2010

Biochemical and Structural Analysis of the Nucleoporin Nup214 and its Involvement in mRNA Export

Johanna Napetschnig

BIOCHEMICAL AND STRUCTURAL ANALYSIS OF THE NUCLEOPORIN NUP214 AND ITS INVOLVEMENT IN mRNA EXPORT

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Johanna Napetschnig

June 2010

Biochemical and Structural Analysis of the Nucleoporin Nup214 and its involvement in mRNA export

Johanna Napetschnig, PhD

The Rockefeller University 2010

The mammalian nuclear pore complex (NPC) is a large proteinaceous assembly consisting of approximately 30 proteins, collectively termed nups. The NPC is the sole gateway between the nucleus and the cytoplasm, and facilitates transport of macromolecules across the nuclear envelope. Key steps in the export of mRNA from the nucleus to the cytoplasm are the transport through the NPC, and the subsequent remodeling of mRNA-protein complexes that occurs at the cytoplasmic side of the NPC. Crucial for these events is the recruitment of the DEAD-box helicase Ddx19 to the NPC. Nup214 is located on the cytoplasmic face of the NPC, is implicated in Ddx19 binding, and is a target for chromosomal translocations involved in leukemogenesis.

In order to gain a deeper understanding of the role of nups in leukemogenesis, and to make sense of the architecture and regulation of the mRNA export machinery at the NPC, I set out to biochemically and structurally characterize Nup214. In this thesis, I present the crystal structure of the human Nup214 N-terminal domain at 1.65 Å resolution. The structure reveals a seven-bladed β-propeller fold followed by a 30-residue C-terminal extended peptide segment (CTE). The CTE folds back onto the β-propeller and binds to its bottom face. Conserved surface patches on the Nup214 NTD reveal putative protein-interaction sites, one of which is crucial for the interaction with Ddx19. Using a comprehensive mutational and biochemical analysis, the interaction between the Nup214 NTD and Ddx19 is dissected.

The structure of the Nup214 NTD•Ddx19 in its ADP-bound state at 2.5 Å resolution reveals the molecular basis for the interaction between the two proteins. A conserved residue of Ddx19 is shown to be crucial for complex formation *in vitro* and *in vivo*. Strikingly, the interaction surfaces exhibit strongly opposing surface potentials, with the helicase surface being positively and the Nup214 surface being negatively charged. Ddx19 is shown to bind RNA only in its ATP-bound state, and the binding of RNA and the Nup214 NTD is mutually exclusive. Finally, I speculate that Nup214 is the ATP-exchange factor for Ddx19, and propose the Ddx19 ATPase cycle as the terminal step in mRNA export.

FÜR MEINE ELTERN

IN LIEBE UND DANKBARKEIT

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to all the people who have supported me throughout my time at The Rockefeller University as mentors, colleagues and friends.

I would first like to thank my advisor Dr. Günter Blobel for his guidance throughout my time at Rockefeller. With his enthusiasm and passion for science he will always be an inspiration and role model for me.

I am grateful to my committee members, Dr. Seth Darst and Dr. Thomas Sakmar, for bearing with my project and me for all these years. Their insightful comments and helpful advice was always motivating.

Special thanks go to Dr. Leemor Joshua-Tor, my external committee member, for taking the time to come all the way to New York.

I also thank Dr. André Hoelz for his guidance in the lab and for teaching me crystallography. Pushing my borders made me realize how much one can achieve. His genuine devotion and commitment to science is something I aspire to.

Thanks to all members of the Blobel lab for sharing advice and helpful discussions. Special thanks go to Vivien Nagy and Martin Kampmann for their

support throughout my PhD. Their friendship and company during late night hours, early mornings, and weekends in the lab meant a lot to me.

I also thank Dr. Daniel Andor, Dr. André Hoelz, Dr. Erik Debler and Dr. Elias Coutavas for critical comments on the thesis. I want to thank Susanne Kassube for technical assistance with the helicase mutants and Dr. Richard Wong for his assistance with light microscopy. Special thanks to Lourdes Quirolgico and Tita Isberto for their administrative support and help

I am grateful for the financial support that I received from the Rockefeller University's Women & Science Program and the David Rockefeller Graduate Program.

A "thank you" goes to the David Rockefeller Graduate Program and the Dean's Office for their unbureaucratic support. Kristen Cullen, Marta Delgado, Michelle Sherman and Christian Rosario deserve special acknowledgements for their loyalty and commitment to the graduate students.

Finally, I want to thank my parents, Johann and Dragi Napetschnig, for their support during all the years of education. I am forever grateful for their love and understanding. I'm grateful to my sister for her support and providing me with good excuses for traveling. Last but not least, I want to thank Daniel, the cornerstone of my life, for philosophical discussions about the English language, for bearing my "Austrian-isms" and his patience during my thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The research reported in this thesis is concerned with the biochemical and structural analysis of the nucleoporin Nup214 and its involvement in mRNA export. This introductory chapter provides an overview of the nuclear pore complex (NPC), of which Nup214 is a component. Sections 1.1 to 1.2 describe the general architecture, composition and transport function of the NPC. Sections 1.3 to 1.6 explain how macromolecules are transported across the nuclear envelope, with a focus on the export of mRNA and the nucleoporins and NPC-associated proteins involved. Finally, section 1.7 details the involvement of nucleoporins in cancer development.

1.1 The nuclear pore complex (NPC)

An important step in the evolution of eukaryotes was the compartmentalization of the cell. Unlike bacteria, which consist of a single compartment surrounded by a plasma membrane, eukaryotic cells are divided into several membrane-enclosed compartments, or organelles, that each have a distinct and specific function in the cell. With compartmentalization, the need for communication between organelles and therefore regulated transport arose. One important compartment of the eukaryotic cell is the nucleus, which harbors the genetic material and is the primary site of DNA and RNA synthesis. The nuclear envelope partitions the nucleus from the cytoplasm, where protein synthesis and degradation take place. The only gate in the nuclear envelope, and thus responsible for all

communication between the cytoplasm and the nucleus, is the nuclear pore complex (NPC).

1.1.1 Architecture of the NPC

Mammalian NPCs were first described as “holes” or “pores” in electron microscopy (EM) thin sections of nuclear membranes[1] (**Figure 1 A** to **C**). The NPC is embedded in the nuclear envelope, which is a double lipid bilayer and consists of an outer (ONM) and inner (INM) nuclear membrane.

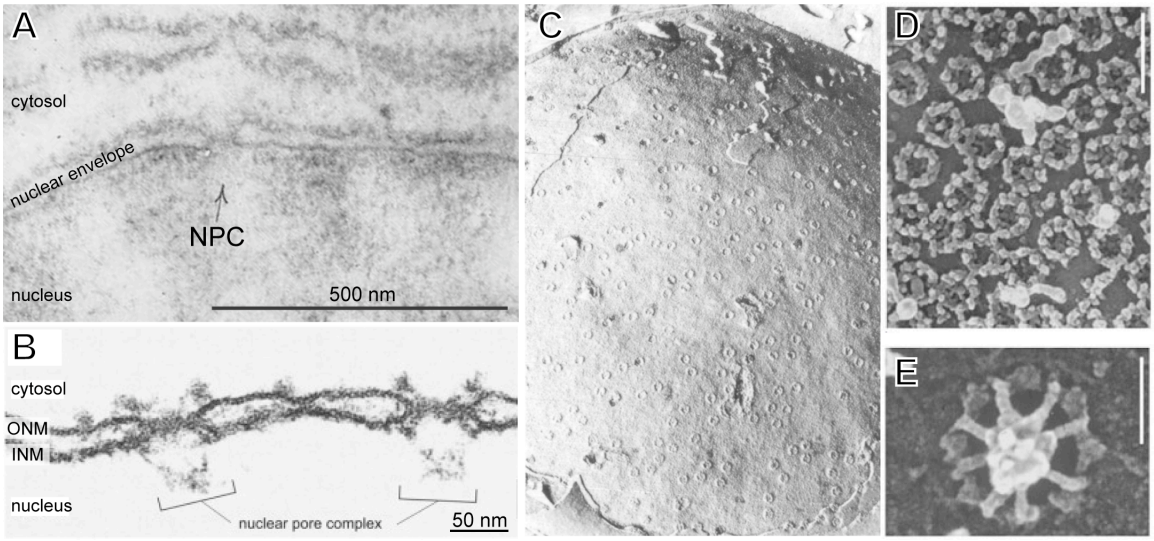


Figure 1: **Electron microscopy images.** EM images of thin sections of (**A**) a mouse pancreas cell[1] and (**B**) an amphibian oocyte nuclear membrane[2]. The cytoplasmic and nuclear sides of the nuclear envelope as well as the NPCs are indicated. (**C**) Freeze fracture micrographs of nuclear membrane from rat liver cells[3]. Field-emission scanning EM image of (**D**) the cytoplasmic face (scale bar = 250 nm) and (**E**) the nuclear face (scale bar = 50 nm) of the nuclear envelope[4].

The NPC is one of the largest proteinaceous assemblies in the cell and has a mass of ~120 MDa in vertebrates, a diameter of ~125 nm and a height of ~110 nm[5] (**Figure 2**). The ~60 MDa yeast NPC is much smaller, with an approximate diameter of ~100 nm and a height of ~30 nm[6, 7]. Despite the differences in molecular mass and size of the mammalian and yeast NPC, the overall architecture of the NPC is conserved.

Electron microscopy images have revealed the NPC to have a doughnut-like shape, with distinct features on the cytoplasmic and nuclear faces, termed the cytoplasmic filaments and the nuclear basket, respectively[8, 9] (**Figure 1 B**). EM investigations in the late 1960's revealed an eight-fold radial symmetry perpendicular to the nuclear envelope as well as a pseudo two-fold symmetry in the plane of the nuclear envelope (**Figure 1 D**, **E** and **Figure 3**)[10]. Recent advances in cryo-electron tomography (cryo-ET) studies have refined our picture of the overall architecture of the NPC[11, 12] (**Figure 3**).

The NPC is divided into three distinct parts: the cytoplasmic filaments, a symmetrical core and the nuclear basket. The cytoplasmic filaments are localized at the cytoplasmic face of the NPC and protrude into the cytoplasm. Although the cytoplasmic filaments are largely flexible, a cytoplasmic ring that is connected to the symmetric core of the NPC has been observed by cryo-ET[11]. The symmetric core of the NPC is embedded in the membranes of the nuclear envelope and consists of the spoke ring and nuclear ring (**Figure 3**). A lumenal connector element spans the space between the INM and ONM. The lumenal connector is

attached to the membrane at the perinuclear side of the cytoplasmic and nuclear rings and is thought to stabilize the NPC[12]. The nuclear basket is located on the nuclear side of the NPC, and consists of the distal ring and nuclear filaments, which attach to the symmetric core. In addition, an electron-dense structure, termed the central plug, can be observed by cryo-ET within the central channel[9, 11, 13]. The central plug was suggested to be cargo in transit[14] and overlaps with the observed location of cargo in the central channel[12].

1.1.2 Composition of the NPC

Despite the large molecular mass of the vertebrate NPC, proteomic studies have shown that the NPC consists of only ~30 different proteins[15, 16]. This seemingly small number can be explained by the internal symmetry of the NPC and by the large mass of the NPC proteins, which are collectively termed nucleoporins (nups). Nups are found in copies of multiples of eight and often consist of several modular domains[17, 18]. In addition to other secondary structure folds, the most prevalent secondary folds of the nups are the β-propeller fold, α-helical repeats, coiled-coils and the natively unfolded FG-domains featuring repeats of the characteristic phenylalanine-glycine (FG) motif[18, 19].

The nups can be classified according to their position in the NPC as determined by immuno-electron microscopy labeling[15, 20], as well as sequence and structural analysis[21-23] (**Figure 4**). The asymmetrically localized nups form the nuclear basket (Tpr, Nup153, Nup50) or the cytoplasmic filaments (Nup88,

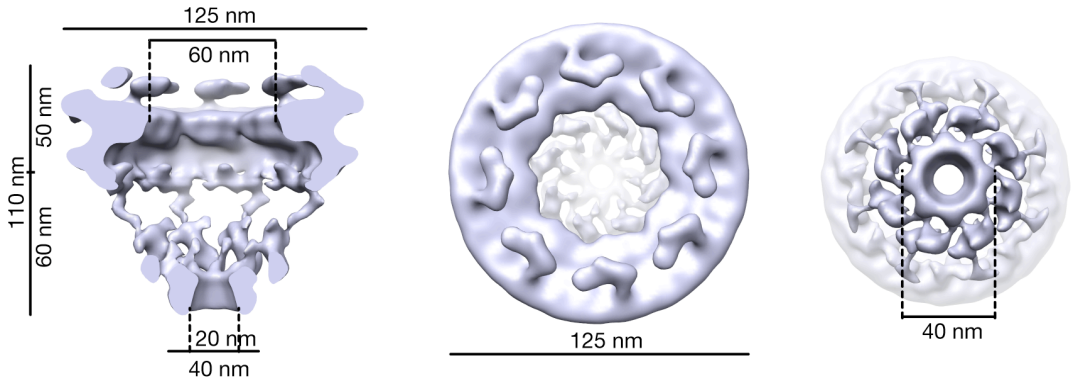


Figure 2: Cryo-electron tomography structure of the NPC. Surface rendition of the structure of *Dictyostelium* NPC (EMD-1097) at a resolution of 85 Å depicting the approximate dimensions of the NPC in a cutaway view (*left*), a cytoplasmic face (*middle*) and nuclear face view (*right*)[11]. For clarity, the cytoplasmic plug was omitted.

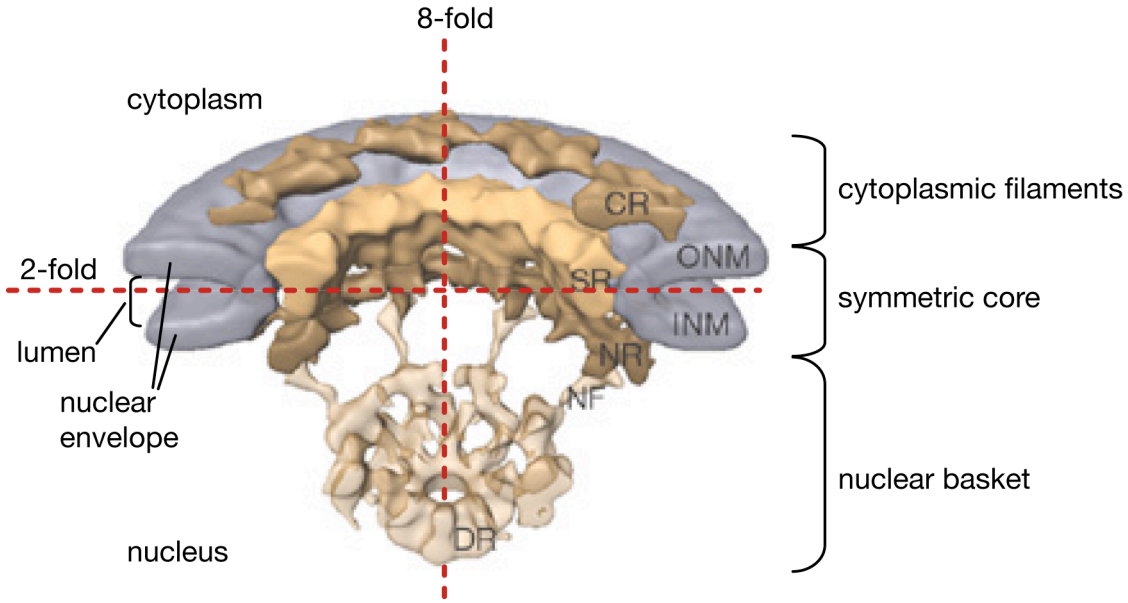


Figure 3: Overall architecture of the NPC. Cut-away view of the cryo-electron tomography structure of the *Dictyostelium* NPC at 58 Å resolution. Subjective segmentation for the nuclear filaments (NF), cytoplasmic ring (CR), spoke ring (SR), nuclear ring (NR) and distal ring (DR) are indicated in yellow. The inner (INM) and outer (ONM) nuclear membrane are shown in violet. Figure adapted from Beck et al[12].

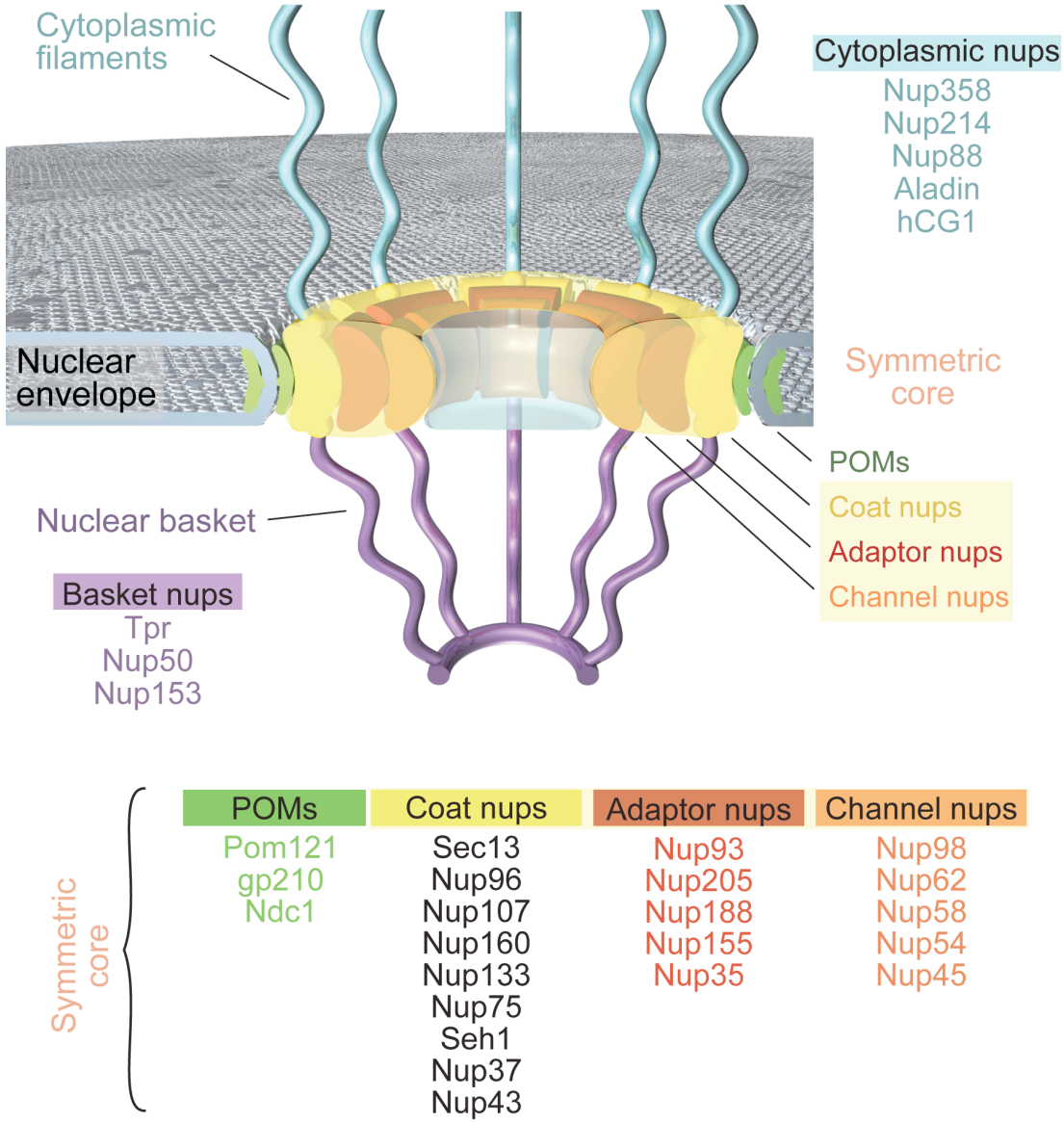


Figure 4: Schematic model of the NPC. The vertebrate nups that make up the cytoplasmic filaments (blue), symmetric core (shades of yellow) and nuclear basket (violet) are listed. The transport barrier, consisting of FG-repeat domains of nups in the central channel of the NPC, is symbolized by the blue, transparent plug. Figure adapted from [24].

hCG1, Aladin, Nup214, Nup358). The symmetrically localized nups are found in the symmetric core of the NPC. The nups of the symmetric core can be further categorized into the integral pore membrane proteins (POMs), the coat, the adaptor and the channel nups[25] (**Figure 4**). The POMs contain transmembrane domains and anchor the NPC in the nuclear envelope. The coat nups are involved in membrane stabilization[19], while the adaptor nups are thought to cushion dynamic changes of the channel nups[24]. The FG-repeat domain containing channel nups line the central transport channel and provide docking sites for the transport receptors[25]. Additionally, the channel nups were proposed to adjust the diameter of the central transport channel by circumferential sliding during transport of large cargo[21].

The modular architecture of the NPC is not only apparent in its internal symmetry and modular nup domain structure, but also in the existence of building blocks or sub-complexes within the NPC. A dramatic change in cellular organization is the nuclear envelope breakdown at the onset of mitosis in higher eukaryotes. This breakdown is marked by an increase in nuclear envelope permeability initiated by NPC disassembly[26]. Destabilization of the nup protein interactions by phosphorylation[26] is followed by dissociation of stable nup sub-complexes from the NPC[27]. These sub-complexes can be reconstituted[28] or isolated from interphase cells using detergents and salts[29, 30], and dissected biochemically[31].

1.1.3 Arriving at an atomic-resolution model of the NPC

Although at ~60 Å resolution the general architecture of the NPC is evident from cryo-EM tomography studies, the molecular details of sub-complexes or individual proteins will remain elusive until higher resolution information about the NPC becomes available[12]. Several X-ray structures of nups or nup complexes have been solved in the last eight years[21, 23, 25, 32-46]. However, the large mass and flexibility of the NPC makes the determination of a high-resolution structure of the entire NPC a seemingly impossible mission.

To overcome this limitation, a molecular "divide-and-conquer" approach is employed. The approach exploits the modular character of the NPC to bridge the resolution gap between the cryo-EM map of the NPC (~60 Å) and crystal structures (~3 Å)[28]. The NPC is first "divided" biochemically into sub-complexes, for which cryo-electron tomography structures at a higher resolution can be obtained. If the resolution is sufficiently high (<20 Å), the NPC can then be "conquered" by fitting atomic resolution structures into the cryo-EM density envelope with high confidence.

This strategy was pioneered using the Y-shaped heptameric complex[22, 23, 28] (**Figure 5**). The Y-shaped complexes are localized close to the nuclear envelope *in vivo*[15] and have been suggested to serve as "membrane-curving modules" to ''coat'' and stabilize the curvature between the inner and outer nuclear membranes[19, 47]. In yeast, the Y-shaped complex, also called the Nup84 complex, consists of Nup133, Nup84, Nup145C, Sec13, Nup85, Seh1 and

Nup120[48], and can be reconstituted from recombinant proteins or isolated using nonionic detergents and salts from budding yeast[28, 30]. The equivalent Y-shaped complex in vertebrates consists of Nup160, Nup133, Nup107, Nup96, Nup75, Nup43, Nup37, Seh1 and Sec13, and can also be isolated using detergents and salts[29, 49, 50]. The three-dimensional structure of the yeast Y-shaped heptameric complex was determined at a resolution of 35 Å[22], and the available crystal structures were docked into the EM envelope[22, 23] (**Figure 5**).

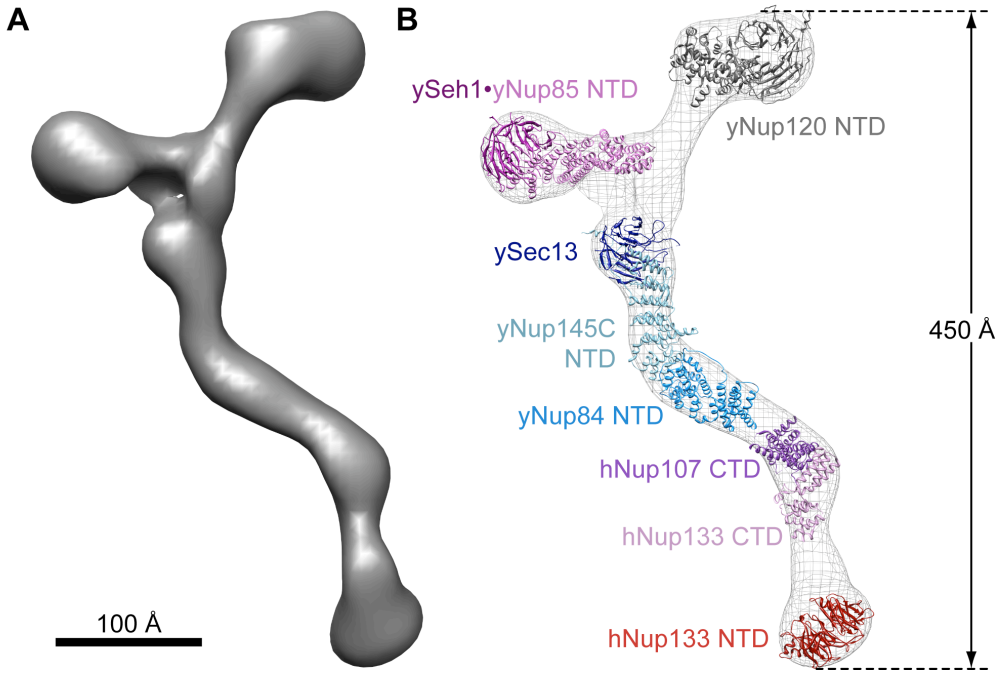


Figure 5: The "divide-and-conquer" strategy is exemplified by the heptameric complex. (**A**) Surface rendition of the EM envelope of the heptameric complex at a resolution of 35 Å[22]. (**B**) Docking of available crystal structures into the EM envelope of the Y-shaped complex[22, 23]. The EM map is shown in mesh representation; crystal structures of ySeh1•Nup85[32, 38], yNup120[40, 51], ySec13•yNup145C•yNup84[23, 25, 52], Nup107 CTD•Nup133 CTD[37] and the Nup133 NTD[34] are shown in ribbon representation. Figure adapted from[22, 23].

1.2 The NPC functions as a selective barrier

All transport of macromolecules in and out of the nucleus is dependent on the NPC, the only opening in the nuclear envelope. The NPC is a doughnut-shape complex with a ~60 nm wide central transport channel[11]. The central transport channel is lined with the FG-repeat domains of the channel nups[53]. While molecules smaller than ~40 kDa can passively diffuse through the central channel of the NPC, larger molecules require facilitated transport[54] **(Figure 6)**. The FG motifs serve as binding sites for nuclear transport receptors called karyopherins or Kaps and facilitate their transport through the NPC[55].

The transport of macromolecules is a four-step process and will be discussed in detail in the next section. In short, (1) the Kaps recognize and bind to their cargo, (2) the cargo-bound Kaps interact with the FG-domains lining the central channel of the NPC, (3) the Kap•cargo complex moves through the central channel and (4) a small GTPase, called Ran, dissociates the Kap•cargo complex at the other side of the nuclear envelope.

The transport barrier of the central channel of the NPC consists of the FG-repeat domains of the channel nups[53]. The FG-repeat domains are natively unfolded and contain multiple phenylalanine-glycine (FG) motifs separated by hydrophilic residues[56, 57]. The transport receptors can bind to the FG motifs and traverse the transport barrier of the NPC[58, 59].

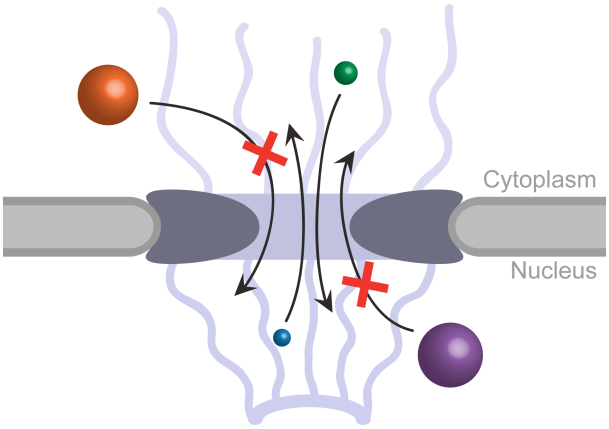


Figure 6: Transport function of the NPC. While molecules smaller than ~40 kDa (green and blue spheres) can freely diffuse through the NPC, larger molecules (orange and violet spheres) require facilitated transport. Image from [24].

The exact mechanism by which the interaction between Kaps and FG-regions facilitates transport and maintains the barrier for non-Kap proteins is still under debate. Transport through the NPC requires the formation and dissociation of Kap•cargo complexes, while transit of the complex through the central channel of the NPC does not require energy *per se*[60]. The direction of transport can be reversed *in vitro*[61], excluding explanations involving a molecular motor or an affinity gradient of Kap-binding sites on the FG-repeat domains within the central channel. Several models have so far been proposed (**Figure 7**).

Rout and coworkers proposed the “virtual gate” model in which the NPC can be viewed as an entropic barrier[15]. The free ends of FG-repeat domains are highly mobile and act as repulsive bristles occluding molecules from the central channel. The entropic repulsion behavior of FG-domains has been observed *in vitro*[56]. Specific binding of Kaps to the FG-domains of the peripheral nups

increases the probability of Kaps entering the channel and allows transport by Brownian motion[62]. Therefore, binding of the Kaps to the FG-domains overcomes the entropic barrier of entering the central channel, allowing Kap•cargo complexes to move through the NPC.

The “selective phase” model was suggested by Görlich and coworkers and proposes that the FG-repeats interact with each other via weak hydrophobic interactions, thereby forming a sieve-like hydrophobic FG-mesh, the “selective phase”[63, 64]. Indeed, FG-mediated cross-linking of FG-repeat regions was found to form an elastic, reversible hydrogel[65]. Transport Kap•cargo complexes can enter the selective phase and transiently open the FG-mesh barrier by binding to the FG-repeats. This allows Kap•cargo complexes to traverse the central channel[66].

Rexach and coworkers proposed the “dual gate” model after an extensive *in vitro* study of FG-domain interactions[67]. To test for interactions between FG-domains, soluble fluorescence-labeled FG-domains were mixed with bead-immobilized GST-FG nups. Binding of the FG-domains was visible as fluorescence around the otherwise dark beads with the GST-FG nups. According to the dual gate model, the FG-repeat domains of the symmetric NPC core are cohesive and function as a selective mesh, whereas FG-domains of the peripheral structures function exclusively as repulsive bristles and form a “second” gate at the nuclear basket[68].

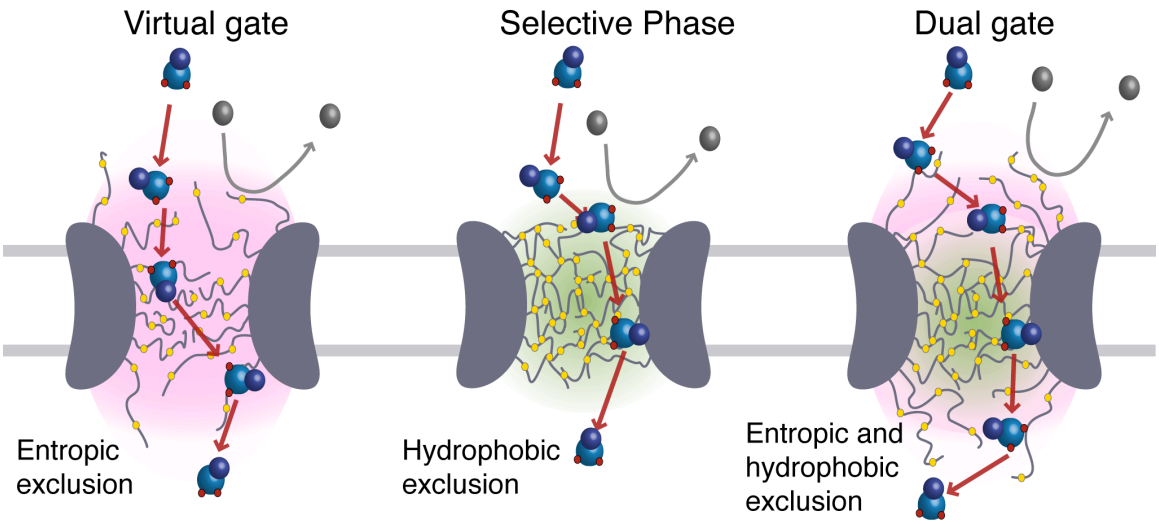


Figure 7: The three models of selective transport. In the virtual gate model the FG-repeat domains act as repulsive "bristles" and occlude the space in the central channel (pink). In the selective phase model the FG-repeats form a sieve-like hydrophobic FG-mesh (green). The dual gate model proposes that the FG-domains at the center of the NPC function as a selective mesh (green), while the peripheral FG-domains function as repulsive "bristles" (pink). The FG-domains are shown as grey lines protruding into the central transport channel of the NPC. Kap•cargo complexes (blue) and molecules not bound to a Kap (grey) are depicted as spheres. The Kap-binding sites on the FG-repeats and FG-binding sites on Kaps are indicated by yellow and red dots, respectively.

1.3 Protein transport

Protein transport through the NPC is dependent on short sequence elements in the cargo proteins that are recognized by karyopherins (Kaps), the transport factors (**Figure 8**). The sequence elements are called nuclear localization sequences (NLS) and nuclear export sequences (NES). Proteins containing a NLS are transported by import Kaps into the nucleus, whereas cargo containing a NES are transported into the cytoplasm by export Kaps[69]. Classical NLSs contain a cluster of three to five positively charged residues and can be monopartite or bipartite[69, 70]. The bipartite NLSs contain an additional cluster of arginine/lysine residues that is spaced 10 to 12 residues from the first cluster. NESs are diverse and less defined, but can contain a short leucine-rich signal[71].

The transport is driven by a small GTPase called Ran. Ran can be found in a GTP-bound and a GDP-bound state in the cell. While Ran-GTP has a high concentration in the nucleus, the cytoplasm has a low Ran-GTP concentration, which results in a Ran-GTP gradient across the nuclear envelope[72]. The two regulators of GTPase activity, RCC1 and RanGAP1, establish the Ran-GTP gradient by their localization to distinct cellular compartments. RCC1 is the Guanine nucleotide exchange factor (GEF) and RanGAP1 is the GTPase-activating protein (GAP) for Ran. While the RanGEF RCC1 is bound to nucleosomes in the nucleus and stimulates nucleotide exchange of Ran-GDP on

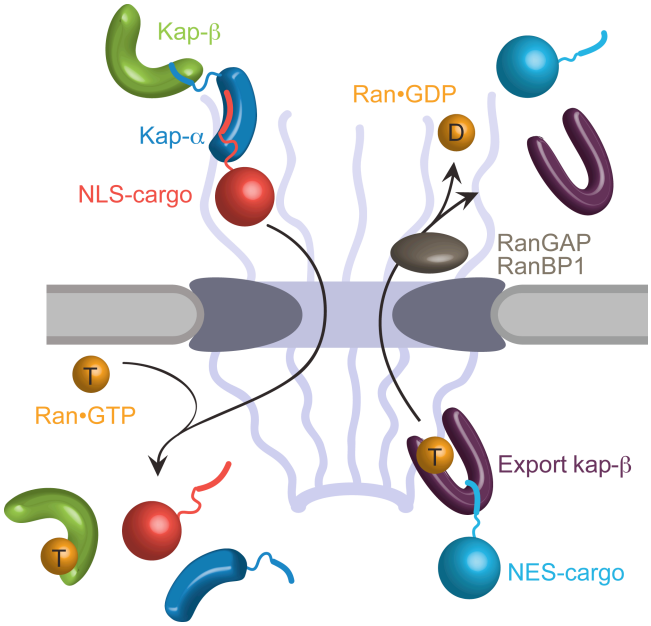


Figure 8: **Protein transport through the NPC.** Nucleocytoplasmic transport is facilitated by nuclear localization (NLS) and nuclear export signals (NES), which are recognized by transport factors, collectively called Kaps. The transport of Kap•cargo complexes is dependent on the Ran gradient, with high Ran-GTP concentration in the nucleus and Ran-GTP hydrolysis in the cytoplasm. Export Kap•cargo complexes can only be formed in the presence of Ran-GTP, while import Kap•cargo complexes are destabilized by it. Image reproduced from [24].

chromatin, RanGAP1 is localized to the cytoplasm[55, 73, 74]. RanGAP1 needs additional activating cofactors, the Ran-binding proteins RanBP1 and RanBP2, to stimulate GTP hydrolysis and disassemble the Kap•cargo complexes[75]. In higher eukaryotes, RanGAP1 becomes sumoylated and binds to the cytoplasmic side of the NPC via interaction with RanBP2, also known as Nup358[76-80]. Therefore, GTP-hydrolysis of Ran is only activated at the cytoplasmic face of the NPC.

Transport through the NPC is driven by preventing the return of transport complexes to the donor compartment and is regulated by Ran. While Ran-GTP dissociates import-cargo•Kap complexes in the nucleus, export-cargo•Kap complexes are only formed in the presence of Ran-GTP and are dissociated upon GTP-hydrolysis in the cytoplasm[55, 81, 82]. When transport complexes are disassembled, new complexes are transported to the target compartment to restore equilibrium. Transport through the pore can therefore be viewed as a form of Brownian motion, where energy is required for transport complex formation and disassembly.

1.3.1 Recognition of cargo by the karyopherin transport factors

The majority of nucleocytoplasmic transport factors belong to the β-karyopherins (Kap-β)[83]. The Kap-β recognize their cargo molecules either directly via a short sequence (the NES or NLS) or via an adaptor karyopherin (**Figure 8**). The Kaps mediating nuclear import are termed import Kap-β (or importins), while Kaps mediating nuclear export are termed export Kap-β (or exportins).

β–Karyopherins are large, HEAT-repeat containing proteins[70]. HEAT-repeats are tandem repeats of about 40 residues that fold into a pair of α-helices. Multiple HEAT-repeats stack together in parallel with a slight right-handed twist. This Karyopherins fold into a superhelical architecture with an inner concave and an outer convex surface.

1.3.1.1 Nuclear import

Proteins carrying classical NLSs are transported by karyopherin-β1 (Kap-β1). Kap-β1 binds not directly to NLS-containing cargo, but instead uses Kap-α as transport adaptor[84]. However, Kap-β1 can also bind to some cargo directly, as is the case with the sterol regulatory element-binding protein (SHREBP-2) or the parathyroid hormone-related protein (PTHrP)[85, 86]. Kap-α cannot interact with the FG-repeats by itself and contains an inhibitory segment at the N-terminus that functions as an 'auto-NLS' [70, 87]. The 'auto-NLS' binds to the NLS-binding site of Kap-α and prevents any interaction with cargo. Binding of Kap-β1 to the 'auto-NLS' (also called importin-β-binding domain, IBB) of Kap-α frees up the NLS-binding site, allowing cargo to bind to the Kap-α•Kap-β complex and transport into the nucleus.

Structures of Kap-β1 have revealed that the convex surface is used for interaction with the FG-repeats[59] (**Figure 9 A**), while the cargo binds at the inner concave surface[84, 88] (**Figure 9 B**).

The Kap-β1 transport adaptor Kap-α contains an 'auto NLS' or IBB domain as well as a NLS-binding domain. The NLS-binding domain consists of up to 10 armadillo repeats[88]. Armadillo repeats are structurally similar to the HEAT repeat fold, but are less variable in the length of their repeat and in amino acid sequence[89]. Similarly to Kap-β, Kap-α folds into an elongated shape with a convex and concave surface[88, 90] (**Figure 9 C**). Kap-α has two clusters of lysine-

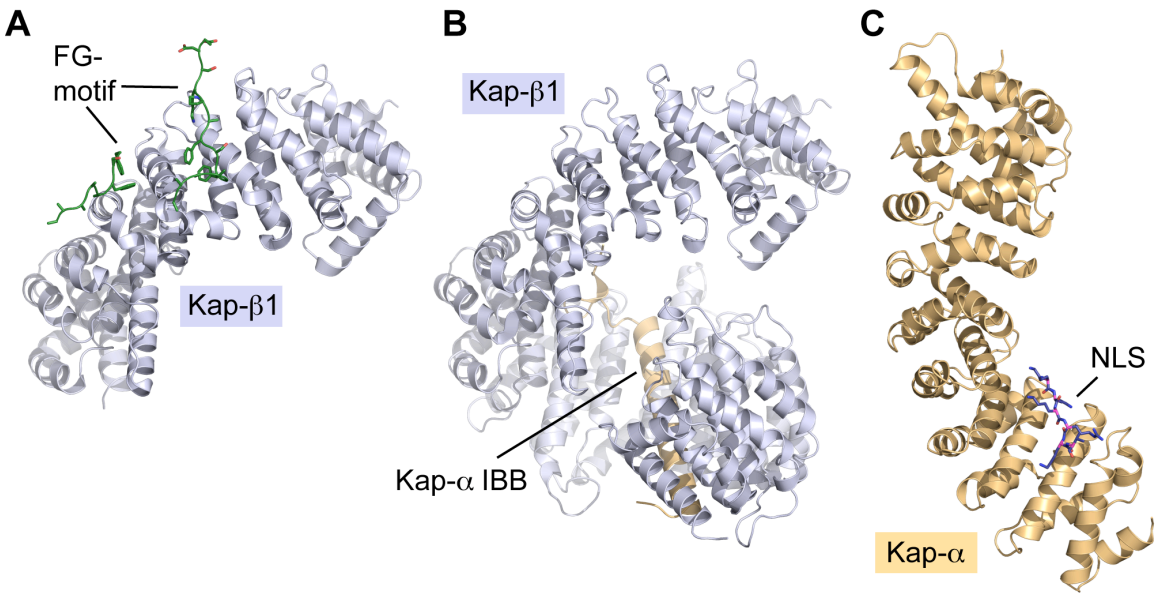


Figure 9: Structures of Karyopherins. Ribbon representation of Kap-β1 bound to (**A**) FG-motif peptides (stick representation, green) (PDB code 1F59) and (**B**) the IBB of Kap-α (light orange) (PDB code 1QGK). **(C)** Kap-α bound to a classical NLS (stick representation, blue) (PDB code 1BK6).

binding pockets on its concave surface and binds NLS peptides in an extended conformation.

After transport into the nucleus, the Kap-β•Kap-α•cargo complex is dissociated by Ran-GTP[55]. Binding of Ran-GTP to Kap-β releases the 'auto-NLS' (IBB) from Kap-β. Intramolecular binding of the auto-NLS in turn induces the release of the cargo-NLS from Kap-α, thereby releasing the cargo in the nucleus. The Ran-GTP-bound Kap-β is then transported back to the cytoplasm to start a new cycle of transport.

1.3.1.2 Nuclear export

While Ran-GTP dissociates import-Kap•cargo complexes in the nucleus, assembly of export-Kap•cargo complexes requires Ran-GTP. The export-Kap Cse1 (CAS in humans) is responsible for the export of Kap-α from the nucleus, and structural analyses have confirmed that its architecture is similar to that of import Kap-β[91] (**Figure 10 A**). In the ternary complex comprising Kap-α, Cse1 and Ran-GTP, Ran-GTP is bound to the N-terminal arch of Cse1. The cargo, Kap-α, is bound by both arches of Cse1, and also interacts with Ran-GTP[92] (**Figure 10 B**).

Leucine-rich NESs are recognized by the export-Kap Crm1 (exportin 1). Crm1 is also responsible for the export of snurportin from the nucleus. Snurportin is the import adaptor for snRNP and and will be discussed in section 1.4.3. The structure of the ternary complex comprising Crm1, snurportin, and Ran-GTP revealed that Crm1 adopts a ring-like, almost planar conformation[93, 94] (**Figure 11**). Ran is bound at the inside of Crm1, while snurportin binds to the outside of the Crm1 ring and not to the concave surface as seen in other cargo-Kap complexes[84, 88].

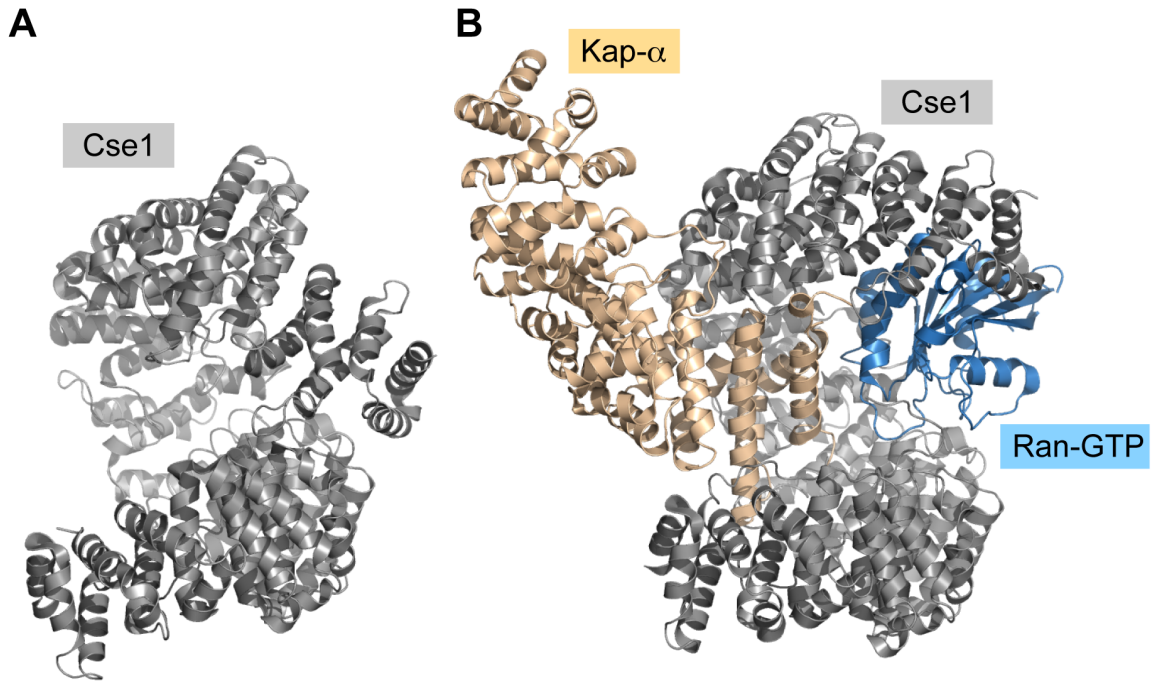


Figure 10: Structure of the export-Kap Cse1. Structural overview of (**A**) Cse1 unbound (grey) and (**B**) in complex with its cargo Kap-α (wheat) and Ran-GTP (blue). (PDB code 1Z3H and 1WA5).

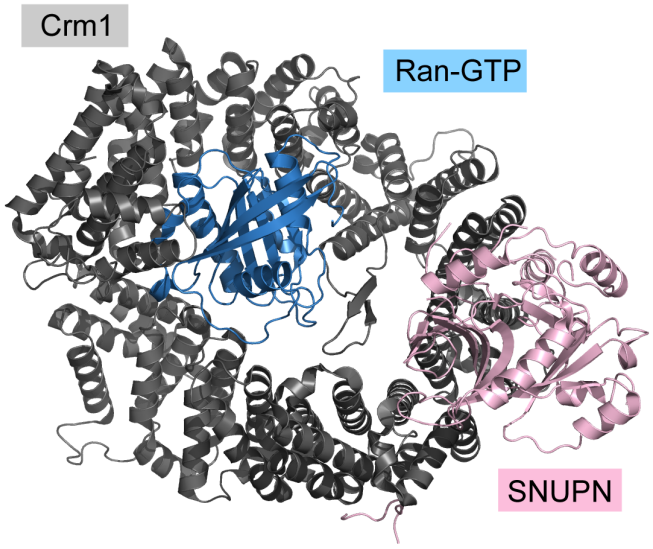


Figure 11: Structure of the export-Kap Crm1 in complex with Ran-GTP and snurportin. Structural overview of Crm1 bound to the cargo snurportin and Ran-GTP. Crm1 is shown in grey, snurportin (SNUPN) in rose and Ran-GTP in blue (PDB code 3GJX).

1.4 RNA transport

The transport of transfer RNA (tRNA), microRNA (miRNA), small nuclear RNA (snRNA) and ribosomal RNA (rRNA) follows the general paradigm of protein transport in the sense that it also depends on transport receptors of the karyopherin family and the Ran-gradient and cycle (**Figure 12**). However, mRNA export uses a transport receptor unrelated to the karyopherin family, which does not bind Ran and will be discussed in section 1.5.

1.4.1 Export of tRNA

There are roughly 40 different tRNAs in eukaryotic cells. tRNAs are transcribed by RNA polymerase III in the nucleus and fold into a clover-leaf like structure[95]. Before tRNAs are transported to the cytoplasm, multiple RNA processing steps have to be completed, including the removal of the 5' and 3' trailers, base modifications and addition of CCA-nucleotides to the 3' end of the tRNA[95]. The class-specific transport receptor for tRNA is the karyopherin exportin-t[96, 97]. Exportin-t binds directly to tRNAs in a Ran-GTP dependent manner and recognizes a tertiary structural element as a nuclear export signal (NES), which consists of properly processed 3' and 5' tRNA termini[98] (**Figure 12 A**).

1.4.2 Export of miRNA

miRNAs are non-coding RNAs that are involved in the gene regulation of a wide range of biological processes including apoptosis, immunity, cell differentiation

and development[99-101]. miRNAs are transcribed by Pol II or III, contain a characteristic stem-loop structure, and initially receive a 5' cap and a poly(A)-tail[102, 103]. This primary transcript is cleaved by the RNAse Drosha to yield the ~65-nucleotide long, export-competent intermediate called pre-miRNA. The pre-miRNA contains a double-stranded stem-loop RNA helix with a short 3' overhang that is recognized by the miRNA karyopherin exportin-5 in a Ran-dependent manner[104] (**Figure 12 B**). After export into the cytoplasm, GTP-hydrolysis disassembles the export complex. The pre-miRNA is further cleaved by Dicer and incorporated into the RNA-induced silencing complex (RISC)[105].

1.4.3 Transport of snRNA

Most spliceosomal RNAs (snRNAs) are synthesized by Pol II and acquire a 5' cap[106]. The snRNAs do not have a distinct karyopherin, but use the general karyopherin Crm1 in combination with the NES-containing adaptor protein named PHAX. Recruitment of Crm1 to the pre-snRNA requires phosphorylation of PHAX in the nucleus as well as the cap-binding complex (CBC) and Ran-GTP[107, 108] (**Figure 12 C**). After export from the nucleus, GTP-hydrolysis of Ran as well as dephosphorylation of PHAX are required to dissociate the export complex and release the pre-snRNA[107]. In the cytoplasm, the pre-snRNA associates with a heptameric ring of Sm proteins, inducing tri-methylation of the RNA and 3' processing. The tri-methyl cap together with the Sm proteins provides a composite NLS that is recognized by the nuclear import-adaptor snurportin and

the karyopherin Kap-β[109]. After re-import into the nucleus the snRNPs assemble into the spliceosome[110].

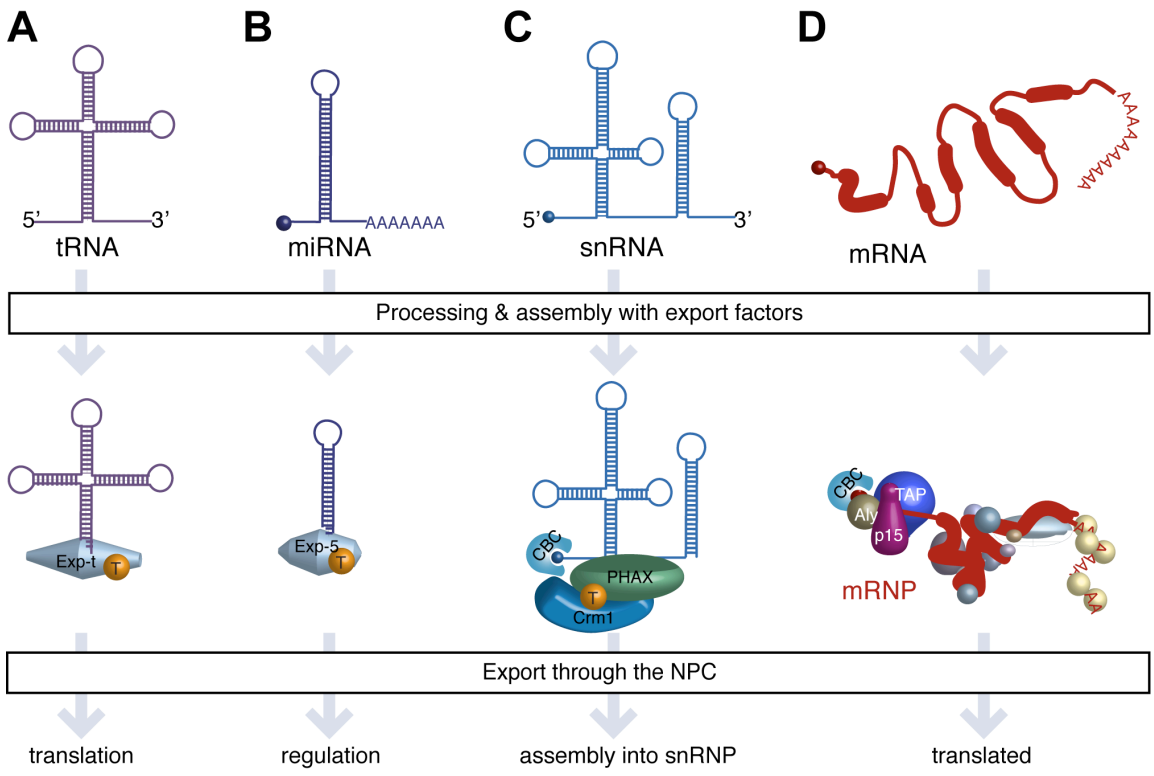


Figure 12: Assembly of export-competent RNA for tRNA, miRNA, snRNA and mRNA. For each RNA class the primary transcript (*top*) as well as the export competent RNA (*bottom*) with assembled transport factors are shown. The transport receptors are (**A**) Exportin-t (Exp-t) for tRNA, (**B**) Exportin-5 (Exp-5) for miRNA, (**C**) Crm1 for snRNA and (**D**) TAP/p15 for mRNA. The transport adaptor proteins are the Cap binding protein complex (CBC), PHAX for snRNA, and CBC and Aly for mRNA. Ran-GTP is shown as an orange sphere.

1.5 mRNA export

While RNA classes with highly related structures, like tRNAs or miRNAs, can be recognized by class-specific transport receptors as described above, mRNAs vary substantially in sequence, length and structure. Typically, eukaryotic mRNAs are synthesized by Pol II as pre-mRNA and undergo rigorous processing in the nucleus. Splicing, 5' capping, editing and polyadenylation deposit a multitude of proteins on the mRNA and generate a messenger ribonucleoprotein (mRNP)[111] (**Figure 12 D**).

The transport process of the mRNP through the NPC can be directly visualized in the salivary gland cells of *Chironomus tentans* using EM. The giant transcript of the Balbiani ring (BR) genes is cotranscriptionally packaged into a large mRNP particle (BR particle). The BR mRNP particle folds into a ring-like structure with four domains and a diameter of 50 nm (**Figure 13 B**)[112, 113]. Domain 1 is synthesized first and contains the 5' end of the mRNA, whereas domain 4 contains the 3' end. When the BR particle reaches the NPC, the structure of the BR mRNP changes dramatically[114] (**Figure 13 A**). The particle unfolds, elongates, and becomes rod shaped. The elongated particle has a diameter of ~25 nm as it travels through the NPC and emerges with the 5' end of the transcript in the lead[115]. Instead of reforming the globular particle, the 5' end of the unfolded fiber associates with ribosomes almost immediately upon entering the cytoplasm[115] (**Figure 13 C**).

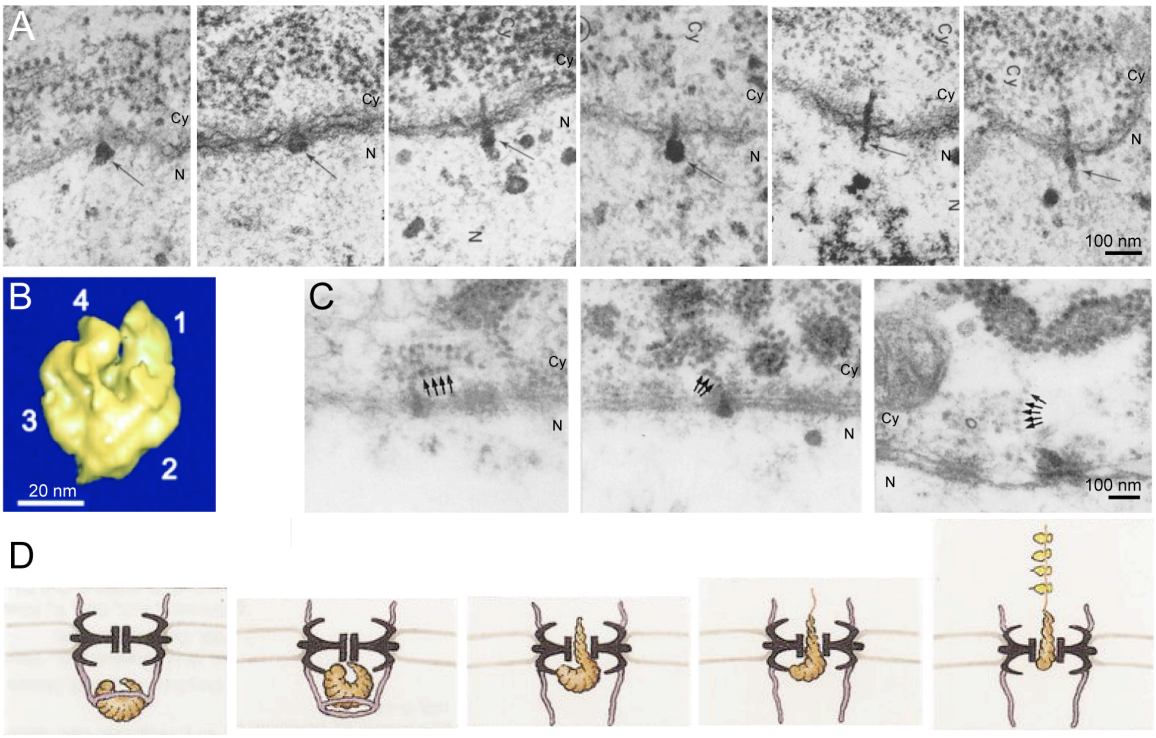


Figure 13: Translocation of the BR mRNP particle through the NPC. (**A**) A series of electron-micrograph shows the movement of the BR mRNP particles (arrow) through the NPC[114]. The BR mRNP docks as a globular particle to the NPC and elongates during transport (*left to right*). (N, nucleus; Cy, cytoplasm). (**B**) Surface rendition of a cryo-ET structure of a free, nuclear BR mRNP particle at 4 nm resolution[112]. The four domains of the BR particle are indicated. (**C**) Electron micrograph of BR mRNP particles with their leading 5' end associating with ribosomes (arrows). (**D**) Schematic representation of the movement of the BR mRNP particle shown in A. The very right picture depicts a particle with associated ribosomes. Figures reproduced from[115, 116].

The four main processing events during formation of a mature mRNP are 5' capping, splicing, 3'-end cleavage and polyadenylation[111]. Each of these processes impacts mRNA export in two ways. First, if the pre-mRNA is not properly processed, the pre-mRNA gets targeted for degradation. Second, during the processing, factors necessary for export are recruited.

After transcription of ~20 to 30 nucleotides, a 7-methylguanosine cap is added to the nascent pre-mRNA. This 5' cap protects the transcript from degradation in the nucleus[117]. Microinjection experiments in Xenopus have shown that uncapped mRNA is poorly exported[118]. The 5' cap is bound by the nuclear cap-binding complex (CBC), which consists of CBC20 and CBC80[119]. The process of splicing is catalyzed by the spliceosome, a complex composed of multiple proteins and RNAs, which removes introns from the pre-mRNA[120, 121]. Splicing deposits a multi-protein complex, the exon-junction complex (EJC), near the junction of exons[122, 123]. Finally the polyadenylation site is recognized in the 3' untranslated region (UTR) and the pre-mRNA cleaved by the cleavage and polyadenylation specificity factor (CPSF). After cleavage, a poly(A) tail is added by the poly(A) polymerase (PAP) and bound by poly(A) binding proteins (PABP)[124].

Some of the proteins in the mRNP are mRNA export adaptors and mRNA export receptors. The mRNA export receptor complex TAP/p15 consists of a large (~70 kDa, TAP) and a small (~15 kDa, p15) subunit that together act as a

single functional unit. TAP contains an N-terminal domain, which interacts with cargo, and a C-terminal domain that is required for interaction with the FG-domains[125, 126] (**Figure 14**). The C-terminal domain consists of the p15 interacting domain and a ubiquitin-associated fold (UBA-like) (**Figure 14 B**). TAP forms a heterodimer with p15 and contains two FG-repeat binding sites: one on the p15 interacting domain and one on the UBA-like fold at the C-terminus[58, 127]. p15 lacks an FG-binding pocket. Although TAP/p15 can interact with the FG-repeats of the nups directly, this mRNA export receptor is structurally unrelated to the karyopherin receptor family and does not interact with Ran[58].

The N-terminal of domain of TAP contains an RNA-recognition motif (RRM) and a leucine-rich region (LRR) (**Figure 14 A**). The RRM domain of TAP is not essential for general mRNA export activity, but can associate with some cargo directly[128]. For instance, the RRM and LRR domains of TAP bind directly to simian virus RNAs containing a constitutive transport element (CTE)[125]. The CTE helps the virus to overcome nuclear retention of unspliced RNAs and allows for nuclear export by recruitment of TAP/p15[129]. However, the bulk of TAP/p15 cargo is not recognized directly, but via transport adaptors.

From the mRNA export adaptors only a few have been identified. mRNA export adaptors are classical RNA binding proteins, and their main function is to recruit the mRNA export receptor complex TAP/p15 to the mRNP[130]. In addition, the mRNA export adaptors also serve as communicators between upstream processing steps and mRNA export.

The main export adaptor protein for TAP/p15 is Aly. Aly contains a conserved RNA-binding domain (RBD)[131, 132] which is flanked by sequences that interact with the N-terminal domain of TAP[133]. Aly is recruited to the mRNA via a direct interaction with the RNA helicase UAP56, which in turn is recruited by the multi-protein subcomplex THO. THO, UAP56 and Aly together form the transcription-coupled export complex (TREX). In yeast and unspliced mRNA transcripts, TREX is continuously loaded onto the emerging transcripts. In spliced transcripts, TREX is recruited to the transcript in a splicing-dependent manner via the EJC, as well as in a cap-dependent manner via CBC80[118]. Aly then facilitates association of the mRNA export receptor TAP/p15 with the mRNP. The cap-dependent recruitment cascade may explain the observation that Balbiani ring mRNPs of *Chironomus tentants* emerge with their 5' cap first in a 5' to 3' direction from the NPC[115, 134].

Two other export adaptors, SRp20 and 9G8, belong to the serine/arginine rich proteins (SR proteins). SR proteins are essential splicing regulators that are important in constitutive and regulated splicing of pre-mRNA[135]. SR proteins associate with the mRNP in a hyperphosphorylated form, but can also associate with the poly(A) tail and via an mRNA export element in intron-less mRNPs[136]. During splicing, the SR proteins are dephosphorylated, which favors recruitment of TAP/p15 to the mRNP. Thus, the phosphorylation status of SR proteins can act as a signal for export competency of spliced mRNPs[137].

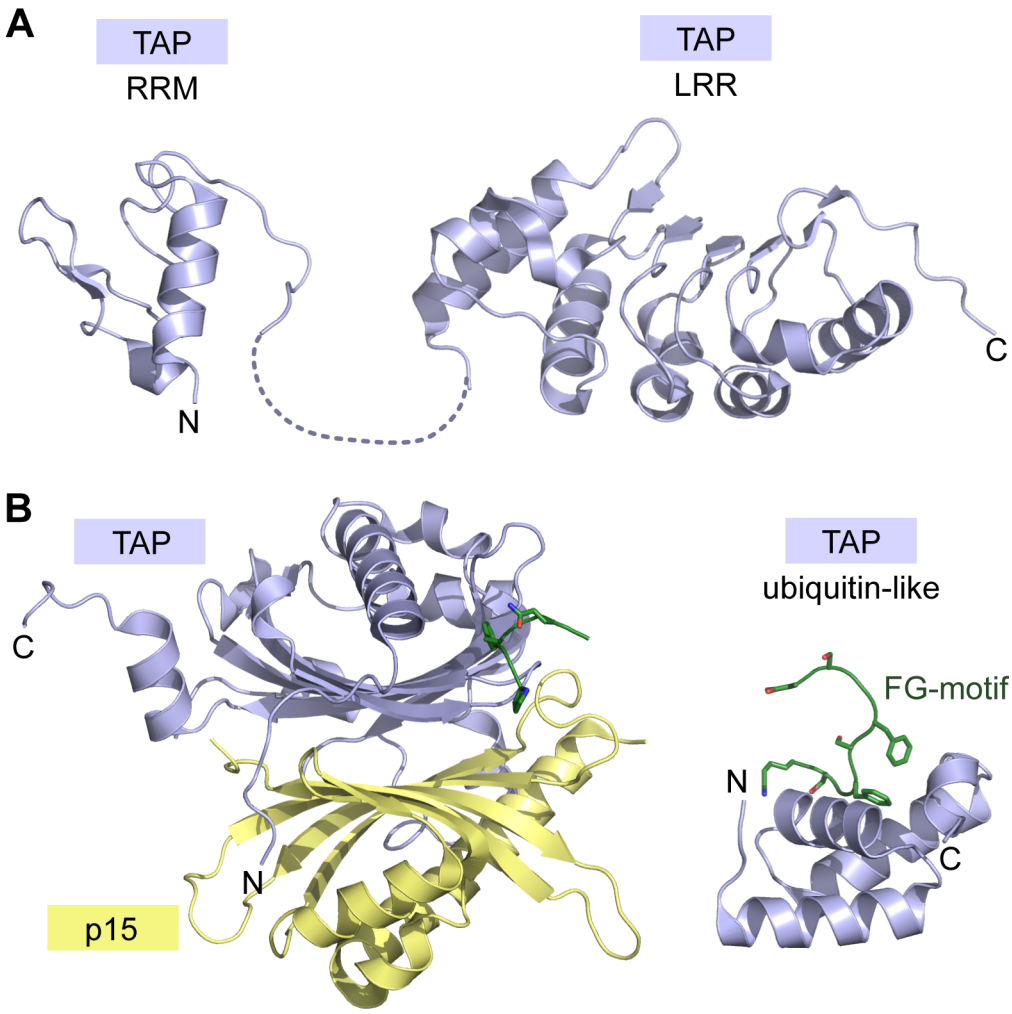


Figure 14: The mRNA export receptor complex TAP/p15. (**A**) Structural overview of the N-terminal domain of TAP containing the RRM (*left*) and LRR fold (*right*). A dashed line indicates the unstructured connector between the two folds. (PDB code 1FO1). (**B**) The C-terminal domain of TAP consists of the p15 interacting domain (*left*) and a ubiquitin-like fold (*right*). FG-motifs bound to Tap are shown in green. (PDB code 1JN5, 1OAI).

1.5.1 Uni-directionality of mRNA export can be achieved by a molecular ratchet mechanism

All macromolecules larger than ~40 kDa require facilitated transport across the central channel of the NPC. mRNAs are synthesized in the nucleus and are packaged into an mRNP that has to be transported to the cytoplasm where translation takes place. Transport of proteins and most classes of RNA is achieved by the karyopherin transport factors. Kaps bind the cargo and interact with the FG-repeat domains of the NPC. Ran regulates formation and disassembly of the transport complexes and determines the direction of transport. However, the transport receptor complex for mRNA, TAP/p15, belongs to a distinct protein family and is unable to interact with Ran. Therefore, the mechanism by which uni-directionality is achieved is still unknown.

Recently a molecular ratchet mechanism has been proposed[138] (**Figure 15**). In this model the mRNP moves through the NPC by thermal motion and by the interaction of its export receptor TAP/p15 with the FG-repeat domains. Once the mRNP arrives on the cytoplasmic face of the NPC, an RNA helicase uses ATP-hydrolysis to remove from the mRNP those molecules that facilitate passage through the NPC. The mRNP is unable to slide backwards and consequently is transported out of the nucleus. Although the RNA helicase and its activating cofactor involved in mRNA export have been identified in yeast, the

energy requirements as well as the exact mechanism by which the displacement of proteins from the mRNP is achieved are still unclear.

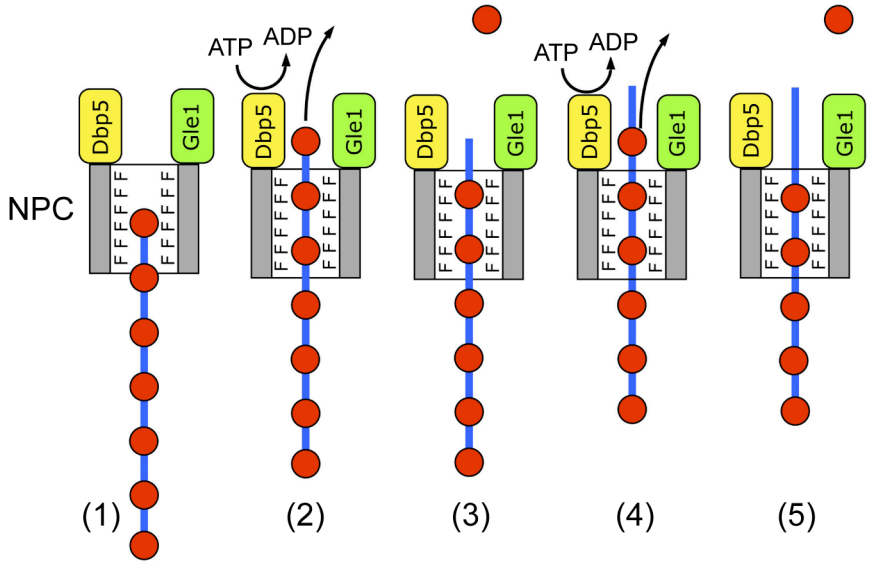


Figure 15: Schematic illustration of transport of large mRNPs by a Brownian ratchet through the NPC. (**1**) Thermal fluctuations move a mature mRNP with several mRNA export receptors bound (blue line with red circles) through the NPC. The transit is facilitated by interaction of the export receptor and the FG-repeat domains. (**2**) When the first export receptor reaches the cytoplasmic side of the NPC, it is removed by the DEAD-box helicase yDbp5. yGle1 stimulates the ATPase activity of yDbp5. (**3**) The removal of the export receptor functions as a molecular ratchet and prevents the mRNP from moving back into the transport channel. (**4**) The next mRNA export receptor is removed by another cycle of ATP-hydrolysis. (**5**) This creates a longer segment of mRNP that is prevented from backsliding into the transport channel and ratchets the mRNP out of the nucleus. Figure reproduced from Stewart[58].

1.6 NPC and NPC associated proteins in mRNA export

To identify proteins involved in mRNA transport, genetic screens for mutants that display accumulation of poly(A) RNA in the nucleus as well as genetic interactions with known mRNA export factors have been analyzed. Several nups and NPC-associated proteins are implicated in mRNA export in yeast[139-142] (**Table 1, Figure 16**). While the exact role of most nups in mRNA export is still unknown, specific functions in mRNA export have been identified for some and will be discussed in this section[142].

Table 1: Human and yeast homologues of nups and NPC-associated proteins involved in mRNA export.

nups		NPC-associated proteins	
human	**yeast**	**human**	**yeast**
Tpr	yMlp1, yMlp2	Ddx19	yDbp5
Nup98	yNup116	Gle1	yGle1
Nup75	yNup85	Rae1	yGle2
Nup62	yNsp1		
Nup88	yNup82		
Nup214	yNup159		
Nup358	-		
hCG1	yNup42		

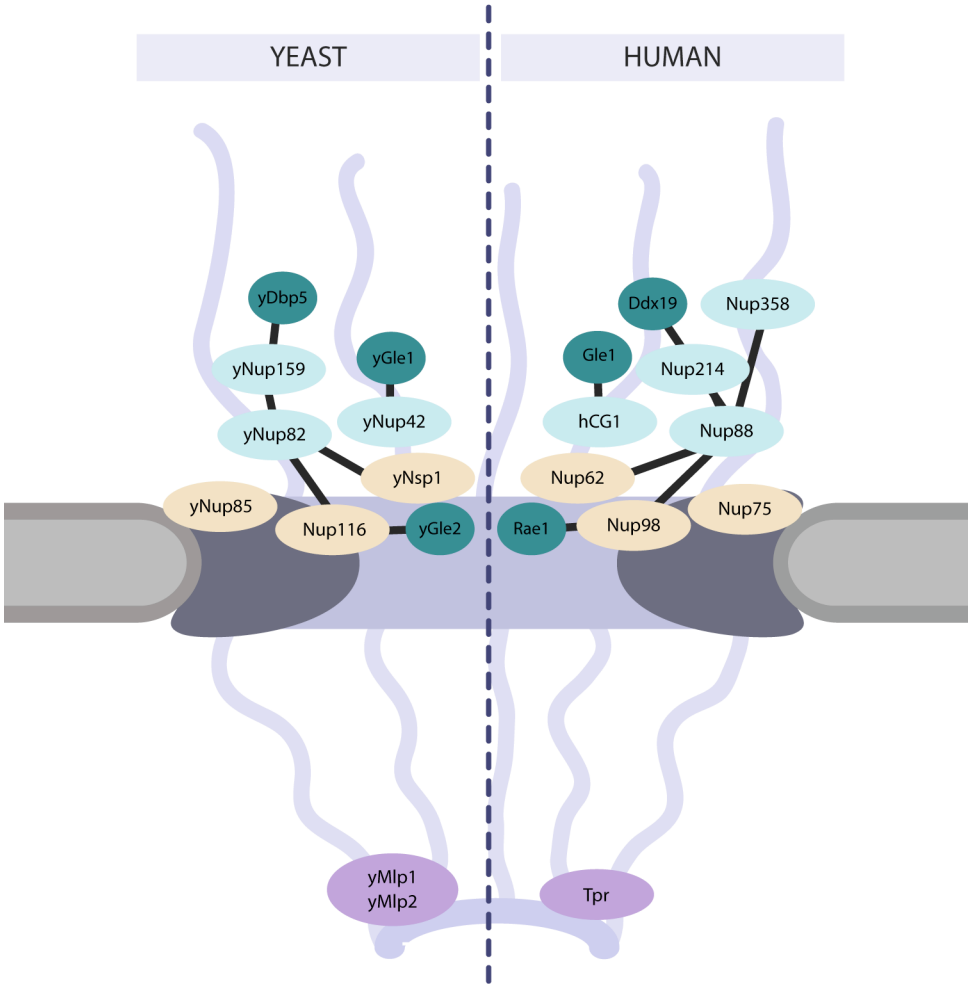


Figure 16: Schematic representation of the nups and NPC-associated proteins implicated in mRNA export. The relative positions of the yeast (left) and human (right) proteins of the cytoplasmic filaments (blue), symmetric core (yellow) and the nuclear basket (violet) as well as the NPC associated proteins (cyan) involved in mRNA export are indicated.

1.6.1 The nuclear basket nups yMlp1 and yMlp2 function in mRNA surveillance

In yeast, the Mlp's are localized to the nuclear site of the NPC and are involved in mRNP quality control at the NPC by assuring that only mature, fully processed mRNPs are transported to the cytoplasm[143]. In a proposed two-step mechanism, Mlp1 first concentrates export-competent mRNPs at the NPC by direct interaction with the hnRNP Nab2[144], which is involved in poly(A)$^+$ length control. Then, incompletely spliced mRNPs are tethered and retained at the nuclear basket of the NPC. This retention is mediated by an RNA-dependent interaction with the branch-point binding protein SF1[145].

1.6.2 The cytoplasmic filaments and mRNA export

1.6.2.1 Nup214 recruits the essential DEAD-box protein Ddx19 to the NPC

In the human NPC, Nup214 is asymmetrically localized to the cytoplasmic filaments of the NPC[146], and together with Nup88 and Nup358 forms the Nup88 complex. The complex is anchored via a conserved interaction with Nup98 and Nup62 to the symmetric core of the NPC[31, 147-149]. Nup214 contains an N-terminal all-β-sheet region, which is followed by an α-helical region, and a C-terminal FG-repeat containing domain (**Figure 17 A**). One of the main roles of the nups involved in mRNA export seems to be to recruit and coordinate essential mRNA export factors at the NPC. In the case of Nup214, an N-terminal region has been

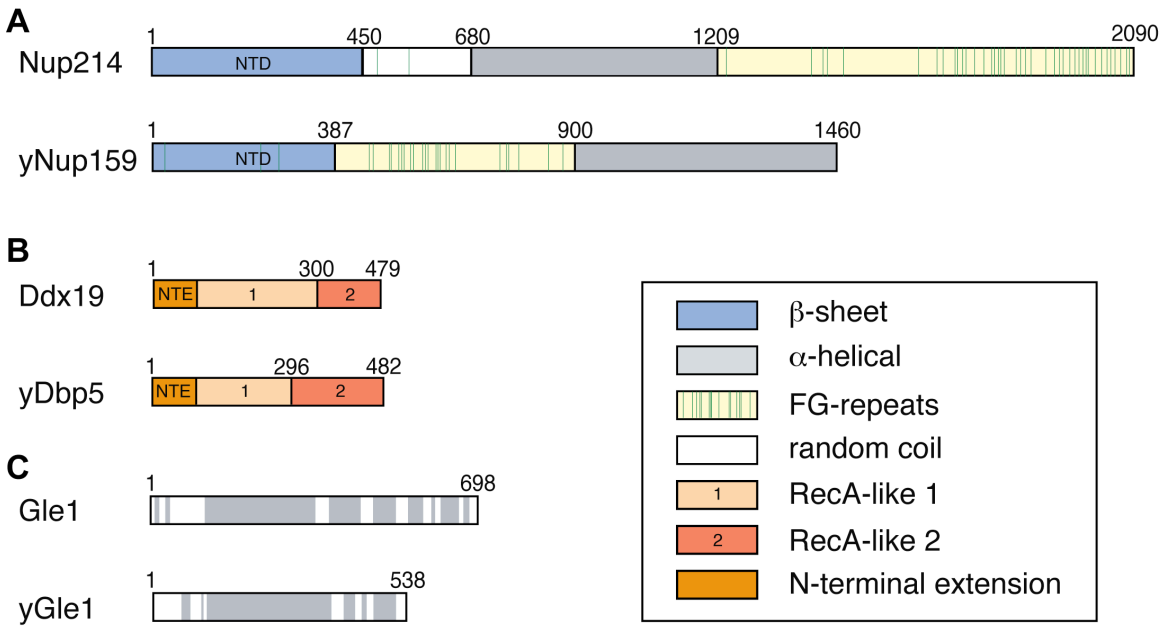


Figure 17: Domain structures of Nup214, Ddx19 and Gle1. Domain structures of (**A**) Nup214 and its yeast homologue yNup159, (**B**) the DEAD-box helicase Ddx19 and its yeast homologue yDbp5, and (**C**) Gle1 and its yeast homologue yGle1. The approximate domain boundaries are based on secondary structure analysis and are indicated above the domains.

found to interact with the DEAD-box helicase Ddx19[150]. Mutations in Ddx19 that disrupt binding to Nup214 inhibit mRNA export and cause poly(A)$^+$ mRNA retention in the nucleus in yeast[35].

1.6.2.2 Ddx19 is a DEAD-box helicase involved in mRNA export

The yeast homologue of Ddx19, yDbp5, is an essential protein that was first discovered in a search for DEAD-box protein genes[140, 151] as well as in a genetic screen for mutants with mRNA export defects[152]. DEAD-box proteins are a large protein family found in all eukaryotes and most prokaryotes[153]. The family is named after the amino acid sequence D-E-A-D of one of the nine conserved

consensus motifs. The conserved motifs are distributed over the entire length of the protein core and are involved in nucleotide coordination or substrate binding (**Figure 18**). The core sequence of DEAD-box proteins is typically ~350 residues long and not conserved at the primary level[154]. Nevertheless, all DEAD-box helicases fold into two globular RecA-like domains that are connected by a flexible linker[155]. The core sequence can be supplemented by N- or C-terminal extensions and it is thought that these flanking regions confer specificity. This may explain why DEAD-box proteins are unable to substitute for each other.

DEAD-box proteins are involved in virtually all processes of mRNA biogenesis: from transcription to splicing, export, translation and decay. Although DEAD-box proteins typically function in larger protein complexes, which hampers their characterization, the few that have been biochemically characterized possess ATPase, RNA binding and unwinding activity, and are therefore also referred to as DEAD-box helicases. Intrinsically, DEAD-box helicases have a low ATPase activity. The ATPase activity is stimulated greatly by RNA as well as ATPase activating cofactors, as exemplified by the eIF4A DEAD-box helicase, which requires binding of the eIF4B and eIF4F cofactors for its function in translation initiation[156].

Although DEAD-box proteins are often referred to as DEAD-box helicases or RNA helicases, helicase activity was demonstrated only for a small subset of DEAD-box proteins and the binding of RNA is unspecific and stimulated by

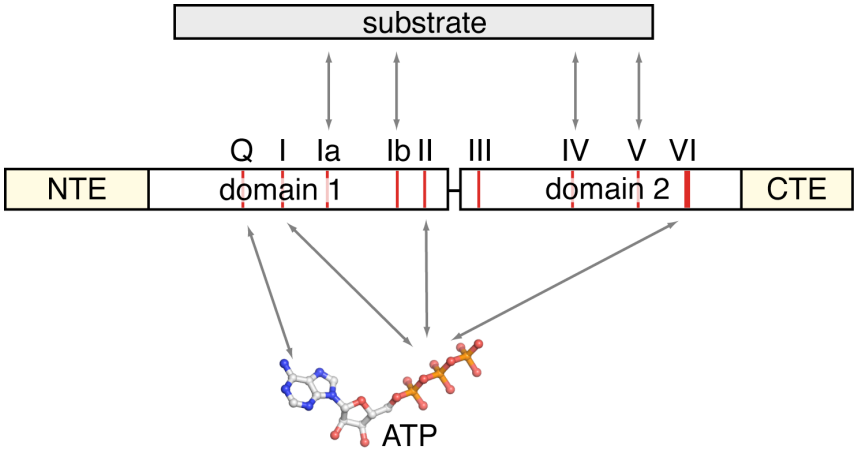


Figure 18: Domain structure of DEAD-box proteins. The N-terminal (NTE) and C-terminal (CTE) extensions (yellow) and the core domain (white) are shown. The nine conserved sequence motifs interacting with nucleotide (shown in ball-and-stick representation) or substrate (grey box) are indicated in red.

ATP[153]. While most of the characterized helicases have been shown to require a ssRNA overhang to unwind short RNA duplexes, some are able to unwind blunt-end dsRNA, e.g. eIF4A or RhlE *in vitro*[157]. The helicase activity of DEAD-box proteins is generally bidirectional, not processive, and highly dependent on the stability of the dsRNA duplex. A “destabilization model” for helicase activity has been proposed[158]. In this model, the helicase binds to a single stranded region of the RNA, after which the ATPase cycle results in a structural rearrangement of the two domains of the helicase. This in turn destabilizes the structure of the bound RNA. However, the molecular details of the RNA duplex destabilization are still unknown.

1.6.2.3 Gle1

Gle1, which stands for glycine-leucine-phenylalanine-glycine mutant 1, is an essential mRNA export factor and was first discovered as a high copy extragenic suppressor of a Nup159 temperature sensitive allele[141]. In yeast, yGle1 was also found in a screen for mutations synthetically lethal with an yNup100 disruption[159]. Additionally, genetic and physical interaction between yGle1 and yNup42, yNup100, yDbp5 implicate Gle1 as an important factor in mRNA export[160].

Gle1 is an all-α-helical protein, and the human and yeast Gle1 homologues are highly conserved (**Figure 17 C**). Although the C-terminal domains are 54% identical the human protein is, however, unable to rescue a yGle1 null mutation in yeast[161]. The human Gle1 gene contains a cryptic splice site in exon 14 and is expressed as two transcript variants, hGle1A and hGle1B that contain distinct C-termini[162]. Gle1A mRNA is ~1000 fold less abundant, and the protein does not localize to the nuclear rim and has not been characterized in detail. Gle1B is the prevalent isoform, and will be subsequently referred to as Gle1.

Gle1 is an NPC-associated protein, and is localized to the cytoplasmic filaments[159, 160] via an interaction with the C-terminal domain of Nup155 as well as the non-FG-repeat domain of hCG1[162, 163]. Gle1 was identified as an ATPase-activating cofactor for yDbp5, which has a low ATPase activity by itself. In addition, Gle1 directly binds to phytic acid, also known as inositol-hexakisphosphate (IP_6). IP_6 was found to potentiate yGle1-mediated stimulation

by increasing the catalytic efficiency of yDbp5 approximately 16-fold *in vivo* in the presence of yGle1 and RNA[164, 165]. Therefore, yGle1-IP_6 controls the activation of yDbp5 in the cell.

1.6.2.4 Ddx19, Nup214 and Gle1 are involved in the terminal export step of mRNA

Little is known about the enzymatic activity of Ddx19. However, it has been shown that Ddx19 can unwind short double stranded RNA duplexes with 5' and 3' overhangs, and that it has a low intrinsic ATPase activity that requires Gle1-IP_6 and RNA for maximum stimulation in yeast[140, 150]. yDbp5 and yGle1 are both recruited to the cytoplasmic filaments of the NPC[47, 150]. Therefore, the ATPase activity of the helicase is spatially controlled and becomes stimulated only at the cytoplasmic side of the NPC[166]. New results suggest that yDbp5 and yGle1-IP_6 may be involved in the terminal step of mRNA export from the nucleus by removing mRNA export receptors from the mRNP[166, 167]. It is predicted that ATP hydrolysis causes a conformational change in the helicase, which results in mRNP remodeling. However, yDbp5-ADP was shown to be sufficient for the removal of RNA-binding proteins[167]. In addition, ATP hydrolysis was shown to be unnecessary for RNA strand separation, but required for RNA release in three different DEAD-box proteins[168]. Because of these conflicting data, the mechanism of RNA remodeling by DEAD-box helicases remains unclear.

1.7 mRNA export nups and human diseases

Many nups that have been shown to have a role in mRNA export were also found to play a significant part in the development of cancer. Chromosomal translocations have been identified for TPR, Nup358, Nup98 and Nup214, while Nup88 overexpression has been linked to tumor progression (**Figure 19**).

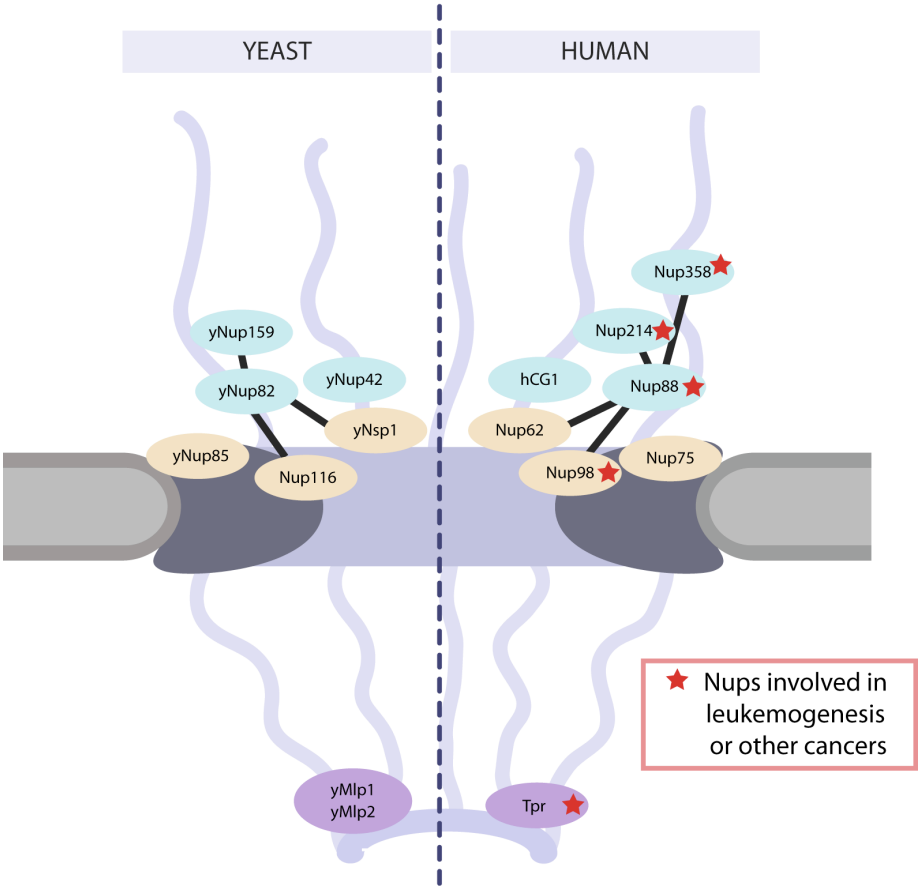


Figure 19: Nuclear pore complex proteins and cancer. Schematic representation of nups involved in mRNA export with nups implicated in cancer highlighted by a red star. The yeast homologues are indicated on the left.

1.7.1 Nup88 as a prognostic marker for tumor state

Nup88 is a non-FG repeat containing nup that is localized exclusively to the cytoplasmic side of the NPC (**Figure 20 A**). The Nup88 complex, consisting of Nup88, Nup214 and Nup358, is anchored via a conserved interaction with Nup98 and Nup62 to the symmetric core of the NPC[31, 147-149]. Nup88 is consistently found to be overexpressed in carcinomas, sarcomas, lymphomas and ovarian tumors[169-171], and the level of overexpression of the Nup88 transcript correlates with the malignant phenotype of the cancer[172]. Staining for Nup88 in tumor cells reveals that Nup88 accumulates in the cytoplasm[170]. The intensity of staining correlates with tumor aggressiveness and invasiveness in breast, colorectal and hepatocellular carcinomas[170, 172-174]. Therefore Nup88 has been proposed as a marker of tumor state and a potential indicator of patient prognosis[175]. The exact nature of the link between Nup88 overexpression and cancer is unknown. In the absence of Nup88, active NF-κB accumulates in the cytoplasm in *Drosophila.* NF-κB is a ubiquitous transcription factor that controls the expression of genes involved in immune responses, apoptosis, and cell cycle. Conversely, Nup88 overexpression may lead to decreased NF-κB export, resulting in an accumulation of NF-κB in the nucleus and aberrant upregulation of target genes[175].

1.7.2 TPR receptor tyrosine kinase fusions involved in cancer

TPR (translocated promoter region) is a 265 kDa protein and is localized exclusively to the nuclear side of the NPC, where it is the major component of the nuclear basket[176]. TPR is found in chromosomal translocations that fuse the N-terminal sequences of *TPR* to the tyrosine kinase domain of the proto-oncogene *MET*, which results in a 65 kDa cytoplasmic TPR-Met fusion protein[177] (**Figure 20 B**). Dimerization through the N-terminal TPR domain leads to constitutive activation of the Met kinase, which is responsible for the activation of the Ras/MAPK and Pl3K pathways. Although TPR-Met translocations are not common, they have been found to be associated with gastric carcinomas and are thought to represent an early step in carcinogenesis[178, 179].

TPR has also been found in chromosome translocations resulting in TPR-TrkA fusion proteins[180]. TrkA is the transmembrane tyrosine kinase receptor for nerve growth factor and the fusion proteins of TPR and other proteins to TrkA are associated with the most common type of thyroid cancer[175, 180].

1.7.3 Nup358 translocations

Nup358 localizes to the cytoplasmic side of the NPC, where it is the main component of the cytoplasmic filaments. It is the largest mammalian nup and contains an N-terminal α-helical region followed by Ran-GTP-binding domains, eight zinc finger motifs, an E3 ligase domain and a C-terminal cyclophilin A domain[76, 181] (**Figure 20 C**).

A chromosomal translocation, found in inflammatory myofibroblastic tumors, fuses the N-terminal region of *NUP358* to the *ALK* gene[182]. The anaplastic lymphona kinase (Alk) is a membrane associated tyrosine kinase receptor, with a role in nervous system development and maintenance[183]. In the fusion protein a ~900 residue long N-terminal region of Nup358 is fused to the cytoplasmic segment of Alk. The Nup358-Alk fusion protein is a constitutively activated tyrosin kinase and localizes to NPCs. It has been hypothesized that the N-terminal domain of Nup358 promotes oligomerization and activation of the Alk kinase domain[182].

1.7.4 Nup98 fuses to homeodomain protein genes

Nup98 is a mobile nup and localizes to both sides of the NPC via interaction with Nup96 on the nuclear side and Nup88 on the cytoplasmic side[147, 184]. Nup98 was first identified in chromosomal translocations that fuse the 5' half of the *NUP98* gene with 3' portions of the *HOXA9* gene in patients with acute myeloid leukemia (AML)[185, 186]. The 98 kDa fusion protein localizes to the nucleoplasm and contains the FG-repeat domain and Rae1-binding site of Nup98, as well as the DNA-binding homeodomain of HoxA9[187] (**Figure 20 D**). Subsequently, Nup98 fusions to a variety of Hox targets have been identified in patients with AML as well as chronic myeloid leukemia and pre-leukemic myelodysplastic syndrome (MDS)[175]. Homeodomain proteins are important transcription factors that function as regulators of development and differentiation[188]. There are four clusters of class I homeobox genes each containing 8 to 11 of 13 paralog gene groups in

humans. The *HOX* genes of the *HOX A*, *B* and *C* cluster are expressed in stem and immature progenitor cell populations, but are downregulated during differentiation to mature hematopoietic cells. The less conserved class II homeobox genes are dispersed throughout the genome and act as cofactors for other Hox transcription factors.

In addition to the Nup98-Hox fusions, fusions of Nup98 to a variety of other nuclear proteins have been identified. These include topoisomerases, histone modifying enzymes like NSD1 and NSD3 methyltransferases, chromatin binding proteins like SetBp1 as well as a multitude of other proteins[175]. The identified fusion partners of Nup98 indicate changes in global transcription as a mechanism for Nup98-induced leukemogenesis. This notion is backed by the fact that gene expression profiling has shown that a number of HOX genes are upregulated in cells expressing Nup98-HoxA9 fusions[189].

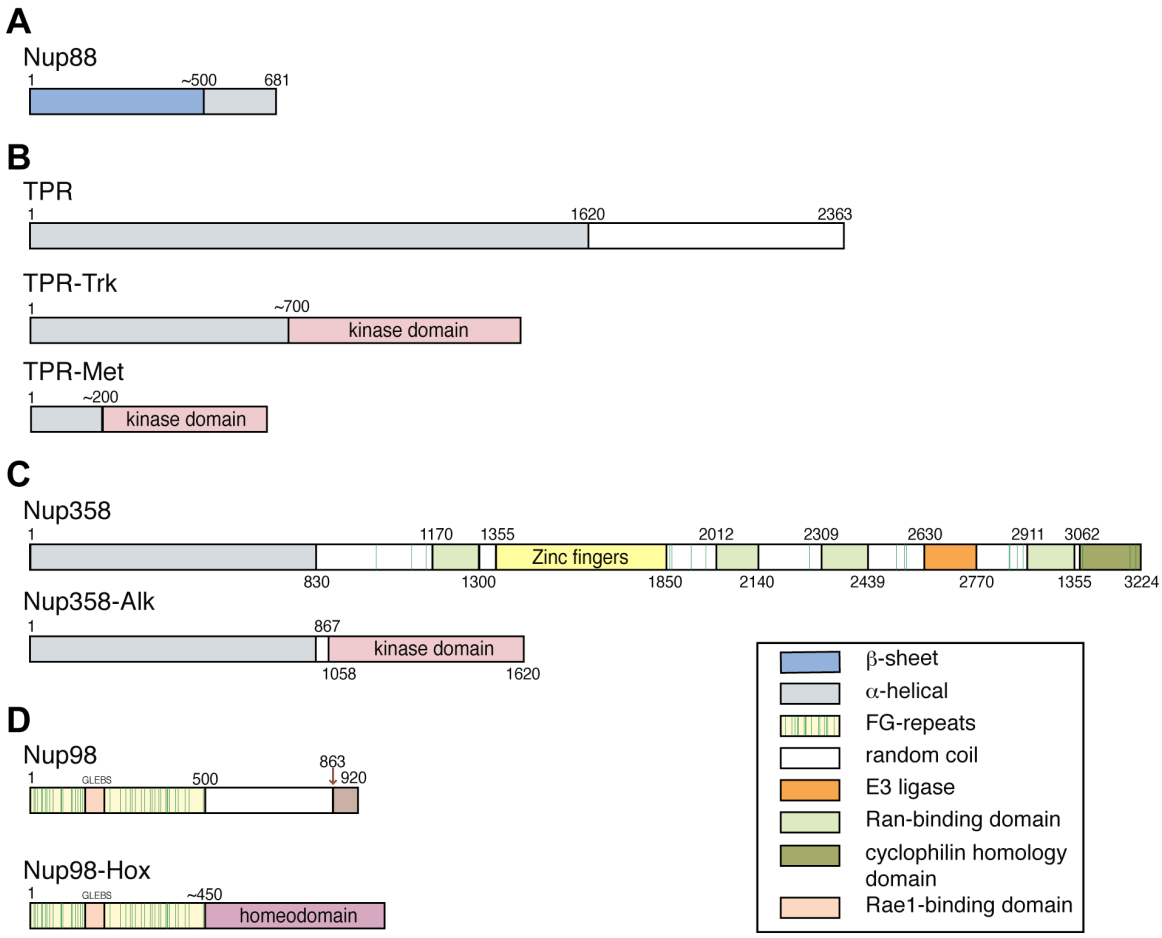


Figure 20. Domains structure of nups and nup fusion proteins. Domain structure of (**A**) Nup88, (**B**) TPR and TPR fusions to the tyrosine kinase domain (light red) of Trk and Met. Domain structure of (**C**) Nup358 and its fusion with the Alk receptor tyrosin kinase, and (**D**) Nup98 and fusions to the homedomain (pink) of Hox proteins.

1.7.5 Nup214 fusions are found in leukemia

The Nup214 gene was first identified as a target for chromosome translocation, and because of its close location to the oncogene c-*ABL* on chromosome 9, named *Cain* (*CAN*)[190]. Nup214 is essential during development, and Nup214 homozygous knockout mouse embryos show protein import and mRNA export defects and die once the maternal Nup214 protein is depleted[191].

Nup214 fusion proteins have been found in at least three different chromosome translocations. The first identified chromosomal translocation fuses the *DEK* and *NUP214* genes, another fuses the *SET* and *NUP214* genes, and an episomal translocation fuses *NUP214* with *ABL1* (**Figure 21**).

The Dek-Nup214 fusions were first reported in a subtype of AML that affects young adults and is characterized by a poor prognosis[190]. The Set-Nup214 fusions were found in patients with undifferentiated leukemia (AUL)[192] and T-cell acute lymphoblastic leukemia (T-ALL). Dek and Set are nuclear phosphoproteins of 45 and 41 kDa, respectively, and are major components of chromatin[193]. Dek induces positive supercoils into closed circular DNA *in vitro*[194] and, in complex with casein kinase 2 (CK2), is thought to serve as a histone chaperone, thereby enabling or inhibiting chromatin accessibility[195, 196]

Dek-Nup214 fusion results in a 165 kDa fusion protein with nuclear localization[190]. The fusion protein includes residues almost the entire sequence of Dek and the C-terminal portion of Nup214, including part of the coiled-coil and the entire FG-repeat region. The Dek-Nup214 fusion protein is unable to bind to

CK2 and it is likely that the chaperone and transcriptional regulation function of Dek is impaired[195]. The exact mechanism of the defect, however, is unknown.

Set is a member of the inhibitor of histone acetylation and transcriptional activation (INHAT) complex[197], but was identified as an activator of transcription *in vitro*[196] as well as on transcriptionally active sites on polytene chromosomes of *Drosophila*[198]. Set was found to remove Dek from chromatin in order to permit access by the transcription machinery. Therefore, Set and Dek may interact functionally as opposing factors in a regulatory cycle to control access to chromatin[199]. In the 155 kDa Set-Nup214 fusion, Set is fused to the exact same C-terminal portion of Nup214 as in the Dek-Nup214 fusion. Set-Nup214 is localized to the nucleus and was indicated by ChIP experiments to bind to a subset of *HOXA* gene promotors, which may contribute to T-ALL pathogenesis by blocking T-cell maturation through transcriptional activation of the *HOXA* genes[200]. Whether the involvement of the Set- and Dek-Nup214 fusion proteins can be explained only by this block of differentiation or whether additional mechanisms are involved in leukemogenesis remains to be seen.

The third chromosomal translocation involving Nup214 is found in ~6% of patients with T-ALL. In these patients, extrachromosomal episomes derived from a region of chromosome 9 are found. The episomes are found in 5 - 50 copies and contain the 3' part of *ABL1*, the *LAMC3*, and the majority of *NUP214*[201]. Circularization generates an *ABL1-NUP214* fusion-gene that lacks the regulatory exon 1 of the Abl1 tyrosine kinase and results in a constitutively activated kinase-

fusion protein. The breakpoint in Nup214 is somewhat variable, between residue 1106 and 1800; however, the fusion-protein always contains the N-terminal β-propeller as well as the majority of the coiled-coil domain (**Figure 21**). The Abl1-Nup214 fusion protein is highly over-expressed and retains Nup214's ability to localize to the NPC, which is required for the transformation potential[202]. It is easy to envision that Abl1-Nup214 competes with Nup214 for binding at the cytoplasmic filaments. The high concentration of Abl1-Nup214 at the pore would increase autophosphorylation and activation of the Abl1 kinase, and provide close contact to kinase substrates as they travel through the pore, and would result in unregulated phosphorylation of proteins at the NPC[175].

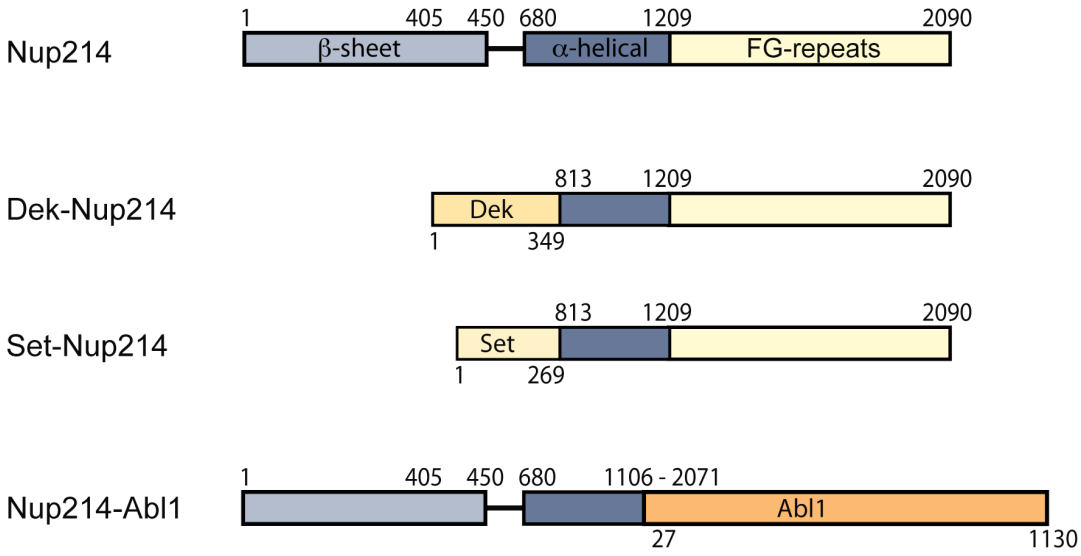


Figure 21: Nup214 is the target of three different chromosomal translocations resulting in fusion proteins. Domain structures of the proto-oncogene Nup214 fusion proteins Dek-Nup214, Set-Nup214 and Nup214-Abl1. The Dek-Nup214 contains residue 1- 349 of 375 from Dek fused to the C-terminal portion, residues 1209 to 2090, of Nup214. Set-Nup214 contains residue 1- 269 of 277 from Set and the exact same C-terminal part of Nup214 as in Dek-Nup214. In the Nup214-Abl1 fusions the breakpoint of Nup214 is between intron 23 and 24, which results in a fusion protein that includes at least the β-propeller and most of the coiled-coil domain.

2 RESULTS AND DISCUSSION

Since high-resolution structural information of the mRNA export machinery at the NPC is unavailable, the principles that govern the terminal step of mRNA export at the NPC remain poorly understood. In order to deepen our understanding of the architecture and the regulation of the mRNA export machinery at the NPC as well as of the role of nups involved in leukemogenesis, I set out to biochemically and structurally characterize Nup214 for my thesis work. My objective was to analyze the interaction of Nup214 with the essential mRNA export factor Ddx19. After first determining the crystal structure of the human proto-oncogene Nup214, I characterized the interaction of Nup214 with Ddx19, and finally solved the structure of the Nup214•Ddx19 complex.

2.1 Biochemical and structural analysis of Nup214

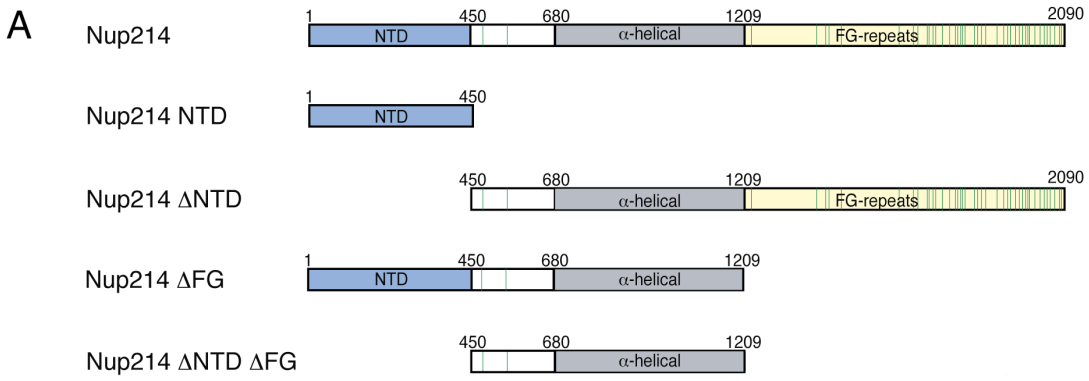
2.1.1 Domain organization and localization

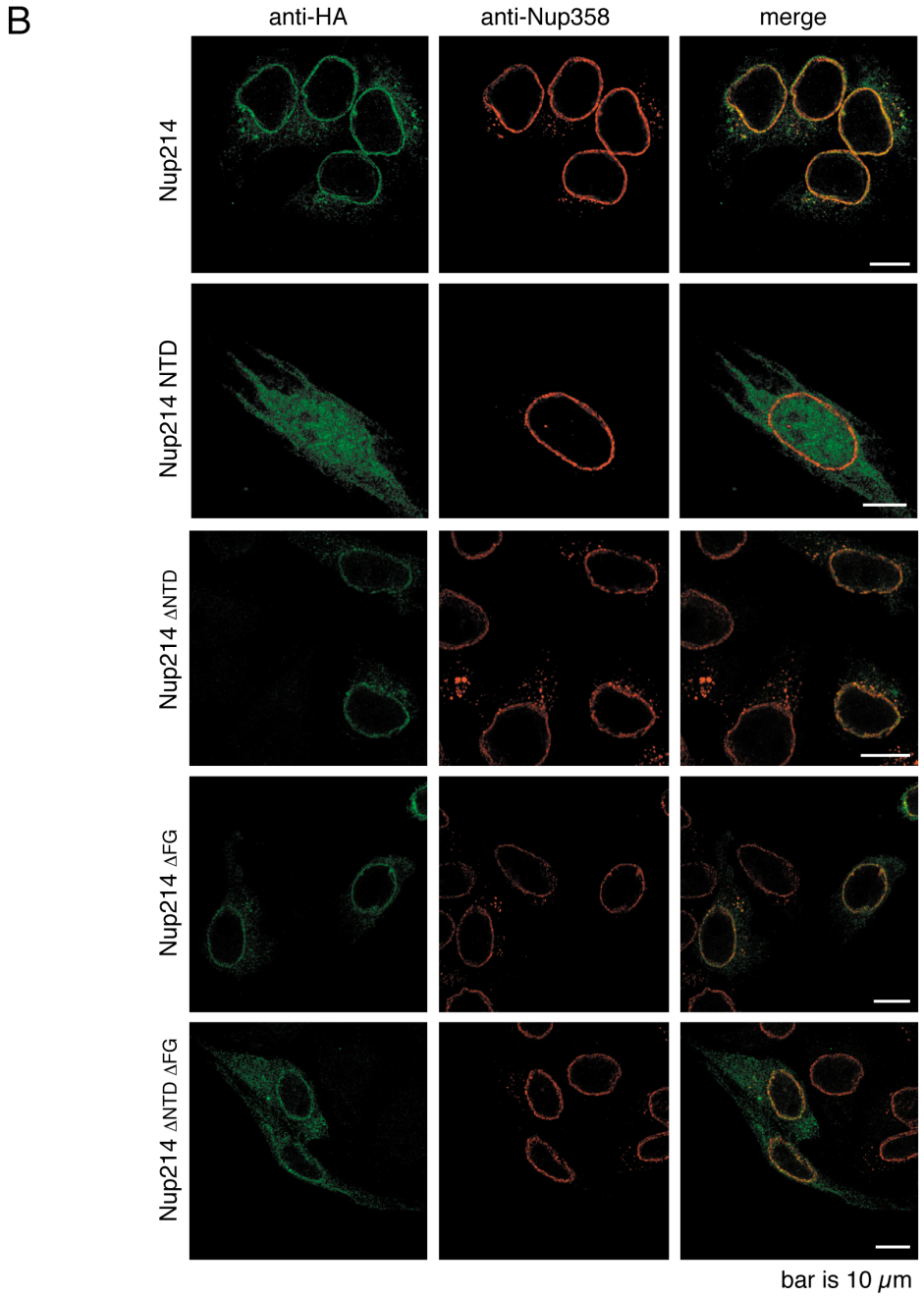
Nup214 is positioned at the cytoplasmic side of the NPC and interacts with proteins of the mRNA export machinery[146, 203, 204]. An N-terminal region of Nup214 has been found to interact with the DEAD-box helicase Ddx19, and mutations in Ddx19 that disrupt binding to Nup214 inhibit mRNA export in yeast[150]. To investigate and clarify the domain organization of Nup214, we performed primary structure and sequence conservation analysis, as well as secondary structure predictions for the 2090-residue human Nup214. We

identified the approximate domain boundaries of three structurally distinct domains: an N-terminal ~400-residue all-β-sheet region which is followed by an ~200-residue unstructured region, an ~500-residue α-helical region, and an ~900-residue C-terminal region (**Figure 22** A, top row). The C-terminal region of Nup214 contains a large number of FG-repeats, which are predicted to be natively unfolded.

Previous studies have shown that Nup214 localizes to the cytoplasmic filaments of the NPC together with Nup88 and Nup358[146, 149]. To identify the functions associated with the various domains of Nup214 with respect to the integration to the NPC, we analyzed the localization of various HA-tagged Nup214 fragments in HeLa cells **(Figure 22, B**). We found that full-length Nup214 is localized to the nuclear envelope in an identical fashion to endogenous Nup358, which served as a control for nuclear rim staining. Furthermore, Nup214 fragments with a deletion of the NTD (Nup214 ΔNTD), or the FG-repeats (Nup214 ΔFG), or both (Nup214 ΔNTD ΔFG) are still able to localize to the nuclear rim. However, HA-tagged Nup214 NTD is dispersed throughout the cell, but not enriched at the nuclear rim, which suggest that the NPC-targeting region resides in the central helical domain of Nup214. The nuclear rim staining of endogenous Nup358 is unaffected by all over-expressed Nup214 fragments.

Figure 22: Localization of human Nup214. (**A**) Domainstructure of Nup 214 and Nup214 deletion constructs used for the *in vivo* localization study. Individual FG-repeats of the FG-repeat domains are indicated by a green line. (**B**) Localization of Ha-tagged Nup214, Nup214 NTD, Nup214 ΔNTD, Nup214 ΔFG and Nup214 ΔNTD ΔFG. The localization of the HA-tagged Nup214 variants is shown in (green), Nup358 localization (red) is shown as reference for rim staining.


A
Nup214
1
450
680
1209
2090
NTD
α-helical
FG-repeats
Nup214 NTD
Nup214 ΔNTD
Nup214 ΔFG
Nup214 ΔNTD ΔFG


B
anti-HA
anti-Nup358
merge
Nup214
Nup214 NTD
Nup214 ΔNTD
Nup214 ΔFG
Nup214 ΔNTD ΔFG
bar is 10 μm

2.1.2 Expression

Since the N-terminal region of Nup214 was shown to bind the mRNA export factor Ddx19 by immuno-precipitation[150], we designed a series of expression constructs for the N-terminal all-β-sheet region of Nup214 (**Table 4**). The constructs were expressed in *E. coli* and tested for solubility. While the shorter Nup214 NTD fragments were largely insoluble, we identified a stable fragment composed of residues 1-450, which we termed Nup214 NTD (**Figure 23**).

2.1.3 Purification and crystallization of Nup214 NTD

Human Nup214 NTD was expressed in BL21(DE3) cells, purified using appropriate affinity and gel filtration chromatography (for details of the purification see Materials and Methods chapter), and concentrated to 50 mg/ml (**Figure 24**). Typically the purification yielded ~15 mg/L of bacterial culture and the purification and purity was monitored by SDS-PAGE.

Crystals of the ~50 kDa Nup214 NTD (50 mg/ml) were obtained at 21°C in hanging drops in several conditions (**Figure 25 A** to **D**). After optimization, large crystals with dimensions of ~350 x 350 x 150 mm^3 grew within 1 week in 0.1M MES pH 6.4 - 6.5 and 23.5 – 27.0% (w/v) PEG 2000 MME (**Figure 25 E**). The crystals belong to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 52.4 Å, b = 81.1 Å, c = 102.6 Å, $\alpha = \beta = \gamma = 90°$ and with one molecule in the asymmetric unit.

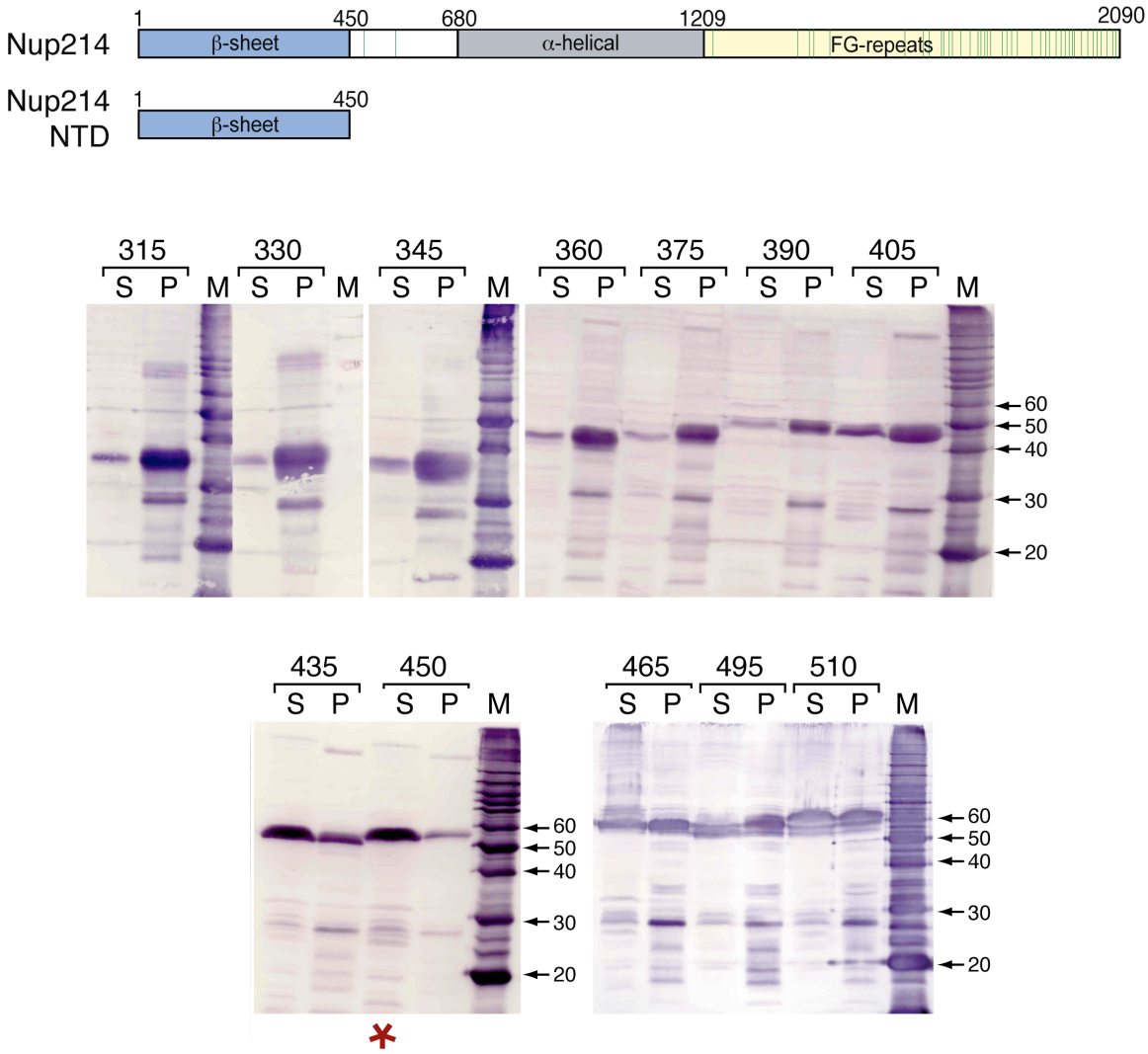


Figure 23: C-terminal deletion constructs reveal a stable fragment of the Nup214 N-terminal all-beta-sheet domain. Anti-hexa-histidine affinity blot of a series of expression and solubility tests of C-terminal deletion constructs identified a stable fragment composed of residues 1-450 (red star). The numbers represents the length of the fragment, S and P denote the soluble and insoluble pellet fraction after cell disruption, respectively.

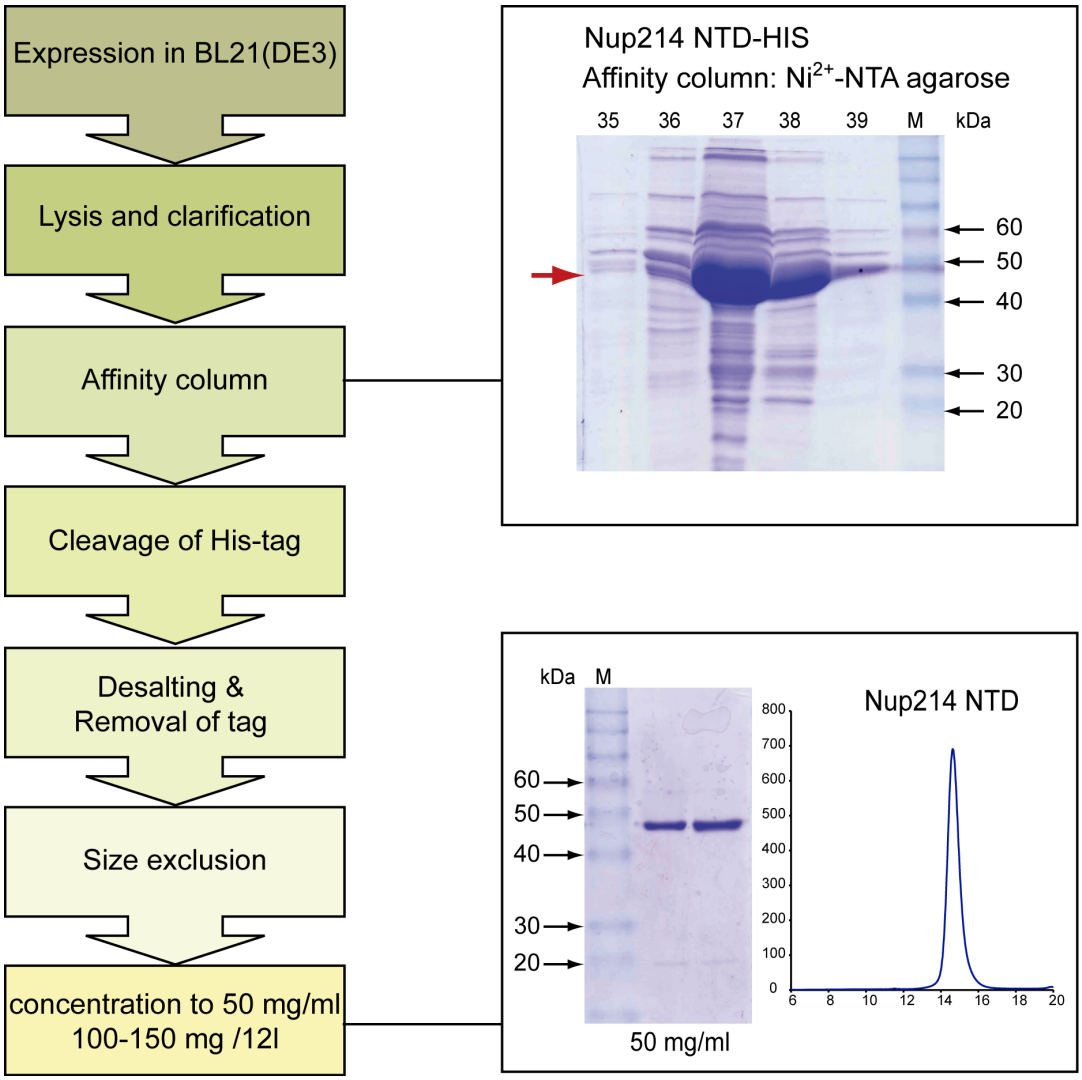


Figure 24: Purification scheme for Nup214 NTD. The flow chart depicts the purification of recombinant Nup214 NTD from *E. coli*. Representative Coomassie brilliant blue stained SDS-PAGE gels of the affinity purification step as well as the final Nup214 NTD protein, which was used for crystallization, are shown on the right.

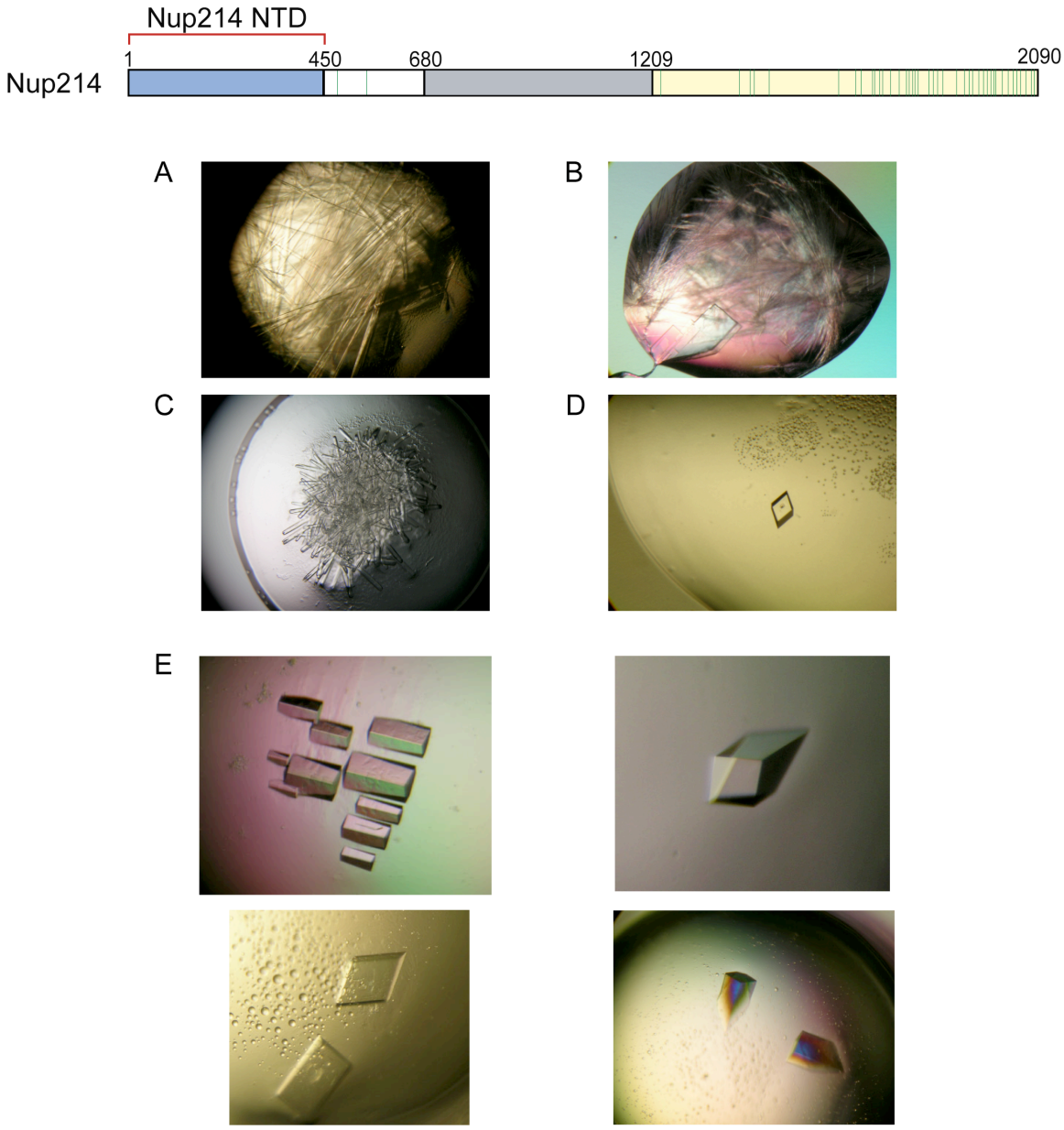


Figure 25: Crystals of the N-terminal domain of Nup241 NTD. Crystals of Nup214 NTD were observed in the following conditions: (**A**) 0.2 M Magnesium Formate, 20% w/v Polyethylene Glycole 3350; (**B**) 0.2 M Potassium Formate, 20% w/v Polyethylene Glycole 3350; (**C**) 0.2 M Sodiumdihydrogen Phosphate monohydrate, 20% w/v Polyethylene Glycole 3350; (**D**) 0.1 M MES pH 6.5, Polyethylene Glycole Monomethyl Ether 2000. (**E**) Crystals obtained after optimization of (**D**).

2.1.4 Structure Determination of the Nup214 NTD

The structure of Nup214 NTD was solved by multiple isomorphous replacement anomalous scattering (MIRAS), using X-ray diffraction data from a native data set and two heavy metal derivatives, Osmiumtetroxid and *p*-Chloromercuribenzoate. No electron density was observed for the 16 residues of the C-terminus of the Nup214 NTD, so these 16 residues were omitted from the final model. The final model contains residues 1-434 and was refined to a 1.65 Å resolution with an R_{work} and an R_{free} of 19.7% and 23.7% respectively (**Table 6**, Material and Methods).

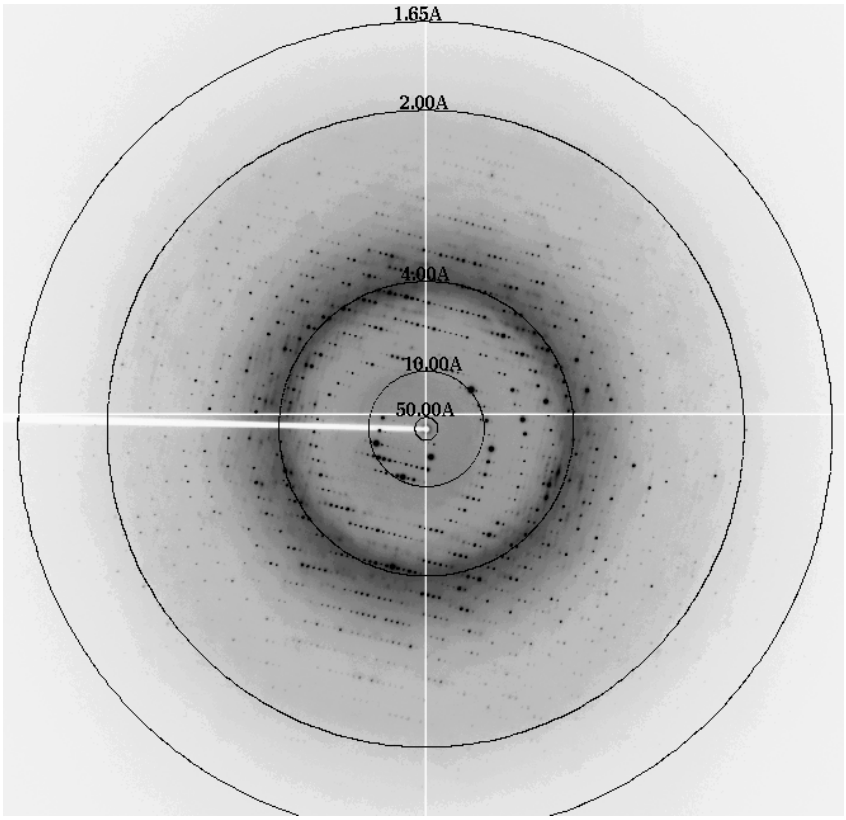
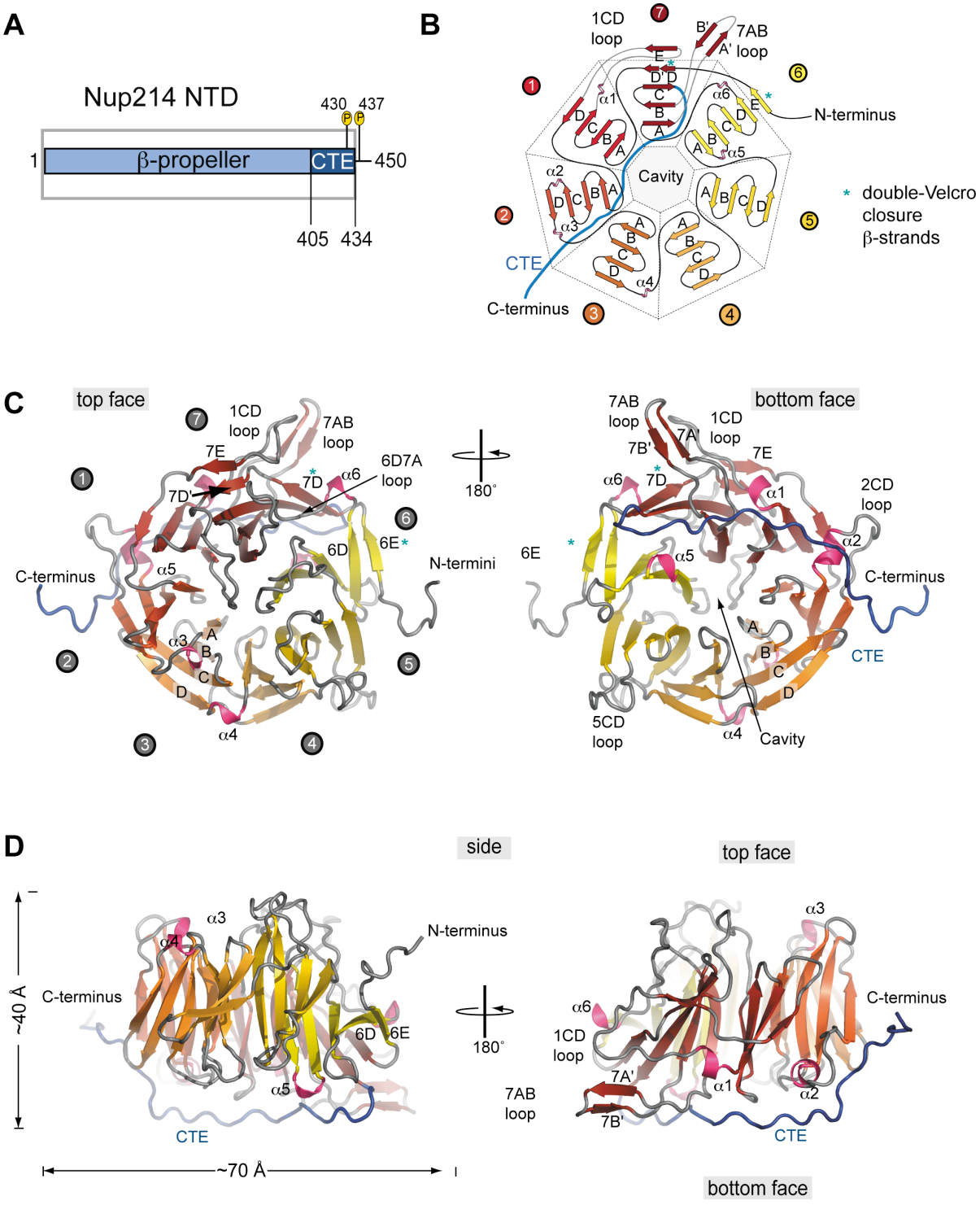


Figure 26: X-ray diffraction pattern of Nup214 NTD obtained on beamline 8.2.1 at the ALS. The outermost resolution ring indicates the diffraction limit of the native Nup214 NTD crystal.

2.1.5 The structure of the Nup214 NTD

The structure of the Nup214 NTD can be divided into two parts: an N-terminal canonical seven-bladed β-propeller domain followed by a C-terminal extended peptide segment (CTE) (**Figure 27**). The Nup214 NTD creates an elliptical, disc-shaped molecule with an overall diameter of ~70 Å and a thickness of ~40 Å (**Figure 27 C**, **D**). The canonical core of the β-propeller is generated by four anti-parallel β-strands in each of the seven blades, which by convention are termed A to D from the inside to the outside of the β-propeller (**Figure 27 B**, **C**). The loops are named according to the β-strands they connect, with the face that contains the N-terminal ends of the innermost β-strands defined as the top face. The blades are pseudo-symmetrically arranged around a seven-fold axis and create a central cavity at the bottom face of the β-propeller (**Figure 27 B**, **C**). Typically, the β-propeller of many proteins are generated by conserved sequence motifs, such as the ~40 residue long WD repeat, in which very short loops connect the β-strands. In the Nup214 β-propeller, however, long insertions between the β-strands form extensive loops protruding on both faces of the molecule (**Figure 27 C**, **D**). The lack of such short connecting loops significantly loosens the restraints on the primary sequence, and consequently no sequence repeat motif can be identified in the β-propeller of the Nup214 NTD. The 30-residue CTE (residues 405-434) folds back onto the β-propeller, adopts an extended conformation and is bound to the bottom face.

Figure 27: The structure of the NTD of human Nup214. (**A**) Crystallization construct of Nup214 NTD. The β-propeller, the CTE as well as the two phosphorylation sites are indicated. Residues observed in the crystal structure are boxed in grey. (**B**) Schematic representation of the Nup214 NTD structure. The blades of the β-propeller are labeled from 1 to 7. The CTE is shown in blue, and β-strands forming the double-Velcro closure are indicated with an asterisk. (**C**) Ribbon representation of the Nup214 NTD structure. A 180°-rotated view is shown on the right. As a reference, the strands of blade 3 are labeled A–D. The blades of the β-propeller and the CTE are labeled as in *B*. The helical insertions are shown in pink. (**D**) Ribbon representation of side views of the structure of the Nup214 NTD.


A
Nup214 NTD
430 437
1
β-propeller
CTE
450
405 434
B
1CD loop
7AB loop
N-terminus
Cavity
CTE
C-terminus
* double-Velcro closure β-strands
C
top face
1CD loop
7AB loop
7E
6D7A loop
7D'
7D
α6
6E*
6D
N-termini
C-terminus
α5
α3
α4
180˚
bottom face
7AB loop
1CD loop
7B'
7A'
7E
α6
7D
α1
2CD loop
6E
α5
α2
C-terminus
CTE
5CD loop
α4
Cavity
D
side
~40 Å
α3
α4
N-terminus
C-terminus
6D
6E
α5
CTE
~70 Å
180˚
top face
α3
α6
C-terminus
1CD loop
7AB loop
7A'
7B'
α1
α2
CTE
bottom face

2.1.6 Unique structural features of the Nup214 NTD

The canonical features of the Nup214 β-propeller, as first seen in the structure of methylamine dehydrogenase β-propeller[205], are complemented by special features, such as the central cavity, structural additions to the β-propeller fold, and a double-Velcro closure (**Figure 27)**. The circular packing of the blades forms a central, water-filled tunnel connecting the top and the bottom in the vast majority of β-propellers, as seen in the $G_{t\beta}$ subunit of the hetero-trimeric G-protein complex transducin[206, 207]. However, in the Nup214 β-propeller, the channel is closed at the top face, generating a ~30 Å-deep conical cavity with a diameter of ~15 Å at its opening that is filled with ~50 well-ordered water molecules.

The structural additions to the canonical β-propeller fold of Nup214 include extensive loops measuring up to 23 residues in length and containing helical turns and non-canonical β-strands. The remarkably long loops include the inter-blade connector 6D7A, as well as loop 7AB and the loops that connect strands C and D of blades 1, 2, and 5 (**Figure 27 C**)**.** Two α-helices, α1 and α2, and four 3_{10}-helices, α3 to α6, can be found in these loops. Helix α1 is inserted between strands C and D of blade 1 and is located on the side of the β-propeller. Helix α6 is also located on the side and inserted in the 6D7A loop. The helices α3 and α4 are located on the top face within loops 2D3A and 3D4A, respectively, whereas helices α2 and α5 are located on the bottom face of the β-propeller and are part of the 2CD and 6AB loop (**Figure 27 B** to **D**). In the Nup214 NTD β-propeller, two five-stranded blades are created by non-canonical β-strands. Binding of the very

N-terminal β-strand 6E to 6D in a parallel fashion forms the five-stranded blade 6. This interaction creates the non-canonical double-Velcro closure (**Figure 27 B, C**). Blade 7 is the second five-stranded blade. The β-strand 7E, which is located in the 23-residue long 1CD loop, binds in a parallel fashion to strand 7D' (**Figure 25 C**, left). Instead of forming part of the bottom face of the β-propeller, this interaction causes the 1CD loop to be pushed towards the side. Similarly, the 19-residue 7AB loop, which is stabilized by two short accessory β-strands (7A' and 7B'), flips to the side of the β-propeller. Almost parallel, both loops protrude ~25 Å from the side of the β-propeller and form hydrogen bonds to each other (**Figure 27 C**, **D**, right views).

2.1.7 The sequence repeat of the Nup214 NTD β-propeller

In the vast majority of cases, the pseudo-seven-fold symmetry of β-propeller fold-containing proteins is reflected in the primary sequence[208]. For example, the primary sequence of the β-propeller of the heterotrimeric G-protein subunit $G_{t\beta}$ contains seven WD repeats[209]. Each of the WD repeats contains a conserved GH dipeptide ~20 residues from its N-terminus and a conserved WD dipeptide at its C-terminus. The conserved tryptophan of the WD dipeptide is buried in a hydrophobic environment between β-sheets and usually interacts with conserved aspartate, histidine, and serine or threonine residues within the same repeat. The hydrogen-bond arrangement of these residues stabilizes the interactions between the β-strands in each β-sheet and creates a rigid molecule [208]. In

addition to the WD repeat, other clearly defined sequence motifs, such as the kelch motif, the AxSPD, or the YWTD repeats, can be found in other β-propeller proteins[206].

In order to identify structurally equivalent positions within the blades and to determine whether the Nup214 β-propeller contains a sequence repeat, we superimposed all seven blades of the β-propeller (**Figure 28**). We found that while the canonical β-strands of the seven blades align very well structurally with a root-mean square deviation (RMSD) of ~1.2 Å, the loops connecting the β-strands are structurally diverse (**Figure 28 B**). Extensive insertions between the β-strands cause the length of the sequence repeat to vary between 43 and 64 residues. Apart from hydrophobic side chains in the central residues of β-strands A, B, and C, no recognizable sequence motif can be identified by this structural alignment (**Figure 28 C**).

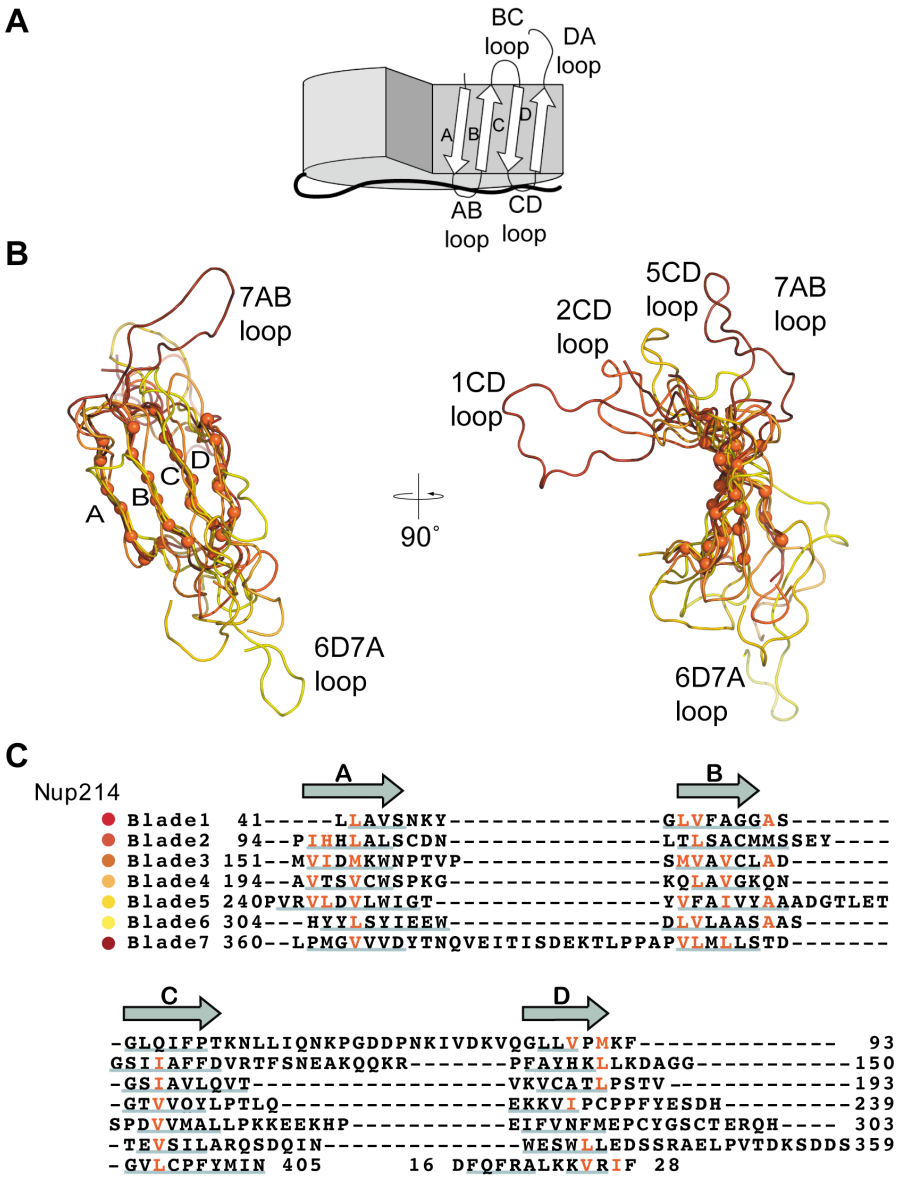


Figure 28: Superposition of the Nup214 NTD β-propeller blades. (**A**) Schematic drawing of a β-propeller fold indicating the β-strands and loops of one β-propeller blade. (**B**) Coil representation of the structural alignment of the seven blades of the β-propeller. Blades are colored as in Figure 27. As a reference, the C_α atoms of blade 2 are shown as orange spheres. A 90°-rotated view is shown on the right. (**C**) Structure-based sequence alignment of the blades. The β-strands are indicated above the sequence. Similar residues are shown in red, and the residues of each blade that participate in β-sheet hydrogen bonds are underlined in gray.

2.1.8 The C-terminal extension

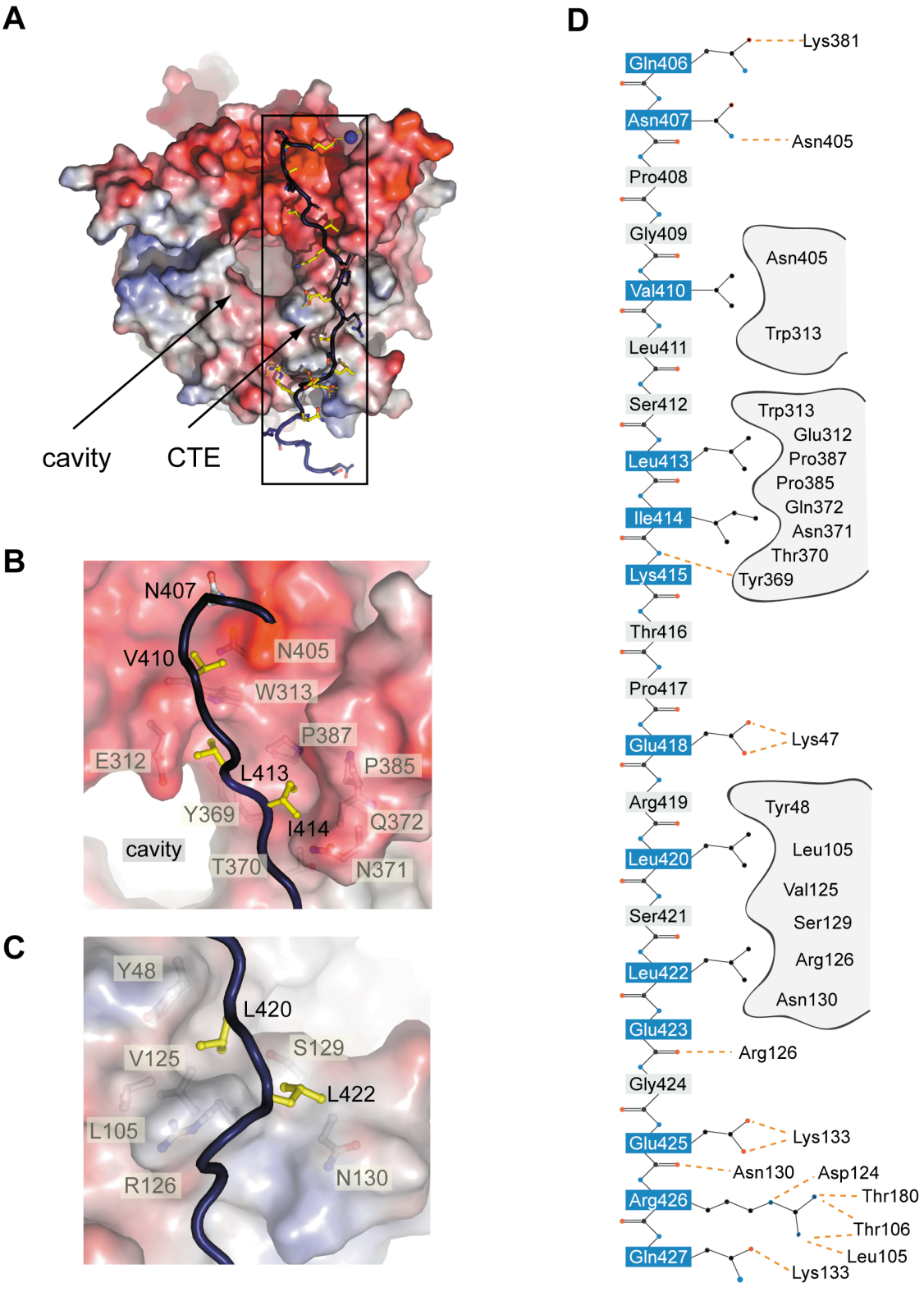
The 30-residue CTE binds in an extended conformation to the bottom face of the β-propeller and buries ~15% (~2,500 $Å^2$) of the surface area of the β-propeller (**Figure 29**). The majority of the interactions between the β-propeller and the CTE are ionic in nature. Eight of the CTE residues form hydrogen bonds with residues of the β-propeller, whereas only five are involved in hydrophobic van der Waals contacts. Hydrogen bonds with the β-propeller are formed by the CTE residues Gln406, Asn407, Lys415, Glu418, Glu423, Glu425, Arg426, and Gln427, while the CTE residues Val410, Leu413, Ile 414, Leu420, and Leu422 form hydrophobic interactions. The side-chains of the remaining residues of the CTE face away from the interface with the β-propeller (**Figure 29** and **Figure 30)**.

The carbonyl group of Gln406 forms a water-mediated hydrogen bond to the ε-amino group of Lys381, and the amide nitrogen atom of Asn407 forms a hydrogen bond with the side-chain carbonyl group of Asn405. Val410 forms close van der Waals contacts with a groove in the β-propeller surface formed by Asn405 and Trp313 (**Figure 29 B**). Leu413 and Ile414 bind to the hydrophobic pockets formed by Glu312, Trp313, Tyr369, and Pro387, and Tyr369, Thr370, Asn371, Gln372, Pro385, and Pro387, respectively. The main-chain amide of Lys415 interacts with the main-chain carbonyl of Tyr369, and the carboxylate group of Glu418 forms a salt bridge with the ε-amino group of Lys47. The hydrophobic residues Leu420 and Leu422 are both lodged into a hydrophobic pocket formed by Tyr48, Leu105, Val125, Arg126, Ser129, and Asn130 (**Figure**

29 C). The main-chain carbonyl moiety of Glu423 forms a hydrogen bond with Arg126, while the side chain faces away from the surface. The carboxylate group of Glu425 interacts with the side-chain amide moiety of Asn130 and forms a salt bridge with the ε-amino group of Lys133. Arg426 forms the most intimate interactions with the β-propeller surface. The guanidinium group of Arg426 interacts with the main-chain carbonyl moiety of Thr180, the hydroxyl group of Thr106, and the main-chain carbonyl of Leu105, as well as with Asp124 via a salt bridge. The side-chain carbonyl group of Gln427 forms a hydrogen bond to the ε-amino group of Lys133 (**Figure 29 D**).

C-terminal extensions have been found in other β-propeller-containing proteins. For example, in Coronin-1, a regulator of actin cytoskeletal dynamics in lymphocytes, the binding of the Coronin-1 CTE to the bottom face of its β-propeller is primarily achieved by strong hydrophobic packing that involves a tryptophan and a tyrosine residue of the CTE[210]. This interaction has been shown to be crucial for the stability of the protein[210, 211]. In order to determine whether the Nup214 CTE is required for the stability of the Nup214 NTD, we expressed a C-terminal truncation mutant, which contains the β-propeller (residues 1-405) but lacks the CTE (Nup214 NTD1-405). Nup214 NTD1-405 was indistinguishable from the crystallized Nup214 NTD in its expression level, stability, localization and behavior on a gel filtration column. This observation strongly suggests that the CTE is not required for the architectural integrity of the Nup214 NTD.

Figure 29: The CTE binds to the bottom face of the Nup214 β-propeller. (**A**) The surface of the Nup214 β-propeller is colored according to the electrostatic potential from –10 k_BT/e_c (red) to + 10 k_BT/e_c (blue). The CTE is shown in blue coil representation with the side chains in ball-and-stick representation. The black box indicates the region magnified in Figure 30. (**B**) Hydrophobic interactions of CTE residues Val-410, Leu-413, and Leu-414 (yellow). (**C**) Interactions of Leu-420 and Leu-422 (yellow) with residues of the β-propeller. Hydrophobic pocket-forming residues are shown in gray. The surface of the β-propeller is colored as in A. (**D**) Schematic representation of the contacts between the β-propeller and the CTE. Hydrogen and ionic bonds are indicated by orange dashed lines and van der Waals contacts with gray grooves.


A
cavity
CTE
B
N407
V410
N405
W313
P387
E312
L413
P385
Y369
I414
Q372
cavity
T370
N371
C
Y48
L420
V125
S129
L422
L105
R126
N130
D
Lys381
Gln406
Asn407
Asn405
Pro408
Gly409
Asn405
Val410
Trp313
Leu411
Ser412
Trp313
Glu312
Leu413
Pro387
Pro385
Gln372
Ile414
Asn371
Thr370
Lys415
Tyr369
Thr416
Pro417
Glu418
Lys47
Arg419
Tyr48
Leu420
Leu105
Val125
Ser421
Ser129
Arg126
Leu422
Asn130
Glu423
Arg126
Gly424
Glu425
Lys133
Asn130
Asp124
Arg426
Thr180
Thr106
Gln427
Leu105
Lys133

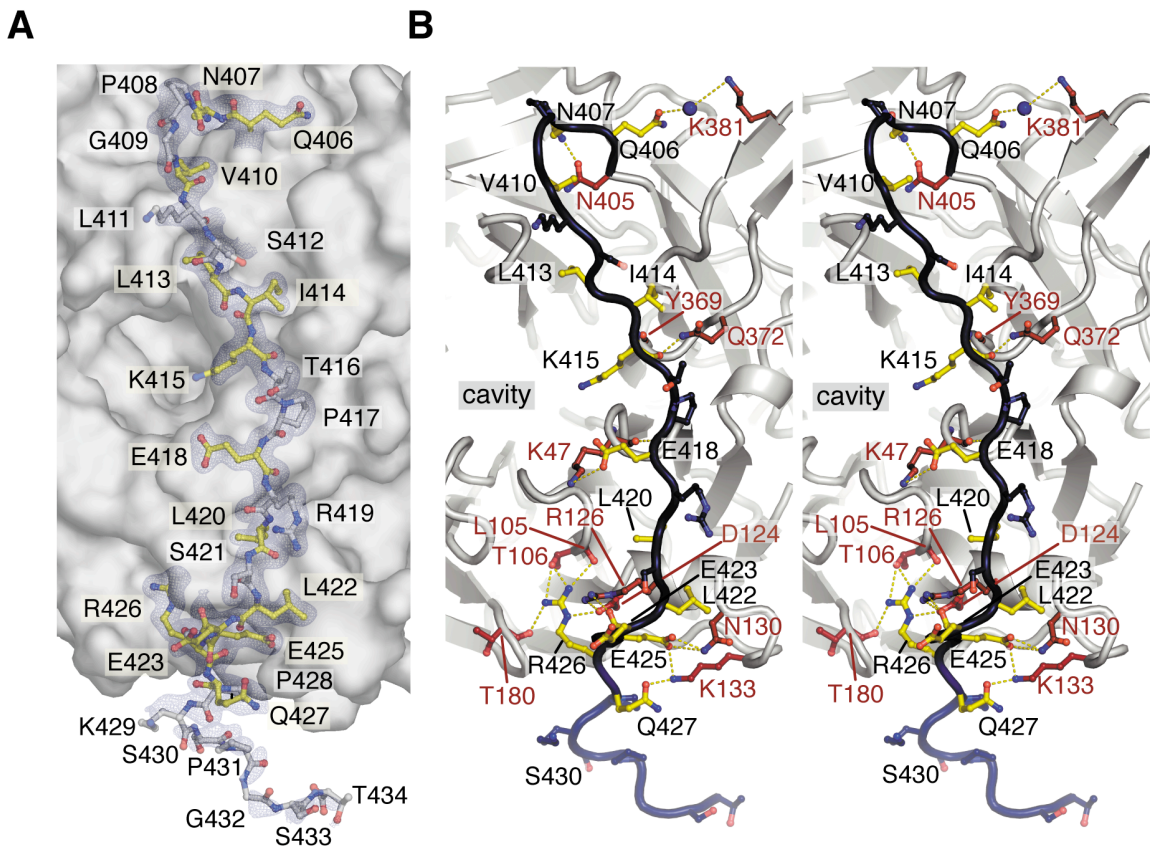


Figure 30: The CTE of Nup214 NTD. (**A**) A simulated-annealed $2|F_o|-|F_c|$ electron density map, contoured at 1.0 σ, is shown for the CTE (ball and stick representation). Residues of the β-propeller domain forming hydrogen bonds with the CTE are colored in ruby, CTE residues that interact with the β-propeller domain are shown in yellow, and CTE residues facing away from the interface are colored blue. (**B**) A stereo-view of the interactions of the CTE with the Nup214 β-propeller domain. The CTE-residues are colored as in Figure 29.

Furthermore, the highly charged nature of the CTE-β-propeller interface does not appear to lock the CTE in a bound conformation. Therefore, we hypothesized that the interaction between the CTE and the β-propeller is dynamic in nature and tested if the CTE is able to dissociate from the β-propeller. We tested the binding of a synthetic 28-residue CTE peptide (residues 405-432) to the Nup214 NTD1-

405 β-propeller on a gel filtration column but were unable to see any interaction. Since isothermal calorimetry (ITC) has generally a much higher sensitivity, we then attempted to measure the binding affinity between the CTE peptide and the Nup214 NTD1-405 with ITC. However, the two domains did not interact with measurable affinity under the condition tested, suggesting that the affinity was below the ITC detection limit. The finding that the CTE is bound to the β-propeller in our crystal structure, together with our inability to measure the inter-molecular binding affinity suggests that the interaction between the β-propeller and the CTE is entropically favored through intra-molecular interactions. In addition, the CTE contains several putative phosphorylation sites, and phosphorylation of Ser430 and Thr437 of the Nup214 CTE was found in a large-scale characterization of HeLa nuclear phosphoproteins[212]. Ser430 is located at the end of the CTE, faces away from the β-propeller and is accessible for phosphorylation (**Figure 30**). Thus, the CTE-β-propeller interaction may be regulated and dynamic in nature.

2.1.9 Comparison to the yeast Nup159 NTD

Although the ~120 MDa vertebrate and the ~60 MDa yeast NPC share a conserved architecture and utilize a similar number of nups[6], several vertebrate nups lack a homologue in yeast, undergo vertebrate specific modifications, or have several homologous nups in yeast[16]. When we compare the sequence of the human Nup214 to its yeast homologue Nup159, it is clear that the various domains are rearranged (**Figure 31 A**).

The crystal structure of the β-propeller of the yeast Nup159 is available[35]. However, while the expression construct of the yeast Nup159 contained additional residues at the C-terminal end of the core β-propeller, those residues were disordered in the crystal structure, suggesting that this region does not interact with the β-propeller. Because of the low sequence conservation between the human and yeast homologues the prediction of the existence of the CTE in lower eukaryotes such as yeast is difficult (**Figure 32**). The comparison between the structures reveals that, apart from the canonical β-propeller folds, both proteins are quite different. The yeast structure lacks the CTE as well as the significantly more extensive and complex decoration of the human NTD. When we compare the β-propeller core of the two homologues, we find that both align reasonably well with an RMSD of ~1.5 Å (**Figure 31 B**). A structural alignment of the Nup214 β-propeller with the only other available nup β-propeller structure, Nup133, shows that they align equally well and also with an RMSD. of ~1.5 Å. We furthermore tested how well other β-propeller proteins that have completely different functions in the cell, such as G_β[209], the murine Coronin-1[210], and the yeast Sif2 WD propeller[213], align. Interestingly, their β-propeller cores align with comparable RMSD of ~1.7 Å. This demonstrates that the human Nup214 β-propeller is not significantly more closely related to the β-propeller core of Nup159 than to the β-propeller of Nup133, G_β, Coronin-1, or Sif2. Although only ~280 residues are required to construct a minimal seven-bladed β-propeller fold, the nup β-propellers have been predicted to contain 400 - 500 residues[34]. Thus,

we speculate that instead of the β-propeller core, the specific functions of β-propeller-containing proteins are conferred by their unique structural features and additional structural elements that dominate the surface properties of the domain.

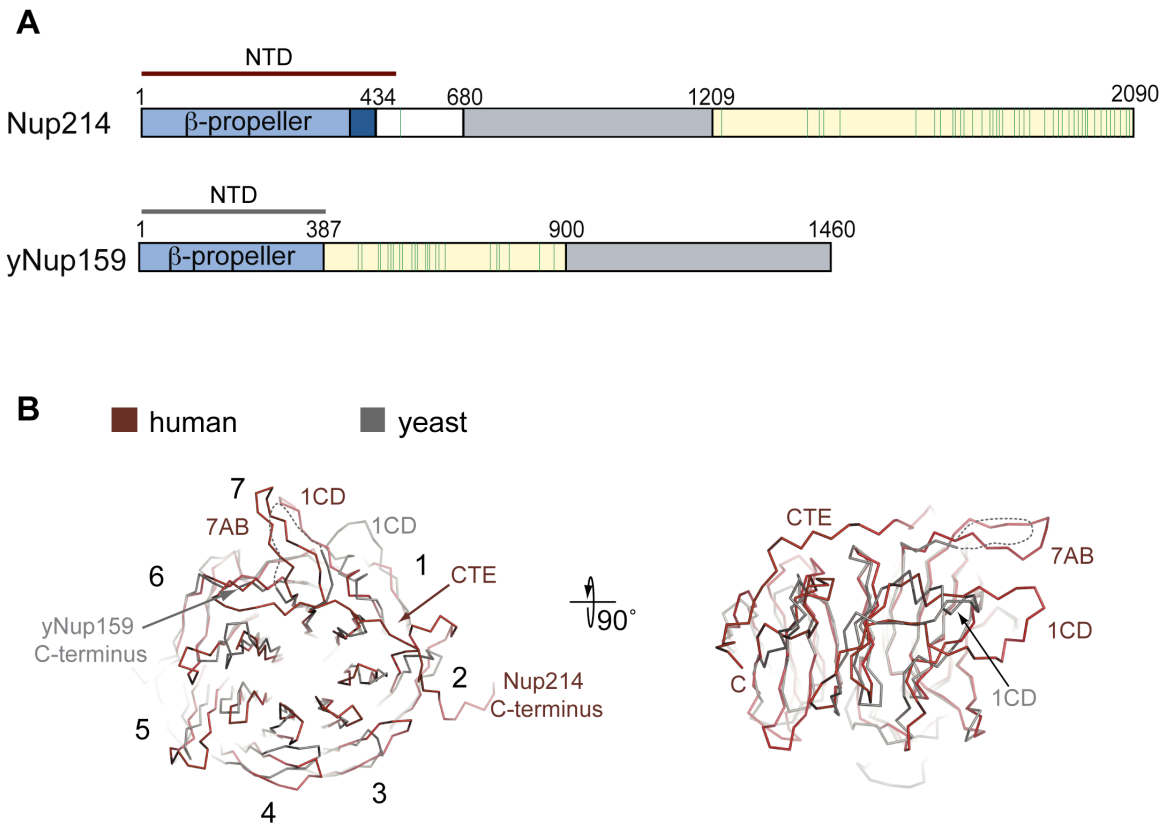


Figure 31: Structural comparison of the NTD of the human Nup214 and its yeast homologue Nup159. (**A**) Domain structure of Nup214 and Nup159. The construct used for crystallization is indicated above the sequence. The domains are colored as in Figure 22. (**B**) C_α trace of a structural superposition of the Nup214 NTD (ruby) and the Nup159 β-propeller (gray).

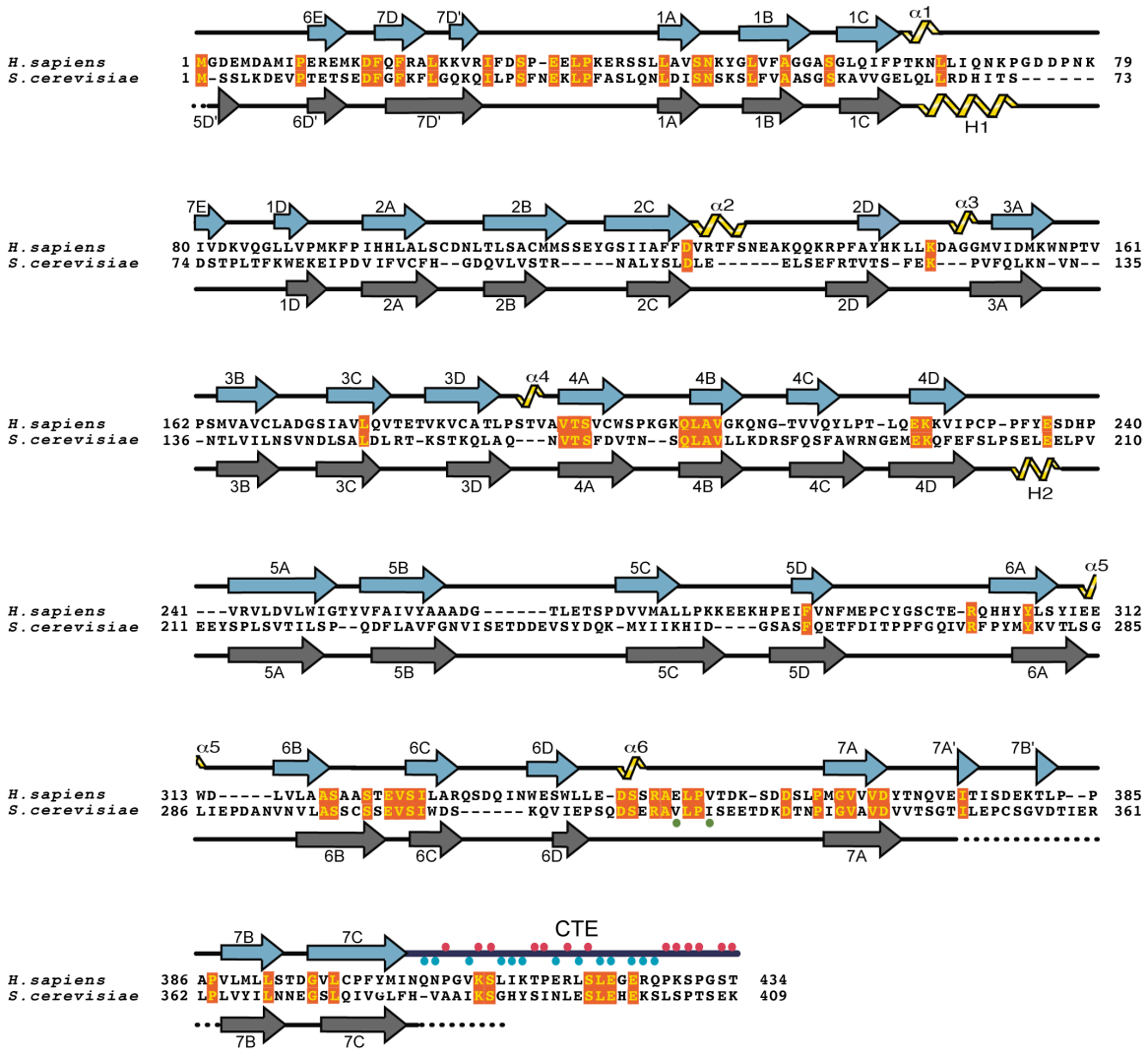


Figure 32: Structure-based alignment of the primary sequence of the human Nup214 and the yeast Nup159 N-terminal β-propeller domains. Secondary structure elements are indicated above (Nup214) and below (Nup159) the aligned sequences. Residues of the CTE that either interact with the β-propeller domain (blue dots) or face away from the interface (red dots) are indicated. The positions of the yeast Nup159 mutations V323E and I326E, which disrupt binding of the yeast Dbp5 helicase, are indicated with a green dot. Blue, β-sheets; black, coil regions; yellow, α-helices; black dots, disordered regions.

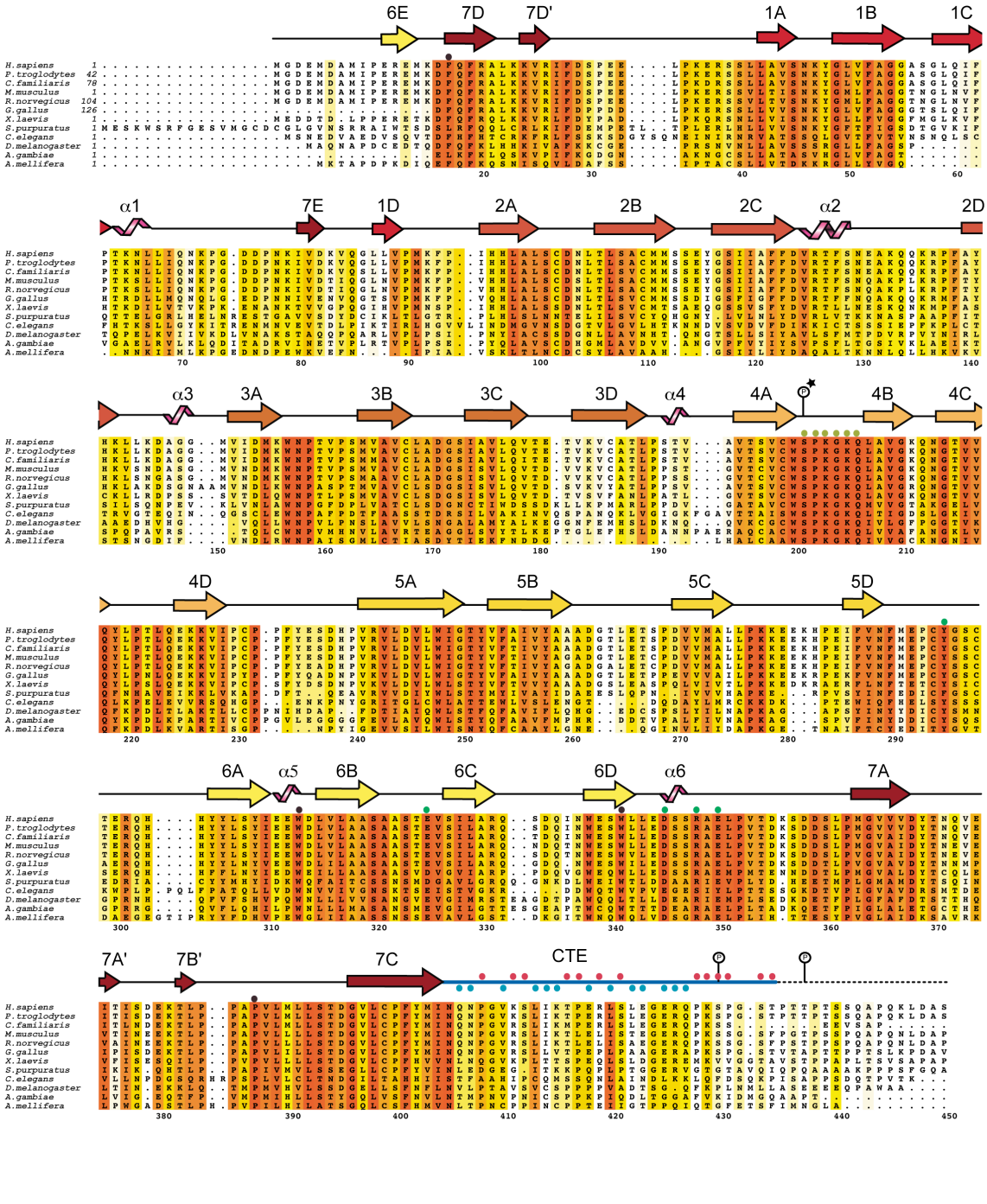
2.1.10 Conserved features of the Nup214 NTD

The vast majority of β-propeller folds are known to mediate protein-protein interactions[206]. Several surface areas have been found to serve as docking platforms. For example, the inter-blade grooves on the side of the β-propeller are often utilized in peptide interactions, while the central channel can bind ligands or prosthetic groups[206]. The β-propeller fold imposes few restraints on its primary sequence and is able to tolerate insertions between the canonical β-strands. These extended loops protrude from the top and bottom face of the β-propeller and can create a variety of binding and regulation sites.

Several proteins have been found to interact directly with Nup214. The Nup88-Nup214 NPC sub-complex is involved in the anchoring of the cytoplasmic filaments[214], whereas the recruitment of Ddx19 by Nup214 is a crucial step of the mRNA export pathway[150]. In order to identify possible protein interaction sites on the surface of the Nup214 NTD, we analyzed its conservation and electrostatic potential. While the sequences of Nup214 homologues are not very conserved within eukaryotes, sequences can be aligned reliably within higher eukaryotes and the majority of the conserved residues were found to be part of the core β-strands (**Figure 33**). Nevertheless, three regions on the Nup214 NTD surface were pinpointed.

Two conserved regions, region 1 and region 2, are located at the bottom face of the β-propeller (**Figure 34).** Region 1 is formed by the 4AB loop, has a

Figure 33: Sequence alignment of Nup214 NTD within higher eukaryotes. The secondary structure is indicated above the sequence as arrows (β-strands), black lines (coil regions), and pink helices (α-helices). The numbering below the alignment is relative to the human Nup214 sequence. The overall sequence conservation at each position is shaded in a color gradient from white (<40% identity) to red (100% identity). Residues of region 1, region 2, and region 3 as well as phosphorylation sites of the Nup214 NTD are indicated above the sequences; a putative phosphorylation site is marked with an asterisk.



- CTE residues interacting with the β-propeller domain
- CTE residues not participating in interactions
- conserved residues of region 1
- conserved residues of region 2
- conserved residues of region 3
- phosphorylation site
- putative phosphorylation site

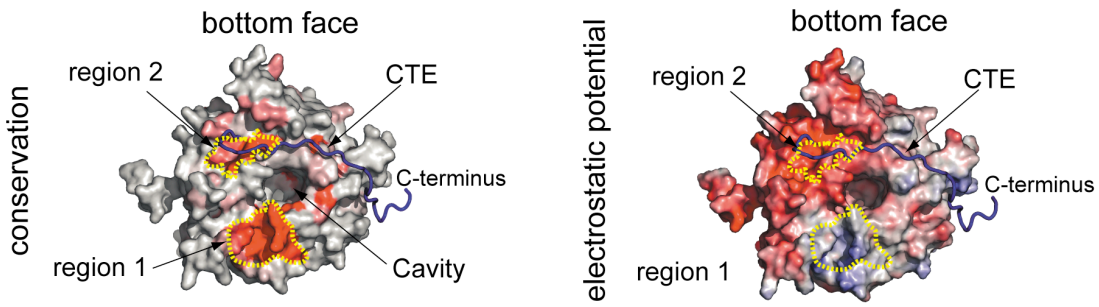


Figure 34: Conserved regions of the Nup214 NTD. Surface representation of the Nup214 NTD showing the surface conservation within higher eukaryotes (*left*) and electrostatic potential (*right*). The conserved surface is shaded from gray (<70% identity) to red (100% identity) according to the alignment in Figure 33. Conserved region 1 and 2 are located at the bottom side of the β-propeller. The β-propeller domain (residue 1 - 405) is shown in surface representation and the C_α trace of the CTE is shown in a blue coil representation.

positive electrostatic potential and is the most conserved region within higher eukaryotes, whereas region 2 is formed by Phe17, Trp313, Trp341, and Pro387 and has a negative electrostatic potential (**Figure 34**). Interestingly, the 4AB loop within region 1 is lysine-rich and invariant within higher eukaryotes but shares no sequence conservation with any of its yeast homologues (**Figure 32** and **Figure 33**). Therefore, we speculate that a vertebrate-specific protein might bind to this positively charged, highly conserved 4AB loop. In addition, Nup214 has been found to be highly post-translationally modified and, in particular, to be phosphorylated during interphase and hyperphosphorylated during mitosis[215]. The sequence of the 4AB loop (W**S**PKGKQL) resembles the consensus site for a number of Ser/Thr protein kinases, and we identify Ser201 within this loop as a

putative serine phosphorylation site using a bio-informatics analysis with NetPhos[216] (**Figure 33**).

The third conserved region, region 3 covers about one third of the side surface of the Nup214 NTD (**Figure 35**) and has an invariant core formed by the 5D6A loop (Tyr295), the 6BC loop (Glu325), and the 6D7A loop (Asp345, Arg348, and Glu350). The 6D7A loop of the yNup159 β-propeller has been shown to directly interact with the DEAD-box helicase yDbp5, and a mutation of two residues within this loop (V323E/I326E) has been shown to disrupt the formation of a complex between the two proteins[35]. Although the yeast and human homologues display considerable sequence conservation within this loop, residues that have been identified to be crucial for yNup159-yDbp5 complex formation are not conserved in Nup214. In fact, one of the two yNup159 residues (V323E) that disrupt the yNup159-yDbp5 complex when mutated is already a glutamate in the human Nup212 β-propeller (Glu350) (**Figure 32**). The primary sequence of yeast yDbp5 and its human homologue Ddx19 are 45% identical, and while it seems plausible that the same region of the surface of the Nup214 β-propeller binds to Ddx19, the requirement of different residues for complex formation appears necessary.

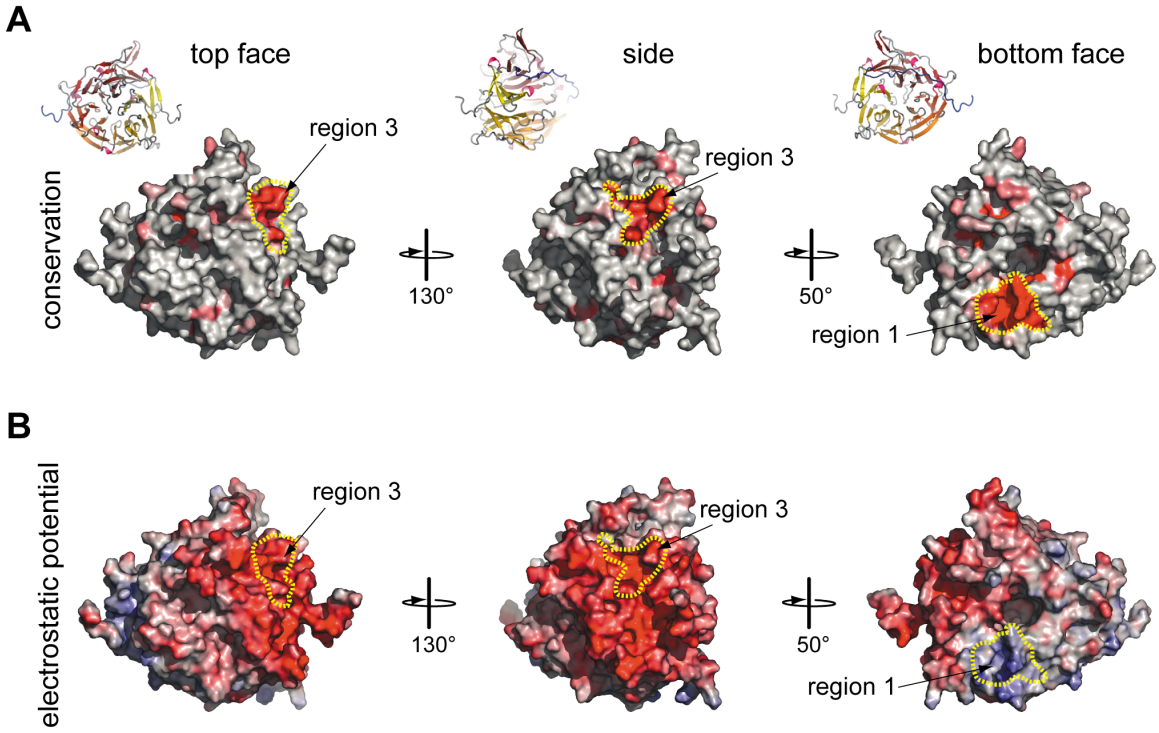


Figure 35: Conserved region 3 is located on the side surface of the Nup214 NTD . (**A**) Surface representation showing conservation of residues within higher eukaryotes. The conserved surface is shaded from gray (<70% identity) to red (100% identity) according to the alignment in Figure 33. (**B**) Electrostatic potential of the Nup214 NTD surface (colored as in Figure 34, A).

2.2 Analysis of the interaction of Nup214 with the DEAD-box helicase Ddx19

In yeast, the DEAD-box helicase yDbp5 localizes to the cytoplasmic side of the NPC by binding to yNup159, and, by removing proteins from mRNA plays an essential role in remodeling of mRNP complexes. Mutants of yDbp5 that interfere with yDbp5's NPC localization or with its mRNP remodeling activity result in mRNA accumulation in the nucleus[35, 166, 167]. Although the exact mode of binding may differ between the yeast and human homologues, targeting of the DEAD-box helicase to the NPC appears to be conserved throughout evolution. Ddx19 and Nup214 were shown to interact when co-immunoprecipitated from rabbit reticulolysate[150]. To gain deeper insight into the Nup214-dependent recruitment of Ddx19 to the cytoplasmic side of the NPC, we undertook an extensive biochemical analysis of the evolutionarily conserved Ddx19-Nup214 interaction.

2.2.1 Purification of Ddx19

Full length human Ddx19 was expressed in *E. coli* BL21 (DE3) cells and purified using appropriate affinity and gel filtration chromatography (**Figure 36**). To overcome stability and aggregation problems during purification, the purification buffers contained 10% glycerol and the purification protocol was optimized for speed (for details of the purification see Materials and Methods chapter). Typically the purification yielded ~10 mg/L of bacterial culture and the purified Ddx19 was concentrated to 40 mg/ml.

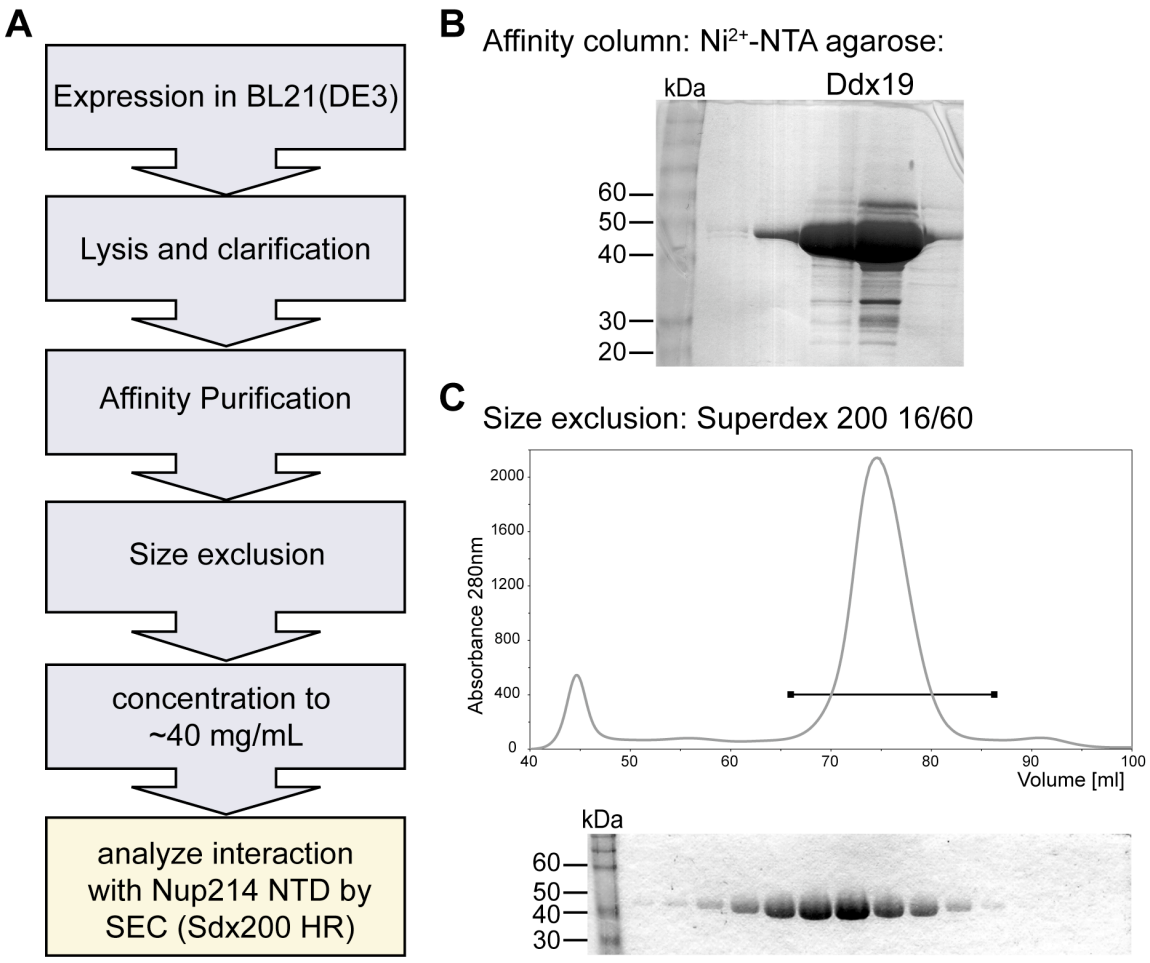


Figure 36: Purification of Ddx19. (**A**) Purification scheme for Ddx19. The flow chart depicts the purification of recombinant full length Ddx19. (**B**) Representative Coomassie brilliant blue stained SDS-PAGE gel of the affinity purification step. (**C**) Representative chromatogram (*top*) and Coomassie brilliant blue stained SDS-PAGE gel (*bottom*) of the gel filtration purification step of Ddx19. The black line in the chromatogram indicates fractions visualized on the SDS-PAGE gel.

2.2.2 Identification of the Ddx19 binding region on Nup214

The structure of Nup214 NTD revealed an N-terminal β-propeller domain followed by a ~35 residue C-terminal extension (CTE). The conserved surface features suggested several possible protein-protein interaction surfaces[46]. In order to identify the surface that mediates the interaction with the DEAD-box helicases Ddx19, we performed two types of experiments. Initially, we generated Nup214 NTD mutants that lacked the previously suggested protein-protein interaction regions[35, 46]. Then we performed additional mutagenesis to identify key residues of Nup214 that mediate the interaction with Ddx19.

2.2.2.1 Nup214 interacts with Ddx19 via its conserved 6D7A loop.

To identify the interface between Ddx19 and Nup214, we tested several putative protein interaction sites on the Nup214 NTD surface for Ddx19 binding. We designed Nup214 variants, including Nup214 NTD 1–405 in which the CTE is deleted and Nup214 NTD Δ6D7A in which a flexible linker replaces the highly conserved surface loop 6D7A. The recombinant proteins were expressed, purified to homogeneity, and monitored for complex formation with Ddx19 using size exclusion chromatography (SEC) (**Figure 37**). While the CTE-deletion mutant Nup214 1–405 was capable of complex formation, the loop-deletion mutant Nup214 NTD Δ6D7A resulted in a protein that was incapable of binding to Ddx19 (**Figure 37 B** to **D**)

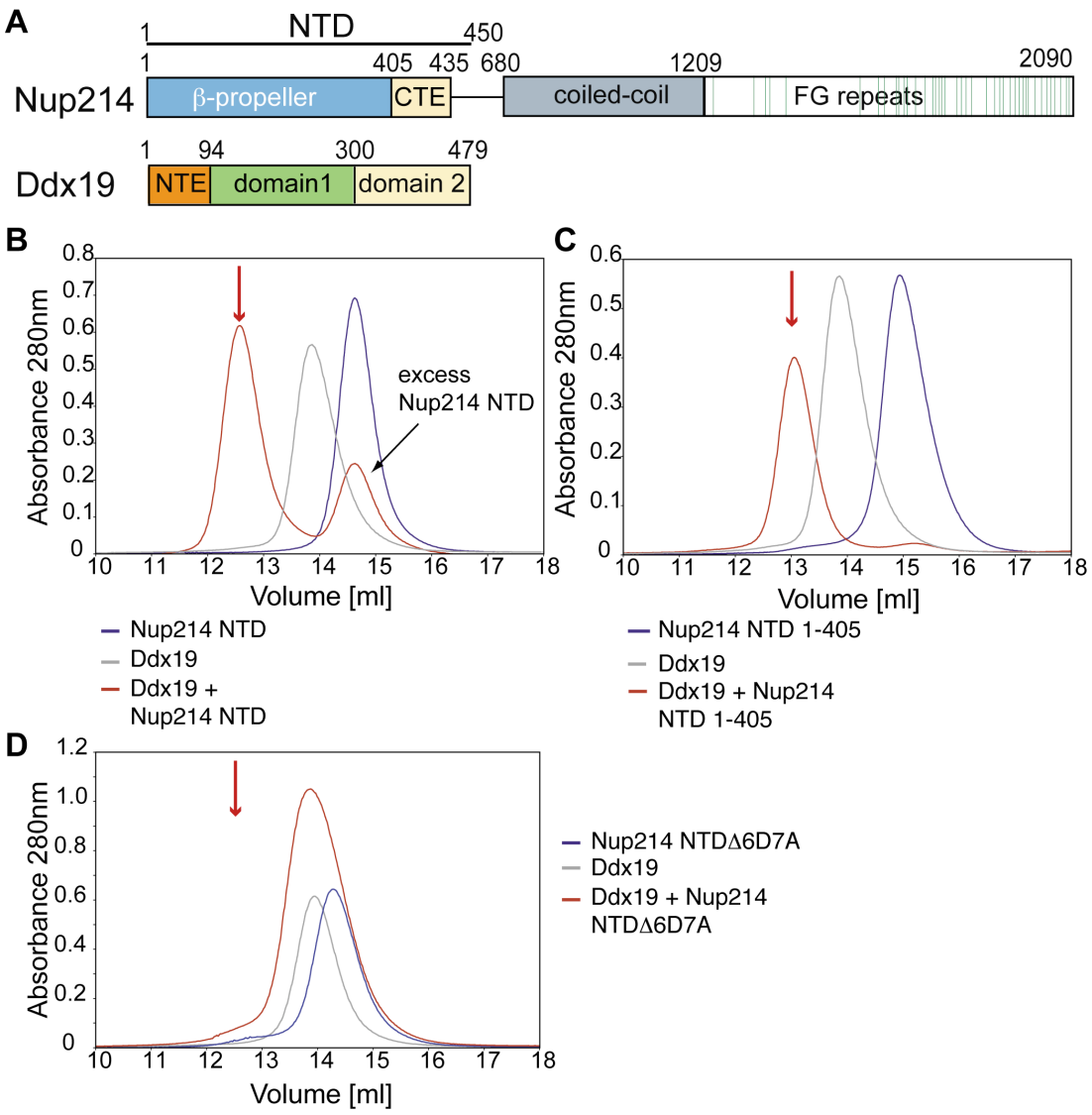
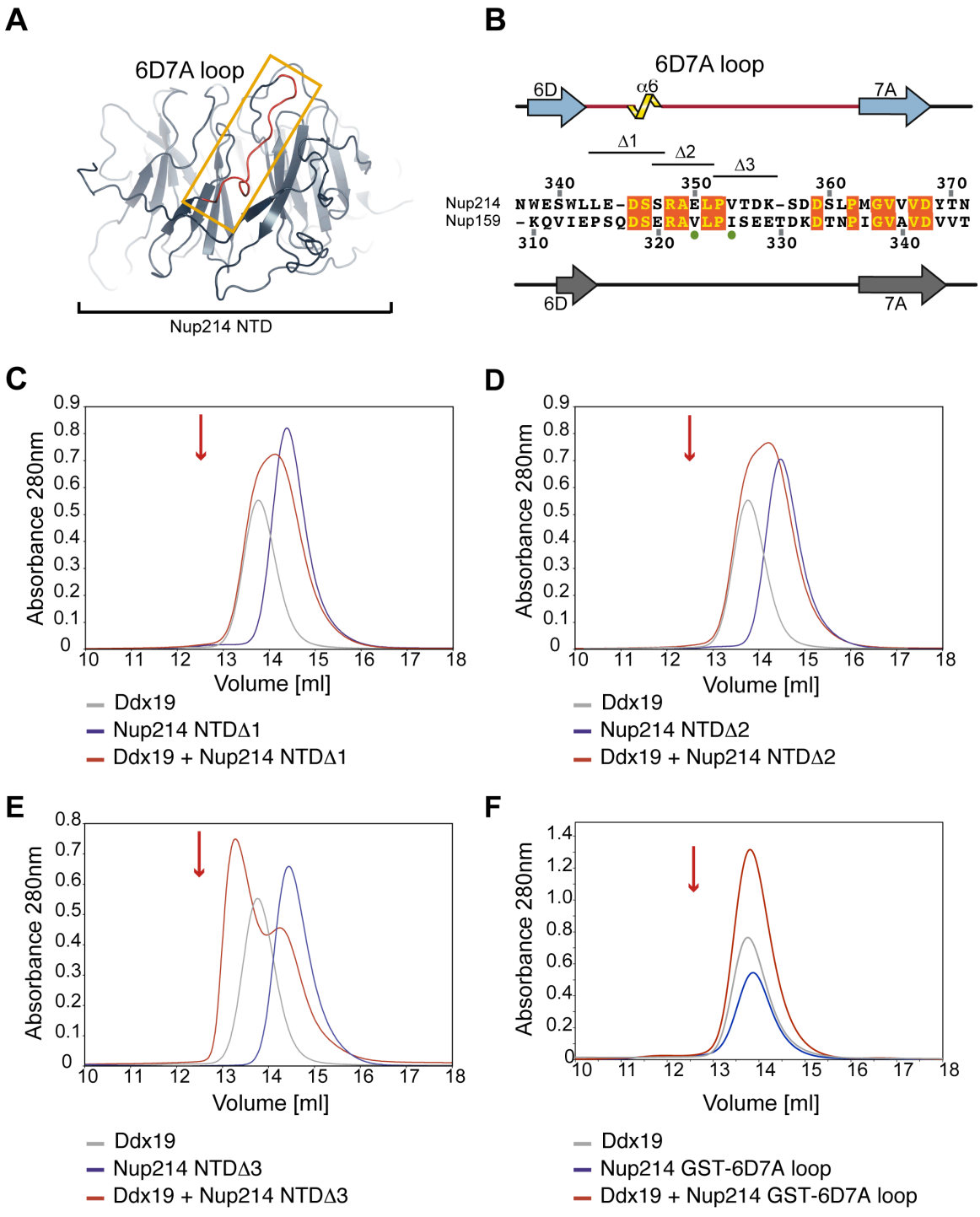


Figure 37: The 6D7A loop of the Nup214 NTD is essential for Ddx19 binding. (**A**) Domain organization of Nup214 and Ddx19. For Ddx19, the N-terminal extension (NTE) and the two RecA-like domains are indicated. For Nup214, the β-propeller domain and its C-terminal extension (CTE), the coiled-coil domain, and the FG-repeats are indicated. Gel filtration profiles of wild-type Ddx19 incubated with the (**B**) Nup214 NTD, (**C**) Nup214 NTD 1–405, or (**D**) Nup214 NTDΔ6D7A. Gel filtration profiles of Ddx19 are colored in gray, the Nup214 NTD variants in blue, and the elution profile resulting from incubating Ddx19 with Nup214 proteins in red. The red arrow indicates the expected elution volume of the Nup214•Ddx19 complex.

The 6D7A interblade connector loop is located on the side surface of the Nup214 β-propeller domain, encompasses 20 residues and connects the outermost β-strand 6D of blade 6 with the innermost β-strand 7A of blade 7 (**Figure 29 A, B**). In yeast, it has been shown that two mutations within the 6D7A loop, V323E and I326E of yNup159 (the yeast homologue of Nup214), disrupt the interaction with yDbp5 (the yeast homologue of Ddx19)[35]. However, while the yeast Nup159 and the human Nup214 display considerable sequence conservation in this loop, one of the two mutated Nup159 residues is a glutamate (E350) in the human Nup214 β-propeller domain (**Figure 38 B**).

To address the differences between the yeast and human Nup214 homologues in helicase binding, we performed an alanine-scanning mutagenesis of the 6D7A loop. A series of 13 single alanine mutants covering the exposed residues of the 6D7A loop was generated (**Table 2,** Material and Methods section). In addition, we constructed the V353E mutant and the E350A/V353E double mutant, since these residues were found to be critical for complex formation in the yeast homologue. All 15 mutants were purified to homogeneity in milligram amounts and behaved indistinguishably from the wild-type protein as judged by gel filtration. The ability of all 15 mutants to bind to full-length Ddx19 was analyzed by SEC. However, none of the 15 Nup214 NTD mutants tested markedly disrupted complex formation, suggesting that multiple residues of the 6D7A loop are involved in the binding of Ddx19.

Figure 38: A 9-residue region in the 6D7A loop of the Nup214 NTD is essential for Ddx19 binding. (**A**) Location of the 6D7A loop (red) of Nup214 NTD. (**B**) Structure-based sequence alignment of the Nup214 and yNup159 6D7A loops. The secondary structure and residue numbering are indicated above and below the sequence for Nup214 and yNup159, respectively. The localization of the Nup159 double mutant V323E/I326E is labeled by green dots, and conserved residues are highlighted. The residues of the 6D7A loop that are deleted in the Nup214 NTDΔ1, Δ2, and Δ3 variants are indicated. Gel filtration profiles of full-length wild-type Ddx19 incubated with the deletion mutants (**C**) Nup214 NTDΔ1, (**D**) Nup214 NTDΔ2, (**E**) Nup214 NTDΔ3 or (**F**) Nup214 GST-loop-6D7A. Gel filtration profiles are colored as in Figure 37 and the red arrow indicates the expected elution volume of the Nup214•Ddx19 complex.


A
6D7A loop
Nup214 NTD
B
6D7A loop
6D
α6
7A
Δ1
Δ2
Δ3
340
350
360
370
Nup214 NWESWLLE-DSSRAELPVTDK-SDDSLPMGVVVDYTN
Nup159 -KQVIEPSQDSERAVLPISEETDKDTNPIGVAVDVVT
310
320
330
340
6D
7A
C
Absorbance 280nm
Volume [ml]
Ddx19
Nup214 NTDΔ1
Ddx19 + Nup214 NTDΔ1
D
Absorbance 280nm
Volume [ml]
Ddx19
Nup214 NTDΔ2
Ddx19 + Nup214 NTDΔ2
E
Absorbance 280nm
Volume [ml]
Ddx19
Nup214 NTDΔ3
Ddx19 + Nup214 NTDΔ3
F
Absorbance 280nm
Volume [ml]
Ddx19
Nup214 GST-6D7A loop
Ddx19 + Nup214 GST-6D7A loop

To identify a region of the 6D7A loop that is required for Ddx19 binding, we constructed three deletion mutants, Nup214 NTDΔ1, Δ2, and Δ3, covering residues 343 to 356 (**Figure 38 B**). For each mutant, a stretch of 5 consecutive residues was deleted from the solvent-exposed part of the 6D7A loop, as seen in the Nup214 NTD crystal structure[46]. Again, the behavior of all three mutants on a gel filtration column was indistinguishable from that of the wild-type protein. We tested all three proteins for their ability to interact with full-length Ddx19 by SEC. While the Nup214 NTDΔ3 was capable of interacting with Ddx19, the complex peak displayed a substantial shift toward lower molecular weight, indicating a weakened, more dynamic interaction between the two proteins (**Figure 38 E**). The two remaining deletion mutants, Nup214 NTDΔ1 and Δ2, abolished complex formation with Ddx19 (**Figure 38 C**, **D**). To determine whether the 6D7A loop was not only necessary, but also sufficient for complex formation, we tested a GST-6D7A-His_6 loop fusion peptide for complex formation with Ddx19 (**Figure 38 F**). We found that the fusion peptide was unable to bind to Ddx19. In summary our data demonstrate that multiple residues of the 6D7A loop of the Nup214 NTD are involved in Ddx19 binding and that the 6D7A loop of the Nup214 NTD is required, but not sufficient for complex formation with Ddx19.

2.2.3 Binding surface on Ddx19

While the structures of the human Nup214 NTD and yeast Nup159 β-propeller domain have been solved, no structural data about the DEAD-box helicases Ddx19 or yDbp5 are available[35, 46]. However, as previously described in section 1.6.2.2, all DEAD-box helicases share a common architecture in which two RecA-like globular domains are connected by a short flexible linker[153, 155]. In addition, all DEAD-box helicases contain nine conserved sequence motifs that are dispersed over the entire protein sequence (**Figure 18**). Therefore we were able to generate a homology model for human Ddx19. The homology model is based on a sequence alignment of 10 eukaryotic Ddx19 homologues with the sequence of the *Methanococcus jannaschii* Dead-box helicase MjDEAD, for which a crystal structure is available[217]. While the central core region that contains the conserved sequence motifs aligned without major sequence gaps, the Ddx19-specific N- and C-terminal sequence overhangs are not conserved in MjDEAD. Hence, we have high confidence in the accuracy of our structural homology model for the core region, but we are unable to deduce structural information for the unique N- and C-terminal extensions of Ddx19.

To detect Ddx19 peptides that are involved in Nup214 binding, we probed a Ddx19 peptide array with radioactively labeled Nup214 NTD (**Figure 39**). This approach identified several consecutive Ddx19 peptide fragments that fall into two major regions when mapped onto the surface of a Ddx19 homology model (**Figure 39 C**). The major interaction surface is found within domain 1, occupying

approximately half of the surface area. A second, substantially smaller surface patch is found within domain 2, closely localized to the surface cleft between the two domains (**Figure 39**). These data suggest that the primary interaction between Ddx19 and Nup214 is mediated via the surface of domain 1.

To identify Ddx19 residues that are critical for mediating complex formation, we performed alanine-scanning mutagenesis. We selected 60 residues of Ddx19 that were dispersed over the entire Ddx19 sequence based on their predicted surface localization, their charge, and their conservation for mutagenesis. The surface localization was determined based on our Ddx19 homology model. We were able to purify 54 of these mutants to homogeneity, and their capability to bind to the Nup214 NTD was measured by SEC. The behavior of the 54 individual mutants on a gel filtration column was indistinguishable from that of the wild-type protein, demonstrating that the introduced mutations did not interfere with the folding of the protein. Notably, we identified a single mutant of Ddx19, Ddx19 R259A that abolishes the interaction between the two proteins (**Figure 40 A**).

Based on this finding, we analyzed additional mutations of R259, R259K, R259Q, and R259E, as well as five alanine mutants of residues surrounding R259 (L222A, D255A, Q256A, I258A, and R262A). The Ddx19 R259K mutant as well as the five surrounding mutants had no effect on the ability of Ddx19 to bind to the Nup214 NTD (**Figure 40 B**). However, the R259Q mutant was only

capable of forming a weak complex with the Nup214 NTD, while the R259E mutant abolished complex formation (**Figure 40 C, D**).

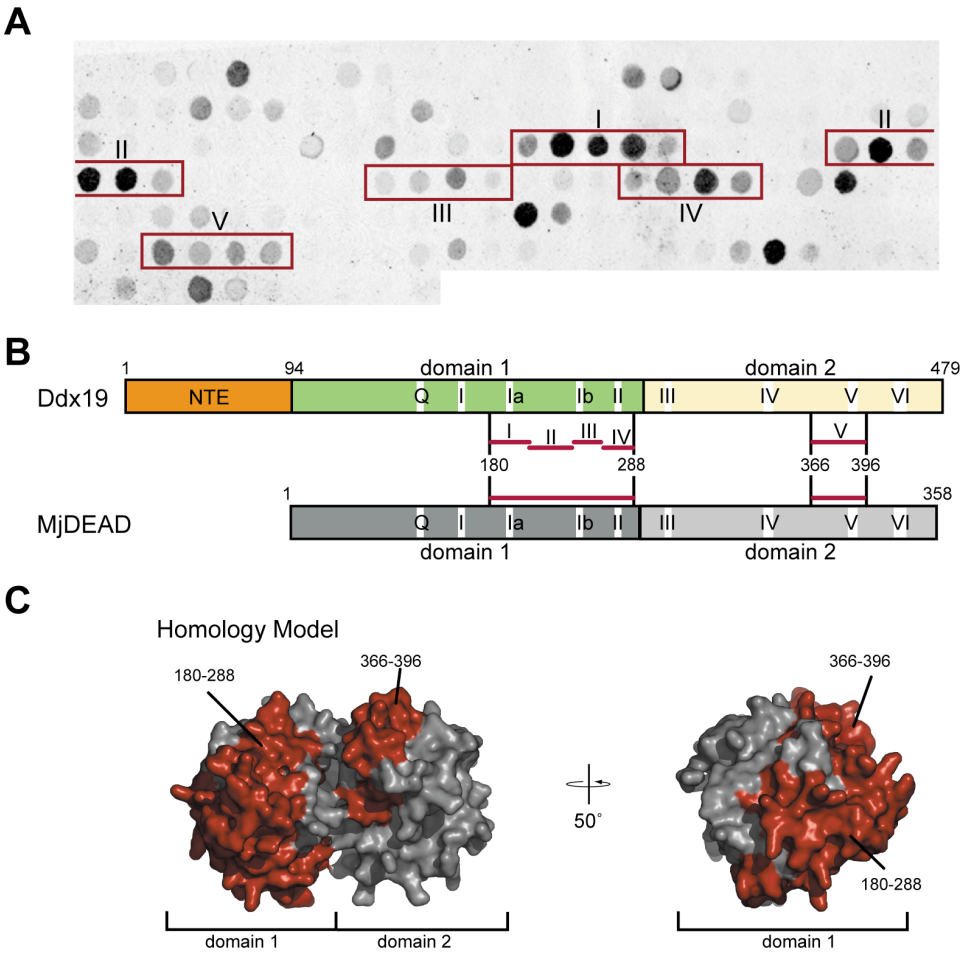


Figure 39: Analysis of the Nup214•Ddx19 interaction by peptide array. (**A**) Ddx19 peptide array probed with the radioactively labeled Nup214 NTD. The five-identified regions containing at least four consecutive peptide spots are numbered I to V and are indicated by a red box. (**B**) Domain structures of Ddx19 and the *M. jannaschii* DEAD-box protein MjDEAD. The five identified regions and their corresponding residues are indicated. (**C**) Homology model of Ddx19 in surface representation based on MjDEAD (PDB code 1HV8). Identified peptides were mapped onto the surface and colored in red. The residue numbering is relative to human Ddx19.

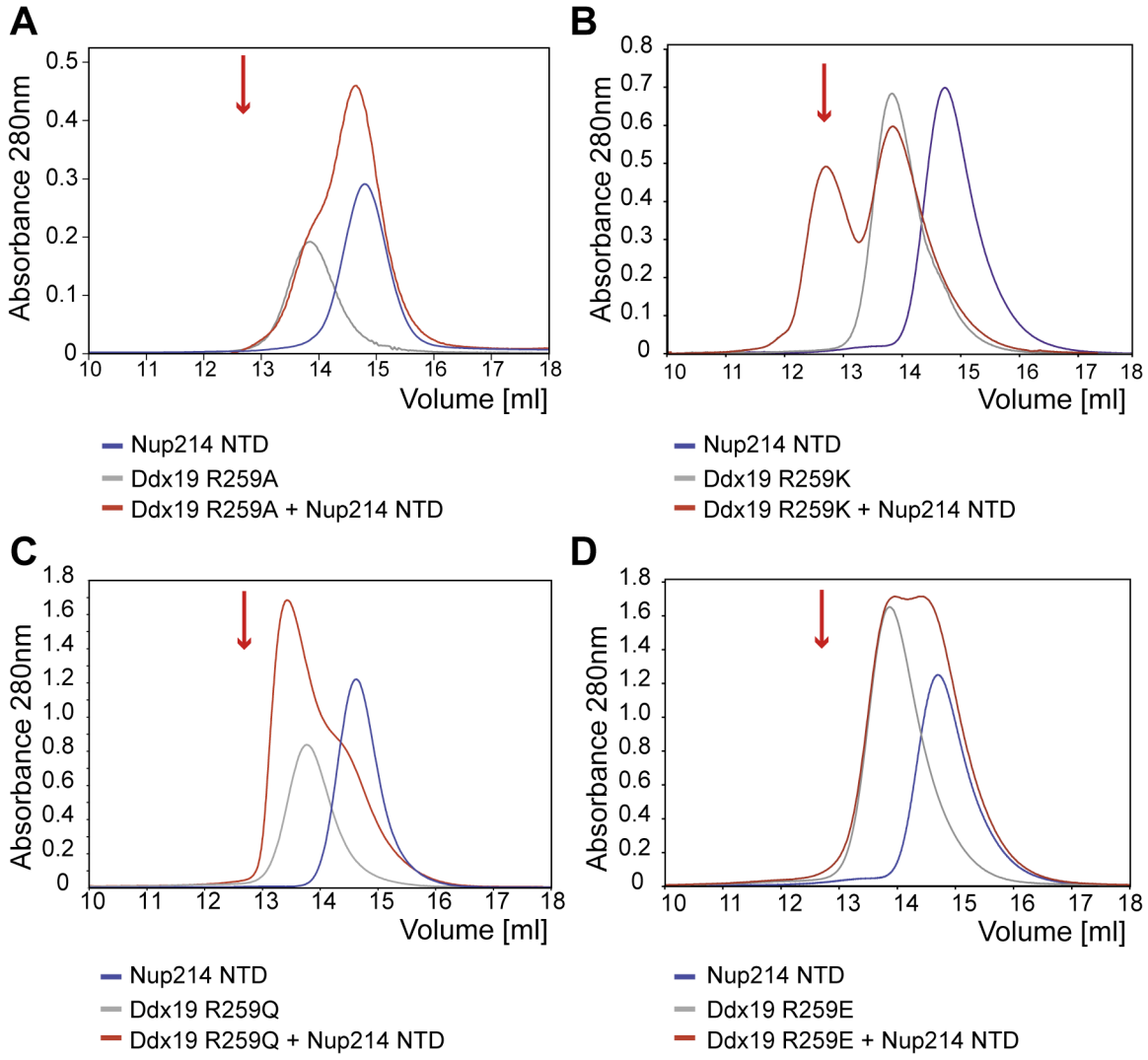


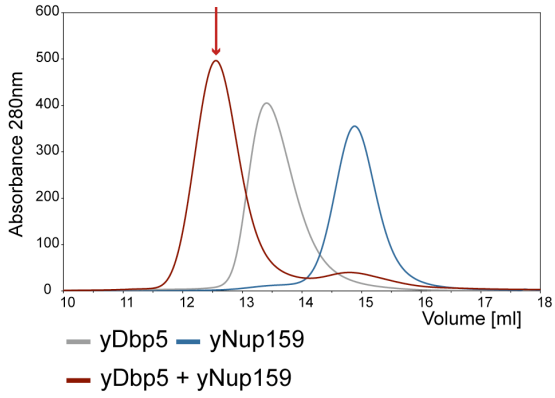
Figure 40: **Arginine 259 of Ddx19 is crucial for binding to Nup214 NTD.** Gel filtration profiles of the Nup214 NTD (blue) and the Ddx19 mutants (grey) (**A**) R259A, (**B**) R259Q, (**C**) R259K and (**D**) R259E. The elution profile resulting from incubation of Nup214 NTD with Ddx19 (red) are indicated. The red arrow indicates the expected elution volume of the Nup214 NTD•Ddx19 complex.

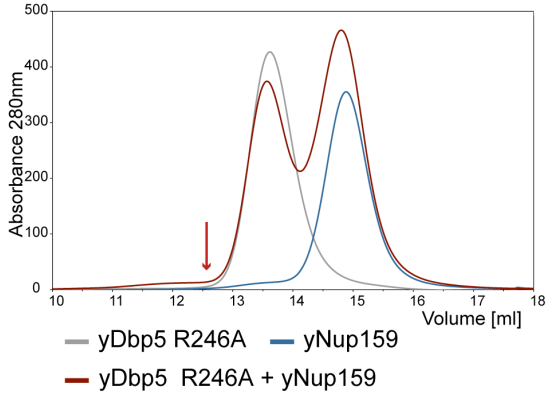
2.2.4 The interaction between Nup214 and Ddx19 is conserved in yeast

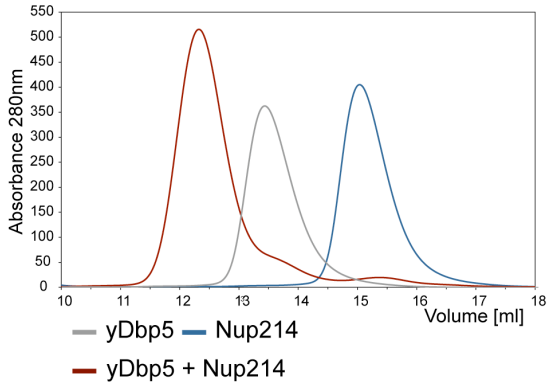
The interaction of Nup214 with Ddx19 has been suggested to be conserved from man to yeast, and genetic and physical interactions have been reported for the yeast homologues of Nup214 and Ddx19, yNup159 and yDbp5, respectively[150]. In order to validate the functional conservation, we first tested for the ability of the yeast proteins to substitute for their human counterparts. Then based on our finding that arginine 259 is essential for complex formation between Nup214 NTD and Ddx19, we tested whether arginine 256, the equivalent residue of the crucial arginine 259 in human Ddx19, is essential for complex formation in yeast.

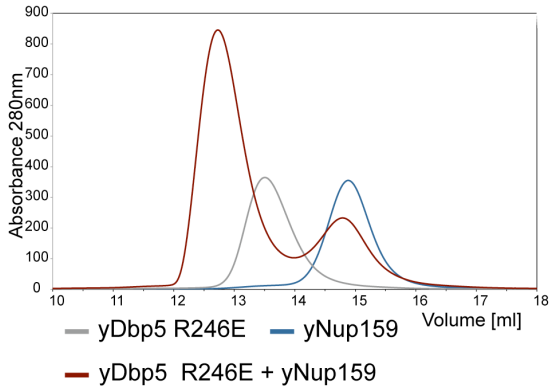
yNup159, yDbp5 and the yDbp5 mutants R256A, R256E and R256K were cloned, expressed, purified to homogeneity and tested for interaction on SEC (**Figure 41**) We found that both yDbp5 and yNup159 are capable of forming a hybrid human-yeast Nup•helicase complex, suggesting a high conservation in the recruitment of the essential DEAD-box helicase to the NPC (**Figure 41 B, C**). We furthermore analyzed a series of yDbp5 mutants for interaction with Nup159. Similar to the human homologues, the yDbp5 R256A and R256E mutants were unable to form a complex with Nup159, while the yDbp5 R259K mutant restored complex formation (**Figure 41 D** to **F**). This behavior is reminiscent of the human proteins and underlines the evolutionary and functional conservation of the interaction between the two homologous proteins.

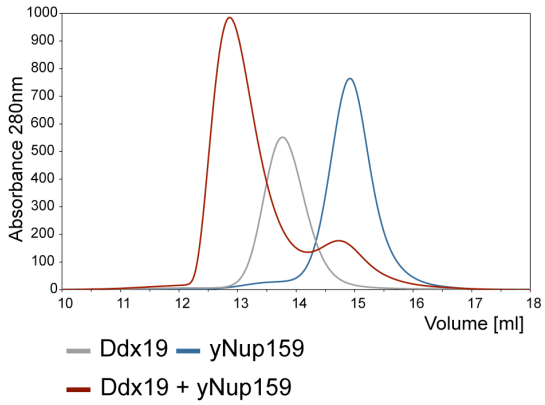
Figure 41: The interaction of Ddx19 with Nup214 is conserved from yeast to human. Gel filtration profiles of yDbp5 incubated with the (**A**) Nup159 NTD or (**B**) Nup214 NTD and the yNup159 NTD incubated with (**C**) Ddx19, (**D**) yDbp5 R256A, (**E**) yDbp5 R256E, (**F**) yDbp5 R256K. Elution profiles of Nup214 NTD or Nup159 are shown in blue; Ddx19, yDbp5 or yDbp5 mutants R256A, R256E and R25K in gray, and the elution profile resulting from incubation in red. The red arrow indicates the expected elution volume of the yNup159 NTD•yDbp5 complex.

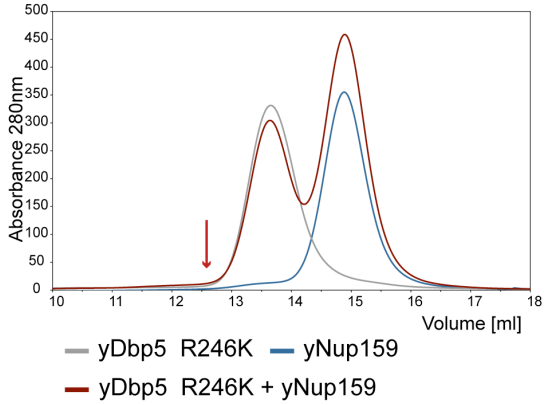

Absorbance 280nm
Volume [ml]
yDbp5
yNup159
yDbp5 + yNup159


Absorbance 280nm
Volume [ml]
yDbp5 R246A
yNup159
yDbp5 R246A + yNup159


Absorbance 280nm
Volume [ml]
yDbp5
Nup214
yDbp5 + Nup214


Absorbance 280nm
Volume [ml]
yDbp5 R246E
yNup159
yDbp5 R246E + yNup159


Absorbance 280nm
Volume [ml]
Ddx19
yNup159
Ddx19 + yNup159


Absorbance 280nm
Volume [ml]
yDbp5 R246K
yNup159
yDbp5 R246K + yNup159

2.3 Crystal Structure of the Nup214 NTD•Ddx19 complex

2.3.1 Purification and Crystallization of the Nup214 NTD•Ddx19 complex

mRNA is transcribed, processed, and packaged into an mRNP in the nucleus. Subsequently, the mRNP is transported through the NPC to the cytoplasm, where translation takes place. A major step in the transport across the NPC is the removal of mRNA export receptors by the DEAD-box helicase Ddx19[166, 167]. DEAD-box helicases are enzymes that utilize the free energy change of ATP binding and hydrolysis to dissociate or unwind RNA duplexes, and/or displace RNA-bound proteins[153]. During this process, the DEAD-box helicases are thought to cycle through distinct conformational states[218]

In order to crystallize the Nup214 NTD•Ddx19 complex, we screened several nucleotide states including ADP, or the non-hydrolysable ATP-analogs AMPPNP and ATP-γ-S, of Ddx19. For complex formation, Nup214 NTD was incubated with an equimolar amount of full length Ddx19, and purified by size exclusion chromatography without, or in the presence of, nucleotide (**Figure 42)**. The fractions containing the complex were pooled and concentrated to 40 mg/ml.

The Nup214 NTD•Ddx19-ADP, Nup214 NTD•Ddx19-ATP-γ-S and Nup214 NTD•Ddx19-AMPPNP as well as a Nup214 NTD•Ddx19-nucleotide-free complexes were screened for crystal formation. Crystals of Nup214 NTD•Ddx19 were obtained only in the presence of ADP. The best crystals grew with fresh protein at a concentration of 15 to 20 mg/ml in 30% (w/v) PEG 3000, 0.1M CHES

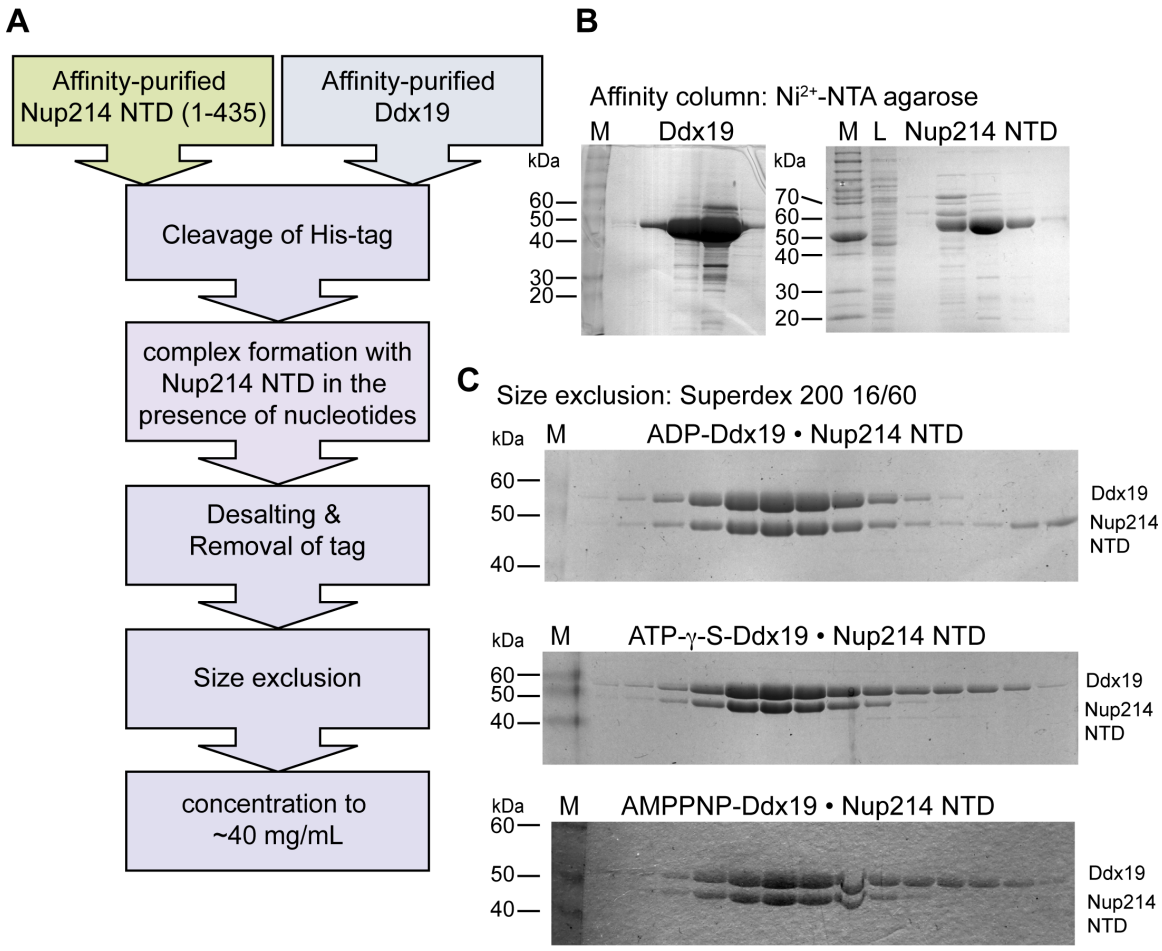


Figure 42: Purification of the Nup214•Ddx19 complex. (**A**) Purification scheme for the Nup214 NTD•Ddx19 and Nup214 NTD•Ddx19 NTD complex. Representative Coomassie brilliant blue stained SDS-PAGE gels for the (**B**) affinity purification step of full length Ddx19 and Nup214 NTD and (**C**) gel filtration purification step for the Nup214 NTD•Ddx19 complex in the presence of ADP (*top*), ATP-γ-S (*middle*) and AMPPNP (*bottom*).

pH 9.5 (**Figure 43**). However, close inspection of the crystals revealed surface imperfections and minor growth defects, which suggested an underlying crystal-packing problem.

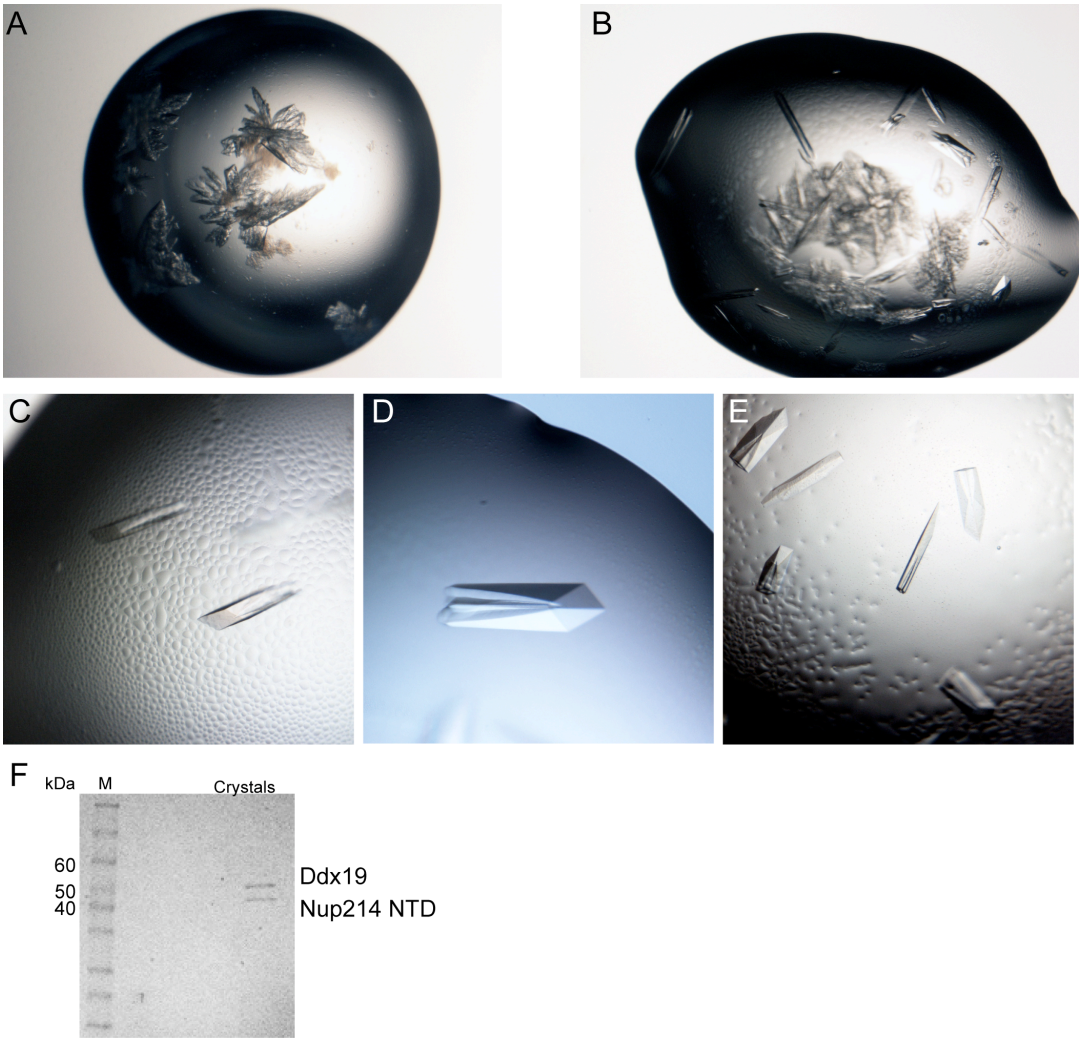


Figure 43: Crystals of the Nup214 NTD•Ddx19-ADP complex. Initial screening hits in (**A**) 10% (w/v) PEG 8000, 0.1M HEPES pH 7.5, 8% ethylene glycol and (**B**) 30% (w/v) PEG 3000, 0.1M CHES pH 9.5. (**C**) to (**E**) Crystal obtained after optimization in 30% (w/v) PEG 3000, 0.1M CHES pH 9.5. (**F**) Crystals obtained in the 30% (w/v) PEG 3000, 0.1M CHES pH 9.5 condition were washed and analyzed with a SDS-PAGE and detected by Coomassie brilliant blue staining.

The crystals were cryo-protected in 10% glycerol, 30% (w/v) PEG 3000, 0.1 M CHES pH 9.5 and flash-frozen in liquid nitrogen-cooled liquid propane and the data collected at the Advanced Light Source at the Lawrence Berkeley National Laboratories. Although the diffraction of the crystals was anisotropic, we were able to collect a full dataset to a 3.2 Å resolution. The crystals belong to the monoclinic space-group $P2_1$ with two complexes in the asymmetric unit and the structure was solved by molecular replacement using the coordinates of Nup214 NTD as a search model in Phaser and docked the N-terminal domain of MjDEAD into the density. The structure resolved the Nup214 NTD bound to the N-terminal domain 1 of Ddx19 bound to ADP. Surprisingly no electron density was observed for the smaller C-terminal domain 2 of Ddx19.

To improve the diffraction data of the crystals, we modified our crystallization construct of Ddx19 to contain only the NTE and domain 1 (residues 1 to 300) of Ddx19 (Ddx19 NTD). The complex containing Nup214 NTD•Ddx19 NTD-ADP was purified similar to the Nup214 NTD•Ddx19 complex and screened for crystallization conditions (**Figure 44**). Crystals of the complex grew readily without visible imperfections (**Figure 44 D**, **E**). The Nup214 NTD•Ddx19 NTD crystals belong to the orthorhombic space group $P2_12_12_1$ with one complex in the asymmetric unit and diffracted to 2.5 Å resolution with substantially improved diffraction data quality (**Figure 45**). The final model contains residues 8 to 428, and 69 to 300 of Nup214 and Ddx19, respectively and was refined to an R_{work} factor of 19.9% and an R_{free} factor of 24.2% (**Table 7**, Material and Methods).

Figure 44: The Nup214 NTD•Ddx19 NTD complex. (**A**) Domain structure of Nup214 and Ddx19 indicating the crystallization fragments of Nup214 NTD and Ddx19 NTD (red brackets). (**B**) Representative Coomassie brilliant blue stained SDS-PAGE gel for the affinity purification step of Ddx19 NTD. (**C**) Typical gel filtration profile of the Nup214 NTD•Ddx19 NTD complex on a 16/60 Superdex 200 column. (**D**) Crystallization screening hits in a 1.0 M sodium citrate, 0.1 M imidazole, pH 8.0 condition (top), and 0.8 M NaH_2PO_4, 1.2 M K_2HPO_4, 0.1 M acetate, pH 4.5 condition (bottom). (**E**) Optimized crystals of the Nup214 NTD•Ddx19 ADP complex grown in 0.8 M NaH_2PO_4, 1.2 M K_2HPO_4, 0.1 M sodium acetate, pH 4.5.

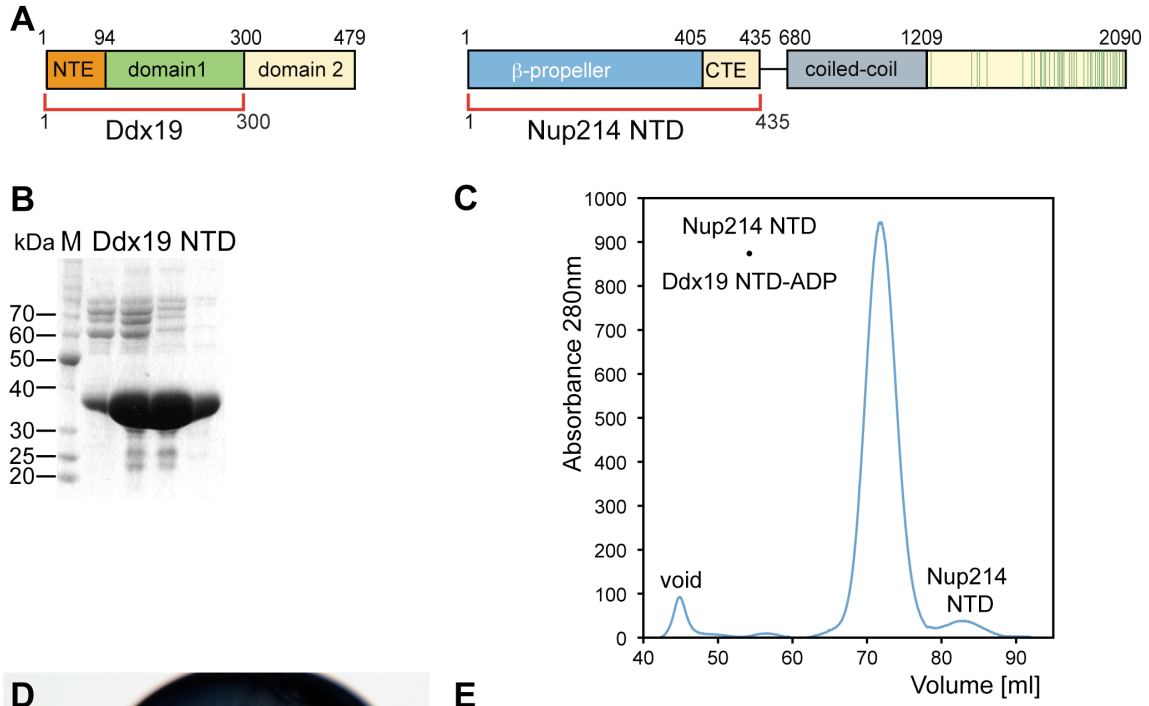
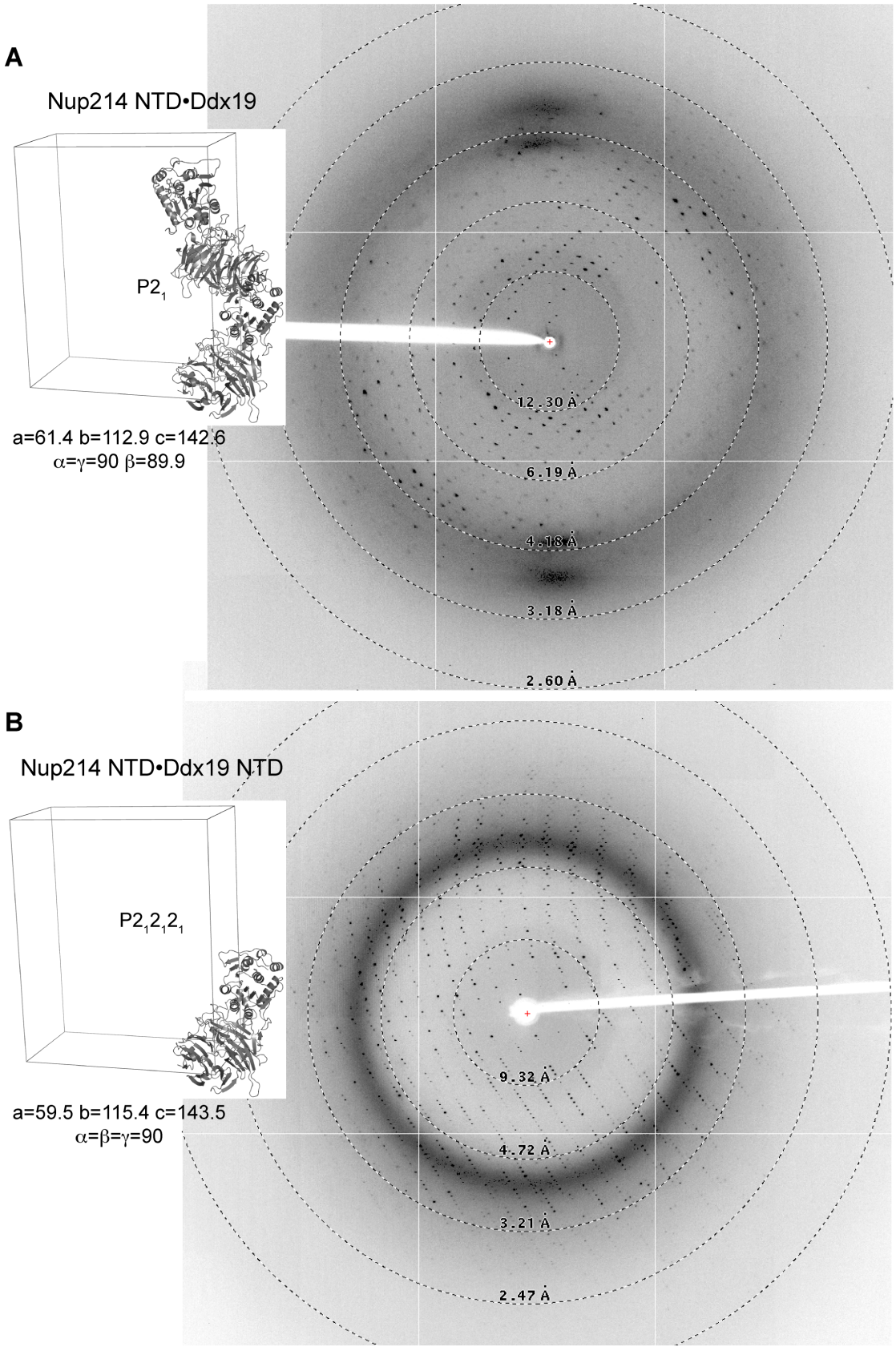

A
1
94
300
479
NTE
domain1
domain 2
1
Ddx19
300
1
405
435
680
1209
2090
β-propeller
CTE
coiled-coil
1
Nup214 NTD
435
B
kDa M Ddx19 NTD
70
60
50
40
30
25
20
C
Nup214 NTD
Ddx19 NTD-ADP
Absorbance 280nm
1000
900
800
700
600
500
400
300
200
100
0
void
Nup214
NTD
40
50
60
70
80
90
Volume [ml]
D
E

Figure 45: Unit cell and diffraction images of crystals of (**A**) Nup214 NTD•Ddx19 and (**B**) Nup214 NTD•Ddx19 NTD. The unit cell and unit cell parameters as well as an asymmetric unit are shown on the left. Note the anisotropic diffraction of the Nup214 NTD•Ddx19 crystals (**A**) and the improved diffraction quality of the Nup214 NTD•Ddx19 NTD crystals (**B**). Diffraction images with resolution circles are shown on the right and the unit cell is indicated on the left.


A
Nup214 NTD•Ddx19
P2$_1$
a=61.4 b=112.9 c=142.6
α=γ=90 β=89.9
12.30 Å
6.19 Å
4.18 Å
3.18 Å
2.60 Å
B
Nup214 NTD•Ddx19 NTD
P2$_1$2$_1$2$_1$
a=59.5 b=115.4 c=143.5
α=β=γ=90
9.32 Å
4.72 Å
3.21 Å
2.47 Å

2.3.2 Crystal structure of the Nup214 NTD•Ddx19 NTD complex

The structure of the Nup214 NTD•Ddx19 NTD complex reveals that the Nup214 NTD, similarly to what was observed for Nup214 NTD in isolation, forms a seven-bladed β-propeller domain that is extensively decorated by long inter-blade connector loops, and contains a C-terminal extension (CTE) bound to the bottom face of the β-propeller[46] (**Figure 46**). For Ddx19 the structure resolves domain 1 and 27 residues of the NTE. The polypeptide chain of domain 1 of Ddx19 folds into the canonical RecA-like domain found in other DEAD-box helicases, such as MjDEAD and eIF4A[155, 219]. This globular domain is formed by an extensive, centrally located β-sheet that is composed of seven parallel β-strands (β2-β8) sandwiched between four α-helices on each side (αA-αD on one side, and αE-αH on the other) (**Figure 46**). In our structure, the unique N-terminal extension of Ddx19 (NTE) is primarily unstructured, but the 26 residues that are resolved (residues 69 to 94) form an extensive loop, and supplement the β-sheet core of domain 1 with an additional anti-parallel β-strand, β1. Moreover, the N-terminal part of the NTE forms a lid over the nucleotide-binding pocket, encapsulating the bound ADP. Strikingly, the binding of Ddx19 to the extensively decorated Nup214 NTD results in only minor conformational changes in the Nup214 NTD, as indicated by an root-mean square deviation (RMSD) of ~0.6 Å over all atoms between the isolated and Ddx19-bound Nup214 NTD structures.

Overall, the structure is in excellent agreement with our biochemical characterization of the interaction between Nup214 and Ddx19. The Nup214 NTD

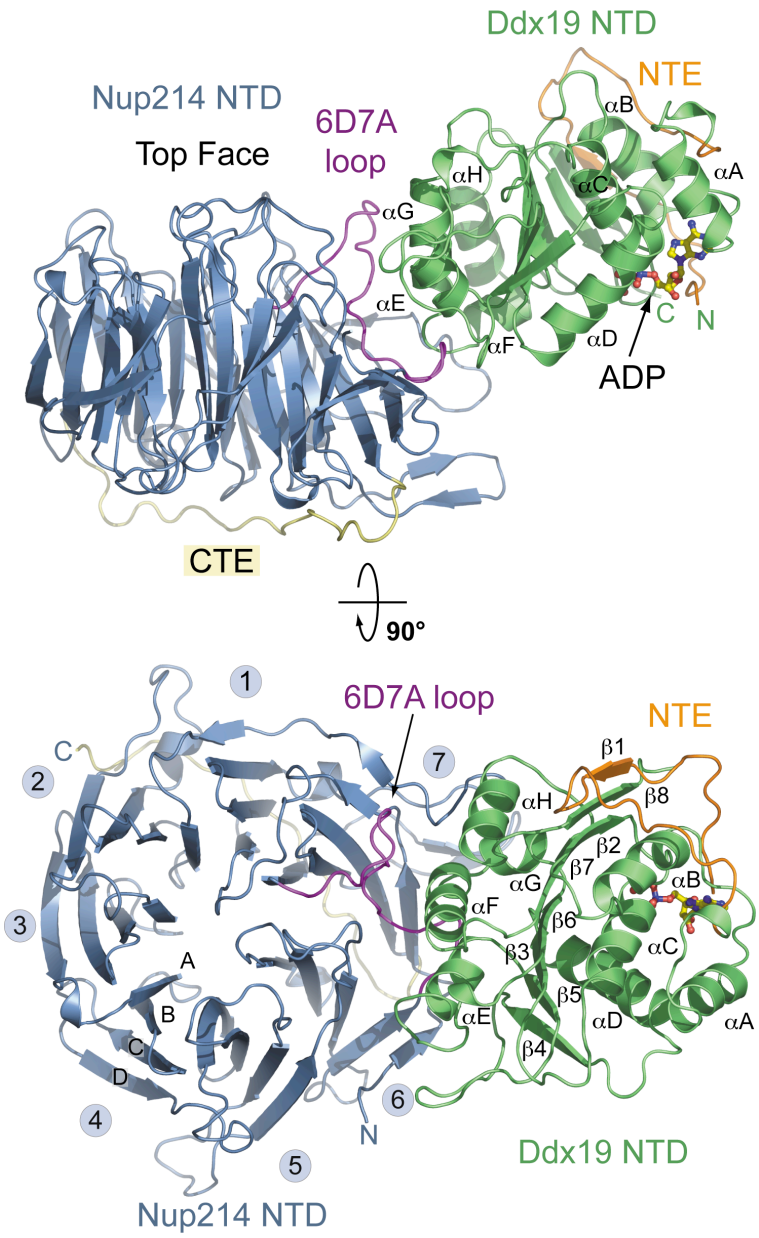


Figure 46: Overview of the Nup214 NTD·Ddx19 NTD structure. Ribbon representation of the Nup214 NTD•Ddx19 NTD complex from the side (*top*) and top view (*bottom*), rotated by 90°. For the Nup214 NTD, the β-propeller domain (blue), the 6D7A loop (violet), the C-terminal extension (CTE; yellow), and the blade numbers are indicated. For the Ddx19 NTD, the domain 1 (green) and the unique N-terminal extension (NTE; orange) are indicated. The ADP molecule bound to the Ddx19 NTD is shown in ball-and-stick representation

binds with its side surface to the N-terminal domain 1 of Ddx19. The interaction between the two proteins is primarily mediated by the interblade connector loop 6D7A with minor contributions from residues of the β-strand 6E, which forms the double Velcro closure, and the 1CD, 5D6A, and 6BC loops of the Nup214 β-propeller domain. In Ddx19, the 4 helices αE–αH and their interhelical loops contribute to the interface.

Domain 1 of Ddx19 provides a deep groove for the ADP cofactor that is partially covered by the unique NTE. The adenine ring of the ADP molecule is specifically recognized by several hydrogen-bond interactions with Q119 and the backbone carbonyl of R114 (**Figure 47**). In addition, the adenine ring is tucked into a hydrophobic slot between F112 of the αA-αB loop and L70 of the NTE. Altogether, the binding of the ADP cofactor is similar to that observed in the structure of UAP56[220]. In contrast to UAP56, no Mg^{2+} ion is found to coordinate the pyrophosphate moiety of ADP.

Strikingly, the association between the two proteins is predominantly mediated by surfaces with strongly opposing electrostatic potentials that are evolutionarily highly conserved (**Figure 48**, **Figure 49**). While the binding surface of Nup214 features a highly negative surface potential, the binding surface of Ddx19 is highly positively charged. In fact, the highest positive surface potential of domain 1 of Ddx19 is buried at the interface with Nup214 (**Figure 49**). Importantly, 22 out of 51 residues are charged and form numerous salt bridges at

the interface, dominating the association of the two proteins via electrostatic interactions.

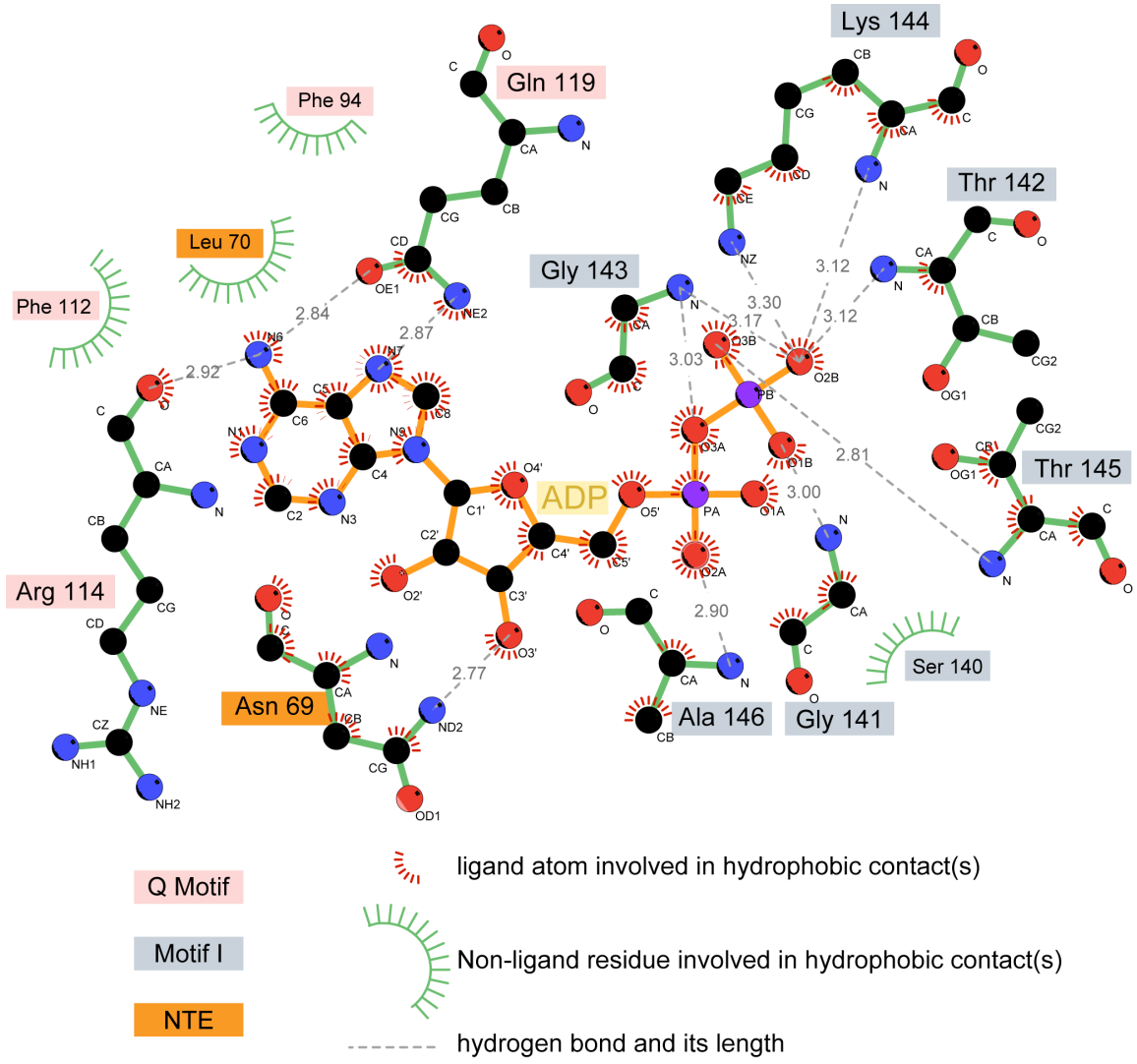


Figure 47: **Schematic diagram of Ddx19-ADP interactions showing details of hydrogen bonding pattern and non-bonded hydrophobic contacts between Ddx19 and ADP.** The bonds between Ddx19 atoms are shown in green and ADP in orange. Ddx19 residues of the Q motif and Motif I participating in ADP binding are highlighted in magenta and violet, respectively, while residues of the NTE are highlighted in orange.

Figure 48: Multi-species sequence alignment of Ddx19 homologues. The overall sequence conservation at each position is shaded in a color gradient from yellow (60% similarity) to dark red (100% identity) using the Blosum62 weighting algorithm. The secondary structure of Ddx19 as observed in the Nup214 NTD•Ddx19 structure is shown above the alignment. The residue numbering is relative to *H. sapiens* Ddx19 and grey boxes below the alignment indicate the position of the nine conserved DEAD-box helicase motifs. As a reference, the primary sequence of the *M. janaschii* DEAD-box protein MjDEAD is aligned with the Ddx19 homologues.

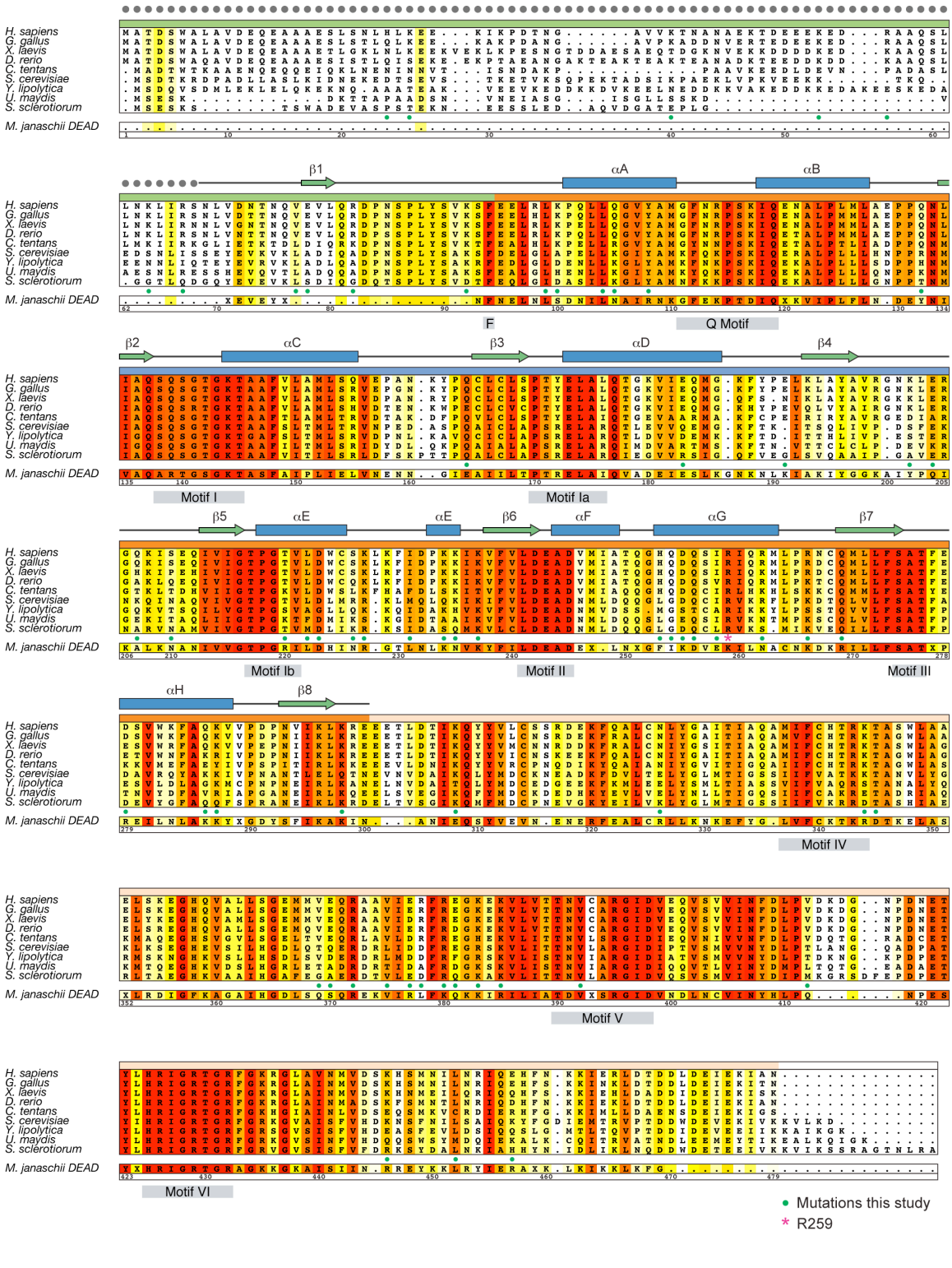
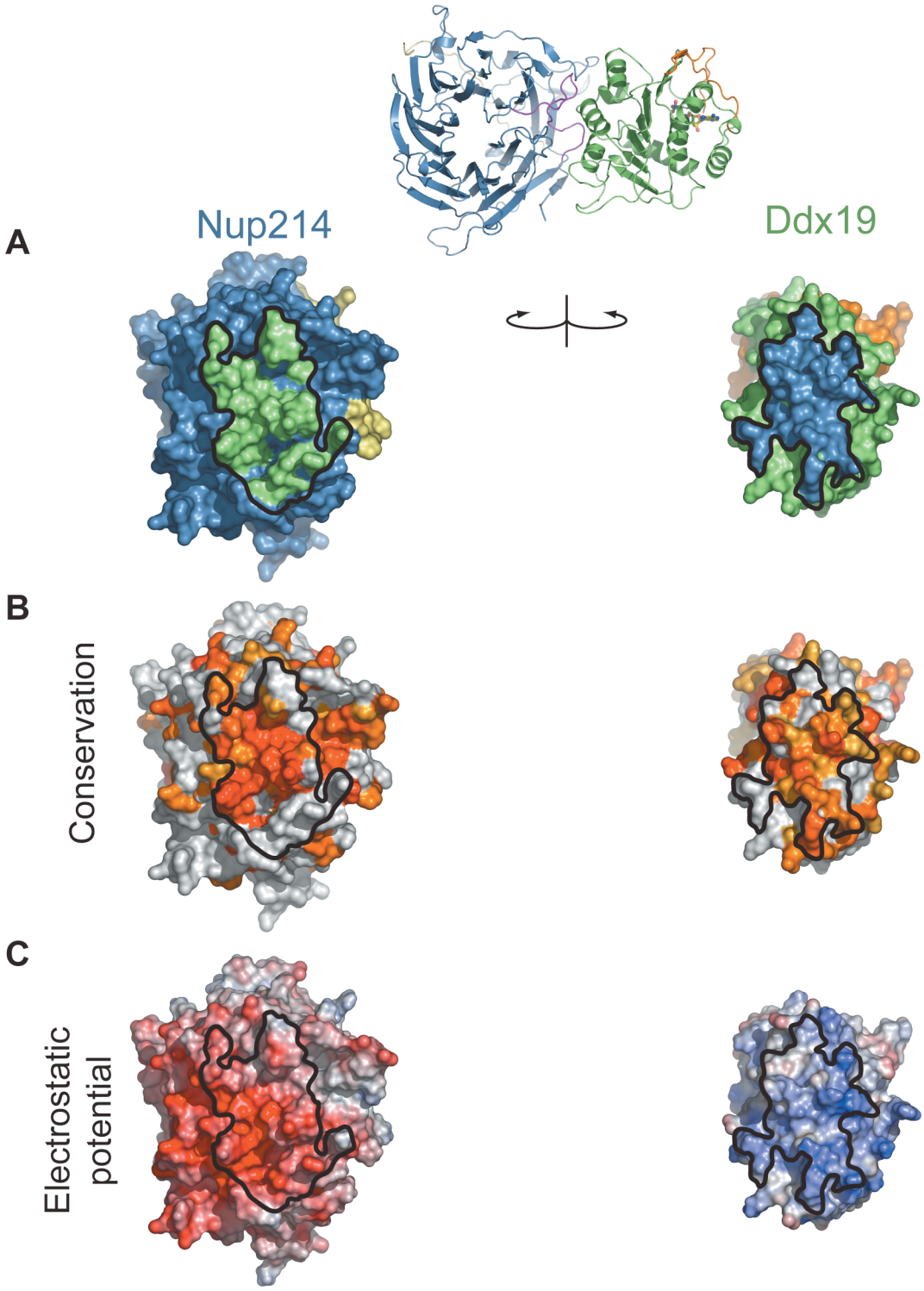

H. sapiens
G. gallus
X. laevis
D. rerio
C. tentans
S. cerevisiae
Y. lipolytica
U. maydis
S. sclerotiorum
M. janaschii DEAD
β1
αA
αB
F
Q Motif
β2
αC
β3
αD
β4
Motif I
Motif Ia
β5
αE
αE
β6
αF
αG
β7
Motif Ib
Motif II
Motif III
αH
β8
Motif IV
Motif V
Motif VI
Mutations this study
R259

Figure 49: Surface properties of the Nup214 NTD-Ddx19 NTD interaction. (**A**) Surface renditions of the NTDs of Nup214 and Ddx19 in an open book representation colored according to the participation of the various domains. Surfaces that mediate the association between the two proteins are indicated in green (Ddx19) and blue (Nup214). As a reference, a ribbon representation of the complex is shown in its original orientation. (**B**) Surface representation colored according to a multispecies sequence alignment (Figure 48). The conservation at each position is mapped onto the surface and is shaded in a color gradient from light yellow (60% similarity) to dark red (100% identity). (**C**) Surface representation colored according to the electrostatic potential. The electrostatic potential is plotted onto the surface and colored in a gradient from red (–10 k_BT/e_c) to blue (+10 k_BT/e_c). Black lines indicate the interface borders.


Nup214
Ddx19
A
B
Conservation
C
Electrostatic potential

These electrostatic interactions appear to be further reinforced by additional hydrophobic van der Waals contacts. In total, a surface area of ~1900 $Å^2$ is buried at the interface between the two proteins. Interestingly, such intermolecular associations governed primarily by electrostatic interactions were also observed in other nucleoporin structures, namely the Nup58/45 tetramer and the Seh1-Nup85 hetero-octamers[21, 221].

2.3.3 The interaction of Nup214 and Ddx19 is electrostatic in nature.

Given the strong electrostatic potential of the Nup214 and Ddx19 surfaces that mediates the interaction between the two proteins, we examined the influence of high-salt buffer conditions on the stability of the Nup214 NTD•Ddx19 complex. We found that the apparent molecular weight of the complex decreased with increasing salt concentrations (**Figure 50**). While the complex eluted with an apparent molecular mass of ~235 kDa in a buffer containing 50 mM KCl (**Figure 50** *top*), the apparent molecular mass decreased to ~105 kDa when examined in a buffer containing 1 M KCl (**Figure 50** *bottom*). Hence the association between the Nup214 NTD and Ddx19 is governed by electrostatic interactions that can be substantially weakened in high-salt conditions.

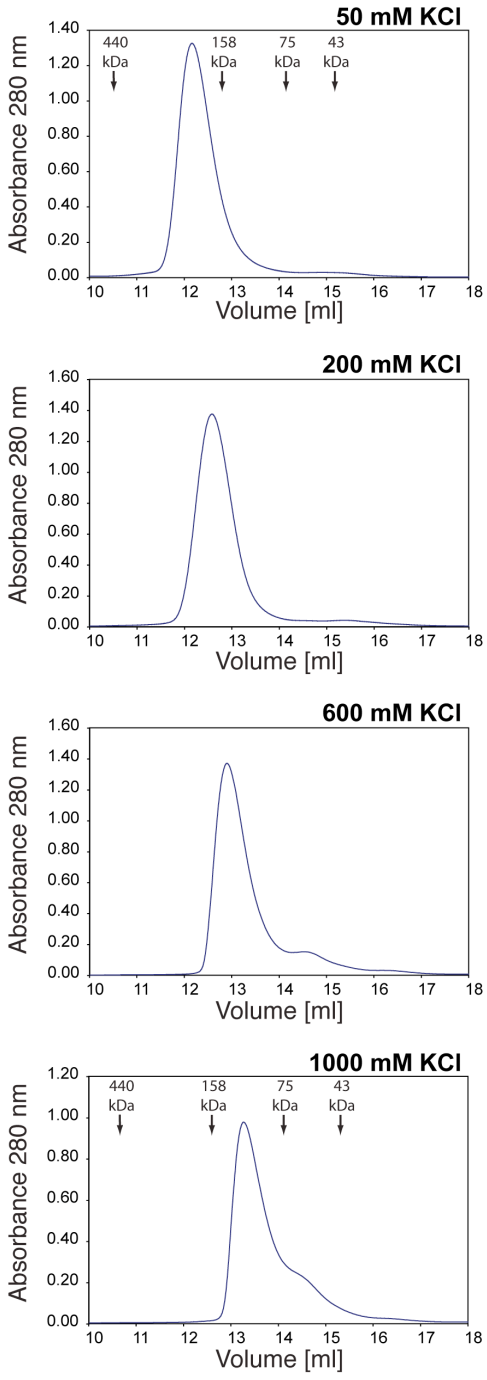


Figure 50: Salt dependence of the Nup214-Ddx19 interaction. Gel filtration profiles of the Nup214 NTD•Ddx19 complex at identical protein complex concentrations of 50 μM and the indicated salt concentrations. The elution positions for molecular weight standards are indicated.

2.3.4 R259 of Ddx19 is a key residue for complex formation

In our biochemical analysis of the Nup214 NTD•Ddx19 complex we found that a single charged and conserved residue in domain 1 of Ddx19, R259, is critical for the association with Nup214, which supports the conclusion that the association is primarily governed by electrostatic interactions. The crystal structure of the complex now reveals that R259 is in fact located in the center of the interface and forms a critical salt bridge with D359, a water-mediated salt bridge with E350, and an additional hydrogen bond with the backbone carbonyl of L351 (**Figure 51**). All of the residues that R259 of Ddx19 interacts with are located within the 6D7A loop of the Nup214 β-propeller domain. However, in addition to the salt bridges formed by R259, several additional salt bridges are formed in the interface, providing the molecular basis for our biochemical data that the two proteins can be dissociated in high-salt conditions (**Figure 50**).

Ddx19 is dispersed throughout the cytoplasm and enriched at the nuclear envelope, whereas Nup214 is exclusively localized at the cytoplasmic face of the NPC[146, 150]. In yeast, the deletion of the β-propeller domain of yNup159 prevents recruitment of yDbp5 to the nuclear envelope[35]. To test whether mutations of R259 not only disrupt the interaction between Nup214 and Ddx19 *in vitro*, but also inhibit the localization of Ddx19 at the nuclear envelope, we transiently transfected HeLa cells with wild-type Ddx19, and with the R259A and R259K mutants. While the wild-type Ddx19 and the Ddx19 R259K mutant were recruited to the NPC as predicted by our biochemical experiments, the R259A mutant

failed to colocalize at the nuclear envelope (**Figure 52**). Our data therefore demonstrate that R259 of Ddx19 is critical for the interaction between the two proteins not only *in vitro*, but also *in vivo*.

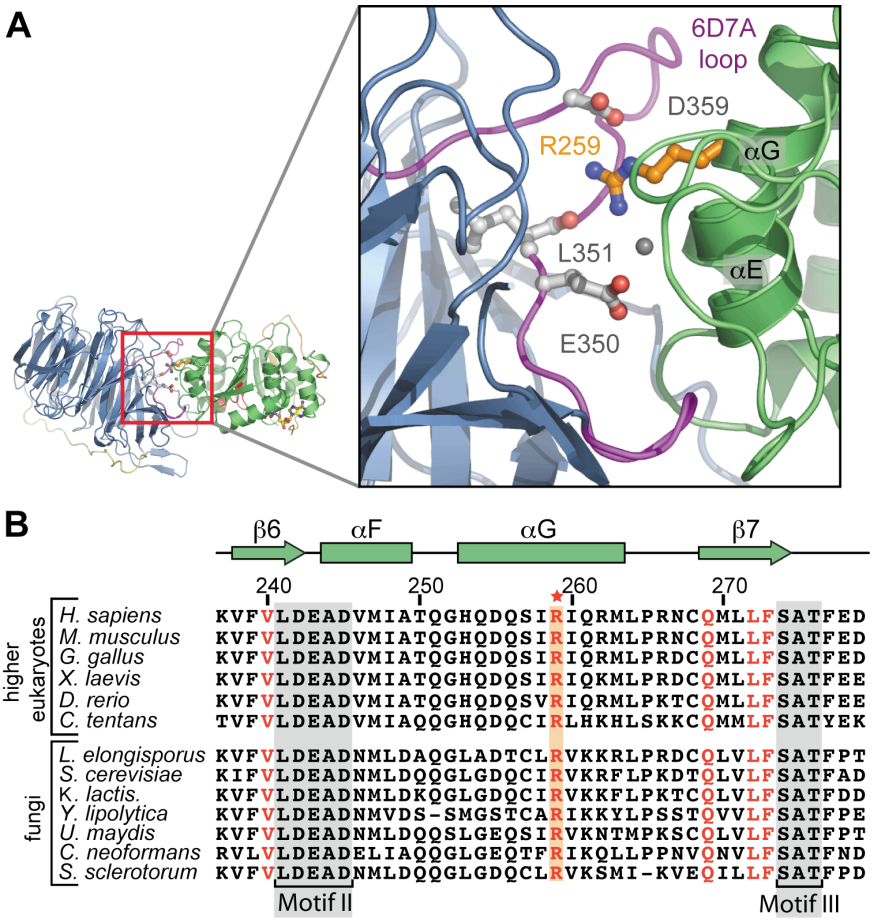


Figure 51: **The conserved arginine 259 of Ddx19 is a key residue for complex formation.** (**A**) Details of the interaction between the NTDs of Nup214 and Ddx19. The ribbon representation is colored according to Figure 46. (**B**) Multispecies sequence alignment of Ddx19 homologues. The red asterisk indicates the location of the invariant R259. The conserved sequence motifs II and III are highlighted in gray boxes and invariant residues outside of the conserved sequence motifs illustrated in red. The residue numbering is relative to human Ddx19 and the secondary structure of Ddx19 is shown above the sequence alignment.

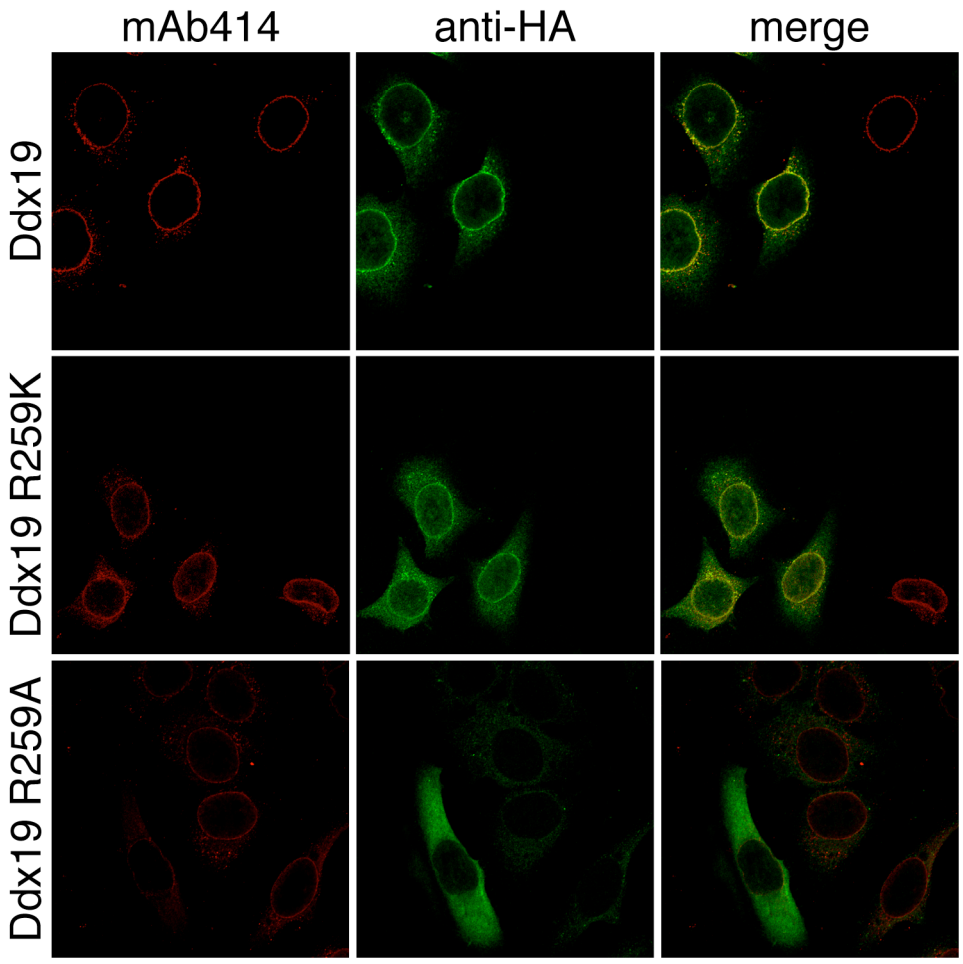


Figure 52: ***In vivo*** **localization of Ddx19 and Ddx19 mutants.** HeLa cells were transfected with Ddx19 and Ddx19 mutants containing a C-terminal HA-tag and analyzed with confocal microscopy. The monoclonal antibody mAb414 was used as a reference for nuclear envelope staining (red, left panel). The cellular localization of HA-tagged proteins was detected with an anti-HA antibody (green, middle panel). The merged image (right panel) reveals the colocalization of wild type Ddx19 and Ddx19 R259K with the nuclear envelope, while Ddx19 R259A displays no detectable nuclear envelope staining.

2.3.5 Nucleotide dependent interaction of Ddx19 with RNA

A comparison of different DEAD-box helicases has revealed that nucleotide binding induces conformational changes that affect the association of the two RecA-like domains and their spatial arrangement with respect to each other as well as RNA substrate binding[155, 219, 220, 222]. In order to assess the nucleotide requirements of RNA binding for Ddx19, we tested the binding of Ddx19 to a degenerate single-stranded RNA oligonucleotide in the presence of ADP or AMPPNP, a non-hydrolysable ATP-analog (**Figure 53**). Binding of the RNA oligonucleotide to Ddx19 was confirmed by SDS-PAGE and UREA-PAA gel electrophoresis. While Ddx19 was unable to bind to RNA in the presence of ADP (**Figure 53 A**, **C**), Ddx19 formed a complex with RNA in the presence of AMPPNP (**Figure 53 B, D**).

We next investigated whether the Nup214 NTD•Ddx19 complex was able to interact with RNA (**Figure 54**) and analyzed binding of RNA to the complex by SEC. As was the case with Ddx19, we did not observe binding of the RNA to the Nup214 NTD•Ddx19 complex in the presence of ADP (**Figure 54 A**). In addition, we were unable to observe a shift in the elution volume that would indicate binding of RNA to the Nup214 NTD•Ddx19 complex in the presence of AMPPNP. Instead, we detected peaks that eluted at a similar elution volume to a Ddx19•RNA complex (blue arrows) and Nup214 NTD (magenta arrow) (**Figure 54 B**). Analysis of the SDS-PAGE and UREA-PAA gels confirmed that a portion of the Nup214 NTD•Ddx19 complex was disassembled upon incubation with

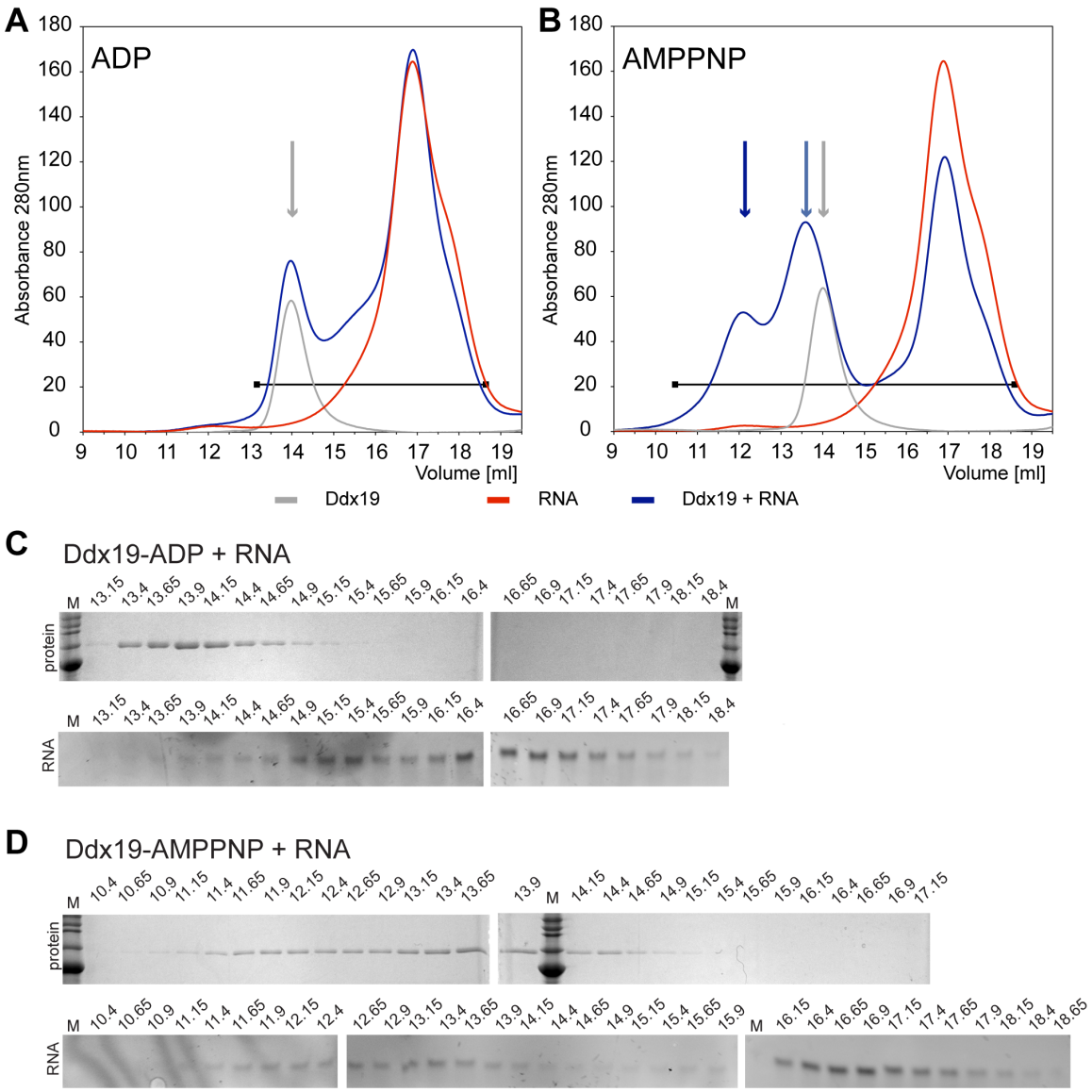


Figure 53: RNA-binding activity of Ddx19 requires ATP. Gel filtration profiles of RNA incubated with (**A**) Ddx19-ADP and (**B**) Ddx19-AMPPNP. The elution profiles of RNA and Ddx19 are shown in red and grey, respectively. The elution profiles resulting from incubation of RNA and Ddx19 are shown in blue. The grey arrow indicates the elution volume of Ddx19 and the blue arrows indicate peaks containing Ddx19-AMPPNP and RNA. Coomassie brilliant blue stained SDS-PAGE gels (*top*) and ethidium-bromide stained PAA-UREA gels (*bottom*) of (**C**) Ddx19-ADP and (**D**) Ddx19-AMPPNP incubated with RNA.

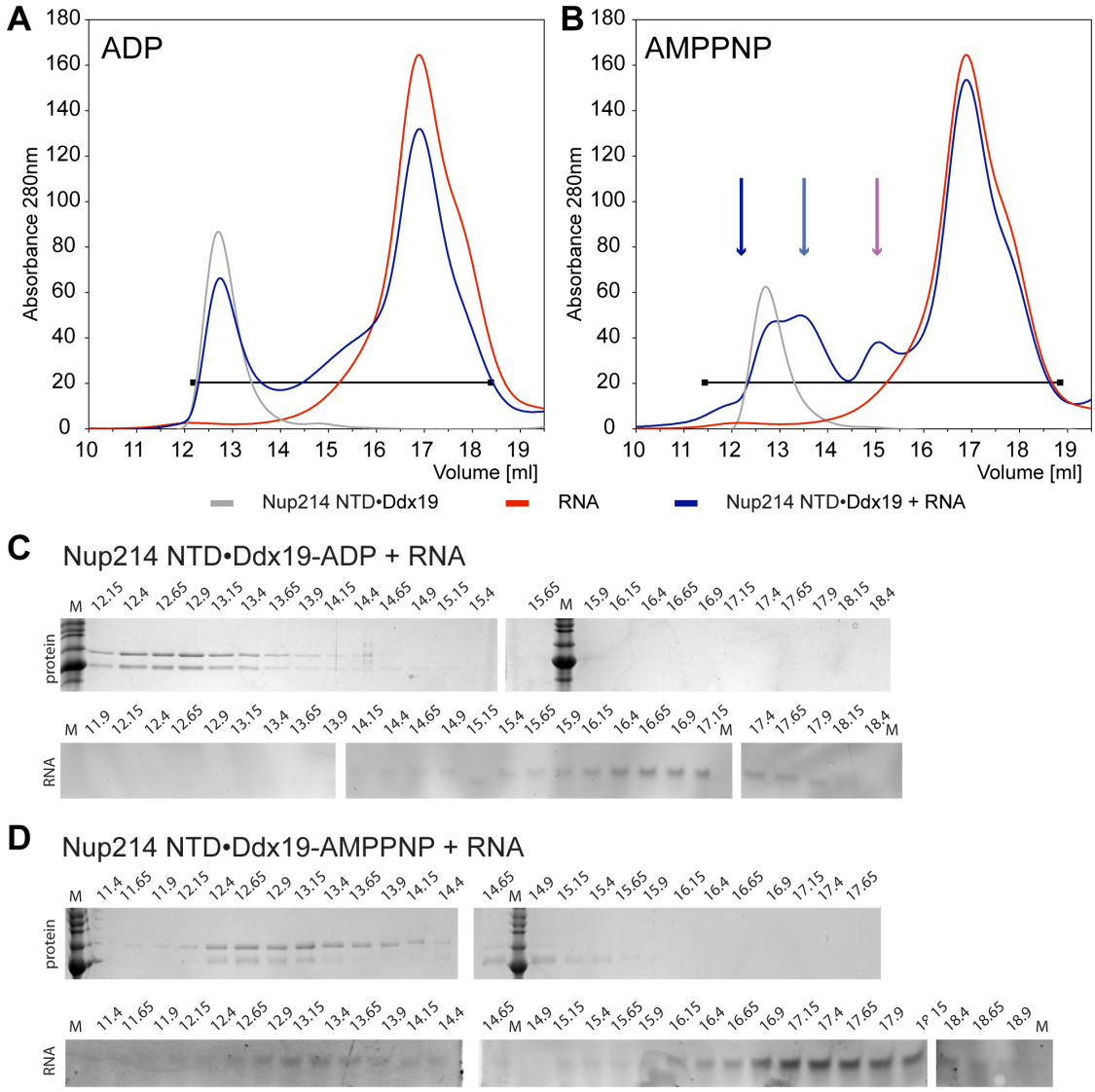


Figure 54: Nup214 NTD and RNA binding of Ddx19 are mutually exclusive. Gel filtration profiles of RNA incubated with the Nup214 NTD•Ddx19 complex in its (**A**) ADP-bound and (**B**) AMPPNP-bound form. The elution profiles of RNA and Nup214 NTD•Ddx19 complex are shown in red and grey, respectively. The elution profiles resulting from incubation of RNA and Nup214 NTD•Ddx19 complex are shown in blue. The magenta arrow indicates the elution volume of Nup214 NTD and the blue arrows indicate peaks containing Ddx19-AMPPNP and RNA as observed in Figure 53. Coomassie brilliant blue stained SDS-PAGE gels (*top*) and ethidium-bromide stained PAA-UREA gels (*bottom*) of (**C**) Nup214 NTD•Ddx19-ADP and (**D**) Nup214 NTD•Ddx19-AMPPNP incubated with RNA.

RNA (**Figure 54 D**). These results indicate that Ddx19 dissociates from Nup214 and form a complex with RNA. The fact that we were unable to detect a ternary complex consisting of Nup214 NTD, Ddx19 and RNA, suggests furthermore that the binding of Nup214 NTD and RNA to Ddx19–AMPPNP is mutually exclusive and that the RNA and Nup214 may use similar binding surfaces on Ddx19.

2.3.6 Interaction with Gle1

The terminal step of mRNA export through the NPC is the removal of transport factors from the mRNA. In yeast, the DEAD-box helicase yDbp5 is crucial for removing the mRNA export receptor as well as other proteins from the mRNP[166, 167]. An NPC associated protein, yGle1, binds directly to inositol-hexakisphosphate (IP_6), and together with RNA, stimulates the ATPase activity of yDbp5[164, 165]. Ddx19 and Gle1 are recruited to the cytoplasmic face of the NPC and are localized in close proximity to each other[47, 150, 160].

In order to test whether the human homologues Gle1 and Ddx19 interacted directly, we aimed to express and purify Gle1. However, all human Gle1 constructs tested were insoluble when expressed by themselves or coexpressed with known interaction partners in *E. coli* (**Table 5,** Material and Methods section). Since the interaction between Gle1 and Ddx19 is conserved in yeast, we therefore tested whether recombinant yGle1 could be obtained. We tested a series of deletion constructs for expression and coexpression with yDbp5 or yNup42 (**Table 5**). Although expression levels were low with less than 0.5 mg/10L of bacterial culture, we were able to purify an N-terminal deletion

fragment of yGle1 (yGle1ΔN100), which included residues 101 to 538, and test for its ability to interact with yDbp5 by SEC (**Figure 55**). yGle1ΔN100 was capable of interacting with yDbp5 (**Figure 55 C**), but not with the human homologue Ddx19 (**Figure 55 B**). In addition, we were able to detect formation of a complex between yGle1ΔN100 and yNup159 NTD•yDbp5, which strongly suggests that yGle1 and yNup159 use distinct binding sites on the yDbp5 surface (**Figure 55 D**).

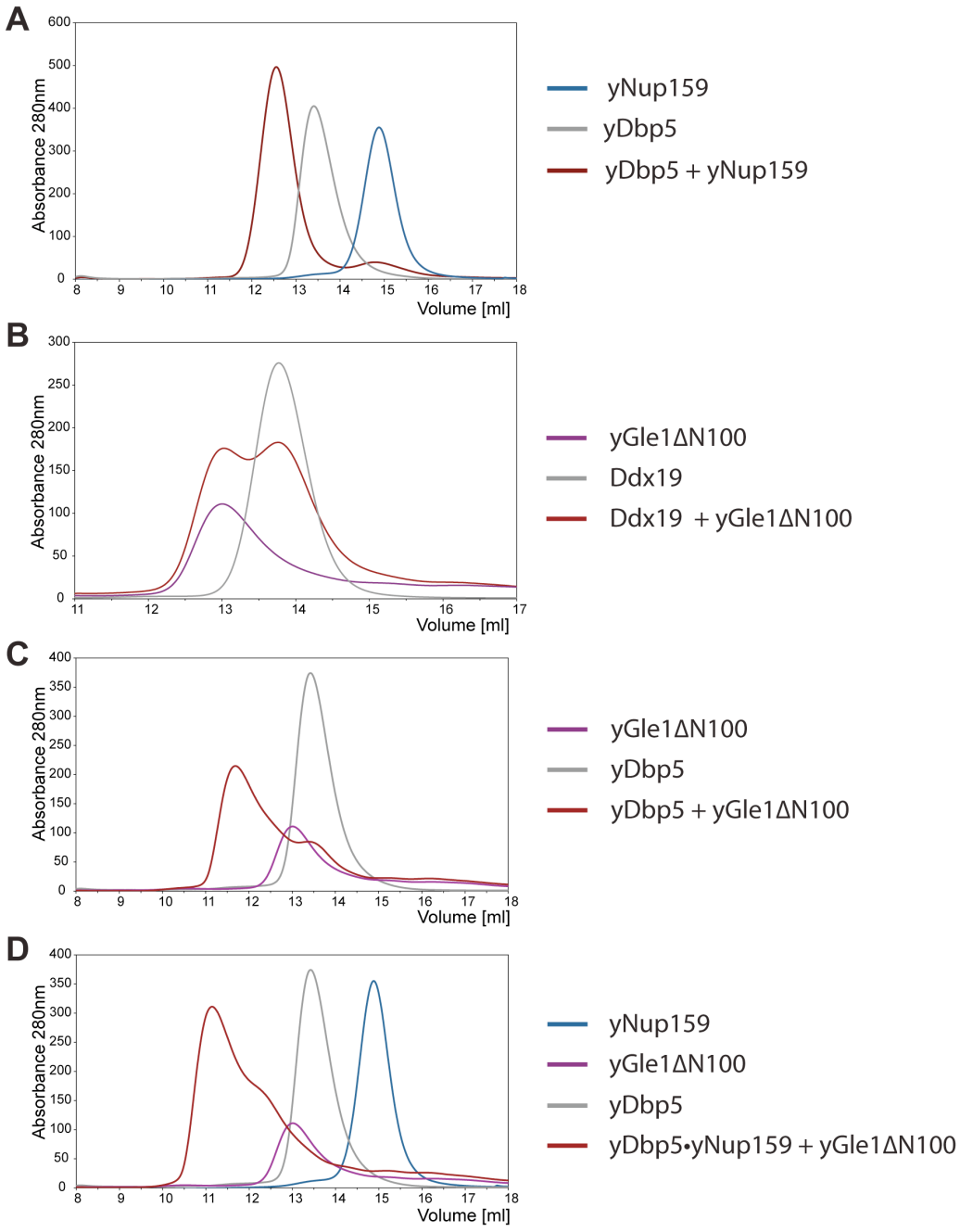


Figure 55: **yGle1 forms a triple complex with yDbp5 and yNup159.** Gel filtration profiles of (**A**) yDbp5 incubated with the yNup159 NTD, and yGle1ΔN100 incubated with (**B**) Ddx19, (**C**) yDbp5 or (**D**) yNup159 NTD•yDbp5. Gel filtration profiles of yDbp5 are colored in grey, Nup159 NTD in blue, yGle1ΔN100 in violet and the elution profile resulting from preincubated proteins in red.

3 CONCLUSION AND FUTURE DIRECTIONS

The NPC is one of the largest known proteinacous assemblies in the cell and we are just beginning to understand the architectural principles that govern the assembly of the NPC from the structural domains of the nups. About two thirds of the 30 nups are predicted to contain β-propeller or α-solenoid domains or a combination of both, and almost all of these nups are predicted to contain unstructured regions directly adjacent to the structured domains[223]. Based on their rigid fold but high concentration of variable interaction surfaces, β-propellers have frequently been found to act as interaction platforms with other proteins in a variety of biological contexts[206]. Therefore, the predicted β-propeller-containing nups are of particular interest in the study of the dynamic assembly of the NPC as well as for the recruitment of NPC associated factors involved in essential processes of the cell.

3.1 The structure of the Nup214 NTD

The structure of the Nup214 NTD presented in this thesis suggests that, while the nup β-propeller domains share a conserved structural design, each one is uniquely decorated with elements that dictate its surface properties and may be important for protein interactions. Furthermore, the finding that the nup β-propeller repeats lack conserved sequence motifs makes it difficult during structure predictions to discriminate between the core elements and the unique

features. This level of information can only be provided by high-resolution structural analysis. The Nup214 NTD structure revealed features of the Nup214 NTD that lead us to two key observations. First, based on the observation that two highly conserved surface patches and a potential phosphorylation site are located at the bottom face of the Nup214 propeller, bound also by an extended peptide (the CTE), we propose a model for the involvement of flexible peptides in the assembly of the NPC. Second, the Nup214 NTD structure provided the basis for the biochemical and structural analysis of the Nup214 NTD's interaction with an essential mRNA export factor Ddx19, revealing key insights into mRNA transport mediated by the NPC.

3.1.1 Implications of the Nup214 CTE for the dynamic reassembly of the NPC

In the cell, many examples have been identified in which the interaction of two proteins is facilitated by an unstructured peptide segment[87]. Intrigued by the location of conserved regions 1 and 2 and the CTE at the bottom face of the β-propeller domain of Nup214, we derived a possible function for the CTE by comparing Nup214 to karyopherins (Kaps), the mobile transport factors that facilitate nucleo-cytoplasmic transport. The binding of the NLS-containing cargo-protein to Kap-α is regulated by the hetero-dimerization of Kap-α and Kap-β. In the absence of Kap-β, the N-terminal peptide segment of Kap-α (auto-NLS) binds *intra*-molecularly to its NLS-binding site, thereby preventing the NLS of the cargo

protein from binding. In the presence of Kap-β, the hetero-dimerization between Kap-α and Kap-β, facilitated by the auto-NLS of kap-α, frees the binding site for the NLS of the cargo protein. After transport, the binding of Ran-GTP to Kap-β triggers the dissociation of the hetero-trimeric complex by displacing the auto-NLS of Kap-α from Kap-β (**Figure 56 A**)[87].

Based on this system of regulation, we propose a model for the involvement of flexible peptide segments, such as the Nup214 CTE, in the assembly of the NPC (**Figure 56 B**). As the CTE binds with low affinity to the essentially invariant bottom face of the β-propeller, we speculate that the Nup214 NTD exists in two states: a closed state that we observe in our crystal structure with the CTE bound to the β-propeller, and a presumed open state, in which the CTE is released from the β-propeller. The binding of another protein to the CTE would release the CTE from the β-propeller and expose the highly conserved surface of the bottom face. In this open state, the conserved region 1 and 2 would then be available for the binding of yet another protein. In contrast to the dissociation of the nuclear import complex through Ran-GTP binding, the complex could be dissociated by post-translational modifications. Support for this idea comes from the observation that the CTE of Nup214 has been found to be phosphorylated at two sites (Ser430 and Thr437) *in vivo*[215]. Furthermore, the bottom face of the β-propeller contains a putative phosphorylation site within the 4AB loop, suggesting that the hetero-trimeric complex could be dissociated by Ser/Thr phosphorylation.

We suggest that unique structural features, such as the CTE, exist in a number of nups. Thus, it appears plausible that the fully assembled NPC is strapped together via numerous phosphorylation-dependent auto-NLS-like interactions. Such a model could help explain the close correlation of the reversible disassembly of the NPC to the cell cycle, which itself is tightly controlled by phosphorylation through an interplay of several Ser/Thr kinases.

Alternatively, the α-helical domain and FG-repeat region of Nup214, which were not included in the crystallization construct, could stabilize the interaction of the β-propeller domain with the C-terminal extended peptide in the cell. However, further structural studies of nups and nup complexes as well as cell biological studies of cell-cycle-dependent post-translational modifications of Nup214 and other nucleoporins are required to validate this model.

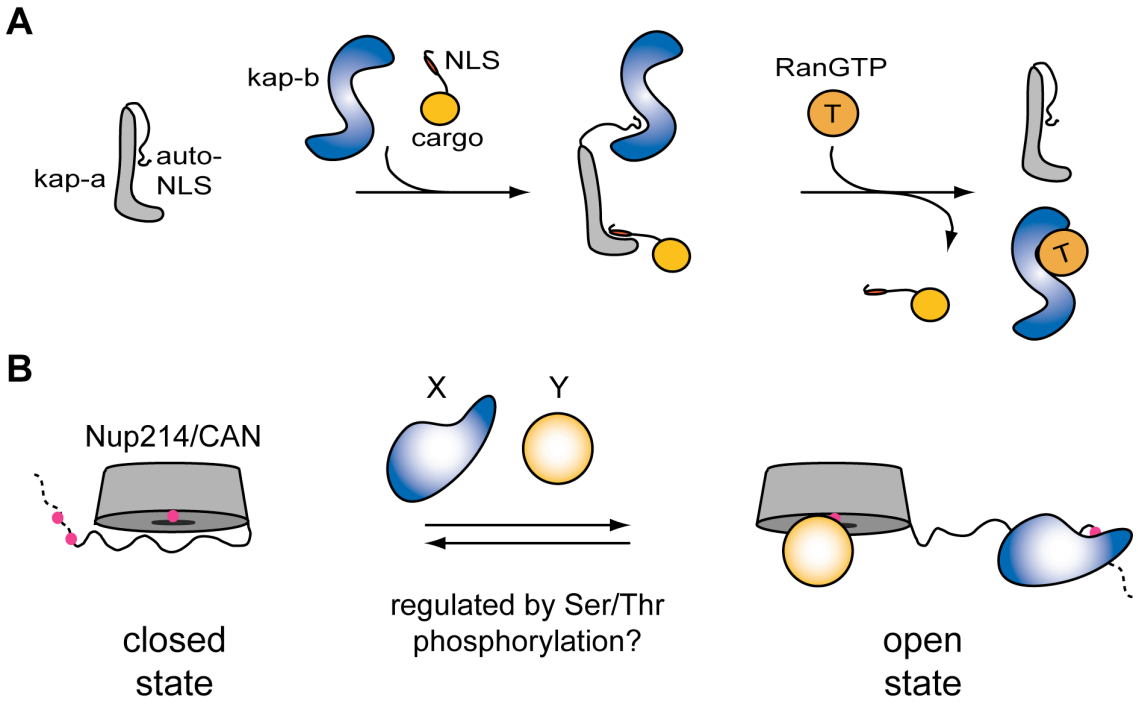


Figure 56: (**A**) Schematic representation of the regulation of the mobile phase of nuclear-cytoplasmic transport. (**B**) Hypothetical model illustrating the role of auto-NLS-like peptide segments, such as the CTE of Nup214, in nup interactions.

3.2 mRNA export through the NPC

mRNAs are synthesized in the nucleus and have to be transported through the NPC to cross the nuclear envelope and be translated in the cytoplasm. The central transport channel of the NPC is lined with the FG-domains of the channel nups[53]. These FG-domains act as a transport barrier. Molecules smaller than ~40 kDa can passively diffuse through the transport barrier of the NPC[24]. However, larger molecules require transport factors that can bind to the FG-domains to cross the NPC[55]. Most transport factors also interact with a small GTPase called Ran[72]. Ran promotes the assembly and disassembly of transport complexes in a location specific manner and drives the import and export of macromolecules.

In the case of mRNA macromolecules, the nascent mRNA associates with a number of proteins during transcription and processing, thereby creating an mRNP[111]. One of the proteins that associate with the mRNA is the TAP/p15 complex. TAP/p15 is the dedicated general mRNA export receptor and is able to bind to the FG-domains of the NPC transport barrier[58]. Therefore, TAP/p15 is able to ferry mRNPs across the NPC into the cytoplasm. However, TAP/p15 is unable to interact with Ran and the questions of how export complexes are disassembled and the transport made irreversible are still under debate. The current model states that mRNP remodeling events at the cytoplasmic side of the NPC displace TAP/p15 and possibly other mRNP export factors from the mRNP[138]. The displacement of TAP/p15 would result in an mRNP that is unable

to interact with the FG-domains of the central transport channel and would therefore implement mRNP export by a molecular ratchet.

Nup214 is an asymmetrically localized nup that can only be found at the cytoplasmic side of the NPC[146]. A conserved interaction was found to recruit the DEAD-box helicase Ddx19 to the NPC[150]. Mutations that abolish recruitment of yDbp5, the yeast homologue of Ddx19, cause mRNA export defects[35]. In addition, yDbp5 was found to be essential for the displacement of TAP/p15 in vivo[166]. Therefore, both Nup214 and Ddx19 are involved in a key step of mRNP export from the nucleus.

3.2.1 The structure of the Nup214 NTD•Ddx19 complex

Our Nup214•Ddx19 structure reveals that the interaction between the two proteins is mediated to a large extent via electrostatic interactions. In fact, we identified the highly conserved R259 in the N-terminal domain of Ddx19 as a key residue for the interaction *in vitro* and for nuclear envelope localization *in vivo*.

Furthermore, we showed that the ATP-bound form of Ddx19 is capable of binding to RNA, and that the interaction with RNA and Nup214 is likely to be mutually exclusive. Indeed, structural investigations of Ddx19 have now revealed the molecular basis of RNA-binding by Ddx19[224, 225]. The structure shows that AMPPNP-Ddx19 is bound in a closed conformation to RNA (**Figure 57 A**). Similar to other DEAD-box proteins, both RecA-like domains of Ddx19 are involved in the binding of RNA[222, 226]. In the Ddx19 structure, a six nucleotide-long single-stranded RNA molecule is bound in a sequence-independent manner with

its 3' end at domain 1 and its 5' end at domain 2 of Ddx19. Similar to other structures of RNA-bound DEAD-box proteins, Ddx19 introduces a kink in the RNA backbone between bases 4 and 5[222, 226]. Additionally, a region of the NTE folds onto domain 1 and contributes to the ATP-binding site before bridging across the top of the cleft to domain 2 (**Figure 57** A). The RNA-bound structure of the Ddx19 is in excellent agreement with our biochemical data, and provides the molecular basis for our observation that Nup214 NTD and RNA binding are mutually exclusive, as the binding sites for RNA and Nup214 NTD on Ddx19 partially overlap (**Figure 58**).

Recently, another snapshot of the ATPase cycle of Ddx19 was revealed by the structure of Ddx19 in its ADP-bound state. This is the first and only structure of both domains of any ADP-bound DEAD-box protein and revealed that a significant section of the NTE folds into an α-helix, α1, and wedges into the central cleft between the two core domains[225] (**Figure 57 B**). This interaction causes the two domains to move apart and results in a ~12 Å shift of the conserved motif VI away from the nucleotide. Consequently, helix α1, rather than motif VI, contacts the phosphate groups of the ADP. Helix α1 has therefore been hypothesized to negatively regulate ATPase activity by displacing the arginine finger of motif VI, essential for ATP hydrolysis, from the nucleotide-binding pocket[225].

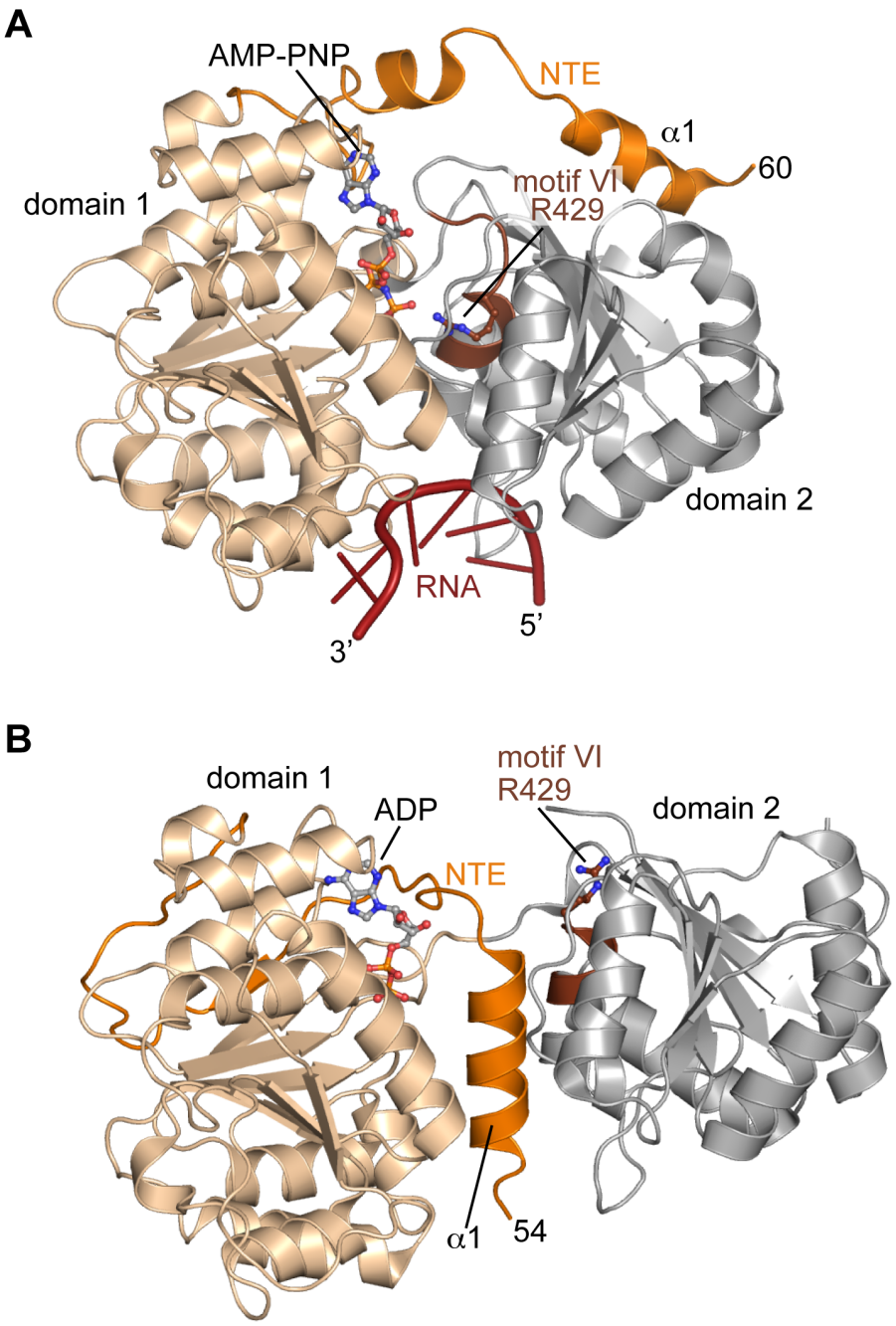


Figure 57: Structures of Ddx19. (**A**) Overview of the structure of Ddx19-AMPPNP in complex with RNA[225]. (**B**) Overview of the ADP-bound structure of Ddx19[225]. Ddx19 is shown in ribbon representation with domain 1 (wheat), domain 2 (grey), and the N-terminal extension (NTE) with helix α1 (orange). The conserved Motif VI with R429 is indicated in brown and the RNA is shown in red.

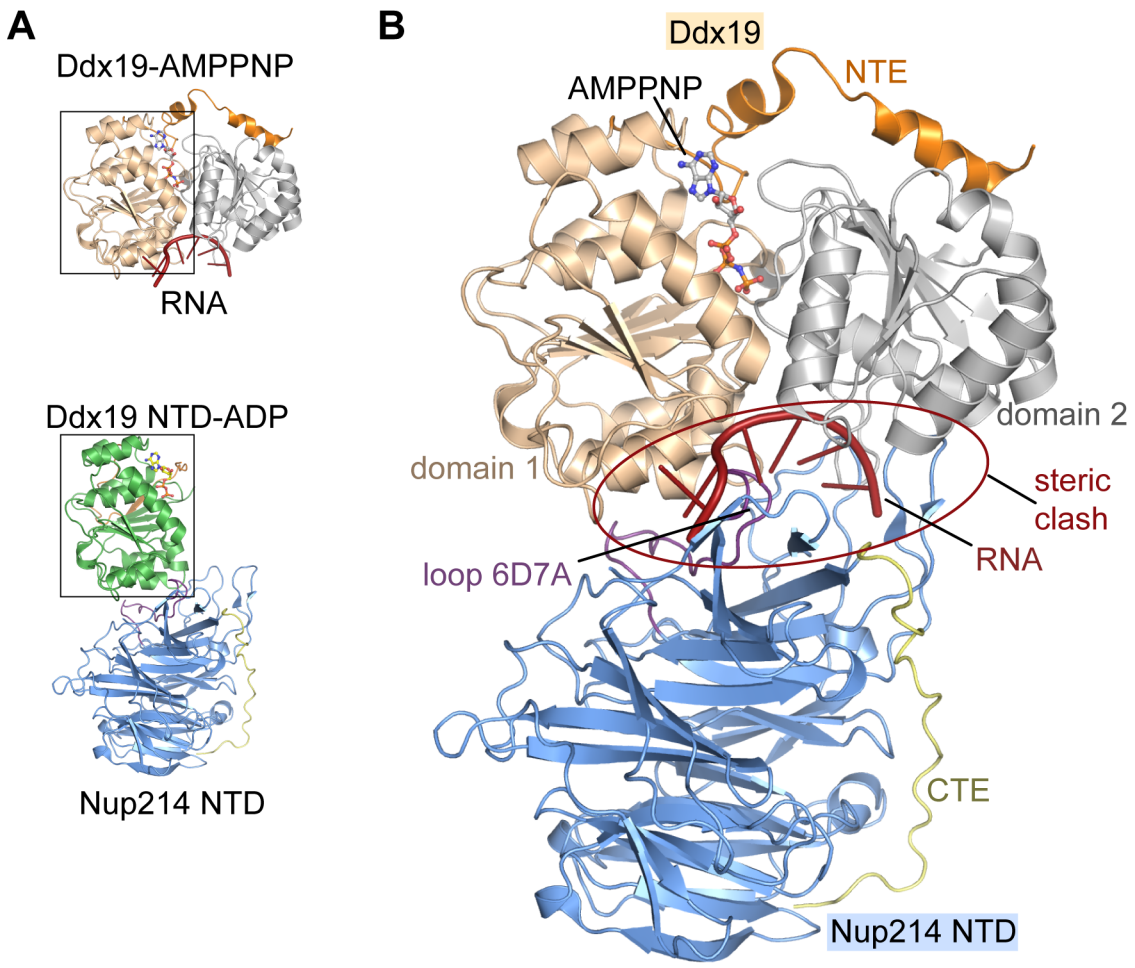


Figure 58: Binding of Nup214 NTD and RNA on Ddx19 are mutually exclusive. (**A**) Overview of the structure of Nup214 NTD•Ddx19 NTD-ADP and Ddx19-AMPPNP in complex with RNA used for the structural alignment[225, 227]. The structural alignment is based on domain 1 of Ddx19, indicated by a black box. (**B**) A structural alignment of the Ddx19-AMPPNP•RNA and Nup214 NTD•Ddx19 NTD-ADP structures reveals overlapping binding sites for Nup214 and the RNA (highlighted by the red ellipse). For clarity, the Ddx19 domain 1 of the Nup214 NTD•Ddx19 NTD-ADP structure is omitted. The ribbon representation is colored according to Figure 44 and Figure 57.

3.2.2 The ATPase cycle of Ddx19

A key step in the transport of nuclear mRNPs to the cytoplasm is the transit through the NPC. This transit involves mRNP remodeling, whereby proteins bound to nuclear mRNA are dissociated. The displacement of mRNA transport factors was shown to be essential for mRNA transport into the cytoplasm[166]. In yeast, the DEAD-box helicase Ddx19 displaces the hnRNP protein Nab2 from RNA *in vitro* and was shown to be essential for the dissociation of the mRNP export factor TAP/p15 *in vivo*[166, 167]. The localization of Ddx19 to the NPC is essential for this process, and Ddx19 is recruited via a conserved interaction with Nup214 to the cytoplasmic filaments of the NPC[35, 150].

The general mechanism of mRNP remodeling by DEAD-box proteins involves binding of the RNA substrate, unwinding, ATP hydrolysis, and concomitant dissociation of RNA-bound proteins[228]. The activity of DEAD-box proteins can be modulated in *cis* by N- or C-terminal additions to the core protein fold, which consists of two RecA-like globular domains. Alternatively, DEAD-box proteins can be regulated in *trans* by interactions with other protein cofactors. DEAD-box proteins have been shown to destabilize and “melt” short RNA duplexes and to require ATP hydrolysis for substrate release *in vitro*[168]. However, the mechanism for the RNA remodeling step, including RNA unwinding and removal of proteins from the RNA, remains unclear.

To date, several structures of snapshots of the Ddx19 that capture various states of the Ddx19 ATPase cycle have been determined, including Ddx19-ADP,

Ddx19-ADP•Nup214, Ddx19•Nup214 and Ddx19-AMPPNP•RNA, which give us insight into the mRNA remodeling step at the NPC[224, 225, 227].

3.2.2.1 RNA-binding and unwinding activity of Ddx19

Unwinding activity for yDbp5 has only been demonstrated for short RNA duplexes *in vitro*[140, 164]. The structure of Ddx19-AMPPNP bound to single-stranded RNA showed that Ddx19 associates with RNA in a closed conformation. Binding of Ddx19 introduces a kink in the RNA backbone (**Figure 57 A**). As this kink is incompatible with a perfect double-stranded RNA helix, one hypothesis is that the kink constitutes a first step toward helicase binding and represents a partially opened duplex[226]. The structure of the AMPPNP-Ddx19 bound to a single-stranded RNA fragment can be considered as the "post-unwinding" complex and a structure of a DEAD-box protein bound to a double-stranded RNA is not available. The question of whether Ddx19 could locally open duplex strands by capturing transiently frayed base pairs and/or by actively inducing helix opening, remains open. Further structural studies are necessary to elucidate the mechanism of duplex unwinding by DEAD-box proteins.

3.2.2.2 The mechanism of ATP-hydrolysis for Ddx19

Another gap in our knowledge of the ATPase cycle of Ddx19, one that is essential for the ratchet function, is the mechanism of ATP-hydrolysis activation. By analogy to GTPases, it is conceivable that ATP hydrolysis and ADP/ATP exchange is enhanced by a specific activating protein (ATPase activating protein, AAP) and/or exchange factor (ATP exchange factor, AEF). Indeed, interaction

partners extensively regulate the DEAD-box helicase eIF4A of the eIF4F complex that is important for recognizing the 5' cap of cytoplasmic mRNPs[229]. eIF4F is a heterotrimer consisting of eIF4A, eIF4G and eIF3. The helicase activity of eIF4A is stimulated by additional factors, including eIF4G[230]. The structure of eIF4A in complex with the central domain of eIF4G reveals that eIF4G holds the crucial DEAD-box sequence motifs in a semi-closed conformation thereby stimulating the ATPase activity of the DEAD-box helicase (**Figure 59**)[231].

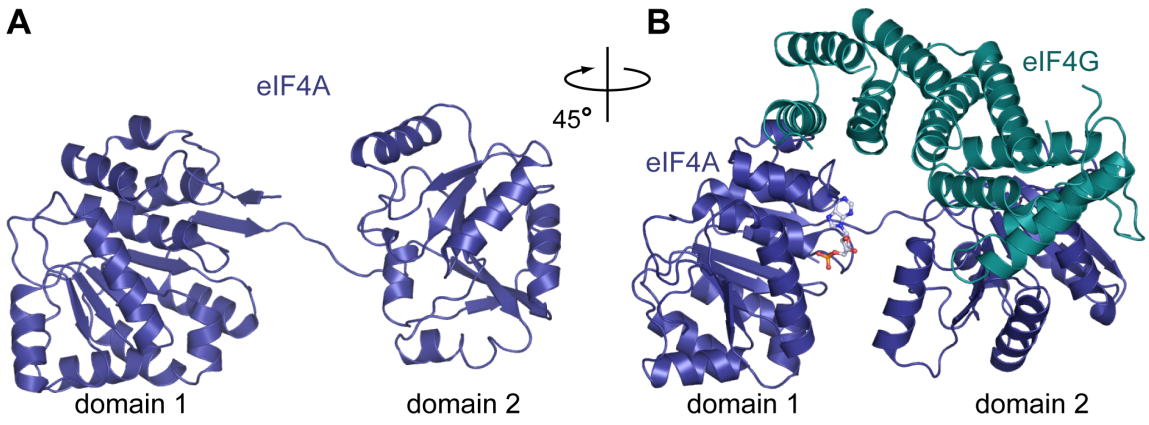


Figure 59: Structures of the initiation factor eIF4A. Overview of the structure of eIF4A in isolation (PDB code 1FUU) (**A**) or in complex with the ATPase activating protein eIF4G (PDB code 1VSU) (**B**). eIF4A is shown in blue and eIF4G in green. Binding of eIF4G induces a large conformational shift of domain 2 with respect to domain 1 of eIF4A. For clarity, the view in (B) is rotated by 45˚.

Similarly, the ATPase activity of Ddx19 is modulated by protein cofactors, and is maximally stimulated in the presence of Gle1 and RNA[164, 165]. A structure of Ddx19 in complex with Gle1 would provide invaluable insights into the mechanistic details of the ATP hydrolysis activation step. We have shown that binding of yGle1 and yNup159 to yDbp5 is not mutually exclusive *in vitro*, suggesting distinct binding sites. Although highly likely, it remains to be proven whether this is true for RNA and Gle1 as well. Support for this notion comes from recent work which showed that four single point mutants of Ddx19 exhibited reduced ATPase stimulation by Gle1 *in vitro* as well as reduced interaction in a yeast-2-hybrid assay[232]. The mutated residues are located in domain 2 distal from the interdomain cleft and RNA binding site. It is conceivable that binding of RNA and Gle1 to the helicase would dislodge helix α1 of Ddx19 from its interdomain cleft and consequently activate ATP hydrolysis. Further support for such a mechanism comes from the observation that a deletion construct of Ddx19 lacking the entire NTE exhibits an ATPase activity that is independent of RNA and Gle1[225].

3.2.2.3 The Nup214 NTD as a possible ATP-exchange factor for Ddx19

A comparison of the DEAD-box helicases UAP56 and eIF4A has shown that nucleotide binding induces conformational changes that affect the association of the two domains and their spatial arrangement[155, 220, 231]. Indeed, binding of Ddx19 to RNA and ATP stabilizes a closed conformation of Ddx19, while binding to Nup214 NTD promotes mobility of domain 2 as demonstrated by our Nup214

NTD•Ddx19-ADP crystal structure that lacks density for domain 2[225, 227]. Since an AEF has not yet been identified for Ddx19, we hypothesize that Nup214 could act as AEF. Similar to the pattern of interaction of GTPases with their GAPs, nucleotide free Ddx19 binds with high affinity to Nup214, while Ddx19 in its ATP-bound state has the lowest affinity for Nup214[224]. Superposition of the Ddx19-ADP structure with our Nup214 NTD•Ddx19 NTD-ADP structure indicates steric clashes of Nup214 NTD with Ddx19 (**Figure 60**). In particular, loop 1CD and 7AB of the Nup214 NTD clash with domain 2 and helix α1 of Ddx19. Previously, we had identified the 1CD and 7AB loops as unique features of the Nup214 β-propeller. Instead of protruding to the bottom face of the propeller domain, the loops protrude to the side of the β-propeller for up to ~25 Å (**Figure 27 C, D**).

Therefore one possible mechanism would be that the interaction of Ddx19 with Nup214 destabilizes the closed conformation of Ddx19 and promotes mobility of helix α1 and domain 2, promoting in turn nucleotide release. Subsequent ATP binding decreases the affinity of Ddx19 for Nup214 NTD substantially, and increases the affinity of Ddx19 for RNA, leading to the release of Nup214 by Ddx19-ATP in favor of RNA.

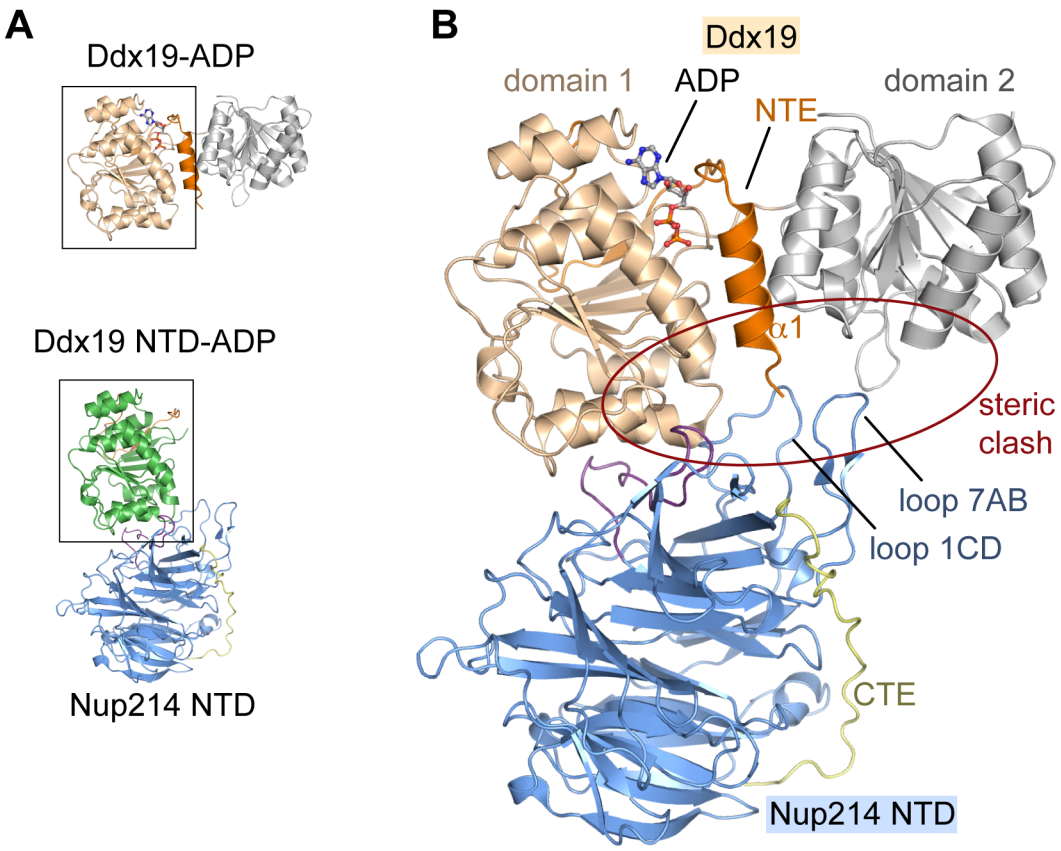


Figure 60: Nup214 may function as an AEF for Ddx19. (**A**) Overview of the structure of Ddx19-ADP and Nup214 NTD•Ddx19 NTD-ADP. The structural alignment is based on domain 1 of Ddx19, indicated by a black box. (**B**) A structural alignment of the Ddx19-ADP and Nup214 NTD•Ddx19 NTD-ADP structures reveals clashes between the Nup214 NTD and domain 2 and the NTE of Ddx19 (highlighted by the red ellipse). For clarity, the Ddx19 domain 1 of the Nup214 NTD•Ddx19 NTD-ADP structure is omitted. The ribbon representation is colored according to Figure 44 and Figure 57.

3.2.3 Transport of mRNA through the NPC

Considering the available biochemical and structural information, a plausible sequence of events for the transport of mRNA through the NPC might be as follows (**Figure 61**): Linearized mRNP (red) moves through the central channel toward the cytoplasmic side of the NPC. Nup214 (blue) recruits Ddx19 to the NPC and destabilizes the closed conformation of ADP-Ddx19, thereby promoting nucleotide exchange. (2) ATP-Ddx19 dissociates from Nup214 and binds strongly to the mRNA. (3) Binding of the ATP-Ddx19 to the RNA destabilizes and displaces RNA-bound proteins (transport factors) from the mRNA. Subsequent ATP hydrolysis (red star), stimulated by Gle1 and mRNA, weakens the interaction of Ddx19 with the mRNA (4) and allows Ddx19 to bind to Nup214 again, thereby completing one cycle of protein removal from a segment of mRNA.

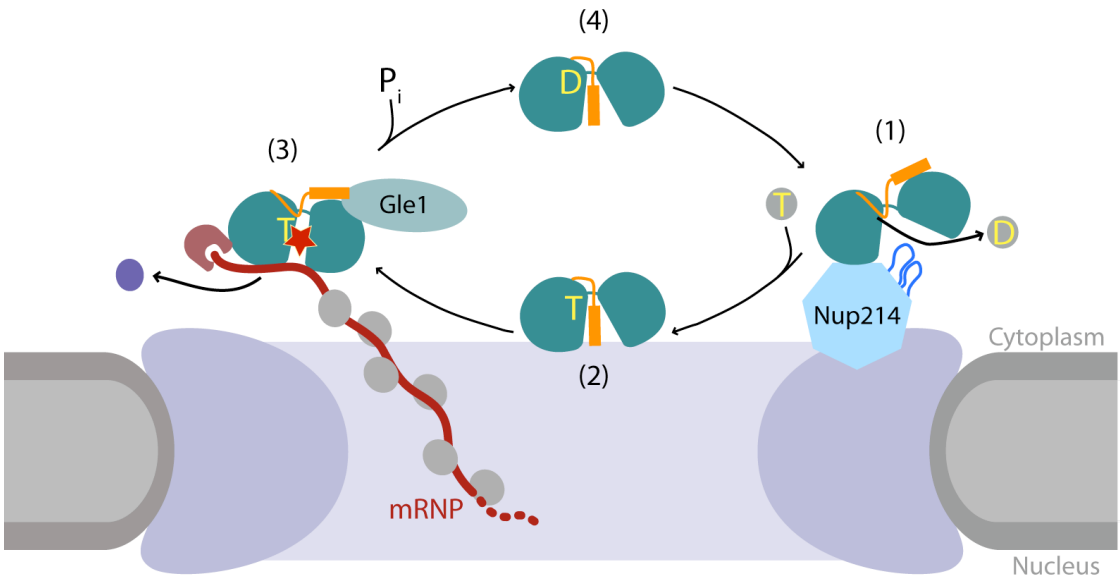


Figure 61: Schematic view of a potential ATPase cycle of Ddx19. The mRNP is shown schematically as a red line with mRNA export receptor complex (grey) attached, although it will have extensive secondary structure and many other mRNA binding proteins in the cell. For clarity only the central core of the NPC is shown. (**1**) Nup214 acts as AEF and promotes nucleotide exchange by destabilizing the closed conformation of Ddx19. After nucleotide exchange, ATP-Ddx19 dissociates from Nup214. (**2**) Helix α1 wedges between the two core domains of Ddx19 and prevents premature ATP-hydrolysis. (**3**) Binding of Ddx19 to mRNA displaces mRNA-export factors from the mRNP, and Gle1 and mRNA stimulate ATP hydrolysis. (**4**) ADP-Ddx19 dissociates from the mRNA and binds to Nup214, thereby completing one ATPase cycle. Nup214 is shown in blue with loop 1CD and 7AB in dark blue; Gle1 is shown in light green, a displaced export receptor complex in violet, and Ddx19 with helix α1 in green and orange, respectively. ATP and ADP are indicated as yellow T and D, and ATP hydrolysis with a red star.

3.2.4 Summary and future directions

A key step in export of mRNP from the nucleus is the transport through the NPC. The transport includes mRNP-remodeling steps to assure a proper sequence of events. Premature remodeling in the nucleus would prevent or impede mRNA export, whereas newly exported mRNP are not structured optimally for translation. Furthermore, the removal of mRNA transport receptors from the mRNP has been shown to be crucial for mRNA export by a molecular ratchet mechanism. Therefore it is of utmost importance for the cell to spatially control mRNP remodeling. Genetic screens and interactions in yeast have implicated the nucleoporin Nup214, the DEAD-box protein Ddx19, and the NPC-associated protein Gle1 in the spatial control of Ddx19 ATPase activation in the cell.

My thesis work focused on the biochemical and structural analysis of Nup214 and Ddx19, and provided the first molecular view of this crucial interaction. Together with the work of other groups we can now start to paint a picture in which the ATPase cycle of Ddx19 drives unidirectional mRNP export through the NPC. However, several snapshots are still missing. These include the “pre-unwinding” state of Ddx19, the ATP-bound state of Ddx19, as well as Ddx19 in complex with its ATPase activating factor Gle1. As described in the results, I have already started to analyze the interaction of the helicase with Gle1 biochemically. However, structural characterization is hampered by the poor solubility of Gle1. Therefore further coexpression tests, as well as expression

construct modifications and screening of homologues from other species seem inevitable.

In addition to the pools of Ddx19 and Gle1 that are associated with the NPC, both proteins are also found in the cytoplasm[141, 150]. Two new studies have implicated Ddx19 and Gle1 in translation termination[233, 234]. Translation termination occurs when the polypeptide release factor, eRF1, recognizes a stop codon in the A-site of the ribosome. Together with eRF3, polypetidyl-tRNA hydrolysis is stimulated and the polypeptide released from the ribosome. New evidence suggests that Ddx19 and Gle1 associate with translating ribosomes and are required for efficient stop codon recognition in yeast. Ddx19 and Gle1 physically interact with eRF1 and are required for loading of eRF3 into the termination complex[233, 234]. One function performed by Ddx19 in translation termination could be that Ddx19's remodeling activity triggers an event on the mRNP that promotes association of eRF3. The exact location of this remodeling step could be either close to the stop codon, or alternatively on the ribosome close to the A-site. Translation termination can be dissected *in vitro* using recombinant proteins and purified ribosomes, which should shed light on the exact mechanism in the near future.

Finally, the number of ATP-hydrolysis cycles of Ddx19 that occur per exported mRNP is not known. Studies of Balbiani ring mRNA showed that the mRNPs emerge with their 5' end first from the NPC and that the CBC is replaced concomitantly with translation initiation factors. Furthermore, it was observed that

ribosomes bind to the mRNP at the cytoplasmic side of the NPC and start translation before the entire mRNP has been exported[235]. This association of ribosomes with mRNA that is still being transported through the NPC could provide an additional pulling force for mRNP export.

3.3 Conclusion

Gene expression is one of the mechanisms necessary for a cell to be able to play its part in an organism's developmental program, to achieve homeostasis, or to respond to external stimuli. The central dogma, that "DNA makes RNA makes protein", describes the information flow from the genetic material to proteins[236]. An essential step in this process involves the biogenesis of mRNA. In eukaryotes, mRNA is transcribed in the nucleus and translated into proteins in the cytoplasm. The life cycle of an mRNA molecule includes transcription, processing, transport through the NPC into the cytoplasm, translation, and degradation[237].

The mechanisms of transcription and translation are reasonably well understood. *In vitro* assays to study RNA synthesis are established and structures of RNA polymerase from different species have provided invaluable insights into the mechanisms of transcription[238, 239]. Likewise, the study of the structure and function of the translational apparatus has been a research interest since the mid-twentieth century, and has resulted in a molecular and structural understanding of the translation mechanism[240, 241].

Less is known about the mechanism of mRNA processing and transport, probably because these processes cannot be studied in prokaryotes. Processing of pre-mRNA involves 5' capping, splicing, and polyadenylation, and is tightly coupled to transcription[111]. Most factors involved in processing have been identified and biochemically characterized[122, 135]. Splicing involves a complex consisting of multiple proteins and RNAs, the spliceosome, and remains a major challenge for structural biology[120, 121].

Our current understanding of mRNP packaging and transport is that during transcription, the nascent mRNA transcript associates with a number of factors, including proteins from the family of heterogeneous nuclear ribonucleoproteins (hnRNPs). hnRNPs are nuclear RNA-binding proteins that are essential for various steps in the mRNA life cycle, including packaging, export and translation[113]. Other factors, like the mRNA transport adaptor Aly and the mRNA export receptor complex TAP/p15, are also recruited to the mRNP. The packaged mRNP is targeted to the NPC, where TAP/p15 mediates interaction with the FG-domains that line the central channel of the NPC. This allows transit to the cytoplasmic side, where Ddx19 removes TAP/p15 from the mRNP, and the mRNP concludes its journey into the cytoplasm.

We have yet to characterize the rearrangements of the mRNP that occur during packaging and transport. One of the few systems studied is the giant transcript of *Chironomus* Balbiani ring (BR) mRNA. In this case, the mRNA is immediately packaged into a 7 nm mRNP fiber during transcription[113], and folds

into a well defined mRNP particle that docks to the NPC[116]. At the NPC, the BR mRNP undergoes conformational rearrangements that facilitate transport through the NPC[235]. Clearly, the giant Balbiani ring transcripts of *Chironomus* have been a valuable tool for the study of mRNA export. However, the question of whether the observed structure of the BR mRNP particles are the norm or, because of the large size of the transcript, are a special case, is still unanswered.

We also need to decipher the biochemical determinants of the mRNP in order to identify what constitutes an export-competent mRNP. Are all mRNPs made equal? Which mRNP-binding proteins are needed to dock to the NPC? Which accompany the mRNP through the pore? Which are removed after transport?

During transcriptional elongation, the nascent mRNA transcript is immediately bound by hnRNPs, and the subsequent processing steps deposit additional factors on the transcript. There are ~30 different hnRNPs in humans and ~10 in yeast[111]. The different hnRNPs can bind to and disassociate from the mRNA at distinct stages of the mRNP biogenesis[111]. For instance, two of the best studied hnRNP proteins are the poly(A)-binding proteins Nab2 and UAP56, which both associates with the mRNP during processing[220]. While UAP56 dissociates from the mRNP in the nucleus, Nab2 accompanies it through the NPC[242]. Further studies on nuclear mRNP proteins are needed to generate a dynamic map of mRNP composition during different stages of mRNA biogenesis.

Finally, the lack of a suitable *in vitro* system hampers functional studies of mRNP transport. Artificial nanopores that contain FG-repeat domains have been shown to mimic the transport selectivity of the NPC, and may be the first step toward this goal[243]. Alternatively, a reconstitution of isolated NPC subcomplexes into a double lipid bilayer *in vitro* may be achievable.

4 MATERIALS AND METHODS

4.1 Molecular cloning

The sequence of the N-terminal domain deletion constructs were amplified by PCR from a human Nup214 cDNA construct (KIAA0023) and cloned into a pET28a vector (GE Bioscience) modified to contain a PreScission protease-cleavable N-terminal hexa-histidine tag[244]. For Nup214 NTDΔ1 residues 343 to 374, for Nup214 NTDΔ2 residues 347 to 351 and for Nup214 NTDΔ3 residues 352 to 356 were deleted and cloned in the pET28a-PreS vector, as described above. For Nup214 NTDΔ6D7A, residues 343 to 361 were replaced by a flexible linker with the sequence GGSGG and cloned in the pET28a-PreS vector, as described above. A DNA fragment encompassing residues 343 to 361 of the Nup214 NTD 6D7A loop and followed by a hexa-histidine tag was cloned into the pGEX-6P1 vector (GE Healthcare).

DNA fragments of the full-length human Ddx19, and a Dxd19 fragment, encompassing residues 1 to 300 (Ddx19 NTD), were amplified from HeLa cell cDNA, and cloned in the pET28a-PreS vector (GE Bioscience) modified to contain a PreScission protease-cleavable N-terminal hexa-histidine tag[244]. Mutants of the Nup214 NTD and Ddx19 were generated using Quickchange site-directed mutagenesis (Stratagene). The details of the expression constructs and the mutants of Nup214 and Ddx19 are listed in **Table 2** to **Table 5.**

Various DNA fragments of the human Gle1 and hCG1 were amplified by PCR from HeLa cell cDNA and cloned in the pET28a-PreS and pGEX-6P1 vectors (GE Healthcare) vector. For the yeast proteins, the DNA sequences of a N-terminal domain deletion construct of yNup159 and full-length yDbp5 were amplified by PCR from the yeast ORF clones YIL115C and YOR046C (Open Biosystems) and cloned into a pET28a vector (GE Bioscience) modified to contain a PreScission protease-cleavable N-terminal hexa-histidine tag[244]. DNA fragments of yGle and yNup42 were amplified by PCR from the yeast ORF clones YDL207W and YDR192C (Open Biosystems) and cloned in the pET28a-PreS and pGEX-6P1 vectors (GE Healthcare) vectors. The details of the expression constructs are listed in **Table 4** and **Table 5**. Mutants of the yDbp5 were generated using Quickchange site-directed mutagenesis (Stratagene).

Table 2: Nup214 Mutants

Deletion mutants	**Point mutants**	
Nup214 NTD	Nup214 NTD L343A	Nup214 NTD D355A
Nup214 NTD 1-405	Nup214 NTD D345A	Nup214 NTD K356A
Nup214 NTD Δ6D7A	Nup214 NTD S346A	Nup214 NTD S357A
Nup214 NTD Δ1	Nup214 NTD R348A	Nup214 NTD D358A
Nup214 NTD Δ2	Nup214 NTD E350A	Nup214 NTD D359A
Nup214 NTD Δ3	Nup214 NTD V353A	Nup214 NTD V353E
GST-6D7A-His_6	Nup214 NTD T354A	Nup214 NTD E350A/V353E

Table 3: Ddx19 mutants

NTE mutants	Domain 1 mutants	Domain 2 mutants
H24A	L99A	R262A
K26A	K100A	R266A
K40A	L104A*	Q269A
K53A	Q105A	D279A*
R56A	Y108A	Q286A
K64A	Q132A	K287A
R67A	Q164A	K298A
V77A	E183A	K308A
E78A	E191A	N326A
R82A	K202A	K344A
K92A	E204A	T345A
	Q207A	V369A
	S210A	E370A
	T220A	R372A
	L222A	V375A*
	D223A	E377A*
	S226A	R378A
	K227A	R380A
	I231A	E381A
	K234A	K383A
	K235A	K385A
	K237A*	V392A
	H253A	V412A
	Q254A	K446A
	D255A	L452A*
	Q256A	E457A
	I258A	

Table 3	continued	
NTE mutants	**Domain 1 mutants**	**Domain 2 mutants**
	R259A	
	R259E	
	R259Q	
	R259K	

Table 4: Expression constructs I

Protein	Residues	Vector	Restriction sites 5', 3'	Tags N / C
Nup214 N315	1-315	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N330	1-330	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N345	1-345	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N360	1-360	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N375	1-375	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N390	1-390	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N405	1-405	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N435	1-435	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N450	1-450	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N465	1-465	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N495	1-495	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 N510	1-510	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 NTD	1-450	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 mutants	1-450	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 NTD 1-405	1-405	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 NTD Δ6D7A	1-342, GGSGG, 362-450	pET28a-PreS	NdeI, NotI	His_6 / none

Table 4	continued			
Protein	**Residues**	**Vector**	**Restriction sites 5', 3'**	**Tags N / C**
Nup214 NTDΔ1	1-342, 348-450	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 NTDΔ2	1-347, 352-450	pET28a-PreS	NdeI, NotI	His_6 / none
Nup214 NTDΔ3	1-352, 357-450	pET28a-PreS	NdeI, NotI	His_6 / none
GST-6D7A-His_6	343-361	pGEX-6P1	EcoRI, NotI	GST / His_6
Ddx19	1-479	pET28a-PreS	NdeI, NotI	His_6 / none
Ddx19 NTD	1-300	pET28a-PreS	NdeI, NotI	His_6 / none
Ddx19 mutants	1-479	pET28a-PreS	NdeI, NotI	His_6 / none
yNup159	1-388	pET28a-PreS	NdeI, NotI	His_6 / none
yDbp5	1-466	pET28a-PreS	NdeI, NotI	His_6 / none
yDbp5 mutants	1-466	pET28a-PreS	NdeI, NotI	His_6 / none

Table 5: Expression constructs II

Protein	Residues	Vector	Restriction sites 5', 3'	Tags N / C
Gle1	1-450	pGEX-6P1	BamHI, NotI	GST / none
Gle1	1-526	pGEX-6P1	BamHI, NotI	GST / none
Gle1	1-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	8-450	pGEX-6P1	BamHI, NotI	GST / none
Gle1	8-526	pGEX-6P1	BamHI, NotI	GST / none
Gle1	8-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	80-450	pGEX-6P1	BamHI, NotI	GST / none
Gle1	115-450	pGEX-6P1	BamHI, NotI	GST / none
Gle1	115-526	pGEX-6P1	BamHI, NotI	GST / none
Gle1	115-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	265-450	pGEX-6P1	BamHI, NotI	GST / none
Gle1	265-526	pGEX-6P1	BamHI, NotI	GST / none
Gle1	265-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	1-698	pET28a-PreS	NheI, NotI	His6 / none
Gle1	201-698	pET28a-PreS	NheI, NotI	His6 / none
Gle1	301-698	pET28a-PreS	NheI, NotI	His6 / none
Gle1	401-698	pET28a-PreS	NheI, NotI	His6 / none
Gle1	501-698	pET28a-PreS	NheI, NotI	His6 / none
Gle1	201-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	301-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	501-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	531-698	pGEX-6P1	BamHI, NotI	GST / none
Gle1	265-698	pGEX-6P1	BamHI, NotI	GST / none
yGle1	1-538	pET28a-PreS	NdeI, NotI	His6 / none
yGle1	50-538	pET28a-PreS	NdeI, NotI	His6 / none

Table 5	continued			
Protein	**Residues**	**Vector**	**Restriction sites 5', 3'**	**Tags N / C**
yGle1	101-538	pET28a-PreS	NdeI, NotI	His6 / none
yGle1	275-538	pET28a-PreS	NdeI, NotI	His6 / none
yGle1	350-538	pET28a-PreS	NdeI, NotI	His6 / none
yGle1	410-538	pET28a-PreS	NdeI, NotI	His6 / none
yGle1	50-538	pGEX-6P1	EcoRI, NotI	GST / none
yGle1	101-538	pGEX-6P1	EcoRI, NotI	GST / none
yGle1	275-538	pGEX-6P1	EcoRI, NotI	GST / none
yGle1	350-538	pGEX-6P1	EcoRI, NotI	GST / none
hCG1	1-423	pET28a-PreS	NdeI, NotI	His6 / none
hCG1	106-409	pET28a-PreS	NdeI, NotI	His6 / none
hCG1	106-210	pET28a-PreS	NdeI, NotI	His6 / none
hCG1	1-423	pGEX-6P1	EcoRI, NotI	GST / none
hCG1	106-409	pGEX-6P1	EcoRI, NotI	GST / none
yNup42	1-430	pET28a-PreS	NdeI, NotI	His6 / none

4.2 Solubility analysis

Proteins were expressed and lysed as described below. The soluble and insoluble fractions were separated by centrifugation at 40,000 g for 60 min. The pellet was resuspended in lysis buffer in a volume equal to the volume of the supernatant. Aliquots of the pellet (P) and supernatant (SN) fractions analyzed by SDS-PAGE gel electrophoresis followed by transfer onto PVDF membranes (Millipore) using a semi-dry electrophoretic transfer cell (Bio-Rad). Membranes were probed with appropriate primary antibodies, penta-histidine antibody (Qiagen) or anti-GST antibody (GE Healthcare). After incubation with appropriate secondary antibodies coupled to alkaline phosphatase (GE Healthcare), proteins were visualized using Sigmafast BCIP/NBT and BZIP tablets (Sigma-Aldrich).

4.3 Expression and purification

4.3.1 Expression and purification of Nup214 and yNup159

Nup214 variants and yNup159 NTD were expressed in *E. coli* BL21 (DE3) cells (Stratagene). Cells were grown at 37°C to an OD_{600} of 0.6, induced with 450 μM isopropyl-β-D-thio-galactoside (IPTG), at 18°C for 9 h. Cells were harvested by centrifugation, resuspended in lysis buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, 5 mM β-mercaptoethanol, 1 mM PMSF, and Complete EDTA-free protease inhibitor cocktail tablets (Roche). The cells were lysed with a cell disrupter

(Avestin), and the lysate was clarified by centrifugation at 40,000 g for 90 min. The clarified lysate was then loaded onto a Ni-NTA column (Qiagen) and eluted via an imidazole gradient. Fractions containing His_6-Nup214 NTD variants were pooled, desalted using a HiPrep 26/20 desalting column (GE Healthcare), and the hexahistidine-tag was cleaved with PreScission protease (GE Healthcare) for 12 h. The protein was further purified by size exclusion chromatography on 16/60 Superdex 200 and Superdex 75 columns (GE Healthcare) in a buffer containing 20 mM Tris pH 8.0, 100 mM NaCl, and 5 mM DTT. Nup214 NTD was concentrated to 50 mg/ml for crystallization.

4.3.2 Expression and purification of Ddx19 and yDbp5

For expression of Ddx19, Ddx19 NTD, Ddx19 mutants as well as yDbp5 and yDbp5 mutants, *E. coli* BL21 (DE3) RIL cells (Stratagene) were grown at 37°C to an OD_{600} of 0.6 and induced with 450 μM IPTG at 17°C for 9 h. Cells were harvested by centrifugation and resuspended in a lysis buffer containing 20 mM Tris, pH 8.0, 500 mM NaCl, 5 mM β-mercaptoethanol, 1 mM PMSF, and Complete EDTA-free protease inhibitor cocktail tablets (Roche). The cells were lysed with a cell disrupter (Avestin), and the lysate was clarified by centrifugation at 40,000 g for 90 min. The lysate was then loaded onto a Ni-NTA column (Qiagen) and eluted via an imidazole gradient. Fractions containing hexa-histidine tagged Ddx19 were pooled, and glycerol was added to a final concentration of 10% (v/v). After cleavage of the hexa-histidine tag, the protein was further purified with PreScission protease (GE Healthcare) by size exclusion

chromatography on a 16/60 Superdex 200 column (GE Healthcare) that was equilibrated in a buffer containing 20 mM Tris, pH 8.0, 100 mM KCl, 10% glycerol and 5 mM DTT.

4.3.3 Purification of the Nup214•Ddx19 complex

For complex formation, Nup214 NTD was incubated with an equimolar amount of either Ddx19 or Ddx19 NTD and purified by size exclusion chromatography on a 16/60 Superdex 200 column (GE Healthcare) that was equilibrated in a buffer containing 20 mM Tris pH 8.0, 100 mM KCl, 5% glycerol, 5 mM DTT and either 0.5 mM Mg-ADP (Spectrum Chemical), Mg-AMPPNP (Sigma-Aldrich) or Mg-ATP-γ-S (Sigma-Aldrich). The fractions containing the complex were pooled and concentrated to 40 mg/ml for crystallization.

4.3.4 Expression and purification of yGle1

For expression of yGle1ΔN100 *E. coli* BL21(DE3) RIL cells (Stratagene) were grown at 37°C to an OD_{600} of 0.6 and induced with 450 μM IPTG at 17°C for 9 h. Cells were harvested by centrifugation and resuspended in a lysis buffer containing 50 mM Tris, pH 8.0, 150 mM NaCl, 5 mM β-mercaptoethanol, 1mM IP_6 (Sigma), 10% glycerol, 1 mM PMSF, and Complete EDTA-free protease inhibitor cocktail tablets (Roche). The cells were lysed with a cell disrupter (Avestin), and the lysate was clarified by centrifugation at 40,000 g for 90 min. The lysate was then loaded onto a GSTrap HP column (GE Healthcare) and eluted via a glutathione gradient. Fractions containing GST-tagged yGle1ΔN100 were pooled, and glycerol was added to a final concentration of 10% (v/v). After

cleavage of the GST tag with PreScission protease (GE Healthcare), the protein was further purified by HiTrap Q FF (GE Healthcare) and size exclusion chromatography on a 16/60 Superdex 200 column (GE Healthcare) that was equilibrated in a buffer containing 20 mM Tris, pH 8.0, 100 mM KCl, 10% glycerol, 1mM IP_6 and 5 mM DTT.

4.4 Crystallization and structure determination

4.4.1 Crystallization and structure determination of Nup214

Crystals of Nup214 NTD (50 mg/ml) were grown at 21°C in hanging drops containing 1 μl of the protein and 1 μl of a reservoir solution consisting of 100 mM MES pH 6.5 and 23-27% (w/v) PEG 2,000 MME. Large crystals with dimensions of 350 x 350 x 150 μm^3 grew within one week and were flash frozen in liquid nitrogen-cooled liquid propane. Crystals belong to the orthorhombic space group $P2_12_12_1$ (unit cell dimensions are a=52.4 Å, b=81.1 Å, c=102.6 Å, $\alpha=\beta=\gamma=90°$), with one molecule in the asymmetric unit. Heavy-atom-derivatized crystals were prepared by soaking native crystals for several days with 10 mM p-chloromercuribenzoic acid (PCMB) or osmiumtetroxide (OsO_4).

X-ray diffraction data were collected at the National Synchrotron Light Source (NSLS), Brookhaven National laboratory (BNL), beamline X9a, and the Advanced Light Source (ALS), Lawrence Berkeley National laboratory (LBNL), beamline 8.2.1. X-ray intensities were processed using the HKL2000 denzo/scalepack package[245] and the CCP4 program package[246] was used for

subsequent calculations. Phases were determined using native and anomalous X-ray diffraction data derived from native and heavy atom derivativatized crystals in SHARP[247]. This was followed by density modification in DM (CCP4) with solvent flattening and histogram matching, and yielded an electron density map of excellent quality (**Figure 62**).

A model was build with the program O[248], refined with CNS[249], and the stereochemical quality of the model was assessed with PROCHECK[250]. The refined model includes residues 1 to 434 with no residues in the disallowed regions of the Ramachandran plot (**Figure 63**). Data collection and refinement statistics are summarized in **Table 6**.

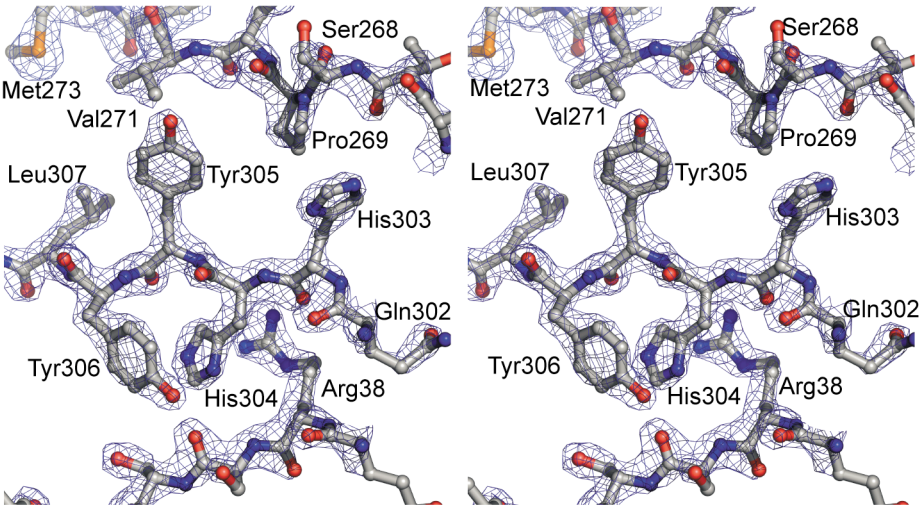
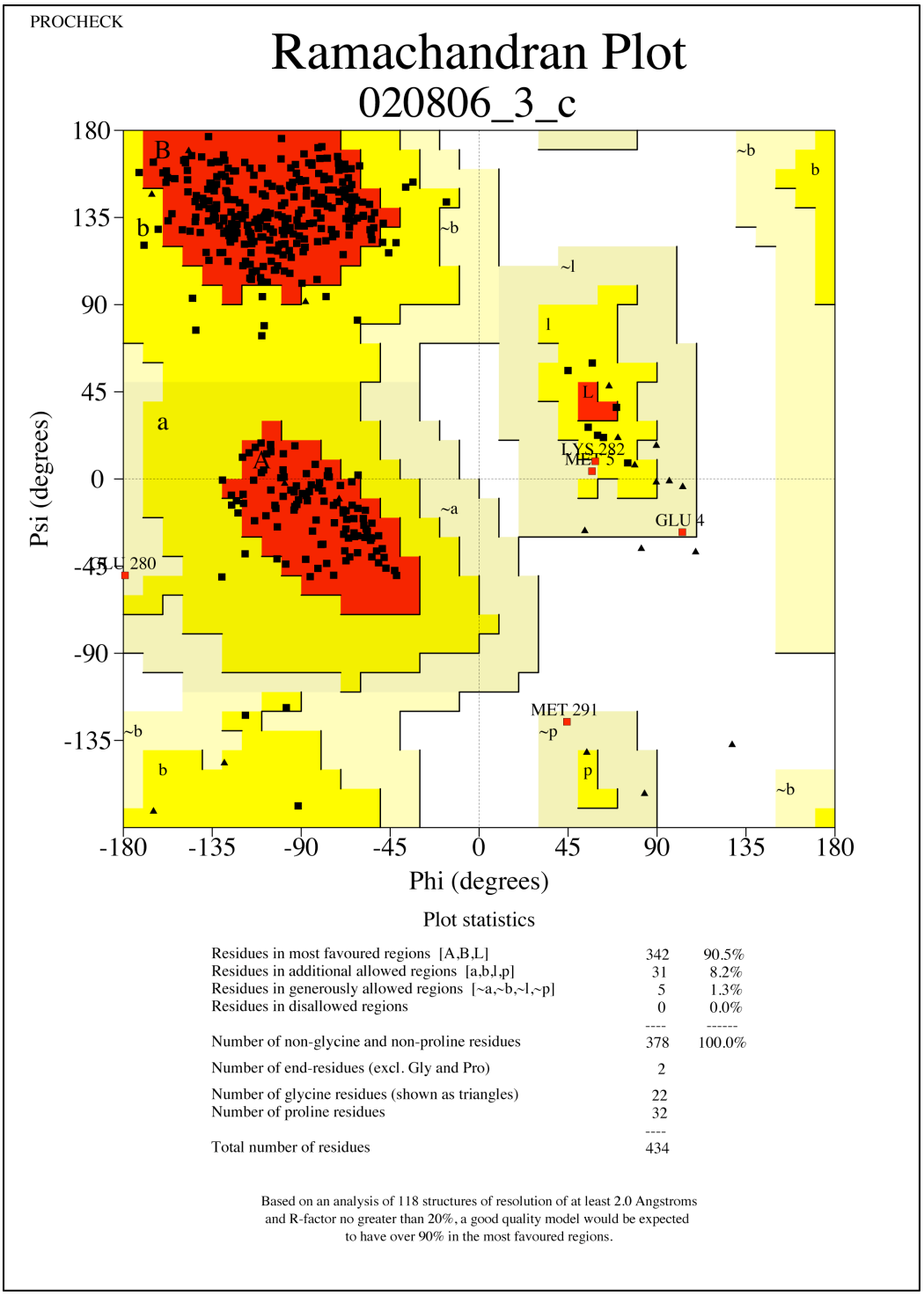


Figure 62: Stereo-view of a representative region of the experimental electron density map contoured at 2.0 σ (blue). The final Nup214 NTD model is shown in ball-and-stick format.

Table 6: Data collection and refinement statistics for the Nup214 NTD

	Crystal 1 native	Crystal 2 PCMB		Crystal 3 PCMB	Crystal 4 OsO_4
Data collection					
Space group	$P2_12_12_1$	$P2_12_12_1$		$P2_12_12_1$	$P2_12_12_1$
Cell dimensions					
a, b, c (Å)	*a* = 52.4, *b* = 81.1, *c* = 102.6	*a* = 52.4, *b* = 81.1, *c* = 102.6		*a* = 52.4, *b* = 81.1, *c* = 102.6	*a* = 52.4, *b* = 81.1, *c* = 102.6
α, β, γ (°)	$\alpha=\beta=\gamma=90°$	$\alpha=\beta=\gamma=90°$		$\alpha=\beta=\gamma=90°$	$\alpha=\beta=\gamma=90°$
		Peak	*Inflection*	*Peak*	*Peak*
Wavelength (Å)	1.0000	1.0089	1.0056	1.0080	1.1402
Resolution (Å)	20.0-1.65	20.0-1.95	20.0-1.88	20.0-1.95	20.0-1.97
R_{sym}	6.4 (84.0)	7.8 (93.6)	4.7 (64.2)	7.5 (66.8)	6.6 (60.8)
$I / \sigma I$	52.3 (2.8)	26.7 (2.3)	40.0 (2.1)	31.4 (2.5)	28.1 (2.6)
Completeness (%)	99.3 (93.9)	100.0 (100.0)	97.3 (72.7)	100.0 (100.0)	100.0 (100.0)
Redundancy	20.3	7.5	7.4	7.7	7.5
Refinement					
Resolution (Å)	20.0-1.65				
No. reflections total	49,618				
No. reflections test set	5,021 (9.4%)				
R_{work} / R_{free}	19.7 / 23.7				
No. atoms		*B*-factors			
Protein	3,372	Protein	37		
Ligand/ion	12	Ligand/Ion	87		
Water	522	Water	55		
R.m.s deviations					
Bond lengths (Å)	0.007				
Bond angles (°)	1.48				

*Highest-resolution shell is shown in parentheses.



Plot statistics

Residues in most favoured regions [A,B,L]	342	90.5%
Residues in additional allowed regions [a,b,l,p]	31	8.2%
Residues in generously allowed regions [~a,~b,~l,~p]	5	1.3%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	378	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	22	
Number of proline residues	32	
Total number of residues	434	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 63: Ramachandran blot of the Nup214 NTD structure.

4.4.2 Nup214 NTD•Ddx19 complex

4.4.2.1 Crystallization and Data collection

Crystals of human Nup214 NTD•Ddx19 and Nup214 NTD•Ddx19 NTD were grown at 21°C in hanging drops containing 1 μl of the protein complex (40 mg/ml), which contained 0.5 mM Mg-ADP and 1 μl of a reservoir solution. Crystals of Nup214 NTD•Ddx19 and Nup214 NTD•Ddx19 NTD grew in 0.1 M CHES, pH 9.5, 30% (w/v) PEG 3000, and in 0.1 M sodium acetate, pH 4.2-4.5, 0.8 M sodium di-hydrogen phosphate, and 1.2 M di-potassium hydrogen phosphate, respectively. For cryoprotection, Nup214 NTD•Ddx19 and Nup214 NTD•Ddx19 NTD crystals were stabilized and cryo-protected in 0.1 M CHES, pH 9.5, 30% (w/v) PEG 3000, supplemented with 10% (v/v) glycerol, and in 0.1 M sodium acetate, pH 4.2, 0.8 M sodium di-hydrogen phosphate, 1.2 M di-potassium phosphate, supplemented with 20% (v/v) glycerol, respectively, and flash frozen in liquid nitrogen-cooled liquid propane. X-ray diffraction data was collected at beamlines 8.2.1 and 8.2.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory. X-ray intensities were processed and integrated using the HKL2000 denzo/scalepack package[245], and the CCP4 program package was used for subsequent calculations[246].

4.4.2.2 Structure determination

The structure of the Nup214 NTD•Ddx19 was solved by molecular replacement using the coordinates of the Nup214 NTD[46] as a search model in PHASER[251]. The resulting electron density map was of high quality and clearly

revealed additional density for the N-terminal domain of Ddx19. A model was built with the program O[248] and Coot[252] and refined with CNS[249]. No electron density was observed for the N-terminal 8 and C-terminal 22 residues of Nup214, and the N-terminal 69 and C-terminal 181 residues that encompass the entire C-terminal RecA-like domain (domain 2) of Ddx19. These residues are presumed to be disordered, and, therefore, have been omitted from the final model. The final model contains residues 9 to 428 and 70 to 298 of Nup214 and Ddx19, respectively, and has been refined to 2.9 Å resolution with an R_{work} factor of 24.9% and an R_{free} factor of 28.3%.

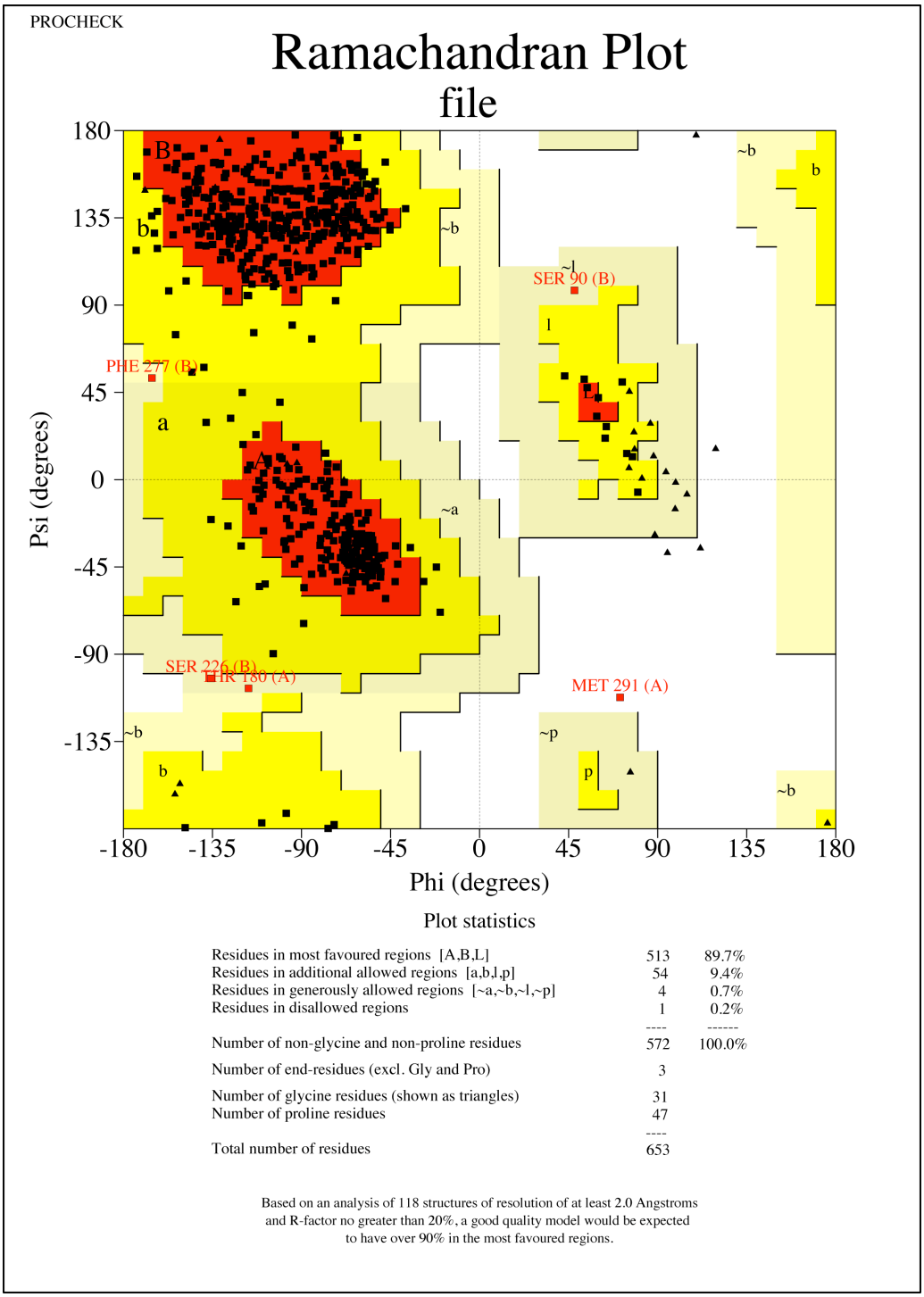
The structure of the Nup214 NTD•Ddx19 NTD complex was determined using the coordinates of the refined Nup214 NTD•Ddx19 complex as a search model in PHASER[251]. The final rounds of refinement were carried out with REFMAC using the TLS option[253]. The final model contains residues 8 to 428 and 69 to 300 of Nup214 and Ddx19, respectively, and has been refined to 2.5 Å resolution with an R_{work} factor of 19.9% and an R_{free} factor of 24.2%. The stereochemical quality of the final models was assessed with PROCHECK[250] and Molprobity[254]. Met291 in Nup214 is the only outlier in the Nup214 NTD•Ddx19 NTD structure (**Figure 64**). However, this residue is well defined in the electron density map and is located in a canonical β-turn type II'[255]. Data collection and refinement statistics are shown in the **Table 7**. The structure factors and atomic coordinates of the Nup214 NTD•Ddx19 and Nup214 NTD•Ddx19 NTD have

been deposited to the Protein Data Bank with the accession codes 3FMO and 3FMP.

Table 7: Data collection and refinement statistics for the Nup214 NTD•Ddx19 complex

	Crystal 1 Native	Crystal 2 Native
Contents		
Nup214	Nup214 NTD	Nup214 NTD
Ddx19	Ddx19, residues 1-300	Ddx19, residues 1-479
Data collection		
Synchrotron	ALS	ALS
Beamline	BL8.2.2	BL8.2.1
Space group	$P2_12_12_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	a=59.5, b=115.4, c=143.5	a=61.4, b=112.9, c=142.6
a, b, g (°)	$\alpha=\beta=\gamma=90$	$\alpha=\gamma=90$, $\beta=89.9$
Wavelength (Å)	1.0000	0.9795
Resolution (Å)	50.0-2.5	50.0-3.2
R_{sym} (%)	10.5 (68.7)	11.3 (39.5)
<*I* / s*I*>	25.0 (2.7)	14.3 (3.6)
Completeness (%)	98.6 (89.9)	91.3 (70.9)
Redundancy	13.8 (10.8)	6.8 (5.9)
Refinement		
Resolution (Å)	50.0-2.5	50.0-3.2
No. reflections		
total	29,350	28,125
test set	1,528 (4.9%)	1,504 (5.1%)
R_{work} / R_{free} (%)	19.9 / 24.2	24.9 / 28.3
No. atoms	5,189	10,218
Protein	5,124	10,164
Ligand/ion	33	54
Water	32	0
B-factors	59	86
Protein	58	86
Ligand/ion	60	89
Water	49	N.A.
R.m.s deviations		
Bond lengths (Å)	0.014	0.013
Bond angles (°)	1.5	1.5
Ramachandran Statistics		
Most favored (%)	89.7	82.1
Additionally allowed (%)	9.4	16.5
Generously allowed (%)	0.7	1.3
Disallowed (%)	0.2	0.0

*Highest-resolution shell is shown in parentheses.



file_01.ps

Figure 64: Ramachandran plot of the Nup214 NTD•Ddx19 NTD complex at 2.5 Å resolution.

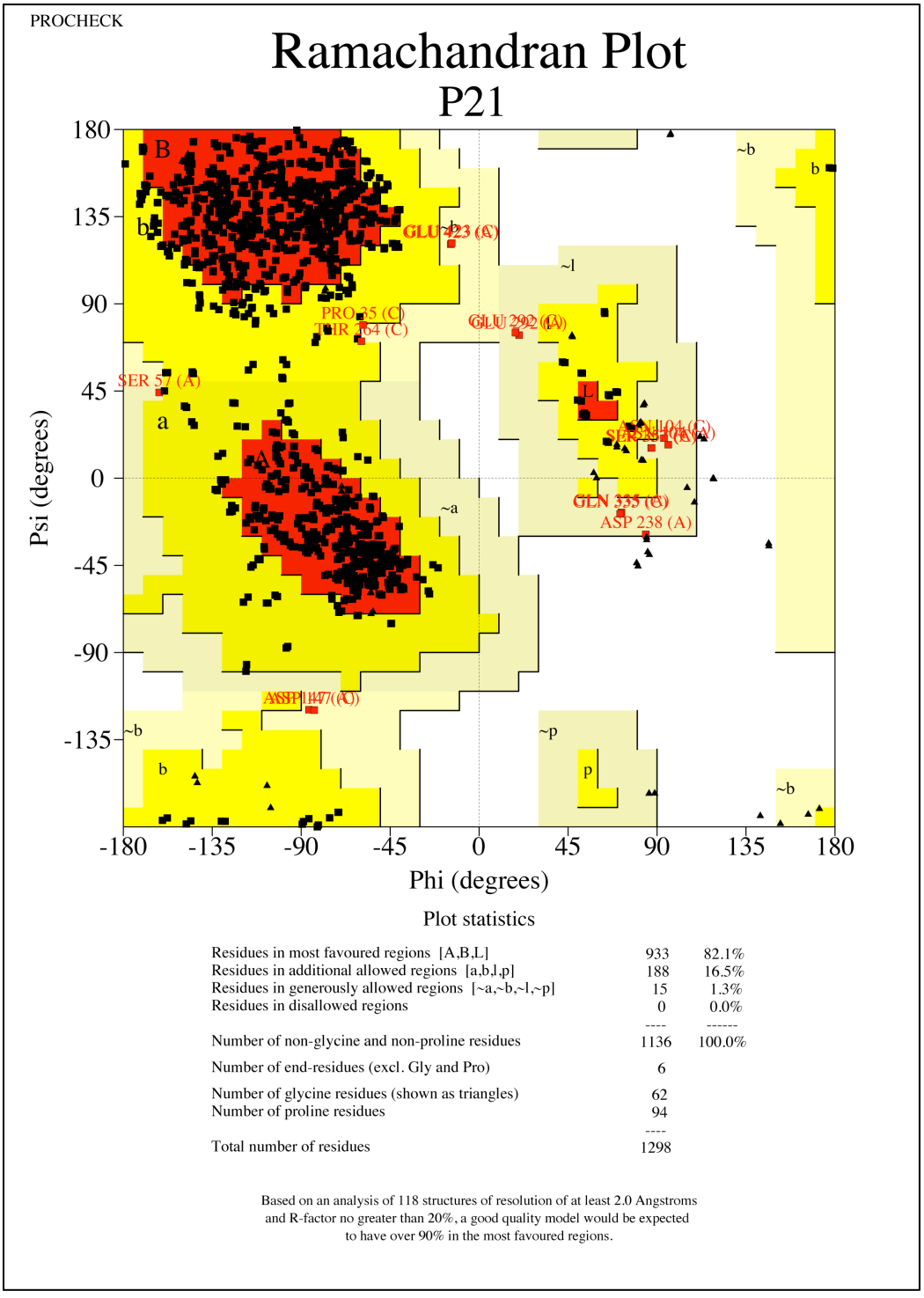


Figure 65: Ramachandran plot of the Nup214 NTD•Ddx19 complex structure at 3.2 Å.

4.5 Generation of a Ddx19 homology model.

Based on the structure of the *M. jannaschii* DEAD-box helicase MjDEAD (PDB code 1HV8)[219] and a sequence alignment of 10 eukaryotic Ddx19 homologues (Figure 48), we generated a homology model for human Ddx19 using MODELLER[256]. While the central core region that contains the ten conserved sequence motifs aligned without major sequence gaps, the Ddx19-specific N-terminal sequence is not conserved in MjDEAD.

4.6 Protein interaction analysis.

Protein interaction experiments were carried out on a Superdex 200 10/300 GL gel filtration column (GE Healthcare) that was equilibrated in a buffer containing 20 mM Tris, pH 8.0, 100 mM KCl, 10% (v/v) glycerol, and 5 mM DTT. Complexes were formed by incubating 1 mg of the various purified proteins for 30 min at 4°C. Complex formation was monitored by injection of the pre-incubated proteins as well as the recombinant purified proteins in isolation. All proteins were analyzed under the same buffer conditions, and complex formation was confirmed by SDS-PAGE of the eluted fractions followed by Coomassie brilliant blue staining.

4.7 Peptide binding assay.

A peptide array composed of 20-mers with an offset of three residues covering the entire sequence of human Ddx19 was synthesized and immobilized

onto a nitrocellulose membrane (Rockefeller University Proteomics Resource Center). The peptide array was blocked for 7 h in TBS-Tween buffer supplemented with 5% (w/v) milk powder and washed 3 times for 10 min in TBS-Tween buffer followed by a single wash step in transport buffer (20 mM Hepes-KOH, pH 7.5, 110 mM potassium acetate, 2 mM $MgCl_2$, 0.1% Tween-20). The peptide array was then incubated with the *in vitro* transcribed, translated, and [^{35}S]-methionine-labeled Nup214 NTD (Promega, TnT quick coupled Transcription/Translation system) in transport buffer at 4°C. Before detection by autoradiography, the membrane was washed 4 times for 10 min with transport buffer.

4.8 Immunofluorescence confocal microscopy.

For immunofluorescence, Ddx19, Nup214 NTD, and their mutants were cloned into the pCMV-HA vector (Clontech) containing a C-terminal HA-tag. The details of the Nup214 and Ddx19 constructs are listed in **Table 8.** Mammalian cell culture and transfection of HeLa cells were performed as described previously[257]. In short, HeLa cells were grown in DME medium (Invitrogen) supplemented with 10% FBS (GibcoBRL). For transient transfection and immunofluorescence microscopy cells were grown on coverslips in 12-well dishes and transfected with Lipofectamine 2000 (Invitrogen) following the manufacturers protocol. HeLa cells were washed in PBS and fixed for 20 min in 4% PFA (Wako) for Ddx19 variants or 2% formaldehyde for Nup214 variants at

room temperature (RT) and then washed twice with PBS again. Cells were then permeabilized with 0.3% Digitonin (Electron Microscope Sciences) or 0.2% Triton-X-100 and 1% BSA (Sigma) in PBS for 10 min at RT. Cells were blocked with 3% BSA in PBS for 20 min.

For Ddx19 variants, the cells were incubated with α-HA antibody (at a dilution of 1:2000, Abcam) for 1 h at RT. The cells were washed twice with PBS and were incubated with secondary antibody (goat α-rabbit Alexa Fluor 488, at a dilution of 1:2000, Invitrogen) for 1 h at RT. The cells were further incubated with α-mAb414 (at a dilution 1:1000, Covance) for 1 h at RT, washed twice with PBS, incubated with secondary antibodies (goat α-mouse Rodamine-Red X, at a dilution of 1:1000, Invitrogen) for 1 h at RT and mounted onto cover slips with ProLong Gold Antifade reagent (Invitrogen). Samples were examined on a Zeiss LSM5 EXCITER confocal microscope, and all images acquired by using an aplan-Apochromat 63 X with a 1.4-n.a. objective.

For Nup214 variants, the cells were co-stained with a mouse α-HA antibody (at a dilution of 1:1000, Covance) and a rabbit α-Nup358 antibody (at a dilution of 1:500)[76] for 1 h at RT. The cells were washed three times with PBS, incubated with secondary antibodies for 1 h at RT (Alexa Fluor 488 goat α-mouse IgG1, Alexa Fluor 647 goat α-rabbit IgG; both at a dilution of 1:200, Molecular Probes), washed three times with PBS, and mounted in Fluoromount G (Electron Microscopy Sciences). The samples were examined using a Leica spectral confocal microscope (model TCS SP).

Table 8: Immunofluorescence microscopy constructs

Protein	Residues	Vector	Restriction sites 5', 3'	Tag
Nup214	1-2091	pCMV	EcoRI, NotI	HA
Nup214 NTD	1-450	pCMV	EcoRI, NotI	HA
Nup214 NTD1-405	1-405	pCMV	EcoRI, NotI	HA
Nup214 ΔNTD	451-2091	pCMV	EcoRI, NotI	HA
Nup214 ΔFG	1-1209	pCMV	EcoRI, NotI	HA
Nup214 ΔNTDΔFG	451-1209	pCMV	EcoRI, NotI	HA
Ddx19 Full-length	1-479	pCMV	EcoRI, NotI	HA
Ddx19 R259A	1-479	pCMV	EcoRI, NotI	HA
Ddx19 R259K	1-479	pCMV	EcoRI, NotI	HA

4.9 Illustrations and figures.

Molecular graphics images were produced using PyMOL (www.pymol.org) and the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco[258]. Schematic diagrams of protein-ligand interactions were done with LIGPLOT[259]. The molecular surfaces were calculated using MSMS[260], and the electrostatic potential was calculated using APBS[261]. Sequence alignments were generated using ClustalX[262] and colored with Alscript[263]. Figures were generated using Adobe Illustrator.

References

1. Watson, M.L. Pores in the mammalian nuclear membrane. *Biochim Biophys Acta* **15**, 475-479 (1954).

2. Franke, W.W. & Scheer, U. The ultrastructure of the nuclear envelope of amphibian oocytes: a reinvestigation. I. The mature oocyte. *J Ultrastruct Res* **30**, 288-316 (1970).

3. Nicolini, C., Vernazza, G., Chiabrera, A., Maraldi, I.N. & Capitani, S. Nuclear pores and interphase chromatin: high-resolution image analysis and freeze etching. *J Cell Sci* **72**, 75-87 (1984).

4. Allen, T.D. *et al.* A protocol for isolating Xenopus oocyte nuclear envelope for visualization and characterization by scanning electron microscopy (SEM) or transmission electron microscopy (TEM). *Nat Protoc* **2**, 1166-1172 (2007).

5. Reichelt, R. *et al.* Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. *J Cell Biol* **110**, 883-894 (1990).

6. Rout, M.P. & Blobel, G. Isolation of the yeast nuclear pore complex. *J Cell Biol* **123**, 771-783 (1993).

7. Yang, Q., Rout, M.P. & Akey, C.W. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol Cell* **1**, 223-234 (1998).

8. Allen, T.D., Cronshaw, J.M., Bagley, S., Kiseleva, E. & Goldberg, M.W. The nuclear pore complex: mediator of translocation between nucleus and cytoplasm. *J Cell Sci* **113 (Pt 10)**, 1651-1659 (2000).

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

10. Gall, J.G. Octagonal nuclear pores. *J Cell Biol* **32**, 391-399 (1967).

11. Beck, M. *et al.* Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306**, 1387-1390 (2004).

12. Beck, M., Lucic, V., Forster, F., Baumeister, W. & Medalia, O. Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449**, 611-615 (2007).

13. Kiseleva, E., Goldberg, M.W., Allen, T.D. & Akey, C.W. Active nuclear pore complexes in Chironomus: visualization of transporter configurations related to mRNP export. *J Cell Sci* **111 (Pt 2)**, 223-236 (1998).

14. Fahrenkrog, B. & Aebi, U. The nuclear pore complex: nucleocytoplasmic transport and beyond. *Nat Rev Mol Cell Biol* **4**, 757-766 (2003).

15. Rout, M.P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol* **148**, 635-651 (2000).

16. Cronshaw, J.M., Krutchinsky, A.N., Zhang, W., Chait, B.T. & Matunis, M.J. Proteomic analysis of the mammalian nuclear pore complex. *J Cell Biol* **158**, 915-927 (2002).

17. Wente, S.R., Rout, M.P. & Blobel, G. A new family of yeast nuclear pore complex proteins. *J Cell Biol* **119**, 705-723 (1992).

18. Schwartz, T.U. Modularity within the architecture of the nuclear pore complex. *Curr Opin Struct Biol* **15**, 221-226 (2005).

19. Devos, D. *et al.* Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol* **2**, e380 (2004).

20. Vasu, S.K. & Forbes, D.J. Nuclear pores and nuclear assembly. *Curr Opin Cell Biol* **13**, 363-375 (2001).

21. Melcak, I., Hoelz, A. & Blobel, G. Structure of Nup58/45 suggests flexible nuclear pore diameter by intermolecular sliding. *Science* **315**, 1729-1732 (2007).

22. Kampmann, M. & Blobel, G. Three-dimensional structure and flexibility of a membrane-coating module of the nuclear pore complex. *Nat Struct Mol Biol* **16**, 782-788 (2009).

23. Nagy, V. *et al.* Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proc Natl Acad Sci U S A* **106**, 17693-17698 (2009).

24. Debler, E.W., Hsia, K.C., Nagy, V., Seo, H.S. & Hoelz, A. Characterization of the membrane-coating Nup84 complex. *Nucleus* **1**, 8 (2010).

25. Hsia, K.C., Stavropoulos, P., Blobel, G. & Hoelz, A. Architecture of a coat for the nuclear pore membrane. *Cell* **131**, 1313-1326 (2007).

26. Lenart, P. *et al.* Nuclear envelope breakdown in starfish oocytes proceeds by partial NPC disassembly followed by a rapidly spreading fenestration of nuclear membranes. *J Cell Biol* **160**, 1055-1068 (2003).

27. Dultz, E. *et al.* Systematic kinetic analysis of mitotic dis- and reassembly of the nuclear pore in living cells. *J Cell Biol* **180**, 857-865 (2008).

28. Lutzmann, M., Kunze, R., Buerer, A., Aebi, U. & Hurt, E. Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *EMBO J* **21**, 387-397 (2002).

29. Loiodice, I. *et al.* The entire Nup107-160 complex, including three new members, is targeted as one entity to kinetochores in mitosis. *Mol Biol Cell* **15**, 3333-3344 (2004).

30. Siniossoglou, S. *et al.* Structure and assembly of the Nup84p complex. *J Cell Biol* **149**, 41-54 (2000).

31. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683-694 (2007).

32. Debler, E.W. *et al.* A fence-like coat for the nuclear pore membrane. *Mol Cell* **32**, 815-826 (2008).

33. Brohawn, S.G., Partridge, J.R., Whittle, J.R. & Schwartz, T.U. The nuclear pore complex has entered the atomic age. *Structure* **17**, 1156-1168 (2009).

34. Berke, I.C., Boehmer, T., Blobel, G. & Schwartz, T.U. Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J Cell Biol* **167**, 591-597 (2004).

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

36. Hodel, A.E. *et al.* The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Mol Cell* **10**, 347-358 (2002).

37. Boehmer, T., Jeudy, S., Berke, I.C. & Schwartz, T.U. Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Mol Cell* **30**, 721-731 (2008).

38. Brohawn, S.G., Leksa, N.C., Spear, E.D., Rajashankar, K.R. & Schwartz, T.U. Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* **322**, 1369-1373 (2008).

39. Whittle, J.R. & Schwartz, T.U. Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element. *J Biol Chem* **284**, 28442-28452 (2009).

40. Leksa, N.C., Brohawn, S.G. & Schwartz, T.U. The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture. *Structure* **17**, 1082-1091 (2009).

41. Vetter, I.R., Nowak, C., Nishimoto, T., Kuhlmann, J. & Wittinghofer, A. Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* **398**, 39-46 (1999).

42. Schrader, N. *et al.* The crystal structure of the Ran-Nup153ZnF2 complex: a general Ran docking site at the nuclear pore complex. *Structure* **16**, 1116-1125 (2008).

43. Partridge, J.R. & Schwartz, T.U. Crystallographic and biochemical analysis of the Ran-binding zinc finger domain. *J Mol Biol* **391**, 375-389 (2009).

44. Handa, N. *et al.* The crystal structure of mouse Nup35 reveals atypical RNP motifs and novel homodimerization of the RRM domain. *J Mol Biol* **363**, 114-124 (2006).

45. Sun, Y. & Guo, H.C. Structural constraints on autoprocessing of the human nucleoporin Nup98. *Protein Sci* **17**, 494-505 (2008).

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

47. Drin, G. *et al.* A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat Struct Mol Biol* **14**, 138-146 (2007).

48. Siniossoglou, S. *et al.* A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. *Cell* **84**, 265-275 (1996).

49. Glavy, J.S. *et al.* Cell-cycle-dependent phosphorylation of the nuclear pore Nup107-160 subcomplex. *Proc Natl Acad Sci U S A* **104**, 3811-3816 (2007).

50. Belgareh, N. *et al.* An evolutionarily conserved NPC subcomplex, which redistributes in part to kinetochores in mammalian cells. *J Cell Biol* **154**, 1147-1160 (2001).

51. Seo, H.S. *et al.* Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc Natl Acad Sci U S A* **106**, 14281-14286 (2009).

52. Brohawn, S.G. & Schwartz, T.U. Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat Struct Mol Biol* **16**, 1173-1177 (2009).

53. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695-701 (2007).

54. Debler, E.W., Hsia, K.C., Nagy, V., Seo, H.S. & Hoelz, A. Characterization of the membrane-coating Nup84 complex. *Nucleus* **1**, 8 (2009).

55. Rexach, M. & Blobel, G. Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell* **83**, 683-692 (1995).

56. Lim, R.Y. *et al.* Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc Natl Acad Sci U S A* **103**, 9512-9517 (2006).

57. Denning, D.P., Patel, S.S., Uversky, V., Fink, A.L. & Rexach, M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc Natl Acad Sci U S A* **100**, 2450-2455 (2003).

58. Fribourg, S., Braun, I.C., Izaurralde, E. & Conti, E. 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**, 645-656 (2001).

59. Bayliss, R., Littlewood, T., Strawn, L.A., Wente, S.R. & Stewart, M. GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta. *J Biol Chem* **277**, 50597-50606 (2002).

60. Weis, K., Dingwall, C. & Lamond, A.I. Characterization of the nuclear protein import mechanism using Ran mutants with altered nucleotide binding specificities. *EMBO J* **15**, 7120-7128 (1996).

61. Nachury, M.V. & Weis, K. The direction of transport through the nuclear pore can be inverted. *Proc Natl Acad Sci U S A* **96**, 9622-9627 (1999).

62. Rout, M.P., Aitchison, J.D., Magnasco, M.O. & Chait, B.T. Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol* **13**, 622-628 (2003).

63. Ribbeck, K. & Gorlich, D. The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *EMBO J* **21**, 2664-2671 (2002).

64. Ribbeck, K. & Gorlich, D. Kinetic analysis of translocation through nuclear pore complexes. *EMBO J* **20**, 1320-1330 (2001).

65. Frey, S., Richter, R.P. & Gorlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815-817 (2006).

66. Frey, S. & Gorlich, D. A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* **130**, 512-523 (2007).

67. Patel, S.S. & Rexach, M.F. Discovering novel interactions at the nuclear pore complex using bead halo: a rapid method for detecting molecular interactions of high and low affinity at equilibrium. *Mol Cell Proteomics* **7**, 121-131 (2008).

68. Patel, S.S., Belmont, B.J., Sante, J.M. & Rexach, M.F. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* **129**, 83-96 (2007).

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

70. Chook, Y.M. & Blobel, G. Karyopherins and nuclear import. *Curr Opin Struct Biol* **11**, 703-715 (2001).

71. Wen, W., Meinkoth, J.L., Tsien, R.Y. & Taylor, S.S. Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82**, 463-473 (1995).

72. Mattaj, I.W. & Englmeier, L. Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem* **67**, 265-306 (1998).

73. Nemergut, M.E., Mizzen, C.A., Stukenberg, T., Allis, C.D. & Macara, I.G. Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B. *Science* **292**, 1540-1543 (2001).

74. Pu, R.T. & Dasso, M. The balance of RanBP1 and RCC1 is critical for nuclear assembly and nuclear transport. *Mol Biol Cell* **8**, 1955-1970 (1997).

75. Bischoff, F.R., Krebber, H., Smirnova, E., Dong, W. & Ponstingl, H. Co-activation of RanGTPase and inhibition of GTP dissociation by Ran-GTP binding protein RanBP1. *EMBO J* **14**, 705-715 (1995).

76. Wu, J., Matunis, M.J., Kraemer, D., Blobel, G. & Coutavas, E. Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J Biol Chem* **270**, 14209-14213 (1995).

77. Matunis, M.J., Coutavas, E. & Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol* **135**, 1457-1470 (1996).

78. Delphin, C., Guan, T., Melchior, F. & Gerace, L. RanGTP targets p97 to RanBP2, a filamentous protein localized at the cytoplasmic periphery of the nuclear pore complex. *Mol Biol Cell* **8**, 2379-2390 (1997).

79. Mahajan, R., Delphin, C., Guan, T., Gerace, L. & Melchior, F. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**, 97-107 (1997).

80. Wilken, N., Senecal, J.L., Scheer, U. & Dabauvalle, M.C. Localization of the Ran-GTP binding protein RanBP2 at the cytoplasmic side of the nuclear pore complex. *Eur J Cell Biol* **68**, 211-219 (1995).

81. Moroianu, J., Blobel, G. & Radu, A. Nuclear protein import: Ran-GTP dissociates the karyopherin alphabeta heterodimer by displacing alpha from an overlapping binding site on beta. *Proc Natl Acad Sci U S A* **93**, 7059-7062 (1996).

82. Floer, M., Blobel, G. & Rexach, M. Disassembly of RanGTP-karyopherin beta complex, an intermediate in nuclear protein import. *J Biol Chem* **272**, 19538-19546 (1997).

83. Mosammaparast, N. & Pemberton, L.F. Karyopherins: from nuclear-transport mediators to nuclear-function regulators. *Trends Cell Biol* **14**, 547-556 (2004).

84. Cingolani, G., Petosa, C., Weis, K. & Muller, C.W. Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature* **399**, 221-229 (1999).

85. Cingolani, G., Bednenko, J., Gillespie, M.T. & Gerace, L. Molecular basis for the recognition of a nonclassical nuclear localization signal by importin beta. *Mol Cell* **10**, 1345-1353 (2002).

86. Lee, S.J. *et al.* The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor. *Science* **302**, 1571-1575 (2003).

87. Hoelz, A. & Blobel, G. Cell biology: popping out of the nucleus. *Nature* **432**, 815-816 (2004).

88. Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell* **94**, 193-204 (1998).

89. Groves, M.R. & Barford, D. Topological characteristics of helical repeat proteins. *Curr Opin Struct Biol* **9**, 383-389 (1999).

90. Fontes, M.R., Teh, T. & Kobe, B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *J Mol Biol* **297**, 1183-1194 (2000).

91. Cook, A. *et al.* The structure of the nuclear export receptor Cse1 in its cytosolic state reveals a closed conformation incompatible with cargo binding. *Mol Cell* **18**, 355-367 (2005).

92. Matsuura, Y. & Stewart, M. Structural basis for the assembly of a nuclear export complex. *Nature* **432**, 872-877 (2004).

93. Monecke, T. *et al.* Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* **324**, 1087-1091 (2009).

94. Dong, X. *et al.* Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* **458**, 1136-1141 (2009).

95. Wilusz, J.E. & Spector, D.L. An unexpected ending: noncanonical 3' end processing mechanisms. *RNA* **16**, 259-266.

96. Kutay, U. *et al.* Identification of a tRNA-specific nuclear export receptor. *Mol Cell* **1**, 359-369 (1998).

97. Arts, G.J., Fornerod, M. & Mattaj, I.W. Identification of a nuclear export receptor for tRNA. *Curr Biol* **8**, 305-314 (1998).

98. Arts, G.J., Kuersten, S., Romby, P., Ehresmann, B. & Mattaj, I.W. The role of exportin-t in selective nuclear export of mature tRNAs. *EMBO J* **17**, 7430-7441 (1998).

99. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297 (2004).

100. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-3027 (2004).

101. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419 (2003).

102. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957-1966 (2004).

103. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-4060 (2004).

104. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95-98 (2004).

105. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366 (2001).

106. Medlin, J.E., Uguen, P., Taylor, A., Bentley, D.L. & Murphy, S. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J* **22**, 925-934 (2003).

107. Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I.W. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187-198 (2000).

108. Izaurralde, E. *et al.* A cap-binding protein complex mediating U snRNA export. *Nature* **376**, 709-712 (1995).

109. Huber, J. *et al.* Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J* **17**, 4114-4126 (1998).

110. Rino, J. & Carmo-Fonseca, M. The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends Cell Biol* **19**, 375-384 (2009).

111. Dreyfuss, G., Kim, V.N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3**, 195-205 (2002).

112. Miralles, F. *et al.* Electron tomography reveals posttranscriptional binding of pre-mRNPs to specific fibers in the nucleoplasm. *J Cell Biol* **148**, 271-282 (2000).

113. Lonnroth, A. *et al.* Demonstration of a 7-nm RNP fiber as the basic structural element in a premessenger RNP particle. *Exp Cell Res* **199**, 292-296 (1992).

114. Stevens, B.J. & Swift, H. RNA transport from nucleus to cytoplasm in Chironomus salivary glands. *J Cell Biol* **31**, 55-77 (1966).

115. Mehlin, H., Daneholt, B. & Skoglund, U. Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. *Cell* **69**, 605-613 (1992).

116. Daneholt, B. Packing and delivery of a genetic message. *Chromosoma* **110**, 173-185 (2001).

117. Shatkin, A.J. & Manley, J.L. The ends of the affair: capping and polyadenylation. *Nat Struct Biol* **7**, 838-842 (2000).

118. Cheng, H. *et al.* Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389-1400 (2006).

119. Izaurralde, E. *et al.* A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657-668 (1994).

120. Berget, S.M., Moore, C. & Sharp, P.A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74**, 3171-3175 (1977).

121. Jurica, M.S. Detailed close-ups and the big picture of spliceosomes. *Curr Opin Struct Biol* **18**, 315-320 (2008).

122. Kataoka, N. *et al.* Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* **6**, 673-682 (2000).

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

124. Proudfoot, N. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **16**, 272-278 (2004).

125. Braun, I.C., Rohrbach, E., Schmitt, C. & Izaurralde, E. 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**, 1953-1965 (1999).

126. Bachi, A. *et al.* The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *RNA* **6**, 136-158 (2000).

127. Grant, R.P., Neuhaus, D. & Stewart, M. Structural basis for the interaction between the Tap/NXF1 UBA domain and FG nucleoporins at 1A resolution. *J Mol Biol* **326**, 849-858 (2003).

128. Varani, G. & Nagai, K. RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* **27**, 407-445 (1998).

129. Bray, M. *et al.* A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc Natl Acad Sci U S A* **91**, 1256-1260 (1994).

130. Reed, R. & Hurt, E. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523-531 (2002).

131. Bruhn, L., Munnerlyn, A. & Grosschedl, R. ALY, a context-dependent coactivator of LEF-1 and AML-1, is required for TCRalpha enhancer function. *Genes Dev* **11**, 640-653 (1997).

132. Strasser, K. & Hurt, E. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J* **19**, 410-420 (2000).

133. Zenklusen, D., Vinciguerra, P., Strahm, Y. & Stutz, F. The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Mol Cell Biol* **21**, 4219-4232 (2001).

134. Visa, N. *et al.* A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. *Cell* **84**, 253-264 (1996).

135. Zahler, A.M., Lane, W.S., Stolk, J.A. & Roth, M.B. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev* **6**, 837-847 (1992).

136. Huang, Y., Gattoni, R., Stevenin, J. & Steitz, J.A. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell* **11**, 837-843 (2003).

137. Huang, Y. & Steitz, J.A. SRprises along a messenger's journey. *Mol Cell* **17**, 613-615 (2005).

138. Stewart, M. Ratcheting mRNA out of the nucleus. *Mol Cell* **25**, 327-330 (2007).

139. Goldstein, A.L., Snay, C.A., Heath, C.V. & Cole, C.N. Pleiotropic nuclear defects associated with a conditional allele of the novel nucleoporin Rat9p/Nup85p. *Mol Biol Cell* **7**, 917-934 (1996).

140. Tseng, S.S. *et al.* Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. *Embo J* **17**, 2651-2662 (1998).

141. Del Priore, V., Snay, C.A., Bahr, A. & Cole, C.N. The product of the Saccharomyces cerevisiae RSS1 gene, identified as a high-copy suppressor of the rat7-1 temperature-sensitive allele of the RAT7/NUP159 nucleoporin, is required for efficient mRNA export. *Mol Biol Cell* **7**, 1601-1621 (1996).

142. Kohler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* **8**, 761-773 (2007).

143. Fasken, M.B. & Corbett, A.H. Process or perish: quality control in mRNA biogenesis. *Nat Struct Mol Biol* **12**, 482-488 (2005).

144. Green, D.M., Johnson, C.P., Hagan, H. & Corbett, A.H. The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. *Proc Natl Acad Sci U S A* **100**, 1010-1015 (2003).

145. Galy, V. *et al.* Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* **116**, 63-73 (2004).

146. Kraemer, D., Wozniak, R.W., Blobel, G. & Radu, A. 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**, 1519-1523 (1994).

147. Griffis, E.R., Xu, S. & Powers, M.A. Nup98 localizes to both nuclear and cytoplasmic sides of the nuclear pore and binds to two distinct nucleoporin subcomplexes. *Mol Biol Cell* **14**, 600-610 (2003).

148. Wu, X. *et al.* Disruption of the FG nucleoporin NUP98 causes selective changes in nuclear pore complex stoichiometry and function. *Proc Natl Acad Sci U S A* **98**, 3191-3196 (2001).

149. Bernad, R., van der Velde, H., Fornerod, M. & Pickersgill, H. Nup358/RanBP2 attaches to the nuclear pore complex via association with Nup88 and Nup214/CAN and plays a supporting role in CRM1-mediated nuclear protein export. *Mol Cell Biol* **24**, 2373-2384 (2004).

150. Schmitt, C. *et al.* 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**, 4332-4347 (1999).

151. Chang, T.H., Arenas, J. & Abelson, J. Identification of five putative yeast RNA helicase genes. *Proc Natl Acad Sci U S A* **87**, 1571-1575 (1990).

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

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

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

155. Caruthers, J.M., Johnson, E.R. & McKay, D.B. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A* **97**, 13080-13085 (2000).

156. Rogers, G.W., Jr., Komar, A.A. & Merrick, W.C. eIF4A: the godfather of the DEAD box helicases. *Prog Nucleic Acid Res Mol Biol* **72**, 307-331 (2002).

157. Bizebard, T., Ferlenghi, I., Iost, I. & Dreyfus, M. Studies on three E. coli DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry* **43**, 7857-7866 (2004).

158. Rogers, G.W., Jr., Lima, W.F. & Merrick, W.C. Further characterization of the helicase activity of eIF4A. Substrate specificity. *J Biol Chem* **276**, 12598-12608 (2001).

159. Murphy, R. & Wente, S.R. An RNA-export mediator with an essential nuclear export signal. *Nature* **383**, 357-360 (1996).

160. Strahm, Y. *et al.* The RNA export factor Gle1p is located on the cytoplasmic fibrils of the NPC and physically interacts with the FG-nucleoporin Rip1p, the DEAD-box protein Rat8p/Dbp5p and a new protein Ymr 255p. *Embo J* **18**, 5761-5777 (1999).

161. Watkins, J.L., Murphy, R., Emtage, J.L. & Wente, S.R. The human homologue of Saccharomyces cerevisiae Gle1p is required for poly(A)+ RNA export. *Proc Natl Acad Sci U S A* **95**, 6779-6784 (1998).

162. Kendirgi, F., Barry, D.M., Griffis, E.R., Powers, M.A. & Wente, S.R. An essential role for hGle1 nucleocytoplasmic shuttling in mRNA export. *J Cell Biol* **160**, 1029-1040 (2003).

163. Kendirgi, F., Rexer, D.J., Alcazar-Roman, A.R., Onishko, H.M. & Wente, S.R. Interaction between the shuttling mRNA export factor Gle1 and the nucleoporin hCG1: a conserved mechanism in the export of Hsp70 mRNA. *Mol Biol Cell* **16**, 4304-4315 (2005).

164. Weirich, C.S. *et al.* Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. *Nat Cell Biol* **8**, 668-676 (2006).

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

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

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

168. Liu, F., Putnam, A. & Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A* **105**, 20209-20214 (2008).

169. Martinez, N., Alonso, A., Moragues, M.D., Ponton, J. & Schneider, J. The nuclear pore complex protein Nup88 is overexpressed in tumor cells. *Cancer Res* **59**, 5408-5411 (1999).

170. Gould, V.E. *et al.* Nup88 (karyoporin) in human malignant neoplasms and dysplasias: correlations of immunostaining of tissue sections, cytologic smears, and immunoblot analysis. *Hum Pathol* **33**, 536-544 (2002).

171. Gould, V.E., Martinez, N., Orucevic, A., Schneider, J. & Alonso, A. A novel, nuclear pore-associated, widely distributed molecule overexpressed in oncogenesis and development. *Am J Pathol* **157**, 1605-1613 (2000).

172. Agudo, D. *et al.* Nup88 mRNA overexpression is associated with high aggressiveness of breast cancer. *Int J Cancer* **109**, 717-720 (2004).

173. Knoess, M. *et al.* Nucleoporin 88 expression in hepatitis B and C virus-related liver diseases. *World J Gastroenterol* **12**, 5870-5874 (2006).

174. Zhang, Z.Y. *et al.* Nup88 expression in normal mucosa, adenoma, primary adenocarcinoma and lymph node metastasis in the colorectum. *Tumour Biol* **28**, 93-99 (2007).

175. Xu, S. & Powers, M.A. Nuclear pore proteins and cancer. *Semin Cell Dev Biol* **20**, 620-630 (2009).

176. Krull, S., Thyberg, J., Bjorkroth, B., Rackwitz, H.R. & Cordes, V.C. Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. *Mol Biol Cell* **15**, 4261-4277 (2004).

177. Cooper, C.S. *et al.* Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature* **311**, 29-33 (1984).

178. Soman, N.R., Correa, P., Ruiz, B.A. & Wogan, G.N. The TPR-MET oncogenic rearrangement is present and expressed in human gastric carcinoma and precursor lesions. *Proc Natl Acad Sci U S A* **88**, 4892-4896 (1991).

179. Yu, J. *et al.* Frequency of TPR-MET rearrangement in patients with gastric carcinoma and in first-degree relatives. *Cancer* **88**, 1801-1806 (2000).

180. Pierotti, M.A. & Greco, A. Oncogenic rearrangements of the NTRK1/NGF receptor. *Cancer Lett* **232**, 90-98 (2006).

181. Yokoyama, N. *et al.* A giant nucleopore protein that binds Ran/TC4. *Nature* **376**, 184-188 (1995).

182. Ma, Z. *et al.* Fusion of ALK to the Ran-binding protein 2 (RANBP2) gene in inflammatory myofibroblastic tumor. *Genes Chromosomes Cancer* **37**, 98-105 (2003).

183. ten Dijke, P. *et al.* Activin receptor-like kinases: a novel subclass of cell-surface receptors with predicted serine/threonine kinase activity. *Oncogene* **8**, 2879-2887 (1993).

184. Fontoura, B.M., Blobel, G. & Matunis, M.J. A conserved biogenesis pathway for nucleoporins: proteolytic processing of a 186-kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. *J Cell Biol* **144**, 1097-1112 (1999).

185. Nakamura, T. *et al.* Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nat Genet* **12**, 154-158 (1996).

186. Borrow, J. *et al.* The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nat Genet* **12**, 159-167 (1996).

187. Kasper, L.H. *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol Cell Biol* **19**, 764-776 (1999).

188. Sharkey, M., Graba, Y. & Scott, M.P. Hox genes in evolution: protein surfaces and paralog groups. *Trends Genet* **13**, 145-151 (1997).

189. Argiropoulos, B. & Humphries, R.K. Hox genes in hematopoiesis and leukemogenesis. *Oncogene* **26**, 6766-6776 (2007).

190. von Lindern, M. *et al.* The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA. *Mol Cell Biol* **12**, 1687-1697 (1992).

191. van Deursen, J., Boer, J., Kasper, L. & Grosveld, G. G2 arrest and impaired nucleocytoplasmic transport in mouse embryos lacking the proto-oncogene CAN/Nup214. *EMBO J* **15**, 5574-5583 (1996).

192. von Lindern, M. *et al.* Can, a putative oncogene associated with myeloid leukemogenesis, may be activated by fusion of its 3' half to different

genes: characterization of the set gene. *Mol Cell Biol* **12**, 3346-3355 (1992).

193. Kappes, F., Burger, K., Baack, M., Fackelmayer, F.O. & Gruss, C. Subcellular localization of the human proto-oncogene protein DEK. *J Biol Chem* **276**, 26317-26323 (2001).

194. Alexiadis, V. *et al.* The protein encoded by the proto-oncogene DEK changes the topology of chromatin and reduces the efficiency of DNA replication in a chromatin-specific manner. *Genes Dev* **14**, 1308-1312 (2000).

195. Sawatsubashi, S. *et al.* A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. *Genes Dev* **24**, 159-170 (2010).

196. Gamble, M.J., Erdjument-Bromage, H., Tempst, P., Freedman, L.P. & Fisher, R.P. The histone chaperone TAF-I/SET/INHAT is required for transcription in vitro of chromatin templates. *Mol Cell Biol* **25**, 797-807 (2005).

197. Seo, S.B. *et al.* Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell* **104**, 119-130 (2001).

198. Nowak, S.J., Pai, C.Y. & Corces, V.G. Protein phosphatase 2A activity affects histone H3 phosphorylation and transcription in Drosophila melanogaster. *Mol Cell Biol* **23**, 6129-6138 (2003).

199. Gamble, M.J. & Fisher, R.P. SET and PARP1 remove DEK from chromatin to permit access by the transcription machinery. *Nat Struct Mol Biol* **14**, 548-555 (2007).

200. Van Vlierberghe, P. *et al.* The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood* **111**, 4668-4680 (2008).

201. Graux, C. *et al.* Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. *Nat Genet* **36**, 1084-1089 (2004).

202. De Keersmaecker, K. *et al.* Kinase activation and transformation by NUP214-ABL1 is dependent on the context of the nuclear pore. *Mol Cell* **31**, 134-142 (2008).

203. Fornerod, M. *et al.* The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. *Embo J* **16**, 807-816 (1997).

204. Katahira, J. *et al.* The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *Embo J* **18**, 2593-2609 (1999).

205. Vellieux, F.M. *et al.* Structure of quinoprotein methylamine dehydrogenase at 2.25 A resolution. *Embo J* **8**, 2171-2178 (1989).

206. Paoli, M. Protein folds propelled by diversity. *Prog Biophys Mol Biol* **76**, 103-130 (2001).

207. Wall, M.A. *et al.* The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* **83**, 1047-1058 (1995).

208. Fulop, V. & Jones, D.T. Beta propellers: structural rigidity and functional diversity. *Curr Opin Struct Biol* **9**, 715-721 (1999).

209. Sondek, J., Bohm, A., Lambright, D.G., Hamm, H.E. & Sigler, P.B. Crystal structure of a G-protein beta gamma dimer at 2.1A resolution. *Nature* **379**, 369-374 (1996).

210. Appleton, B.A., Wu, P. & Wiesmann, C. The crystal structure of murine coronin-1: a regulator of actin cytoskeletal dynamics in lymphocytes. *Structure* **14**, 87-96 (2006).

211. Gatfield, J., Albrecht, I., Zanolari, B., Steinmetz, M.O. & Pieters, J. Association of the leukocyte plasma membrane with the actin cytoskeleton through coiled coil-mediated trimeric coronin 1 molecules. *Mol Biol Cell* **16**, 2786-2798 (2005).

212. Beausoleil, S.A. *et al.* Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **101**, 12130-12135 (2004).

213. Cerna, D. & Wilson, D.K. The structure of Sif2p, a WD repeat protein functioning in the SET3 corepressor complex. *J Mol Biol* **351**, 923-935 (2005).

214. Bastos, R., Ribas de Pouplana, L., Enarson, M., Bodoor, K. & Burke, B. Nup84, a novel nucleoporin that is associated with CAN/Nup214 on the cytoplasmic face of the nuclear pore complex. *J Cell Biol* **137**, 989-1000 (1997).

215. Favreau, C., Worman, H.J., Wozniak, R.W., Frappier, T. & Courvalin, J.C. Cell cycle-dependent phosphorylation of nucleoporins and nuclear pore membrane protein Gp210. *Biochemistry* **35**, 8035-8044 (1996).

216. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**, 1351-1362 (1999).

217. Story, R.M. & Steitz, T.A. Structure of the recA protein-ADP complex. *Nature* **355**, 374-376 (1992).

218. Singleton, M.R., Dillingham, M.S. & Wigley, D.B. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**, 23-50 (2007).

219. Story, R.M., Li, H. & Abelson, J.N. Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. *Proc Natl Acad Sci U S A* **98**, 1465-1470 (2001).

220. Shi, H., Cordin, O., Minder, C.M., Linder, P. & Xu, R.M. Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc Natl Acad Sci U S A* **101**, 17628-17633 (2004).

221. Debler, E.W., Ma, Y., Seo, H.S., Hsia, K.C., Noriega, T.R., Blobel, G., Hoelz, A A Fence-Like Coat for the Nuclear Pore Membrane. *Mol. Cell* **in press** (2008).

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

223. Devos, D. *et al.* Simple fold composition and modular architecture of the nuclear pore complex. *Proc Natl Acad Sci U S A* **103**, 2172-2177 (2006).

224. von Moeller, H., Basquin, C. & Conti, E. The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat Struct Mol Biol* **16**, 247-254 (2009).

225. Collins, R. *et al.* The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J Biol Chem* **284**, 10296-10300 (2009).

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

227. Napetschnig, J. *et al.* 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**, 3089-3094 (2009).

228. Hilbert, M., Karow, A.R. & Klostermeier, D. The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol Chem* **390**, 1237-1250 (2009).

229. Dever, T.E. Translation initiation: adept at adapting. *Trends Biochem Sci* **24**, 398-403 (1999).

230. Hinton, T.M., Coldwell, M.J., Carpenter, G.A., Morley, S.J. & Pain, V.M. Functional analysis of individual binding activities of the scaffold protein eIF4G. *J Biol Chem* **282**, 1695-1708 (2007).

231. Schutz, P. *et al.* Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A* **105**, 9564-9569 (2008).

232. Dossani, Z.Y., Weirich, C.S., Erzberger, J.P., Berger, J.M. & Weis, K. Structure of the C-terminus of the mRNA export factor Dbp5 reveals the interaction surface for the ATPase activator Gle1. *Proc Natl Acad Sci U S A* **106**, 16251-16256 (2009).

233. Bolger, T.A., Folkmann, A.W., Tran, E.J. & Wente, S.R. The mRNA export factor Gle1 and inositol hexakisphosphate regulate distinct stages of translation. *Cell* **134**, 624-633 (2008).

234. Gross, T. *et al.* The DEAD-box RNA helicase Dbp5 functions in translation termination. *Science* **315**, 646-649 (2007).

235. Daneholt, B. Assembly and transport of a premessenger RNP particle. *Proc Natl Acad Sci U S A* **98**, 7012-7017 (2001).

236. Crick, F.H. On protein synthesis. *Symp Soc Exp Biol* **12**, 138-163 (1958).

237. Cole, C.N. & Scarcelli, J.J. Transport of messenger RNA from the nucleus to the cytoplasm. *Curr Opin Cell Biol* **18**, 299-306 (2006).

238. Cramer, P. *et al.* Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37**, 337-352 (2008).

239. Boeger, H. *et al.* Structural basis of eukaryotic gene transcription. *FEBS Lett* **579**, 899-903 (2005).

240. Schmeing, T.M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234-1242 (2009).

241. Steitz, T.A. A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* **9**, 242-253 (2008).

242. Strasser, K. *et al.* TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304-308 (2002).

243. Jovanovic-Talisman, T. *et al.* Artificial nanopores that mimic the transport selectivity of the nuclear pore complex. *Nature* **457**, 1023-1027 (2009).

244. Hoelz, A., Nairn, A.C. & Kuriyan, J. Crystal structure of a tetradecameric assembly of the association domain of Ca2+/calmodulin-dependent kinase II. *Mol Cell* **11**, 1241-1251 (2003).

245. Otwinowski, Z. & Minor, W. Processing of X-Ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326 (1997).

246. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763 (1994).

247. Fortelle, E.d.l.a.G.B. Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. . *Methods in Enzymology* **276**, 472-494 (1997).

248. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2)**, 110-119 (1991).

249. Brunger, A.T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921 (1998).

250. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. . *J. Appl. Crystallogr* **26** (1993).

251. McCoy, A.J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41 (2007).

252. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).

253. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. & Dodson, E.J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* **55**, 247-255 (1999).

254. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-383 (2007).

255. Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**, 167-339 (1981).

256. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit 5 6 (2006).

257. Wong, R.W., Blobel, G. & Coutavas, E. Rae1 interaction with NuMA is required for bipolar spindle formation. *Proc Natl Acad Sci U S A* **103**, 19783-19787 (2006).

258. Pettersen, E.F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).

259. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**, 127-134 (1995).

260. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041 (2001).

261. Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. & Gibson, T.J. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23**, 403-405 (1998).

262. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876-4882 (1997).

263. Barton, G.J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* **6**, 37-40 (1993).