proteins interact at which time and space. One example is the transport factor Mex67, which is thought to be removed from the transcript by the helicase. How and if these proteins interact has not been shown so far. It remains to be shown if the two proteins are already in close proximity in the nucleus (on the transcript) or only later at the nuclear pore. More details of the final steps of mRNA export are needed to understand how transcripts are directionally transported and finally released.

Figure 39: mRNA export and release model 2.
The nuclear pore complex (NPC) is shown schematically with the nuclear basket, the central scaffold and the cytoplasmic fibrils. Dbp5 is shown in light blue, with the two RecA domains in a closed conformation (bound to ATP and RNA) and in a completely open conformation (when unbound or bound to Nup214). The cytoplasmic nucleoporin Nup214 is shown in green, with the globular Dbp5-binding domain and the FG-repeat domain. The FG repeats of nucleoporins are shown as zigzagged lines in green (Nup214), grey (hCG1) and in black (other Nups). The cytoplasmic nucleoporin hCG1 is in grey and is shown bound to the Dbp5-activating protein Gle1. The mRNP is shown with proteins bound to it (adaptors, DEAD-box proteins etc.). In the nucleus, the export competent mRNP is shown bound to the transport factor TAP-p15 (in dark gray). TAP-p15 binds the FG-repeats of nucleoporins via the C-terminal domain, and adaptors to bridge to RNA via the N-terminal domain. When the mRNP translocates through the pore and reaches the cytoplasmic Nup214, Dbp5 bound to the nucleoporin is released and binds ATP and the mRNP. The helicase travels further with the mRNP until it approaches Gle1 bound to hCG1. Dbp5 is activated (red asterisk) and displaces the export factor by local unwinding and thereby prevents backsliding of the mRNP. The export factor shuttles back to the nucleus. Free Dbp5 is either further travelling to the cytoplasm - possibly by rebinding the mRNP or is sequestered by Nup214, waiting for the next translocating mRNP.

6   Appendix

6.1   Abbreviations

Å	Ångström (=10^{-10} m)
ADP	adenosine diphosphate
ALS	Advanced Light Source
AMP	adenosine monophosphate
AMPPNP	5´-adenylyl-imido-triphosphate
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
C. tentans	*Chironomus tentans*
dATP	2'-Deoxyadenosine 5'-triphosphate
Dbp5	DEAD-box helicase 5
dCTP	2'-Deoxycytidine 5'-triphosphate
dd	double destilled
DDX19B	DEAD-box polypeptide 19B
dGTP	2'-Deoxyguanosine 5'-triphosphate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dNTP	deoxyribonucleotide
DTT	dithiothreitol
dTTP	2'-Deoxythymidine 5'-triphosphate
E. coli	*Escherichia coli*
e.g.	for example (exempli gratia)
EDTA	ethylenediaminetetraacetic acid
EJC	exon junction complex
EM	electron microscopy
EMBL	European Molecular Biology Laboratories
EMSA	electro mobility shift assay
fl	full length
FPLC	fast protein liquid chromatography
GSH	glutathione
GST	glutathione *S*-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

H. sapiens	*Homo sapiens*
IP$_6$	inositol hexakisphosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
K$_D$	dissociation constant
LB	Luria-Bertani
LIC	ligation independent cloning
LRR	leucine rich region
M	molar
min	minute
MPIB	Max Planck Institute of Biochemistry
MR	molecular replacement
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein particle
MW	molecular weight
MWCO	molecular weight cutoff
NE	nuclear envelope
NEB	New England Biolabs
NLS	nuclear localization sequence
NPC	nuclear pore complex
Nup	nucleoporin
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PSI	Paul Scherrer institute
r.m.s.d	root mean square deviation
RBD	RNA binding domain
R$_G$	radius of gyration
RI	refractive index
RNA	ribonucleic acid
RU	response units
S. cerevisiae	*Saccharomyces cerevisiae*
S200	Superdex 200
SDS	sodium dodecyl sulphate
sls	static light scattering
SLS	Swiss Light Source
SPR	surface plasmon resonance

ss	single stranded
TBE	Tris base, boric acid and EDTA containing buffer
TEV	tobacco etch virus
TLS	translation/libration/screw
TLSMD	translation/libration/screw motion determination
Tm	melting temperature
UV	ultraviolet
v	volume
w	weight
Wt	wild type

Nucleotides

one letter code	nucleobase
A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil

Amino acids

One letter code	Three letter code	Amino acid
A	Ala	alanine
C	Cys	cystein
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

6.2   Purification overview

Figure 40: Purification scheme of Dbp5 and Nup214 constructs.

6.3 Summary of Dbp5 & Nup214/159 mutants

Table 2: summary of mutants and their phenotypes.

Mutant	Phenotype	structural explanation
Human Dbp5		
Dbp5^{D223R}	no binding to Nup214	direct salt bridge with Nup214^{R348}
	no binding to RNA	D223 involved in RNA binding
Dbp5^{R262A}	binds RNA but not Nup214	not involved in RNA binding but direct salt bridge with Nup214
Dbp5^{R259D}	reduced RNA binding no binding to Nup214	partly involved in RNA binding direct salt bridge with Nup214
Dbp5^{I258A}	binds RNA but not Nup214	not involved in RNA binding but hydrophobic interaction with Nup214^{V353}
Dbp5^{E243Q}	ATPase deficient *in vitro*, block of mRNA export *in vivo*	conserved DEAD-box motif, E243 binds γ-phosphate of ATP
Dbp5$^{E243Q+V386N}$	ATPase deficient *in vitro*, block of mRNA export *in vivo*	conserved DEAD-box motif
Dbp5^{R429Q}	reduced ATPase activity *in vitro*, *in vivo* partial inhibition of mRNA export when injected at high concentrations into Xenopus oocytes	R429 involved in ATP coordination
Dbp5$^{M270P+L271R}$	no binding of Nup214 *in vitro*	not involved in binding of Nup214, prob. disrupts the central sheet of the Nterm. domain.
Dbp5$^{M270P+L271R+ E243Q}$	no binding of Nup214 *in vitro*	not involved in binding of Nup214, disrupts the central sheet of the Nterm. domain.
Dbp5^{S138R}	no binding of Nup214 *in vitro*	not involved in binding of Nup214, S138 located in the core
Dbp5$^{S138R+E243Q}$	restores Nup214 binding by 50 %	-
Human Nup214		
Nup214^{V353A}	weak binding to Dbp5	hydrophobic contact to Dbp5^{I258}
Nup214^{D359R}	no binding to Dbp5	direct salt bridge to Dbp5^{R259} and Dbp5^{R262}
Yeast Dbp5		
Dbp5^{L267P}	growth defects at permissive temperature	located in the protein core, corresponds to human Dbp5^{M270}
Dbp5^{I285N}	growth defects at permissive temperature	located in the protein core, corresponds to human Dbp5^{V388}
Dbp5^{P170H}	growth defects at permissive temperature (cold sensitive), modest nuclear accumulation of poly(A)$^{+}$-RNA	located in the protein core
Yeast Nup159		
Nup159^{V323E,I326E}	Dbp5 mislocalization and mRNA export blocked *in vivo*	Effect most likely only due to mutation of V323E - hydrophobic contact to Dbp5^{I326}

6.4   Summary of other attempts

This is a brief overview of other attempts.

Attempt	Experiment	observation	reference
crystallization of yeast Dbp5 in complex with RNA	sitting and hanging drop vapor diffusion	good expression and purification no crystals, no binding of RNA	
crystallization of yeast Dbp5 in complex with Nup214	several constructs tested in sitting and hanging drop vapor diffusion, microbatch	good expression and purification no crystals	
surface engineering of yeast Dbp5	Several mutation of surface lysines	good expression and purification lower solubility no crystals	Derewenda 2004
methylation of surface lysines	chemical reductive methylation of surface lysines of dbp5, nup159 and the complex	aggregation and heterogeneity of samples. no crystals	
expression of yeast Gle1	several constructs with different Nterm.- and Cterm. tags tried in *E. coli.*	all constructs that were expressed were insoluble	
	few constructs tried in *S. cerevisiae*	no expression	
	full-length Gle1 in insect cells	good expression but insoluble	
crystallization of human Dbp5 in complex with different ATP transition state analogues	co-crystallization of Dbp5+RNA together with ADP-BeF$_x$ and ADP-AlF$_3$	up to 2.2 Å resolution diffracting needle shaped crystals	

Before I started working on this project, a *S. cerevisiae* Dbp5-Nup159 complex was crystallized by Judith Ebert in our lab. A few crystals grew after several month in one condition (Hampton cryo screen: 170 mM Na-acetate, 85 mM Tris pH 8.5, 15 % glycerol, 25.5 % PEG4000), with only one of them diffracting to 4.5 Å resolution. The structure was solved by Esben Lorentzen by molecular replacement using a modeled Dbp5 structure based on the UAP56 helicase (27 % sequence identity) structures (pdb codes: 1T5J, 1T6N, 1XTK, 1XTI). A solution was found with the program Phaser (McCoy 2004) with two molecules per asymmetric unit cell. Although a construct of Dbp5 containing both RecA-like domains in complex with the N-terminal domain of Nup159 was used, only one N-terminal RecA-like domain of Dbp5 and 2 molecules of Nup159 were found. The C-terminal domain and the second molecule of the helicase had no visible electron density. Not only the relative low resolution of the data, but high mosaicity, high R factors and partial completeness of the dataset made refinement and model building impossible. Despite Judith's and later my work these crystals were never reproduced.

6.5 Alignment of known S. cerevisiae DEAD-box helicases

```
              10        20        30        40        50        60        70        80        90        100       110       120       130       140       150
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
DBP5   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
PRP5   1   METIDSKQNINRESLLEERRKKLAKWKQKKAQFDAQKEHQTSRNDIVTNSLEGKQTTEKFTERQERVKEELRKRKNEFRKSDEPVSVKPSKKKSKRSKVKKKISFDFSDDDDSEIGVSFRSKEHIQKAPEHDNEKDPLDEFMTSLKEEKM 150
DRS1   1   -----------------------MVVGTKKYSNLDFVPTISDSEDDVPILDSSDDEKVEAKKTTKKRKGKNNKKKVSEGDNLDEDVHEDLDAGFKFDLDADDTTSNFQGWNFLAEGESNKDDAEAFVKKDVDLDKIIRRKGGLVKMAHI 126
DED1   1   -------------------------------------------------------------------------MAELSEQVQNLSINDNNENG------YVPPHLR--GKPRSARNNSSNYNNNNGGYNGGRGGGSFFSN 59
DBP6   1   -----------------------------------------------------------------------MFASRFDPSQLTAPAASAPEGIVGTTPPAIVPLKRQATESDNEEYGSHQDSDESSNSSSEEDEDRMQ 67
DBP2   1   ----------------------------------------------------------------------------MTYGGRDQQYNKTNYKSRGG----------DFR--GGRNSDRN--SYNDRPQGGNYRG--------- 44
DBP1   1   ----------------------------------------------------------------------MADLPQKVSNLSINNKENGGGGGKSSYVPPHLRSRGKPSFERSTPKQEDKVTGGDFFRAGRQTGNN 67
DBP3   1   ----------------------------------------------------------------------MTKEEIADKKRKVVDEEVIE----------------KKKSKKHKKDKKDKKEKKDKKHKKHK----- 46
PRP28  1   -----------------------------------------------------------------------MARPIDVSQLIAGINKKKGLDENTSGKISKPRFLNKQERSKQERLKENEESLTPTQSDSAKVEIKKV 67
MAK5   1   ----------------------------------------------------------------MGKKRAPQKGKTVTKPQEIIVDESKLNWKPVDIPDTLDDFGGFYGLEEIDGVDVKVVDGKVTFVT 65
DBP7   1   ----------------------------------------------------------------------------MSDEDSMLLNFTTNEDTAGSSYKQAAKVTGGRWKDRRR 38
DBP10  1   --------------------------------------------------------------------------------------MAGVQKRKRDLEDQDDNGSEEDDIAFDIANEI 32
MRH4   1   -------------------------------------------------------------------------------------------------MSLFFKPVISPQWSFPVLLKI 21
MS116  1   ----------------------------------------------------------------------------------------------------MLTSILIKGRTPVLASRN 18
RRP3   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
DBP4   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
HAS1   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
DHH1   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
SUB2   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
IF4A   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
FAL1   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
IF4A   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
DBP9   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
SPB4   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1
DBP8   1   ------------------------------------------------------------------------------------------------------------------------------------------------------ 1

              160       170       180       190       200       210       220       230       240       250       260       270       280       290       300
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
DBP5   1   -------------MSDTKRDPADLLASLKIDNEKEDTSEVSTKETVKSQPEKTADSIKPAEKLVPKVEEKKTKQEDSNLISSEYEVKVKLADIQADPNSPLYSAKSEDELG--LAPELLKGLIYAMKGQKPSKIQERALI---LLIHNPP- 131
PRP5   151 SNSKGMYDRGDILDVEDQLFELGGTDDEDVEDNTDNSNIAKIAKLKAKKRVKQIYYSPEELEPFQKNFYIESETVSSMSEMEVEELRLSLDNIKIKGTGCPKPVTKWSQLG-LSTDTMVLITEKLHGGSLTPIQSQALP---AIMSF---293
DRS1   127 DSKQEEETEKEKVEKENDSDDEELAMDGFGMGAPMNNGDENQSEEEEEEEEKEEEEEEEEEQEEMTLEKGGKDDEIDEEDDSEEAKADFYAPETEGDEAKKQMYENFNSLS--LSRPVLKGLASLGQVKPSPIQSATIP---IALLG---268
DED1   60  NRRGGYG---NGGFFGGNN-GGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPNFQSSGINFDNY------------------DDIPVDASGKDVPEPITEFTSP--PLDGLLLENIKLARETKPTPVQKYSVP---IVANS---179
DBP6   68  VDYGASEEDSSEVEEEESKPSTHSTVLSRFKQTVSLQERLGASDIAESKEDEGIEDEAASTHQLKQIPQPEFVKNPMNLNTNSLQFKSTGWLNTEKIYYDNSLAKPESDYANELEAKLLQNICKNFSTNTFPIQSIILDSILPVLNFTLN 217
DBP2   44  ----GFG---GRSNYNQPQELIKPNWDEELPKLPTFEKNFYVEHESVRDRSDSEIAQFR-----------------KENEMTISGHDIPKPITEDEA--GFPDYVINEVKAEGEDKPTGIQCQGWF---MALSF---150
DBP1   68  GGFFGFSKERNGGTSANYNRGGSSNYKSSGNRWVNGKHIPGPKNAKLEAELFGVHDDPDYHSSGIKFDNY------------------DNIPVDASGKDVPEPITLDESSP--PLDELLMENIKLASFTKPTPVQKYSIP---IVTKS---191
DBP3   46  --------------------KEKKGEKEVEVPEKESEKKPEPTSAVASEFYVQSEALTSLPQSDIDEYFK-----------------ENEIAVEDSLDLALRPILSEDYE---SLDSSIQAEIS--KGPKPTPIQAVAWP---YLISS---149
PRP28  68  NSRDDSFFNETNDKKRNPSKQNGSKFHFSWNESEDTLSGYDPIVSTRAIDLLWKGKTPKNAAESSYMGKHWTEKSLHEMNERDWRILKEDYAIVTKGGTVENPLRNWEEINIIPRDLLRVIIQELRFPSPTPIQRITIPNVCNMKQY---214
MAK5   66  KKDSKVLKDSNKEKVGDDQESVENESGSDSESELLEFKNLDDIKEGELSAASYSSSDEDEQGNIESSKLTDPSEDVDEDVDEDVLKENVFNKDINIDDISPVNLPEWTNIAP-LSMTILQSLQNLNGLRPIEIQKKSIP---VIMQF---208
DBP7   39  MKMKLEGKTVSRKRKANTTGDEGIIPGRGENSIKKLHKESSYSSEEQEKYKGRNAHNTQGRTLPADSQFVSSLFTSNREITTAVNTNIHDENVAINPSNAPLKGDQFASLG-VSSLLVSHLEQKMRIKKPTSIQKQAIP---QIIGNAGK-184
DBP10  33  ALNDSESDANDSDSEVEADYGPNDVQDVIEYSSDEEEGVNNKKKAENKDIKKKKNSKKEIAAFPMLEMSDDENNASGKTQTGDDEDDVNEYFSTNNLEKTKHKKGSIPSFG--LSKIVLNNIKRKGRQPTPIQRKTIP---LIIQS---174
MRH4   22  GVRSYAGGPRTKHKGNSPLASVPTGSSNKNRKQKAKGKKGNKKNDPDQAFNFGEYGGLKKDVEMNMDSTNKLIQKISN--------------FDQLLILPPVRDAVKEITSKESLKLQDSRKKTSENIIPSPIQTVAIKRISKNIMDP---155
MS116  19  LLAALSNCNHITWAVSRRLYNDGNRDQRNFGRNQRNNNSNRYRNSRFNSRPRTRSREDDDEVHFDKTTFSKLIHVP------------------KEDNSKEVTLDSLLEEG-VLDKEIHKAITRMEIPGLTPVQKTIK---PIISSED-145
RRP3   1   --------MSKIVKRKEKKANDELTSLAEKIRAKALENQKKLIEAEKEGGSESDSEEDATAEKKKVLKSKSKSTVSTQNENT----------------NEDESFESFSEIN--LVPELIQACKNLNYSKPTPIQSKAIP---PAEES---118
DBP4   1   -----------------------------------MAK-KNRLNTTQRKTLRQKEDEYIENLKTKIDEYDPK---------------------ITKAKFEKDIP--ISDPTLKGTRESSFIKLLEIQADSIP---VSIQF---78
HAS1   1   -------------------------------------MATPSNKRSRDSESTEEPVVDEKSTSKQNNAAPEGEQ-------------------------TTCVEKFEELK--ISQPTLKAIEKMGFTTMQSVQARTIP---PLIAC---79
DHH1   1   ---------------------------------MGSINNNFNTNNNSNTDLDRDWKTALNIPKKDT----------------RPQTDDVLNTKGNTEDFY--LKRELLMGIFEAGFEKPSPIQEEAIP---VAITC---83
SUB2   1   -----------------------------MSHEGEEDLLEYSDNEQEIQIDASKAAEAGETGAATSATEGDNNNNTAAGDKKGSYVGIHSTGFKDFL--LKPELSRAIIDCGFEHPSEVQHTIP---QSIHC---99
IF4A   1   ----------------------------MSEGITDIEESQIQTN---------------------------------YDKVVYKFDDMELDENLLRGVFGYGFEEPSAIQRAIM---PIIEG---59
FAL1   1   -------------------------------MSFDREEDQKLKFKTSK---------------------------KLKVSSTFESVN--IKDDLLRGIYSYGFEAPSSIQSRAIT--QIISC---60
IF4A   1   ----------------------------MSEGITDIEESQIQTN---------------------------------YDKVVYKFDDMELDENLLRGVFGYGFEEPSAIQRAIM---PIIEG---59
DBP9   1   ----------------------------------MSYEKKSVEG---------------------------------AYIDDSTTFEAFH--LDSRLIQAIKNIGFQYTLIQSHAIP---LALQQK--55
SPB4   1   --------------------------------MSKSLEWDNIGFSLLPWIRTGLDVMGFETMQSASTIP-----MIAGN---44
DBP8   1   --------------------------------------MADFIKSLG--LSKWLTESLRAMKITQPTAIQKACIP---KIIEG---39
                                                                                                                                                        Q motif

              310       320       330       340       350       360       370       380       390       400       410       420       430       440       450
```

Motif I

DBP5	131	...RNMIAQSQSRTGKTAAFSITMLTRVN...PEDASPQATCLAPSRELARCTLEVQEMG...KFT...KITSQ--LIVPDSFEKNKQ...INAQVIVGTPTVLDLMRRKLMQ...LQKIKIF	236
PRP5	293	...RDVIGISKTSGKTISYLLPLLRQVKAQRP...LSKHETGPMGLILAPTRELALQIHEEVTKFT...EADTSIRSVCCTGG--SEMKKCITDLKRG...TEIVVATPGRFIDILTLNDGK--LLSTKRITEV	412
DRS1	268	...KDIIAGAVTSGKTAAFMIPIIERLL-YKP...AKIASTRVIVLLPTRELALQIVADVGKQILA...RFVSGITFGLAVG--LNIRQEQMIKSR...PDIVIATPGRFIDIIRNSAS----FNVDSVEIL	382
DED1	179	...RDLMACAQTSGKTLAFLLPVLSESFKTGPSPQPESQGSFYQRKAYPTAVIMAPTRELAIQIFDEAKKFT...YRSWVKACVVYGG--SPIGNCLREIERG...CDLLVATPGRLNDLLERGKIS----IANVKYL	303
DBP6	218	VSKRNFTRRIGDIIVNAATSGKTLAYSIPIVQTFKR...QINRLRCIIVPTKLLINQVYTTLTKLTQG...TSLIVSIAKLEN--SIKDEHKKISNL--EPDILITTPGRLVDHINMKSIN----IANIKFL	338
DBP2	150	...RDMVGIAATSGKTLSYCLPGIVHIN...AQPLLAPGDGPIVLVLAPTRELAVQIQTECSKFG...HSSRIRNTCVYGG--VPKSQCIRDISRG...SEIVIATPGRLIDMLEIGKTN...IKRVTYL	264
DBP1	191	...RDLMACAQTSGKTLGFLFPLFTEIFRSGPSPVPEKAQSFYSRKGYPSALVLAPTRELAIQIFEEARKFT...YRSWVRPCVVYGG--APIGNQMREVDRG...CDLLVATPGRLNDLLERGKVS...IANIKYL	315
DBP3	149	...KDVVGVAETSGKTFAFGVEAISHLMND...QKKRGIQVLVISPTRELAISIYDNLIVLT...DKVGMQCCCVYGG--VPKDECRIQLKK...SQVVVATPGRLLDLLQEGSVD...ISQVNYL	259
PRP28	214	...RDFIGVASTSGKTLAFVIPILIKMSRSPP...RPPSLKIIDGPKALILAPTRELVQQIQKETQKVTKIWSKESNYDCKVISIVGG--HSLEEISFSISEG...CDILVATPGRLIDSIENHLLV...ISQVETL	338
MAK5	208	...VDVLGKASTSGKTLAYGIPIVEKIISNFS...QKNKKPISLIFTPTRELAHQVTDHLKKICEP--VLAKSQYSILSLTGG--LSIQKCQRLIKYD-NSGQIVIATPGRFLELLEKDNTLI-KR-FSKVNTL	330
DBP7	184	...NDFFIHAQTSGKTISYLLPIISTILNMDT...HVDRTSGAFALVIAPTRELASQIYHVCSTLVS...CCHYLVPCLIIG--ERKKSEKARIRKG...CFIIGTPGRVLDHIQNTKVIK-EQLSQSIRYI	304
DBP10	174	...RDIVGMARTSGKTAAFILPMVEKIKSHSG...KIGARAVILSPSRELAMQTFNVFKDFAR...GTELRSVLLTGQ--PDFVIIATPGRFLDLIKVEMNLD...IKSVEVN	285
MRH4	155	...KLQIHAIAAETSGKTMAYLIPLIDYIKRQELETP--ELWETLRKNVLIRSIILVPTHELVDQVYETVSKTK...TLIGLNSFKWDKATSYRDLLENIKNR...IDILVTTPGRLLNIFSIRMITRPDKVLSKVGEV	284
MS116	145	...HDVIARAKTGTGKTFAFLIPIFQHLINTKF...DSQYMVKAVINAPTRDIALQIEAEVKKIHDMN--YGLKKYACVSIVGG--TDFRAAMNKVNKLR--PNIVIATPGRLIDVLEKYSNK...FFRFVDYK	264
RRP3	118	...HDIIGLAQTSGKTAAFAILNRIW-HDQ...EPYYA---CILAPTRELAQQIKETFDSIG...SLMG--VRSTCIVG--MNMMDCARDIMRK...PHIIIATPGRLMDHIENTKG----FSLRKIKEI	228
DBP4	78	...HDVIAAAKTSGKTLAFLVPVIEKLYREKW...TEFDGLGALIISPTRELAMQIYEVLTKIG...SHTS-FSAGLVIGG--KDVKFELERISR...IILLIGTPGRILQHLDQAVG----LNTSNIQML	191
HAS1	79	...RDVIGAAKTSGKTLAFLIEAIELIHSLKF...KPRRNGTGIIVLTPTRELAIQIFGVAREIM...EFHS-QTFGIVIGG--ANRRQEAEKIMKG...VNMLIATPGRLLDHIQNTKG----FVFKNLKAL	193
DHH1	83	...RDIIARAKNTGKTAAFVIETLEKVK...PKLNKIQALIMVPTRELALQTSQVVRTLG...KHC-GISCMVTTG...TNLRDDILRIN---ETVHILVGTPGRVLDIASRKVAD...ISDCSLF	192
SUB2	99	...TDVICQAKSILIKIAVIVISTLQQID...PVPGEVAVVVICNARELAYQIRNEYLRFS...KYMPDVKTAVFYG...TPISKDAELIKNKDTAPHIVVATPRLKAIVREKYID...ISHVKNF	212
IF4A	59	...HDVIAQAQSGTGKTATFAISILQQID...TSVKAPQALMLAPTRELAIQIQKVMALA...FHM-DIKVHACIGG...TSFVEDAEGLRDAQIVVGTPGRVFDNIQRRRFR...TDKIKMF	167
FAL1	60	...KDVIAQAQSGTGKTATFTIGLLQAID...LRKKDLQALILSPTRELASQIGQVVKNIG...DYM-NVNAFAITGG...KTIKDDLKKVQK--HGCQAVSGTPGRVLDMIKKQMLQ...TRNVQML	170
IF4A	59	...HDVIAQAQSGTGKTATFTIGLLQAID...TSVKAPQALMLAPTRELASQIQKVMALA...FHM-DIKVHACIGG...TSFVEDAEGLRDAQIVVGTPGRVFDNIQRRRFR...TDKIKMF	167
DBP9	55	...RDIIAKAATSGKTLAYLIPVIETILEYKKTID...NGEENGTLGIILVPTRELAQQVYNVIEKIVLY...CSKDIRTLNIISSD--MSDVLSTLIMDQ...PEIIVGTPGKLLDILQTKINS...ISINELKFL	176
SPB4	44	...KDVVVDSVTSGKTAAFVIPVLEKVVKEEAN...TSKFKKAHFHSLIIAPTRELSRQIESVLSFLEHYP-SDLFPIKCQLIVGTNEATVRDDVSNFLRN--RPQILIGTPGRVLDFLQMPAVK...TSACSMV	169
DBP8	39	...RDCIGGAKTSGKTIAFGPMLTKWS-EDP...SGMFG---VVLTPTRELAMQAEQFTALG...SSMN-IRVSVIVGG...ESIVQQCALDIQRK--PHFIIATPGRLAHHIMSSGDDT-VGGLMRAKYL	152

Motif I · Motif Ia · GG · Motif Ib

460 470 480 490 500 510 520 530 540 550 560 570 580 590 600

DBP5	237	VLDEADNMLDQQ-GLGDQCIRVKRELP...KDTQLVLFSATFADAVRQYAKKIVPANTIEIQ-TNEVNVDA...IKQLYMDCKNEADKFDV	320
PRP5	413	VMDEADRLFDLIEPQITQIMKTVR...PDKCCVLFSATFPNKLRSFAVRVLHSPISITINSKGMVNENVKQKFRICHSEDE...KFDNLVQLIHERSEFFDEVQSE	513
DRS1	383	VMDEADRMLEEIEQDEINEIMGILP...SNRQNLLFSATMNSKIKSLVSLSIKKPVRIMIDPPKKAAT...KL...TQEFVRIRKRDHLKPALL	467
DED1	304	VLDEADRMLDMIEPQITQIRHIVEDCDMTPV...GERQTLMFSATFPADIQHLARDFISDYLSVGRVG-STSEN...TTQKVIYVENQDKKSAL	390
DBP6	339	IIDEADRLLIQS--IEQGWCPKIMSHIKTDKLDTLP...GNVIKMIFSATLTTNTEKINGLNLYKPKLFLKQTDKLYQLPN...KINEFNINIPTAKSVYKPLILLYS	437
DBP2	265	VLDEADRMLDMIEPQIRKIVDQIRP...DRQTLMWSATWPKEVKQLAADYINDPIQVQVGSLELSASHN...ITCIVEVVSDFEKRDRLNK	350
DBP1	316	VLDEADRMLDMIEPQIRHIVEECDMPSV...ENRQTLMFSATFPVDIQHLARDFID-YIFISVGRVG-STSEN...ITCIVEVVYDDMDKKSAL	402
DBP3	260	VLDEADRMLEKIEEDIKNIIRETDAS...KRQTLMFSATWPKEVRELASTFYNIPIKVSIGNTDQLTANKR...ITCIVEVVDPRGKERKLLE	347
PRP28	339	VLDEADKMIDLIEDQVTNILTKVDINADS...AVNRQTLMFSATMTPVIEKIAAGYMQKPVYATIGVET--GSEPL...IQCVVEYADNDEDKFKK	427
MAK5	331	ILDEADRLLQDS-HIGDFEFEKIIKHLIVERRKNRENSEGSSKIWQTLIFSATFSIDIFDKLSSSRQVKDRRFKNNEDELNAVIQHLMSKIHFN...SKPVLIDTNPESKVSSQIKESLIECPPLER	451
DBP7	305	VLDEADKMLEIGDETIESIEIIKIVHDIPINSEKFPKLPHKLVHMLCSAILTDGVNRLRNVAIKDYKLISNGTKKDSDIVTVAP...DQILQRITIVPPKLRLVTLAATLNN...IT	414
DBP10	286	VFDEADRLFDMIEEQINELLASIP...TTRQTLLFSATLPNSIVDFVKAGIVNPVLVRLDAETKVSENLEMLFLSSKNADREANLLYILQEIIKIPLATSEQLQKLQNSNNEADSDSDDENDRQKKRRNFKKEKFRK	420
MRH4	285	VLDEADTLIDRS--IEELEETHSAIKRIP...NINHLIFCSATIPQEFNKTMQRLFPTVVPIMTPRLHKLPFALDFK...VINSAISPFKGSKIKALAQTLY	376
MS116	265	VLDEADRLIEICIERDDIETISGIINEKNSKS...ADNIKTLLFSATLDDKVQKLANNIYNKKECLFIDTVDKNEPEAHER...IDCSVVISEKFANSIFAAVEH	362
RRP3	229	VMDEADRLLDME--GPVIDRIIKIIP-T...QERTYLLFSATMTSKIDKLQRASITPVKCAVSNKYQTVD...TIVCTLMVVPGGLKNTYL	312
DBP4	192	VLDEADRCLDMIKKTIDAIVSTIS...PSRQTLLFSATQSQSVADILARLSITDYKTYGTHDVMDGSVNKEASTPE...TLQQFYIEVPLADKLDIL	282
HAS1	194	IIDEADRILDICIEEDEMRQIIKILPN...EDRQSMLFSAIQTTKVEDLARISIRPGPLF...INVVPETDNSTAD...GLEQGYIVCDSDKRFLLL	280
DHH1	193	IMDEADKMLSR--DIKTIEQIILSFLP...PTHQSLLFSAIFPLTVKEFMVKHLHKEYEINI--MEELTLKG...ITQYYAFVE-EROKLHC	273
SUB2	213	VIDECDKVLEEL-DMRRDVQEIFRATP...RDKQVMMSAILSQEIRPICRRFLQPIEIFVDDEAKLTLHG...IQQYYIKIE-EREKNRK	296
IF4A	168	ILDEADEMLSS--GFKEQIYQIFTLLP...PTTQVVLLSATMPNDVLEVTTKFMRNPVRILVKDEELTLEG...IKQFYVNVEEEEYKYEC	250
FAL1	171	VLDEADELLSETLG-FKQQIYDIFAKLP...KNCQVVVVSATNKDILEVTRKFYNDPVKILVK-RDEISLEG...IKQYVVNVDKEEWKFDT	255
IF4A	168	ILDEADEMLSS--GFKEQIYQIFTLLP...PTTQVVLLSATMPNDVLEVTTKFMRNPVRILVKDEELTLEG...IKQFYVNVEEEEYKYEC	250
DBP9	177	VVDEVDLVLFTFIEQDDINKIGEYLPLK...KNLQFIMSAINDDIQAIKQKFCRSPAIIKFNDEEINKNQNK...ILQYYKVSEFDKFLLC	263
SPB4	170	VMDEADRLLDMS--IKDTEKILRILP...KQRRIGLFSATMRSAGSDIFKTGRIVPVRITVNSKNQAPSS...IKLNYCVVNPAEKLQLL	252
DBP8	153	VLDEADILITST--ADHIATCISALPPK...DKRQTLLFTATIITDQVKSLQNAPVQKGKPPLFAYQVSVDNVAI--PS...TIKIEYTLVPEHVKEAYL	243

Motif II · Motif III

610 620 630 640 650 660 670 680 690 700 710 720 730 740 750

```
DBP5   321 LTELYGLMTIG--SSIIFVATKKTANVLYGKLK-SEG-----------------------------------------------HEVSILHGDLQTQERDRLIDDLR--------------------EGRSKVLITTN------------------------------------ 388
PRP5   514 NDGQSSDVEEVDAKAIIFVSSQNICDFISKKLL-NAG-----------------------------------------------IVTCAIHAGKPYQERLMNLEKLK--------------------REKNSILLCIE------------------------------------ 583
DRS1   468 FNLIRKLDPTGQKRIVVFVARKETAHRLRIIMG-LLG-----------------------------------------------MSVGELHGSLTQEQRLDSVNKLK--------------------NLEVPVLICTD------------------------------------ 537
DED1   390 -LDILSA--STDGLTLIFVETKRMADQLTDFLI-MQN-----------------------------------------------FRATAIHGDRTQSRRERAILAALR--------------------SGAATLLVATA------------------------------------ 457
DBP6   437 -ICQFMAHSPIAAKILIFVKSNESSIRLSKLLQLICESRSQSSVLKN--------------LQNLAVSINSVSNNSKAENKKIVANSSHHS--------------------ESAGITILITD------------------------------------ 526
DBP2   351 YLETASQ--DNEYKTLIFASTKRMCDDITKYLR-EDG-----------------------------------------------WPALAIHFDKDQERDWVIQELR--------------------NGRSPIMVATD------------------------------------ 418
DBP1   402 -LDILSA--EHKGLTLIFVETKRMADQLTDFLI-MQN-----------------------------------------------FKATAIHGDRTQAERERAILSALK--------------------ANVADILVATA------------------------------------ 469
DBP3   348 LLKKYHSGPKKNEKVLIEALYKKEAARVERNLK-YNG-----------------------------------------------YNVAAIHGDLSQQQRTQAIFNELK--------------------SGKSNILLATD------------------------------------ 417
PRP28  427 -LKPIVA--KYDPPIIIFILYKQTADWLAEKFQ-KETN----------------------------------------------MKVTILHGSKSQEQREHSIQLLR--------------------TNKVQIMIATN------------------------------------ 495
MAK5   452 DLYCYYFLTMFPGTTLIFCAIDSVKKLTVYLN-NLG-----------------------------------------------IPAFQIHSSMTOKNRLKSLERKKQQSAKQKTINHSNPDSVQLSTVLIASD------------------------------------ 537
DBP7   415 KDFILASGQQSKTLRTIVFVSCSDSVEFHYDAFSGSDGHHKNLTGDSVRLLTKGNTMFPCFSDSRDPDVVLYKLHGSLSQQMRTSTLQHGARD----------NEATKGKHLIMFCTD------------------------------------ 521
DBP10  421 QKMPAANELPSEKATILFVPTRHHVEYISQLLR-DCG-----------------------------------------------YLISYIYGTLDHARKRQLYNFR--------------------AGLTSILVVTD------------------------------------ 490
MRH4   377 AISNDDTEPGFEKRCIIFVEKKNVPEIVNLLNKKFG-----------------------------------------------HNAIGLTFEDTFERSEKIMPGLSPP--------------------RPLSEVVAQSTSP------------------------------- 452
MS116  362 -IKKQIKERDSNYKAIIEAPTVKFTSFLCSILKNEFKK-----------------------------------------------DLPILEFHKITQNKRTSLVKRFK--------------------KDESGILVCTD------------------------------------ 434
RRP3   313 IYLLN--EFIG-KTMIIFTRTKANAERLSGLCN-LLE-----------------------------------------------FSATALHGDLNQNQRMGSIDLLK--------------------AGKRSILVATD------------------------------------ 379
DBP4   283 FSFIKSHLKCK---MIVFLSSSKQVHFVYETFR-KMQP----------------------------------------------GISIMHLHGRQKQRARTETIDKLN--------------------RAQQVCLFATD------------------------------------ 351
HAS1   281 FSFIKRNQKKK---IIVFLSSCNSVKYYAELLN-YID-----------------------------------------------LPVMLELHGSKSQEQKQKQQKRTNTFFEC--------------------NAERGILICTD------------------------------------ 347
DHH1   274 LNTIFSKLQIN--QAIIFCNSTNRVELLAKKIT-DLG-----------------------------------------------YSCYYSHARMKQQERNKVFHELR--------------------QGKVRTLVCSD------------------------------------ 341
SUB2   297 LAQLLDDLEFN--QVIIFVKSTTRANELTKLLN-ASN-----------------------------------------------FPAITVHGHMKQEERIARYKALK--------------------DFEKRICVSTD------------------------------------ 364
IF4A   251 LTDLYDSISVT--QAVIFCNTRRKVEELTTKLR-NDK-----------------------------------------------FTVSAIYSDIPQQERDTIVKEFR--------------------SGSSRILISTD------------------------------------ 318
FAL1   256 LCDIYDSLTIT--QCVIFCNTKKKVDWISQRLI-QSN-----------------------------------------------FAVVSMHGNRKQRVNDLR--------------------TGHSRVLISTD------------------------------------ 323
IF4A   251 LTDLYDSISVT--QAVIFCNTRRKVEELTTKLR-NDK-----------------------------------------------FTVSAIYSDIPQQERDTIVKEFR--------------------SGSSRILISTD------------------------------------ 318
DBP9   263 --YVIFKLNLIKGKTLIFVNIDRGYRLKLVME-QFG-----------------------------------------------IKSCILNSEIPVNSRQHIVDQLN---------------------KNVYQLLIATDDTEYIKEEDDEIEEGHNTENQEEKSLEGEPEND 364
SPB4   252 ---VSILNNYKFKKCIVFPTCVSVSYFYSFIQYLGKRNILVN-----------------------------------------EVEIFSLHGKLQTSARTKTITAFT--------------------DSLSNSVLFTT------------------------------------ 327
DBP8   244 YQLITCEEYEN-KTAIIFVGRTMTAEILRRTLK-QLE-----------------------------------------------VRVASLHSQMPQQERTNSIHRLR--------------------ANAARILIATD------------------------------------ 312
                        Motif IV                                                                         QxxR
```

```
              760       770       780       790       800       810       820       830       840       850       860       870       880       890       900
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

DBP5   388 ------------------VLARLIDIPTVSMVVNYDLPTLANGQADPATYIHRIRTGRFGR-KLVAISFVHDKNSFN------ILSAIGKYFGDIEMTRVP----TDDWDEVEKIVKKVLKD------------------------- 482
PRP5   583 ------------------VLSRELNVPEVSLVIIYNAVKTFA------QYVHTTRTARGSR-STAITLLHDELSGAYILSKAVRDEEIKALDPLQAKELQEMSAKFESGMKKGKFRLSKGFGGKGLENIKSKREEAQNKDLELKKNDK 709
DRS1   537 ------------------LASRLIDIPKIEVVINYDMPKSYE------IYLHRVRTARAFR-ERSVTFVGESSQDRSIVR-AAIKSVEENKSLT-QGKAL----GRNVDWVQIEETNKLVESMNDTIEDILVEEKEEKEILRAEMQLRK 657
DED1   457 ------------------VAARLIDINDITQVINYDPPMSQ------DYVHRIRTGRAN-TLSAYNLVGRGERT----FFDDLNKDLDRDGKSVQP-------LELDFTLLESDSELYTSSLESLKNYHNNTAQA---------- 629
DBP6   526 ------------------IMSRLIDINDITQVINYDPPMSSQ------QYVHRVRTARANE-LSAYNLVGRGERT----FFDDLNKDLDRDGKSVQP-------LELDFTLLESDSELYTSSLESLKNYHNNTAQA---------- 629
DBP2   418 ------------------VAARLIDVKGINYVINYDMPGYE------DYVHRIRTGRA-A-TTAISFFTEQ-NKGLGA---KLISIMREANQNIPPEL----KYDRRSYGGGHPRYGGGRGGRGGYG-RGGGYGGGR--GGYGGNRQ 533
DBP1   469 ------------------VAARLIDPNVTHVINYDLPSDTD------DYVHRIRTGRA-N-TVATSFFNSN-NQNIVK--GLMEILNEANQEVPTFLS----DLSRQNSRGGRTRGGGGFFNSRNNGSRDYRKHGGN--GSFGSTRP 585
DBP3   417 ------------------VAARLIDIPNVKTVLTFPLTVE------DYVHRIRTGRAQ-TTAHTLFTEQ-EKHLAG---GTVNVLNGANQPVPEDTI----KFGTHTKKKEHSAYG--SFFKDVDLTKKPKKITFD--------- 523
PRP28  495 ------------------VAARLIDIPENVSLVVNQISKKMD------DYIHRIRTGRAAN-ETAVSFVSAAEDESLIR---ELYKYVRKHDPLNSNIFS----EAVKNKYNVGKQLSNEIIY---------------------- 588
MAK5   537 ------------------VAARFVDIPGVQEVIHYHLPRSTD------IYIHRSRTARA-S-EVSAMICSPQESMGPLRKLRKTLATKNSVSTDLNSRSTNRKPIKWQNTVPLLPIETDILSQLRERSRLAGELADHEIASNSLRKDDN 663
DBP7   521 ------------------VASRLIDLPHVGSVIELDPPFAVE------DHLHRVRTARA-E-KESLLELPGEEEKYMDYIQPYHPMGWELLKFDKEILMPAFKDVNVNRNDKFIRKDEKSSKNKDVGDKEYEWDTNATTWHLNIERRV 647
DBP10  490 ------------------VAARFVDIPMLANVNVTLPGSSK------IFVHRVRTARA-N-KWAYSIVAENELPYLLDL-ELFLGKKILLTPMYDSLVD----VMKKRWIDEGKPEYQFQPPKLSYTKRLVLGSCPRLDVEGLGDLYK 611
MRH4   452 ------------------PTSLKKFEIPDSNIVIGKLKNT-SN------------TAPSNKSLHVLVTLDMARGLNFKGVRNVVLYDVPKTSIDLIHRVGR------TARMKQGGRVFMLTDSKTKSWAKALPKIIKKHQRLS------- 561
MS116  434 ------------------VGARLMDFPNVHEVLQIGVPSELAN----------YIHRIRTARS-K-ESSVLFICKDELPFVREL-EDAKNIVIAKQEKYEPSEE----IKSEVLEAVTEEPEDISDIVISLISSYRSCIKEYRFSERRILPEI 555
RRP3   379 ------------------VAARLIDIPSVDIVVNYDIPVDSK------SYIHRVRTARA-R-SKSISLVSQYDLEL-------TLRIEEV-----LGKKL----PKE--SVDKNIILTLRDSVDKANGEVVMEMNRRNKEKIARGKGRR 487
DBP4   351 ------------------VARRLIDFPAVDWVQVDCPEDVD------TYIHRVRCARYLK-KLKSLIMITPQEQEAFLKR-LNARKIEPGKLNIKQSKKK----SIKPQLQSLLFKDPELKYLGQKAFISYVRSIYVQKDKQVFKFDEL 472
HAS1   347 ------------------VAARLIDIPEAVDWIIQCDPPDDPR----DYIHRVRTARGTKGK-KSLMFITPN-ELGFLRY-LKASKVPLNEYEFPENK----NVQSQLEKLIKSNYYLHQTAKDGYRSYLQAYASHSLKTVYQIDKL 468
DHH1   341 ------------------LLIRLIDIQAVNVVINFDFPKTAE------TYLHRIRSGRF H-LLAINLIN-WNDRF------NLYKIEQEL-GTEIAALP----ATIDKSLYVAENDETVPVPFPIEQQSYHQQAIPQQQLPSQQQFAI 455
SUB2   364 ------------------VFGRLIDIERINLAINYDLTNEAD------QYLHRVRAGRFLT-KLAISFVSSKEDEE----VPAKIGERF-DVKIAEFP----EEGIDPSTYLNN------------------------------- 446
IF4A   318 ------------------LLARLIDVQQVSLVINYDLPANKE------NYIHRIGGRFGR-KVAINFVT-NEDVG------AVRELEKFY-STQIEEFP----SDIATLLN------------------------------------ 395
FAL1   323 ------------------WARLIDVQQVSLVINYDLPEIIE------NYIHRIRGGRFGR-KLAINFVT-KADLA------KURELEKFY-SIKINPVP----ANFAELS------------------------------------ 399
IF4A   318 ------------------LLARLIDVQQVSLVINYDLPANKE------NYIHRIGGRFGR-KVAINFVT-NEDVG------AVRELEKFY-STQIEEFP----SDIATLLN------------------------------------ 395
DBP9   365 KKPSKKKVQVKKDKEYGVSRFVDFKNVACVLNFDLPTTAK------SYHRVRTARGGK-TTAISFVVPLKEFGKHKP--SVLQTAKKDERILSRILK----QQSKLGLELQPYKFDQKQVEAFRYRMEDGFRAVTQVAIREARVKE 501
SPB4   327 ------------------DVAARIDIDVDLVIQLDPPTNTD------MFMHRCRTGRANFV-KAITINEGREEDFIPF-MQVKNVELEELDLEVKGTTNFYEDFRNWILEDRDRFDKGVKAYVAFIKYYSNHSATSIFRLQSLDYV 453
DBP8   312 ------------------VASRLIDIBTVELVVNYDIPSDPD------VFIHRSRTARA-R-IDAISFVTQRDVSR------LQAIEDR-----INKKM----TETNKVHDTAVIRKALTKVTKAKRESLMAMQ---KENFGERK-RQ 418
                          Motif V                         Motif VI
```

```
              910       920       930       940       950       960       970       980       990       1000      1010      1020      1030      1040      1050
              ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
```

```
DBP5    482  ----------------------------------------------------------------------------------------------------------------------------------------------  482
PRP5    710  RSDDLEKKISNPREGHDSVSESSALIPRLNYELFKESTDGSIIFYAKVYINDLPQIVRWEATKNTTLLFIKHETGCSITNKGKFYPEGKEPKNENDEPKLYLLIEGQDEKDIQLSIELLEQKVKEGVVKAASLSLKSTKY----------  849
DRS1    658  GENMLKHKKEIQARPRRTWFQSESDKKNSKVLG---------------ALSRNKKVTNSKKRKREEAKADGNGARSYRKTKTDRIADQERTFKKQKSTNSNKKKGFKSRR-------------------------------------  752
DED1    573  RDN--SFRG--GSGWGSDSKSSGWGNSGGSNNSSWW------------------------------------------------------------------------------------------------------------  604
DBP6    629  ----------------------------------------------------------------------------------------------------------------------------------------------  629
DBP2    534  RDG----------GWGNRGRSNY------------------------------------------------------------------------------------------------------------------------  546
DBP1    586  RNTGTSNWGSIGGGFRNDNEKNGYGNS----NASWW----------------------------------------------------------------------------------------------------------  617
DBP3    523  ----------------------------------------------------------------------------------------------------------------------------------------------  523
PRP28   588  ----------------------------------------------------------------------------------------------------------------------------------------------  588
MAK5    664  WLKKAADELGIDVDSDEDDISKSNSDTFLLKNKNKKMQKTINKDKVKAMRATLNELLSVPIRKDRRQKYLTGGLVNLADNLVKKRGHNSIIGHEKTNALETLKKKKKRNN-------------------------------------  773
DBP7    648  VGDSAFKNLAVKGFISHVRAYATHISQEKKFFNVKFLHLGHLAKSFGLRERPKAMGLQSSKDGNSEKKPTKENSKNKMFRMARMAEKQIASEFNY-------------------------------------------------  742
DBP10   612  NLMSSNFDLQLAKKTAMKAEKLYYRTRTSASPESLKRSKEIISSGWDAQNAFFGKNEEKEKLDFLAKLQNRRNKETVFEFTRNPDDEMAVFMKRRRKQLAPIQRKATERRELLEKERMAGLHILSKMKFWKGDDVETGYTVSEDALKEFE  761
MRH4    561  ----------------------------------------------------------------------------------------------------------------------------------------------  561
MS116   556  ASTYGVLLNDPQLKIPVSRRFLDKLGLSRSPIGKAMFEIRDYSSRDGNNKSYDYDDDSEISFRGNKNYNNRSQNRDYDDEPFRRSNNNRRSFSRSNDKNNYSSRNSNIY-----------------------------------  664
RRP3    488  GRMMTRENMDMGER--------------------------------------------------------------------------------------------------------------------------------  501
DBP4    473  PTEEFAYSLGLPGAPKIKMKGMKTIEQAKERKNAPRQLAFLSKANEDGEVIEDKSKQPRTKYDKMFERKNQTILSEHYLNITKAQAQEDEDDDFISVKRKDHEINEAELPALTLPTSRR----AQKKALSKKASLASKGNASKLIFDDEG  618
HAS1    469  DLAKVAKSYGFPVPPKVNIT-----------------------------IGASGKTPNTK---------------------------------RRKTHK-------------------------------------------  505
DHH1    456  PPQQHHPQFMVPPSHQQQQAYPPPQMPSQQGYPPQQEHFMAMPPGQSQPQY--------------------------------------------------------------------------------------------  506
SUB2    446  ----------------------------------------------------------------------------------------------------------------------------------------------  446
IF4A    395  ----------------------------------------------------------------------------------------------------------------------------------------------  395
FAL1    399  ----------------------------------------------------------------------------------------------------------------------------------------------  399
IF4A    395  ----------------------------------------------------------------------------------------------------------------------------------------------  395
DBP9    502  LKQELLASEKLKRHFEENPKELQSLRHDKELHPARVQQHLKRVPDYLLPESARGNGTKVKFVPFHNAKKRHSHKKGRVSKPKNGKVDPLKNFK----------------------------------------------------  594
SPB4    454  GIAKLYGLFRLPRMPEITKYLATEKQEGIFPGNWLVDPPVNMDEYKYKDKKREKERQETLKNISLINDKKKLKSELKKKNLAWSDKTLTKERKLERKEKMSLKRKAIEEELKAEELDENAEEERIKEDWKEIVLQNKRKKVSSKAIQGNF  603
DBP8    419  QKKKQNDGKSLRS-------------------------------------------------------------------------------------------------------------------------------  431
```

```
                 1060      1070      1080      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
DBP5    482  ----------------------------------------------------------------------------------------------------------------------------------------------  482
PRP5    849  ----------------------------------------------------------------------------------------------------------------------------------------------  849
DRS1    752  ----------------------------------------------------------------------------------------------------------------------------------------------  752
DED1    604  ----------------------------------------------------------------------------------------------------------------------------------------------  604
DBP6    629  ----------------------------------------------------------------------------------------------------------------------------------------------  629
DBP2    546  ----------------------------------------------------------------------------------------------------------------------------------------------  546
DBP1    617  ----------------------------------------------------------------------------------------------------------------------------------------------  617
DBP3    523  ----------------------------------------------------------------------------------------------------------------------------------------------  523
PRP28   588  ----------------------------------------------------------------------------------------------------------------------------------------------  588
MAK5    773  ----------------------------------------------------------------------------------------------------------------------------------------------  773
DBP7    742  ----------------------------------------------------------------------------------------------------------------------------------------------  742
DBP10   762  DAHQLLEAQENENKKKKKPKSFKDPTFFLSHYAPAGDIQDKQLQITNGFANDAAQAAYDLNSDDKVQVHKQTATVKWDKKRKKYVNTQGIDNKKYIIGESGQKIAASFRSGRFDDWSKARNLKPLKVGSRETSIPSNLLEDPSQGPAANG  911
MRH4    561  ----------------------------------------------------------------------------------------------------------------------------------------------  561
MS116   664  ----------------------------------------------------------------------------------------------------------------------------------------------  664
RRP3    501  ----------------------------------------------------------------------------------------------------------------------------------------------  501
DBP4    619  EAHPVYELEDEEEFHKRGDAEVQKTEFLTKESAVMADIDNIDKQVAKEKKQEKKRKRLEAMRREMEAAMEEEISGDEEEGKTVAYLGTGNLSDDMSDGDMPDSEGHLKKKARTVDYSHGHNPSNSVDDDIIEVEEPQTLEDLESLTAKLI  768
HAS1    505  ----------------------------------------------------------------------------------------------------------------------------------------------  505
DHH1    506  ----------------------------------------------------------------------------------------------------------------------------------------------  506
SUB2    446  ----------------------------------------------------------------------------------------------------------------------------------------------  446
IF4A    395  ----------------------------------------------------------------------------------------------------------------------------------------------  395
FAL1    399  ----------------------------------------------------------------------------------------------------------------------------------------------  399
IF4A    395  ----------------------------------------------------------------------------------------------------------------------------------------------  395
DBP9    594  ----------------------------------------------------------------------------------------------------------------------------------------------  594
SPB4    604  DDL-------------------------------------------------------------------------------------------------------------------------------------------  606
DBP8    431  ----------------------------------------------------------------------------------------------------------------------------------------------  431
```

```
                 1210      1220      1230      1240      1250      1260      1270      1280
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
```

```
DBP5    482  ------------------------------------------------------------------  482
PRP5    849  ------------------------------------------------------------------  849
DRS1    752  ------------------------------------------------------------------  752
DED1    604  ------------------------------------------------------------------  604
DBP6    629  ------------------------------------------------------------------  629
DBP2    546  ------------------------------------------------------------------  546
DBP1    617  ------------------------------------------------------------------  617
DBP3    523  ------------------------------------------------------------------  523
PRP28   588  ------------------------------------------------------------------  588
MAK5    773  ------------------------------------------------------------------  773
DBP7    742  ------------------------------------------------------------------  742
DBP10   912  RTVRGKFKHKQMKAPKMPDKHRDNYYSQKKKVEKALQSGISVKGYNNAPGLRSELKSTEQIRKDRIIAEKKRAKNARPSKKRKF  995
MRH4    561  ------------------------------------------------------------------  561
MS116   664  ------------------------------------------------------------------  664
RRP3    501  ------------------------------------------------------------------  501
DBP4    769  QG--------------------------------------------------------------  770
HAS1    505  ------------------------------------------------------------------  505
DHH1    506  ------------------------------------------------------------------  506
SUB2    446  ------------------------------------------------------------------  446
IF4A    395  ------------------------------------------------------------------  395
FAL1    399  ------------------------------------------------------------------  399
IF4A    395  ------------------------------------------------------------------  395
DBP9    594  ------------------------------------------------------------------  594
SPB4    606  ------------------------------------------------------------------  606
DBP8    431  ------------------------------------------------------------------  431
```

6.6 Suppliers

Supplier	Location	Website
Art Robbins Instruments	Sunnyvale, USA	http://www.artrobbinsinstruments.com
Avestin	Mannheim, Germany	http://www.avestin.com/
Beckman Coulter	Fullerton, USA	http://www.beckman.com
Biorad	Hercules, USA	http://www.bio-rad.com
Dharmacon	Lafayette, USA	http://www.dharmacon.com
Dynal	Karlsruhe, Germany	http://www.invitrogen.com/site/us/en/home/brands/Dynal.html
Edmund Bühler	Hechingen, Germany	http://www.edmund-buehler.de/
Eppendorf	Wesseling-Berzdorf, Germany	http://www.eppendorf.de
EUROSCARF	Frankfurt, Germany	http://web.uni-frankfurt.de/fb15/mikro/euroscarf
Finnzymes	Espoo, Finland	http://www.finnzymes.fi
Fluka	Hamburg, Germany	http://www.sigmaaldrich.com/analytical-chromatography.html
GE Healthcare	München, Germany	http://www.gehealthcare.com
Hampton	Aliso Viejo, USA	http://hamptonresearch.com
Invitrogen	Karlsruhe, Germany	http://www.invitrogen.com
Jena Bioscience	Jena, Germany	http://www.jenabioscience.com
Kühner	Birsfelden, Switzerland	http://www.kuhner.com
Memert	Schwabach , Germany	http://www.memmert.com
Molecular Dimensions	Suffolk, UK	http://www.moleculardimensions.com/
New England Biolabs	Ipswich, USA	http://www.neb.com
Novagen	San Diego, USA	http://www.emdbiosciences.com/html/NVG/
Oxford Diffraction	Oxfordshire, UK	http://www.oxford-diffraction.com/
Packard	Waltham, USA	http://www.perkinelmer.de/
Peqlab	Erlangen, Germany	http://www.peqlab.de
Promega	Mannheim, Germany	http://www.promega.com
Qiagen	Hilden, Germany	http://www.qiagen.com
Sartorius	Goettingen, Germany	http://www.sartorius.de
Schott	Mainz, Germany	http://www.schott.com
Sigma-Aldrich	Hamburg, Germany	http://www.sigmaaldrich.com
Sorvall	Waltham, USA	http://www.thermo.com/com/cda/landingpage/0,,128,00.html
Stratagene	La Jolla, USA	http://www.stratagene.com
SwissCI	Zug, Switzerland	http://www.swissci.com/

6.7 Protein sequences

Human Dbp5 (UniprotKB/Swiss-prot entry Q9UMR2)

```
MATDSWALAVDEQEAAAESLSNLHLKEEKIKPDTNGAVVKTNANAEKTDEEEKEDRAAQSLLNKLIRSNLVDNTN
QVEVLQRDPNSPLYSVKSFEELRLKPQLLQGVYAMGFNRPSKIQENALPLMLAEPPQNLIAQSQSGTGKTAAFVL
AMLSQVEPANKYPQCLCLSPTYELALQTGKVIEQMGKFYPELKLAYAVRGNKLERGQKISEQIVIGTPGTVLDWC
SKLKFIDPKKIKVFVLDEADVMIATQGHQDQSIRIQRMLPRNCQMLLFSATFEDSVWKFAQKVVPDPNVIKLKRE
EETLDTIKQYYVLCSSRDEKFQALCNLYGAITIAQAMIFCHTRKTASWLAAELSKEGHQVALLSGEMMVEQRAAV
IERFREGKEKVLVTTNVCARGIDVEQVSVVINFDLPVDKDGNPDNETYLHRIGRTGRFGKRGLAVNMVDSKHSMN
ILNRIQEHFNKKIERLDTDDLDEIEKIAN
```

Human Nup214 (UniprotKB/Swiss-prot entry P35658)

```
MGDEMDAMIPEREMKDFQFRALKKVRIFDSPEELPKERSSLLAVSNKYGLVFAGGASGLQIFPTKNLLIQNKPGD
DPNKIVDKVQGLLVPMKFPIHHLALSCDNLTLSACMMSSEYGSIIAFFDVRTFSNEAKQQKRPFAYHKLLKDAGG
MVIDMKWNPTVPSMVAVCLADGSIAVLQVTETVKVCATLPSTVAVTSVCWSPKGKQLAVGKQNGTVVQYLPTLQE
KKVIPCPPFYESDHPVRVLDVLWIGTYVFAIVYAAADGTLETSPDVVMALLPKKEEKHPEIFVNFMEPCYGSCTE
RQHHYYLSYIEEWDLVLAASAASTEVSILARQSDQINWESWLLEDSSRAELPVTDKSDDSLPMGVVVDYTNQVEI
TISDEKTLPPAPVLMLLSTDGVLCPFYMINQNPGVKSLIKTPERLSLEGERQPKSPGSTPTTPTSSQAPQKLDAS
AAAAPASLPPSSPAAPIATFSLLPAGGAPTVFSFGSSSLKSSATVTGEPPSYSSGSDSSKAAPGPGPSTFSFVPP
SKASLAPTPAASPVAPSAASFSFGSSGFKPTLESTPVPSVSAPNIAMKPSFPPSTSAVKVNLSEKFTAAATSTPV
SSSQSAPPMSPFSSASKPAASGPLSHPTPLSAPPSSVPLKSSVLPSPSGRSAQGSSSPVPSMVQKSPRITPPAAK
PGSPQAKSLQPAVAEKQGHQWKDSDPVMAGIGEEIAHFQKELEELKARTSKACFQVGTSEEMKMLRTESDDLHTF
LLEIKETTESLHGDISSLKTTLLEGFAGVEEAREQNERNRDSGYLHLLYKRPLDPKSEAQLQEIRRLHQYVKFAV
QDVNDVLDLEWDQHLEQKKKQRHLLVPERETLFNTLANNREIINQQRKRLNHLVDSLQQLRLYKQTSLWSLSSAV
PSQSSIHSFDSDLESLCNALLKTTIESHTKSLPKVPAKLSPMKQAQLRNFLAKRKTPPVRSTAPASLSRSAFLSQ
RYYEDLDEVSSTSSVSQSLESEDARTSCKDDEAVVQAPRHAPVVRTPSIQPSLLPHAAPFAKSHLVHGSSPGVMG
TSVATSASKIIPQGADSTMLATKTVKHGAPSPSHPISAPQAAAAAALRRQMASQAPAVNTLTESTLKNVPQVVNV
QELKNNPATPSTAMGSSVPYSTAKTPHPVLTPVAANQAKQGSLINSLKPSGPTPASGQLSSGDKASGTAKIETAV
TSTPSASGQFSKPFSFSPSGTGFNFGIITPTPSSNFTAAQGATPSTKESSQPDAFSSGGGSKPSYEAIPESSPPS
GITSASNTTPGEPAASSSRPVAPSGTALSTTSSKLETPPSKLGELLFPSSLAGETLGSFSGLRVGQADDSTKPTN
KASSTSLTSTQPTKTSGVPSGFNFTAPPVLGKHTEPPVTSSATTTSVAPPAATSTSSTAVFGSLPVTSAGSSGVI
SFGGTSLSAGKTSFSFGSQQTNSTVPPSAPPPTTAATPLPTSFPTLSFGSLLSSATTPSLPMSAGRSTEEATSSA
LPEKPGDSEVSASAASLLEEQQSAQLPQAPPQTSDSVKKEPVLAQPAVSNSGTAASSTSLVALSAEATPATTGVP
DARTEAVPPASSFSVPGQTAVTAAAISSAGPVAVETSSTPIASSTTSIVAPGPSAEAAAFGTVTSGGSSVFAQPPA
ASSSSAFNQLTNNTATAPSATPVFGQVAASTAPSLFGQQTGSTASTAAATPQVSSSGFSSPAFGTTAPGVFGQTT
FGQASVFGQSASSAASVFSFSQPGFSSVPAFGQPASSTPTSTSGSVFGAASSTSSSSSFSFGQSSPNTGGGLFGQ
SNAPAFGQSPGFGQGGSVFGGTSAATTTAATSGFSFCQASGFGSSNTGSVFGQAASTGGIVFGQQSSSSSGSVFG
SGNTGRGGGFFSGLGGKPSQDAANKNPFSSASGGFGSTATSNTSNLFGNSGAKTFGGFASSSFGEQKPTGTFSSG
GGSVASQGFGFSSPNKTGGFGAAPVFGSPPTFGGSPGFGGVPAFGSAPAFTSPLGSTGGKVFGEGTAAASAGGFG
FGSSSNTTSFGTLASQNAPTFGSLSQQTSGFGTQSSGFSGFGSGTGGFSFGSNNSSVQGFGGWRS
```

Yeast Dbp5 (UniprotKB/Swiss-prot entry P20449)

```
MSDTKRDPADLLASLKIDNEKEDTSEVSTKETVKSQPEKTADSIKPAEKLVPKVEEKKTKQEDSNLISSEYEVKV
KLADIQADPNSPLYSAKSFDELGLAPELLKGIYAMKFQKPSKIQERALPLLLHNPPRNMIAQSQSGTGKTAAFSL
TMLTRVNPEDASPQAICLAPSRELARQTLEVVQEMGKFTKITSQLIVPDSFEKNKQINAQVIVGTPGTVLDLMRR
KLMQLQKIKIFVLDEADNMLDQQGLGDQCIRVKRFLPKDTQLVLFSATFADAVRQYAKKIVPNANTLELQTNEVN
VDAIKQLYMDCKNEADKFDVLTELYGLMTIGSSIIFVATKKTANVLYGKLKSEGHEVSILHGDLQTQERDRLIDD
FREGRSKVLITTNVLARGIDIPTVSMVVNYDLPTLANGQADPATYIHRIGRTGRFGRKGVAISFVHDKNSFNILS
AIQKYFGDIEMTRVPTDDWDEVEKIVKKVLKD
```

Yeast Nup159 (Swiss-prot entryP40477)

```
MSSLKDEVPTETSEDFGFKFLGQKQILPSFNEKLPFASLQNLDISNSKSLFVAASGSKAVVGELQLLRDHITSDS
TPLTFKWEKEIPDVIFVCFHGDQVLVSTRNALYSLDLEELSEFRTVTSFEKPVFQLKNVNNTLVILNSVNDLSAL
DLRTKSTKQLAQNVTSFDVTNSQLAVLLKDRSFQSFAWRNGEMEKQFEFSLPSELEELPVEEYSPLSVTILSPQD
FLAVFGNVISETDDEVSYDQKMYIIKHIDGSASFQETFDITPPFGQIVRFPYMYKVTLSGLIEPDANVNVLASSC
SSEVSIWDSKQVIEPSQDSERAVLPISEETDKDTNPIGVAVDVVTSGTILEPCSGVDTIERLPLVYILNNEGSLQ
IVGLFHVAAIKSGHYSINLESLEHEKSLSPTSEKIPIAGQEQEEKKKNNESSKALSENPFTSANTSGFTFLKTQP
AAANSLQSQSSSTFGAPSFGSSAFKIDLPSVSSTSTGVASSEQDATDPASAKPVFGKPAFGAIAKEPSTSEYAFG
KPSFGAPSFGSGKSSVESPASGSAFGKPSFGTPSFGSGNSSVEPPASGSAFGKPSFGTPSFGSGNSSAEPPASGS
AFGKPSFGTSAFGTASSNETNSGSIFGKAAFGSSSFAPANNELFGSNFTISKPTVDSPKEVDSTSPFPSSGDQSE
```

```
DESKSDVDSSSTPFGTKPNTSTKPKTNAFDFGSSSFGSGFSKALESVGSDTTFKFGTQASPFSSQLGNKSPFSSF
TKDDTENGSLSKGSTSEINDDNEEHESNGPNVSGNDLTDSTVEQTSSTRLPETPSDEDGEVVEEEAQKSPIGKLT
ETIKKSANIDMAGLKNPVFGNHVKAKSESPFSAFATNITKPSSTTPAFSFGNSTMNKSNTSTVSPMEEADTKETS
EKGPITLKSVENPFLPAKEERTGESSKKDHNDDPKDGYVSGSEISVRTSESAFDTTANEEIPKSQDVNNHEKSET
DPKYSQHAVVDHDNKSKEMNETSKNNERSGQPNHGVQGDGIALKKDNEKENFDSNMAIKQFEDHQSSEEDASEKD
SRQSSEVKESDDNMSLNSDRDESISESYDKLEDINTDELPHGGEAFKAREVSASADFDVQTSLEDNYAESGIQTD
LSESSKENEVQTDAIPVKHNSTQTVKKEAVDNGLQTEPVETCNFSVQTFEGDENYLAEQCKPKQLKEYYTSAKVS
NIPFVSQNSTLRLIESTFQTVEAEFTVLMENIRNMDTFFTDQSSIPLVKRTVRSINNLYTWRIPEAEILLNIQNN
IKCEQMQITNANIQDLKEKVTDYVRKDIAQITEDVANAKEEYLFLMHFDDASSGYVKDLSTHQFRMQKTLRQKLF
DVSAKINHTEELLNILKLFTVKNKRLDDNPLVAKLAKESLARDGLLKEIKLLREQVSRLQLEEKGKKASSFDASS
SITKDMKGFKVVEVGLAMNTKKQIGDFFKNLNMAK
```

6.8 Vector sequences

pETMC_Ka_TEV_His

```
TTCCCCGTCAAGCTCTAAATCGGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGG
TTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA
ACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAA
AATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTT
CTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATT
ATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCT
GCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTG
AATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAAC
CAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGG
CGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGG
TGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATC
TGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGAT
TGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAA
TATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTT
CTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA
GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC
AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAGCGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT
GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC
GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCG
GTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT
GGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGAC
CGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGC
GATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAA
GGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCT
CACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAA
ATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAA
TGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCA
GCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGG
AGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTT
GAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAG
CGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAGAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTG
ACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTT
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTC
TTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAG
GCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACG
CGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCA
TTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCA
GCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAGTTGGGGTCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCC
AGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCA
CTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACA
GGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGC
GCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCT
CCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGC
ATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTT
TTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGC
ACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGC
TCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGC
CACGATGCGTCCGGCGTAGAGGATCGAGATCTCGGCCGCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATT
CCCCCCTAGGGCTAGCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCACGATTACG
ATATCCCAACGACCGAAAACCTTTACTTCCAGGGCCATATGCTCGAGGGTACCCTTAAGCAATTGGGATCCTAAGGCTGCTAACAAAGCCCGAA
AGGAAGCTGAGTTGGCTGCTGCCACCGCACTAGTATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGCGCGCAG
```

GCCGGCCTGTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCA
ATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

pEC-K.Hi-sumo

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAG
AAGGAGATATACATATGCACCATCACCACCATCACCATGGAATGGAAGGTGAATATATTAAACTCAAAGTCATTGGACAGGATAGCAGTGAGAT
TCACTTCAAAGTGAAAATGACAACACATCTCAAGAAACTCAAAGAATCATACTGTCAAAGACAGGGTGTTCCAATGAATTCACTCAGGTTTCTC
TTTGAGGGTCAGAGAATTGCTGATAATCATACTCCAAAAGAACTGGGAATGGAGGAAGAAGATGTGATTGAAGTCTACCAGGAACAAACCGGCG
GCCGCGGCCGGTGCTTTGCAGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACTGAGATCCGG
CTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAG
GGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCA
AGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGT
GGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCA
ACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGC
GAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACAT
TCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCA
TATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGA
CTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGA
GAATGGCAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTA
TTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACA
CTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCA
TGGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCGTAACATCA
TTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACAT
TATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCAT
AACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGA
CCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTG
GTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAA
GTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTG
GAGCGAACGACCTACACCGAGTCGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGG
TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTG
ACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGC
TGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCA
GCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACA
CCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGC
TGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGG
AGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGA
TGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTT
TTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGG
GTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGG
GTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGG
CGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCG
CTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCA
TGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGC
GTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGC
ACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGA
AGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGG
AAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAG
TGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCC
TGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGG
ACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGT
TTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGA
CGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTAC
CGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGC
AATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACG
CCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGG
CCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGC
CGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCG
ACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATT
CGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCG
CAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCC
GAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGT
CCGGCGTAGAGGATCG

pEGFP-C1

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT
TGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC
AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG
AGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCC

```
TGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTT
CATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCAC
ATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAA
GCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATC
GAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCA
CCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGA
CGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAA
CTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACTCTCCCCCTGAACCTGAAACATAAAATGAATG
CAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACT
GCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTG
TTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCA
GTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCAT
CACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCC
GGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACA
CCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATAC
ATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCAGCTG
TGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGT
GTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCAT
CCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCT
GAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATT
GAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATG
CCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGC
GCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAA
GTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATC
CGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGA
CGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGAT
GCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAG
CGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCG
CATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAG
ATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCAT
GCTGGAGTTCTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGA
CAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACCGAGACCCCATT
GGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGG
CCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGAT
AATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTT
TTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAA
GGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCT
ACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGG
ATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGA
GCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTT
CCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGA
TTCTGTGGATAACCGTATTACCGCCATGCAT
```


pCLneo-lamdaD

```
TCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATCTATATCATAAT
ATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAAT
AATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACAT
CAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACT
TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA
CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAAC
TGCGATCGCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCAC
TAGAAGCTTTATTGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGCTTCTGACACAACAGTCTCGAACTTAAGCTGCAGTGACTCT
CTTAAGGTAGCCTTGCAGAAGTTGGTCGTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGC
TTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCC
AGTTCAATTACAGCTCTTAAGGCTAGAGTACTTAATACGACTCACTATAGGCTAGCACCATGGACGCACAAACACGACGACGTGAGCGTCGCGC
TGAGAAACAAGCTCAATGGAAAGCTGCAAACCCACCGCTCGAGAATTCACGCGTGGTACCTCTAGAGTCGACCCGGGCGGCCGCTTCCCTTTAG
TGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTG
AAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAATTGCATTCATTTTATGTTTCAGGTTCA
GGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCCGATAAGGATCGATCCGGGCTGGCGTAATAGCGAAG
AGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTC
CCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTC
ACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACA
ACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAAT
TTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACGG
GGATCTGCGCAGCACCATGGCCTGAAATAACCTCTGAAAGAGGAACTTGGTTAGGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTG
TCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCC
CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAA
CTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCA
GAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTTGATTCTTCTGACACAACAGTCTCGAACTTAAGGCTAGAGCCACCA
TGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTC
TGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAG
GCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGG
```

```
GCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCT
TGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGAT
CTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATG
GCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGA
CATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCG
CAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATC
ACGATGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGCGATAAGGATCCGCGTATGGTGCAC
TCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCG
GCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCC
TCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGT
ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAA
AAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACG
TTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACAC
TATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAA
CCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGT
AACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTG
CGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGC
GCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGG
TAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTG
ATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGA
TCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACT
CTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTG
TAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACG
ATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATAC
CTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCA
CGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGGCTCGACAGATC
T
```

pGEX-4T

```
ACGTTATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAAT
TCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACA
ATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGTATTCATGTCCCCTATACTAGGTTATTGGA
AAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAA
ATGGCGAAACAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATC
ATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCAAAAGACGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTA
GATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGA
AGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATG
GACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCA
AGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGTTCCGCGTGGATCCCCGGA
ATTCCCGGGTCGACTCGAGCGGCCGCATCGTGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAG
CTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCG
CAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCAT
GATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATG
TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTC
CCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGG
TTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTA
TGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAG
TCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCT
GACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAAT
GAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAG
CTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGA
TAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACG
GGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACT
CATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTA
ACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG
CAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCT
GTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC
CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAAC
GCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCG
CCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTT
TCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAAATTCCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCG
GAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTTATCAGACCGTTTCCCGCG
TGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACA
ACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCT
CGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGC
AACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCT
TGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCAC
CAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTC
AGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGAT
GCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATAC
```

```
GACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGC
TGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAAC
CGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGT
GAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACA
GGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT
GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGC
GCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTG
GCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTAACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGT
TGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTGGAATT
```

pProEx-HTa

```
GTTTGACAGCTTATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATC
ACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGA
GCTGTTGACAATTAATCATCCGGTCCGTATAATCTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGACCATGTCGTACTACCATC
ACCATCACCATCACGATTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGCGCCATGGATCCGGAATTCAAAGGCCTACGTCGACGAGC
TCAACTAGTGCGGCCGCTTTCGAATCTAGAGCCTGCAGTCTCGAGGCATGCGGTACCAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCA
GCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCC
GAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGC
TCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTT
GCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTT
TTGCGTTTCTACAAACTCTTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAA
TATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG
AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAG
TTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAA
CTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAAT
TATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCA
CAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTACA
GCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAG
TTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGC
AGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCT
GAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTA
AAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAA
GATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG
GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCC
ACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACC
TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG
ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCT
CACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCG
AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAATTTT
GTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACC
GAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGG
GCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCG
ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCG
GTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGTATGGTGCACTCTCA
GTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACA
CCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTT
CACCGTCATCACCGAAACGCGCGAGGCAGCAGATCAATTCGCGCGCGAAGGCGAAGCGGCATGCATTTACGTTGACACCATCGAATGGTGCAAA
ACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATG
CCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGA
GCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCG
CCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCT
GTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGC
TGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTG
GGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCT
GGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAAT
GCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGC
GTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCC
TGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAG
AAAAACCACCCTGGCACCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAA
AGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCGCGAATTGATCTG
```

pETM-14

```
ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGC
AGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCTTGTCGACGGAG
CTCGAATTCGGATCCGGTACCTCATTAGTCCATGGGCCCCTGGAACAGAACTTCCAGACCCGCGGAGTGATGGTGATGGTGATGTTTCATGGTA
TATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGA
TCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGA
TGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCC
ATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAG
AGCGTCGAGATCCCGGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATG
```

```
TGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGC
GAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTG
ATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGG
TGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTA
TCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAAC
AGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCAT
TAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTG
GAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTG
GGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTCGGGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATA
TCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAA
GGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCA
TTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCGGGA
TCTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCT
TTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTC
GCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCC
GGCATGGCGGCCCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAAT
CACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTA
AAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCT
GTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTA
ACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTACAC
GGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG
CTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGGTGATGACG
GTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGC
GGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTG
TACTGAGAGTGCACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACT
GACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGC
TCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATC
TCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
TTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA
AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAAACTGTCTGCTTA
CATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTTA
TATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGA
AACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCA
TTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGT
GAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTC
TCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAA
AGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTA
ATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCAT
TACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGAATT
AATTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAAATTGT
AAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAA
TCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAA
AAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCC
TAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCG
CTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCA
```

pYES263

```
TATGCACCATCACCATCACCATGCTAGCTTCATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTG
GAATATCTTGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTC
CCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGG
TTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTT
GAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGATC
ATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCCATGTGCCTGGATGCGTTCCCAAAATTAGTTTG
TTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACG
TTTGGTGGTGGCGACCATCACGAGAATCTTTATTTTCAGGGCGCCATGGTCTAGCATGACTGGTGGACAGCAAATGGGTCGGGGATCCACTAGT
AACGGCCGCCAGTGTGCTGGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCGCATCATGTAATTAGTTAT
GTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTT
ATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACATAGACGCGTGTACGCATGTAACATTATACTGAAA
CCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCGGCCCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTA
TTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACG
GTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGCCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGAC
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
GAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCT
GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAG
ATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTT
TGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
ATAACTACGATACGGGAGCGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAA
ACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATTCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAG
```

119

```
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGGCATTGCTACAGGCATCGTGGTGTCACTCTCGTCGTTTGGTATGGCTTCATTCAGCTCC
GGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGT
TGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTA
CTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATAGTGTATCACATAGCAGAACT
TTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTG
CACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGC
GACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATGGGTAATAACTGATATAATTAAATTGAAGCTCTAATTTGTGAGTTTAGTATAC
ATGCATTTACTTATAATACAGTTTTTTAGTTTTGCTGGCCGCATCTTCTCAAATATGCTTCCCAGCCTGCTTTTCTGTAACGTTCACCCTCTAC
CTTAGCATCCCTTCCCTTTGCAAATAGTCCTCTTCCAACAATAATAATGTCAGATCCTGTAGAGACCACATCATCCACGGTTCTATACTGTTGA
CCCAATGCGTCTCCCTTGTCATCTAAACCCACACCGGGTGTCATAATCAACCAATCGTAACCTTCATCTCTTCCACCCATGTCTCTTTGAGCAA
TAAAGCCGATAACAAAATCTTTGTCGCTCTTCGCAATGTCAACAGTACCCTTAGTATATTCTCCAGTAGATAGGGAGCCCTTGCATGACAATTC
TGCTAACATCAAAAGGCCTCTAGGTTCCTTTGTTACTTCTTCTGCCGCCTGCTTCAAACCGCTAACAATACCTGGGCCCACCACACCGTGTGCA
TTCGTAATGTCTGCCCATTCTGCTATTCTGTATACACCCGCAGAGTACTGCAATTTGACTGTATTACCAATGTCAGCAAATTTTCTGTCTTCGA
AGAGTAAAAAATTGTACTTGGCGGATAATGCCTTTAGCGGCTTAACTGTGCCCTCCATTGAAAAATCAGTCAAGATATCCACATGTGTTTTTAG
TAAACAAATTTTGGGACCTAATGCTTCAACTAACTCCAGTAATTCCTTGGTGGTACGAACATCCAATGAAGCACACAAGTTTGTTTGCTTTTCG
TGCATGATATTAAATAGCTTGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATGATTTATCTTCGTTTCC
TGCAGGTTTTTGTTCTGTGCAGTTGGGTTAAGAATACTGGGCAATTTCATGTTTCTTCAACACTACATATGCGTATATATACCAATCTAAGTCT
GTGCTCCTTCCTTCGTTCTTCCTTCTGTTCGGAGATTACCGAATCAAAAAAATTTCAAAGAAACCGAAATCAAAAAAAAGAATAAAAAAAAAAT
GATGAATTGAATTGAAAAGCTAGCTTATCGATGATAAGCTGTCAAAGATGAGAATTAATTCCACGGACTATAGACTATACTAGATACTCCGTCT
ACTGTACGATACACTTCCGCTCAGGTCCTTGTCCTTTAACGAGGCCTTACCACTCTTTTGTTACTCTATTGATCCAGCTCAGCAAAGGCAGTGT
GATCTAAGATTCTATCTTCGCGATGTAGTAAAACTAGCTAGACCGAGAAAGAGACTAGAAATGCAAAAGGCACTTCTACAATGGCTGCCATCAT
TATTATCCGATGTGACGCTGCAGCTTCTCAATGATATTCGAATACGCTTTGAGGAGATACAGCCTAATATCCGACAAACTGTTTTACAGATTTA
CGATCGTACTTGTTACCCATCATTGAATTTTGAACATCCGAACCTGGGAGTTTTCCCTGAAACAGATAGTATATTTGAACCTGTATAATAATAT
ATAGTCTAGCGCTTTACGGAAGACAATGTATGTATTTCGGTTCCTGGAGAAACTATTGCATCTATTGCATAGGTAATCTTGCACGTCGCATCCC
CGGTTCATTTTCTGCGTTTCCATCTTGCACTTCAATAGCATATCTTTGTTAACGAAGCATCTGTGCTTCATTTTGTAGAACAAAAATGCAACGC
GAGAGCGCTAATTTTTCAAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACGAAGAATCTGT
GCTTCATTTTTGTAAAACAAAAATGCAACGCGACGAGAGCGCTAATTTTTCAAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAA
CGCGAGAGCGCTATTTTACCAACAAAGAATCTATACTTCTTTTTTGTTCTACAAAAATGCATCCCGAGAGCGCTATTTTTCTAACAAAGCATCT
TAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTCTTGATAACTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTT
TGGTGTCTATTTTCTCTTCCATAAAAAAAGCCTGACTCCACTTCCCGCGTTTACTGATTACTAGCGAAGCTGCGGGTGCATTTTTTCAAGATAA
AGGCATCCCCGATTATATTCTATACCGATGTGGATTGCGCATACTTTGTGAACAGAAAGTGATAGCGTTGATGATTCTTCATTGGTCAGAAAAT
TATGAACGGTTTCTTCTATTTTGTCTCTATATACTACGTATAGGAAATGTTTACATTTTCGTATTGTTTTCGATTCACTCTATGAATAGTTCTT
ACTACAATTTTTTTGTCTAAAGAGTAATACTAGAGATAAACATAAAAAATGTAGAGGTCGAGTTTAGATGCAAGTTCAAGGAGCGAAAGGTGGA
TGGGTAGGTTATATAGGGATATAGCACAGAGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGTATTCGCAATGGGAAGCTC
CACCCCGGTTGATAATCAGAAAAGCCCCAAAAACAGGAAGATTGTATAAGCAAATATTTAAATTGTAAACGTTAATATTTTGTTAAAATTCGCG
TTAAATTTTTGTTAAATCAGCTCATTTTTTAACGAATAGCCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGA
GTGTTGTTCCAGTTTCCAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAAGGGTCTATCAGGGCGATGGCCCACT
ACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCAGTAAATCGGAAGGGTAAACGGATGCCCCCATTTAGAGCTTGA
CGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGGGCTAGGGCGGTGGGAAGTGTAGGGGTCACGCTGGGCG
TAACCACCACACCCGCCGCGCTTAATGGGGCGCTACAGGGCGCGTGGGGATGATCCACTAGTACGGATTAGAAGCCGCCGAGCGGGTGACAGCC
CTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCCTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATA
AAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACA
ACCATAGGATGATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAA
ATGCAAAAACTGCATTAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAA
ATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACCCCGGATCGGACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTA
AGCT
```

pYES260

```
GGCCGCAAATTAAAGCCTTCGAGCGTCCCAAAACCTTCTCAAGCAAGGTTTTCAGTATAATGTTACATGCGTACACGCGTCTGTACAGAAAAAA
AAGAAAAATTTGAAATATAAATAACGTTCTTAATACTAACATAACTATAAAAAAATAAATAGGGACCTAGACTTCAGGTTGTCTAACTCCTTCC
TTTTCGGTTAGAGCGGATGTGGGGGGAGGGCGTGAATGTAAGCGTGACATAACTAATTACATGATGCGGCCCTCTAGATGCATGCTCGAGCGGC
CGCCAGTGTGATGGATATCTGCAGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCCCGACCCATTTGCTGTCCACCAGTCATGCTAGAC
CATGGCGCCCTGAAAATAAAGATTCTCGCTAGCATGGTGATGGTGATGGTGCATAAGCTTAATATTCCCTATAGTGAGTCGTATTACAGCTGCT
AGTAGTCCGATCCGGGGTTTTTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTTGATACTTTTATTACATTTGAATAAGAA
GTAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGTGGTTAATGCAGTTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCTT
CGCTGATTAATTACCCCAGAAATAAGGCTAAAAAACTAATCGCATTATCATCCTATGGTTGTTAATTTGATTCGTTCATTTGAAGGTTTGTGGG
GCCAGGTTACTGCCAATTTTTCCTCTTCATAACCATAAAAGCTAGTATTGTAGAATCTTTATTGTTCGGAGCAGTGCGGCGCGAGGCACATCTG
CGTTTCAGGAACGCGACCGGTGAGGACGAGGACGCACGGAGGAGAGTCTTCCTTCGGAGGGCTGTCACCCGCTCGGCGGCTTCTAATCCGTACT
AGTGGATCATCCCCACGCGCCCTGTAGCGCCCCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCCCTACACTTCCCACCGCCCTA
GCCCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCATCCGTTTACCCTTCC
GATTTACTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACCCTTTTTCGCCC
TTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTGGAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAA
GGGATTTTGCCGATTTCGGGCTATTCGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTT
AAATATTTGCTTATACAATCTTCCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGTGGAGCTTCCCATTGCGAATACCGCTTCCACAAACAT
TGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCTATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTC
GACCTCTACATTTTTTATGTTTATCTCTAGTATTACTCTTTAGCAAAAAAATTGTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACG
AAAATGTAAACATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAATTTTCTGACCAATGAAGAATCATCAACGCTATCA
CTTTCTGTTCACAAAGTATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTTTATCTTGAAAAAATGCACCCGCAGCTTCGCTAGT
AATCAGTAAACGCGGGAAGTGGAGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGCCTTCTTCTAACCTTAACGGACCTACAGT
GCAAAAAGTTATCAAGAGACTGCATTATAGAGCGCACAAAGGAGAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAAATAGCGCTCTCGGGATG
CATTTTTGTAGAACAAAAAGAAGTATAGATTCTTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATT
CTTTGTTTGAAAAATTAGCGCTCTCGTCGCGTTGCATTTTTGTTTTACAAAATGAAGCACAGATTCTTCGTTGGTAAAATAGCGCTTTCGCGT
TGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGA
TGCTTCGTTAACAAAGATATGCTATTGAAGTGCAAGATGGAAACGCAGAAAATGAACCGGGGATGCGACGTGCAAGATTACCTATGCAATAGAT
GCAATAGTTTCTCCAGGAACCGAAATACATACATTGTCTTCCGTAAAGCGCTAGACTATATATTATTATACAGGTTCAAATACTATCTGTTT
CAGGGAAAACTCCCAGGTTCGGATGTTCAAAATTCAATGATGGGTAACAAGTACGATCGTAAATCTGTAAAACAGTTTGTCGGATATTAGGCTG
```

```
TATCTCCTCAAAGCGTATTCGAATATCATTGAGAAGCTGCAGCGTCACATCGGATAATAATGATGGCAGCCATTGTAGAAGTGCCTTTTGCATT
TCTAGTCTCTTTCTCGGTCTAGCTAGTTTTACTACATCGCGAAGATAGAATCTTAGATCACACTGCCTTTGCTGAGCTGGATCAATAGAGTAAC
AAAAGAGTGGTAAGGCCTCGTTAAAGGACAAGGACCTGAGCGGAAGTGTATCGTACAGTAGACGGAGTATCTAGTATAGTCTATAGTCCGTGGA
ATTAATTCTCATCTTTGACAGCTTATCATCGATAAGCTAGCTTTTCAATTCAATTCATCATTTTTTTTTTATTCTTTTTTTTGATTTCGGTTTC
TTTGAAATTTTTTTGATTCGGTAATCTCCGAACAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACGCATATGTAGTGT
TGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCATGTCGAAAGCTACATAT
AAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAACTTGTGTGCTTCATTGGATG
TTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTC
AATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACA
GTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAGCG
GTTTGAAGCAGGCGGCAGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATGCAAGGGCTCCCTATCTACTGGAGA
ATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGT
TACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTA
CAGGATCTGACATTATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGGGA
AGCATATTTGAGAAGATGCGGCCAGCAAAACTAAAAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACAAATTAGAGCTTCAATTTAA
TTATATCAGTTATTACCCATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCC
TGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGC
GGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGATACACTATTATCCCGTATTG
ACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGG
CATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAG
CTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGAGTG
ACACCACGATGCCTGTAGCAATGCCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTG
AATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGG
TCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCGCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAAC
GAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAA
ACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCG
TCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAG
CGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGT
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCA
GCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCAC
CTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGGCT
TTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCG
CCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATT
CATTAATGCAG
```

7 References

Akey, C. W. and M. Radermacher (1993). "Architecture of the Xenopus nuclear pore complex revealed by three-dimensional cryo-electron microscopy." J Cell Biol 122(1): 1-19.

Alber, F., S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B. T. Chait, A. Sali and M. P. Rout (2007). "The molecular architecture of the nuclear pore complex." Nature 450(7170): 695-701.

Alcazar-Roman, A. R., E. J. Tran, S. Guo and S. R. Wente (2006). "Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export." Nat Cell Biol 8(7): 711-716.

Andersen, C. B., L. Ballut, J. S. Johansen, H. Chamieh, K. H. Nielsen, C. L. Oliveira, J. S. Pedersen, B. Seraphin, H. Le Hir and G. R. Andersen (2006). "Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA." Science 313(5795): 1968-1972.

Antonin, W., J. Ellenberg and E. Dultz (2008). "Nuclear pore complex assembly through the cell cycle: regulation and membrane organization." FEBS Lett 582(14): 2004-16.

Bachi, A., I. C. Braun, J. P. Rodrigues, N. Pante, K. Ribbeck, C. von Kobbe, U. Kutay, M. Wilm, D. Gorlich, M. Carmo-Fonseca and E. Izaurralde (2000). "The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates." Rna 6(1): 136-58.

Ballut, L., B. Marchadier, A. Baguet, C. Tomasetto, B. Seraphin and H. Le Hir (2005). "The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity." Nat Struct Mol Biol 12(10): 861-869.

Bayliss, R., S. W. Leung, R. P. Baker, B. B. Quimby, A. H. Corbett and M. Stewart (2002). "Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats." Embo J 21(12): 2843-53.

Beck, M., F. Forster, M. Ecke, J. M. Plitzko, F. Melchior, G. Gerisch, W. Baumeister and O. Medalia (2004). "Nuclear pore complex structure and dynamics revealed by cryoelectron tomography." Science 306(5700): 1387-90.

Beck, M., V. Lucic, F. Forster, W. Baumeister and O. Medalia (2007). "Snapshots of nuclear pore complexes in action captured by cryo-electron tomography." Nature 449(7162): 611-5.

Bednenko, J., G. Cingolani and L. Gerace (2003). "Importin beta contains a COOH-terminal nucleoporin binding region important for nuclear transport." J Cell Biol 162(3): 391-401.

Bergfors, T. (1999). "Protein Crystallization." International University Line 1 edition.

Berke, I. C., T. Boehmer, G. Blobel and T. U. Schwartz (2004). "Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex." J Cell Biol 167(4): 591-7.

Blundell, T. L. and L. N. Johnson (1976). Protein crystallography. New York ; London, Academic Press,.

Boehmer, T., S. Jeudy, I. C. Berke and T. U. Schwartz (2008). "Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex." Mol Cell 30(6): 721-31.

Bono, F., J. Ebert, E. Lorentzen and E. Conti (2006). "The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA." Cell 126(4): 713-725.

Braun, I. C., A. Herold, M. Rode, E. Conti and E. Izaurralde (2001). "Overexpression of TAP/p15 heterodimers bypasses nuclear retention and stimulates nuclear mRNA export." J Biol Chem 276(23): 20536-43.

Braun, I. C., E. Rohrbach, C. Schmitt and E. Izaurralde (1999). "TAP binds to the constitutive transport element (CTE) through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus." Embo J 18(7): 1953-65.

Brohawn, S. G., N. C. Leksa, E. D. Spear, K. R. Rajashankar and T. U. Schwartz (2008). "Structural evidence for common ancestry of the nuclear pore complex and vesicle coats." Science 322(5906): 1369-73.

Bruhn, L., A. Munnerlyn and R. Grosschedl (1997). "ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function." Genes Dev 11(5): 640-53.

Caruthers, J. M., E. R. Johnson and D. B. McKay (2000). "Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase." Proc Natl Acad Sci U S A 97(24): 13080-5.

References

Caruthers, J. M. and D. B. McKay (2002). "Helicase structure and mechanism." Curr Opin Struct Biol 12(1): 123-33.

CCP4 (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr 50(Pt 5): 760-3.

Clouse, K. N., M. J. Luo, Z. Zhou and R. Reed (2001). "A Ran-independent pathway for export of spliced mRNA." Nat Cell Biol 3(1): 97-9.

Cole, C. N. and J. J. Scarcelli (2006). "Transport of messenger RNA from the nucleus to the cytoplasm." Curr Opin Cell Biol 18(3): 299-306.

Collins, R., T. Karlberg, L. Lehtio, P. Schutz, S. van den Berg, L. G. Dahlgren, M. Hammarstrom, J. Weigelt and H. Schuler (2009). "The DExD/H-box RNA helicase DDX19 is regulated by an alpha -helical switch." J Biol Chem.

Cook, A., F. Bono, M. Jinek and E. Conti (2007). "Structural biology of nucleocytoplasmic transport." Annu Rev Biochem 76: 647-671.

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

Cordin, O., N. K. Tanner, M. Doere, P. Linder and J. Banroques (2004). "The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity." Embo J 23(13): 2478-87.

Crick, F. (1970). "Central dogma of molecular biology." Nature 227(5258): 561-3.

Crick, F. H. (1958). "On protein synthesis." Symp Soc Exp Biol 12: 138-63.

Cronshaw, J. M., A. N. Krutchinsky, W. Zhang, B. T. Chait and M. J. Matunis (2002). "Proteomic analysis of the mammalian nuclear pore complex." J Cell Biol 158(5): 915-27.

D'Angelo, M. A. and M. W. Hetzer (2008). "Structure, dynamics and function of nuclear pore complexes." Trends Cell Biol 18(10): 456-66.

Debler, E. W., Y. Ma, H. S. Seo, K. C. Hsia, T. R. Noriega, G. Blobel and A. Hoelz (2008). "A fence-like coat for the nuclear pore membrane." Mol Cell 32(6): 815-26.

DeLano, W. L. "The PyMOL Molecular Graphics System."" DeLano Scientific LLC, San Carlos, CA, USA http://www.pymol.org.

Denning, D. P., S. S. Patel, V. Uversky, A. L. Fink and M. Rexach (2003). "Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded." Proc Natl Acad Sci U S A 100(5): 2450-5.

Derewenda, Z. S. (2004). "Rational protein crystallization by mutational surface engineering." Structure 12(4): 529-535.

Drenth, J. (1999). "Principles of protein X-ray crystallography." Heidelberg: Springer-Verlag.

Drude, P. (1900). "History of electromagnetic dispersion equations." Annalen Der Physik 1(2): 437-U15.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1): 2126-2132.

Estruch, F. and C. N. Cole (2003). "An early function during transcription for the yeast mRNA export factor Dbp5p/Rat8p suggested by its genetic and physical interactions with transcription factor IIH components." Mol Biol Cell 14(4): 1664-1676.

Fan, J. S., Z. Cheng, J. Zhang, C. Noble, Z. Zhou, H. Song and D. Yang (2009). "Solution and crystal structures of mRNA exporter Dbp5p and its interaction with nucleotides." J Mol Biol.

Forster, F., B. Webb, K. A. Krukenberg, H. Tsuruta, D. A. Agard and A. Sali (2008). "Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies." J Mol Biol 382(4): 1089-106.

Frey, S. and D. Gorlich (2007). "A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes." Cell 130(3): 512-23.

Frey, S., R. P. Richter and D. Gorlich (2006). "FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties." Science 314(5800): 815-7.

Fribourg, S., I. C. Braun, E. Izaurralde and E. Conti (2001). "Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor." Mol Cell 8(3): 645-656.

Fribourg, S. and E. Conti (2003). "Structural similarity in the absence of sequence homology of the messenger RNA export factors Mtr2 and p15." EMBO Rep 4(7): 699-703.

Fuller-Pace, F. V., S. M. Nicol, A. D. Reid and D. P. Lane (1993). "DbpA: a DEAD box protein specifically activated by 23s rRNA." Embo J 12(9): 3619-26.

Gall, J. G. (1967). "Octagonal nuclear pores." J Cell Biol 32(2): 391-9.

References

Gatfield, D., H. Le Hir, C. Schmitt, I. C. Braun, T. Kocher, M. Wilm and E. Izaurralde (2001). "The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila." Curr Biol 11(21): 1716-21.

Gerace, L. and B. Burke (1988). "Functional organization of the nuclear envelope." Annu Rev Cell Biol 4: 335-74.

Gralla, J. and C. DeLisi (1974). "mRNA is expected to form stable secondary structures." Nature 248(446): 330-2.

Grant, R. P., D. Neuhaus and M. Stewart (2003). "Structural basis for the interaction between the Tap/NXF1 UBA domain and FG nucleoporins at 1A resolution." J Mol Biol 326(3): 849-58.

Gross, T., A. Siepmann, D. Sturm, M. Windgassen, J. J. Scarcelli, M. Seedorf, C. N. Cole and H. Krebber (2007). "The DEAD-box RNA helicase Dbp5 functions in translation termination." Science 315(5812): 646-649.

Hetzer, M. W., T. C. Walther and I. W. Mattaj (2005). "Pushing the envelope: structure, function, and dynamics of the nuclear periphery." Annu Rev Cell Dev Biol 21: 347-80.

Hinshaw, J. E., B. O. Carragher and R. A. Milligan (1992). "Architecture and design of the nuclear pore complex." Cell 69(7): 1133-41.

http://www.originlab.com/.

Huve, J., R. Wesselmann, M. Kahms and R. Peters (2008). "4Pi microscopy of the nuclear pore complex." Biophys J 95(2): 877-85.

Iglesias, N. and F. Stutz (2008). "Regulation of mRNP dynamics along the export pathway." FEBS Lett 582(14): 1987-1996.

Jankowsky, E. and M. E. Fairman (2007). "RNA helicases-one fold for many functions." Curr Opin Struct Biol 17(3): 316-324.

Jeudy, S. and T. U. Schwartz (2007). "Crystal structure of nucleoporin Nic96 reveals a novel, intricate helical domain architecture." J Biol Chem 282(48): 34904-12.

Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." J. Appl. Cryst. 26: 795-800.

Kabsch, W. and C. Sander (1983). "Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features." Biopolymers 22(12): 2577-2637.

Karlsson, R. and A. Falt (1997). "Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors." J Immunol Methods 200(1-2): 121-33.

Kim, Y., P. Quartey, H. Li, L. Volkart, C. Hatzos, C. Chang, B. Nocek, M. Cuff, J. Osipiuk, K. Tan, Y. Fan, L. Bigelow, N. Maltseva, R. Wu, M. Borovilos, E. Duggan, M. Zhou, T. A. Binkowski, R. G. Zhang and A. Joachimiak (2008). "Large-scale evaluation of protein reductive methylation for improving protein crystallization." Nat Methods 5(10): 853-4.

Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography 36: 1277-1282.

Kraemer, D., R. W. Wozniak, G. Blobel and A. Radu (1994). "The human CAN protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm." Proc Natl Acad Sci U S A 91(4): 1519-1523.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4." Nature 227(5259): 680-&.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26.

Liker, E., E. Fernandez, E. Izaurralde and E. Conti (2000). "The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain." EMBO J 19(21): 5587-5598.

Lim, R. Y., U. Aebi and B. Fahrenkrog (2008). "Towards reconciling structure and function in the nuclear pore complex." Histochem Cell Biol 129(2): 105-16.

Lim, R. Y., K. S. Ullman and B. Fahrenkrog (2008). "Biology and biophysics of the nuclear pore complex and its components." Int Rev Cell Mol Biol 267: 299-342.

Linder, P., P. F. Lasko, M. Ashburner, P. Leroy, P. J. Nielsen, K. Nishi, J. Schnier and P. P. Slonimski (1989). "Birth of the D-E-A-D box." Nature 337(6203): 121-2.

Liu, F., A. Putnam and E. Jankowsky (2008). "ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding." Proc Natl Acad Sci U S A 105(51): 20209-14.

References

Lund, M. K. and C. Guthrie (2005). "The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim." Mol Cell 20(4): 645-651.

Luo, M. L., Z. Zhou, K. Magni, C. Christoforides, J. Rappsilber, M. Mann and R. Reed (2001). "Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly." Nature 413(6856): 644-7.

Macara, I. G. (2001). "Transport into and out of the nucleus." Microbiol Mol Biol Rev 65(4): 570-94, table of contents.

Maco, B., B. Fahrenkrog, N. P. Huang and U. Aebi (2006). "Nuclear pore complex structure and plasticity revealed by electron and atomic force microscopy." Methods Mol Biol 322: 273-88.

Mansfeld, J., S. Guttinger, L. A. Hawryluk-Gara, N. Pante, M. Mall, V. Galy, U. Haselmann, P. Muhlhausser, R. W. Wozniak, I. W. Mattaj, U. Kutay and W. Antonin (2006). "The conserved transmembrane nucleoporin NDC1 is required for nuclear pore complex assembly in vertebrate cells." Mol Cell 22(1): 93-103.

Marelli, M., D. J. Dilworth, R. W. Wozniak and J. D. Aitchison (2001). "The dynamics of karyopherin-mediated nuclear transport." Biochem Cell Biol 79(5): 603-12.

McCoy, A. J., L. C. Storoni and R. J. Read (2004). "Simple algorithm for a maximum-likelihood SAD function." Acta Crystallogr D Biol Crystallogr 60(Pt 7): 1220-1228.

McPherson, A. (2001). "Crystallization of Biological Macromolecules." Cold Spring Harbor Laboratory Press.

Meier, I. and J. Brkljacic (2009). "The nuclear pore and plant development." Curr Opin Plant Biol 12(1): 87-95.

Melcak, I., A. Hoelz and G. Blobel (2007). "Structure of Nup58/45 suggests flexible nuclear pore diameter by intermolecular sliding." Science 315(5819): 1729-32.

Miller, A. L., M. Suntharalingam, S. L. Johnson, A. Audhya, S. D. Emr and S. R. Wente (2004). "Cytoplasmic inositol hexakisphosphate production is sufficient for mediating the Gle1-mRNA export pathway." J Biol Chem 279(49): 51022-51032.

Miller, J. H. (1972). "Experiments in molecular genetics." Cold Spring Harbor Laboratory Press.

Napetschnig, J., G. Blobel and A. Hoelz (2007). "Crystal structure of the N-terminal domain of the human protooncogene Nup214/CAN." Proc Natl Acad Sci U S A 104(6): 1783-1788.

Napetschnig, J., S. A. Kassube, E. W. Debler, R. W. Wong, G. Blobel and A. Hoelz (2009). "Structural and functional analysis of the interaction between the nucleoporin Nup214 and the DEAD-box helicase Ddx19." Proc Natl Acad Sci U S A 106(9): 3089-94.

Pai, E. F., U. Krengel, G. A. Petsko, R. S. Goody, W. Kabsch and A. Wittinghofer (1990). "Refined Crystal-Structure of the Triphosphate Conformation of H-Ras P21 at 1.35 a Resolution - Implications for the Mechanism of Gtp Hydrolysis." Embo Journal 9(8): 2351-2359.

Paine, P. L., L. C. Moore and S. B. Horowitz (1975). "Nuclear envelope permeability." Nature 254(5496): 109-14.

Painter, J. and E. A. Merritt (2006). "Optimal description of a protein structure in terms of multiple groups undergoing TLS motion." Acta Crystallogr D Biol Crystallogr 62(Pt 4): 439-450.

Pante, N. and M. Kann (2002). "Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm." Mol Biol Cell 13(2): 425-34.

Pause, A., N. Methot and N. Sonenberg (1993). "The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis." Mol Cell Biol 13(11): 6789-98.

Pause, A. and N. Sonenberg (1992). "Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A." Embo J 11(7): 2643-54.

Peters, R. (2005). "Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality." Traffic 6(5): 421-7.

Peters, R. (2009). "Translocation through the nuclear pore: Kaps pave the way." Bioessays 31(4): 466-77.

Petoukhov, M. V., N. A. J. Eady, K. A. Brown and D. I. Svergun (2002). "Addition of missing loops and domains to protein models by X-ray solution scattering." Biophysical Journal 83(6): 3113-3125.

Putnam, C. D., M. Hammel, G. L. Hura and J. A. Tainer (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining

accurate macromolecular structures, conformations and assemblies in solution." Q Rev Biophys 40(3): 191-285.

Reichelt, R., A. Holzenburg, E. L. Buhle, Jr., M. Jarnik, A. Engel and U. Aebi (1990). "Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components." J Cell Biol 110(4): 883-94.

Reichmann, D., Y. Phillip, A. Carmi and G. Schreiber (2008). "On the contribution of water-mediated interactions to protein-complex stability." Biochemistry 47(3): 1051-1060.

Rhodes, G. (2000). Crystallography made crystal clear : a guide for users of macromolecular models. San Diego, Calif. ; London, Academic,.

Ribbeck, K. and D. Gorlich (2001). "Kinetic analysis of translocation through nuclear pore complexes." Embo J 20(6): 1320-30.

Rocak, S., B. Emery, N. K. Tanner and P. Linder (2005). "Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs." Nucleic Acids Res 33(3): 999-1009.

Rogers, G. W., Jr., N. J. Richter and W. C. Merrick (1999). "Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A." J Biol Chem 274(18): 12236-44.

Rout, M. P. and J. D. Aitchison (2001). "The nuclear pore complex as a transport machine." J Biol Chem 276(20): 16593-6.

Rout, M. P., J. D. Aitchison, M. O. Magnasco and B. T. Chait (2003). "Virtual gating and nuclear transport: the hole picture." Trends Cell Biol 13(12): 622-8.

Rout, M. P., J. D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao and B. T. Chait (2000). "The yeast nuclear pore complex: composition, architecture, and transport mechanism." J Cell Biol 148(4): 635-51.

Rout, M. P. and G. Blobel (1993). "Isolation of the yeast nuclear pore complex." J Cell Biol 123(4): 771-83.

Sauter, N. K., R. W. Grosse-Kunstleve and P. D. Adams (2004). "Robust indexing for automatic data collection." Journal of Applied Crystallography 37: 399-409.

Scarcelli, J. J., S. Viggiano, C. A. Hodge, C. V. Heath, D. C. Amberg and C. N. Cole (2008). "Synthetic genetic array analysis in Saccharomyces cerevisiae provides evidence for an interaction between RAT8/DBP5 and genes encoding P-body components." Genetics 179(4): 1945-1955.

Schmitt, C., C. von Kobbe, A. Bachi, N. Pante, J. P. Rodrigues, C. Boscheron, G. Rigaut, M. Wilm, B. Seraphin, M. Carmo-Fonseca and E. Izaurralde (1999). "Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p." Embo J 18(15): 4332-4347.

Schutz, P., M. Bumann, A. E. Oberholzer, C. Bieniossek, H. Trachsel, M. Altmann and U. Baumann (2008). "Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions." Proc Natl Acad Sci U S A 105(28): 9564-9569.

Schwartz, T. U. (2005). "Modularity within the architecture of the nuclear pore complex." Curr Opin Struct Biol 15(2): 221-6.

Sengoku, T., O. Nureki, A. Nakamura, S. Kobayashi and S. Yokoyama (2006). "Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa." Cell 125(2): 287-300.

Snay-Hodge, C. A., H. V. Colot, A. L. Goldstein and C. N. Cole (1998). "Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export." EMBO J 17(9): 2663-2676.

Stavru, F., B. B. Hulsmann, A. Spang, E. Hartmann, V. C. Cordes and D. Gorlich (2006). "NDC1: a crucial membrane-integral nucleoporin of metazoan nuclear pore complexes." J Cell Biol 173(4): 509-19.

Stewart, M. (2007). "Ratcheting mRNA out of the nucleus." Mol Cell 25(3): 327-330.

Stoffler, D., B. Feja, B. Fahrenkrog, J. Walz, D. Typke and U. Aebi (2003). "Cryo-electron tomography provides novel insights into nuclear pore architecture: implications for nucleocytoplasmic transport." J Mol Biol 328(1): 119-30.

Strasser, K. and E. Hurt (2000). "Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export." Embo J 19(3): 410-20.

Strasser, K. and E. Hurt (2001). "Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p." Nature 413(6856): 648-52.

Strawn, L. A., T. Shen, N. Shulga, D. S. Goldfarb and S. R. Wente (2004). "Minimal nuclear pore complexes define FG repeat domains essential for transport." Nat Cell Biol 6(3): 197-206.

Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." Protein Expr Purif 41(1): 207-234.

References

Stutz, F., A. Bachi, T. Doerks, I. C. Braun, B. Seraphin, M. Wilm, P. Bork and E. Izaurralde (2000). "REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export." Rna 6(4): 638-50.

Sumathi, K., P. Ananthalakshmi, M. N. Roshan and K. Sekar (2006). "3dSS: 3D structural superposition." Nucleic Acids Res 34(Web Server issue): W128-32.

Suntharalingam, M. and S. R. Wente (2003). "Peering through the pore: nuclear pore complex structure, assembly, and function." Dev Cell 4(6): 775-89.

Svergun, D., C. Barberato and M. H. J. Koch (1995). "CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates." Journal of Applied Crystallography 28: 768-773.

Svergun, D. I. (1992). "Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria." Journal of Applied Crystallography 25: 495-503.

Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999)." Biophysical Journal 77(5): 2896-2896.

Svergun, D. I. and M. H. J. Koch (2003). "Small-angle scattering studies of biological macromolecules in solution." Reports on Progress in Physics 66(10): 1735-1782.

Svergun, D. I., M. V. Petoukhov and M. H. J. Koch (2001). "Determination of domain structure of proteins from X-ray solution scattering." Biophysical Journal 80(6): 2946-2953.

Tanner, N. K., O. Cordin, J. Banroques, M. Doere and P. Linder (2003). "The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis." Mol Cell 11(1): 127-38.

Terry, L. J. and S. R. Wente (2007). "Nuclear mRNA export requires specific FG nucleoporins for translocation through the nuclear pore complex." J Cell Biol 178(7): 1121-32.

Tran, E. J., Y. Zhou, A. H. Corbett and S. R. Wente (2007). "The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events." Mol Cell 28(5): 850-859.

Tseng, S. S., P. L. Weaver, Y. Liu, M. Hitomi, A. M. Tartakoff and T. H. Chang (1998). "Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export." <u>EMBO J</u> 17(9): 2651-2662.

van den Bogaart, G., A. C. Meinema, V. Krasnikov, L. M. Veenhoff and B. Poolman (2009). "Nuclear transport factor directs localization of protein synthesis during mitosis." <u>Nat Cell Biol</u> 11(3): 350-6.

Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." <u>Journal of Applied Crystallography</u> 36: 860-864.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (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 fold." <u>Embo J</u> 1(8): 945-51.

Walter, T. S., C. Meier, R. Assenberg, K. F. Au, J. S. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart and J. M. Grimes (2006). "Lysine methylation as a routine rescue strategy for protein crystallization." <u>Structure</u> 14(11): 1617-1622.

Weirich, C. S., J. P. Erzberger, J. M. Berger and K. Weis (2004). "The N-terminal domain of Nup159 forms a beta-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore." <u>Mol Cell</u> 16(5): 749-760.

Weirich, C. S., J. P. Erzberger, J. S. Flick, J. M. Berger, J. Thorner and K. Weis (2006). "Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export." <u>Nat Cell Biol</u> 8(7): 668-676.

Winey, M., D. Yarar, T. H. Giddings, Jr. and D. N. Mastronarde (1997). "Nuclear pore complex number and distribution throughout the Saccharomyces cerevisiae cell cycle by three-dimensional reconstruction from electron micrographs of nuclear envelopes." <u>Mol Biol Cell</u> 8(11): 2119-32.

Yang, Q. and E. Jankowsky (2006). "The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases." <u>Nat Struct Mol Biol</u> 13(11): 981-6.

Yang, Q., M. P. Rout and C. W. Akey (1998). "Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications." <u>Mol Cell</u> 1(2): 223-34.

Zeitler, B. and K. Weis (2004). "The FG-repeat asymmetry of the nuclear pore complex is dispensable for bulk nucleocytoplasmic transport in vivo." J Cell Biol 167(4): 583-90.

Zhao, J., S. B. Jin, B. Bjorkroth, L. Wieslander and B. Daneholt (2002). "The mRNA export factor Dbp5 is associated with Balbiani ring mRNP from gene to cytoplasm." EMBO J 21(5): 1177-1187.

8   Curriculum vitae

Personal Dates

Name	Holger-Markus von Moeller
Date of Birth	12.07.1978
Place of Birth	Leverkusen, Germany
Nationality	German
Marital status	single

Education

1985 - 1989	Primary School 'Theodor-Weinz', Bad Honnef, Germany
1989 - 1998	Grammar School 'Gymnasium am Oelberg', Oberpleis, Germany
	Degree: Diploma (Abitur) Grade: 1.8
1999 - 2002	Ruhr University of Bochum, Germany
	Prediploma in Biochemistry
2002 - May 2005	Free University of Berlin, Diploma in Biochemistry
	Exams (& Grades): Biochemistry (1.3), Molecular Biology and Genetics (1.0), Biophysics (1.7)
	Diploma thesis: Research Group of Prof. Dr. Wolfram Saenger 'Expression, Purification and Structure Analysis of the Protein Conjugated Bile Acid Hydrolase from *C. perfringens'* (Grade: 1.0)
Oct. 05 - Sep. 2009	PhD thesis EMBL Heidelberg (Joint degree with LMU Munich), Research group of Prof. Dr. Elena Conti 'Structural and Biochemical Studies of the Human DEAD-box Helicase Dbp5 and Nucleoporin Nup214 Involved in mRNA Export' (Grade: 1.0)
Since Sep. 2007	Max-Planck Institute of Biochemistry, Department of Structural Cell Biology, Prof. Dr. Elena Conti

Social Service

1998 - 1999	German Red Cross, Emergency Medical Service, Linz, Germany
Jul. - Sep. 1999	University Hospital, Neurosurgical Intensive Care Unit, Bonn, Germany,

Stipends

1999 - 2009	Fellow of the Prof. Rhein Foundation, Königswinter, Germany
2002 - 2003	Fellow of e-fellows.net
2005 - 2009	EMBL International PhD Fellowship