Chapter 1: Introduction and overview

1.1 The nuclear pore complex and the nuclear basket

The nuclear pore complex (NPC; see Figure 1) is one of the largest ordered protein assemblies in the eukaryotic cell. NPCs are embedded in the nuclear envelope (NE), which confines the majority of cellular DNA. The NE is a lipid bilayer that is continuous with the endoplasmic reticulum on the cytoplasmic side and bounded on the nuclear side by the lamina network in higher eukaryotes. Lamina-free holes formed by the fusion of the outer and inner NE membranes represent a continous connection between cytoplasm and nucleoplasm. These holes are established by the assembly of various proteins, called nucleoporins or NUPs (Davis & Blobel, 1986), which in turn recruit yet other NUPs. Altogether, these NUPs, representing more than 30 different proteins (Rout *et al*., 2000; Cronshaw *et al*., 2002), form the NPCs in a highly modular manner, with most of the NUPs being part of different subcomplexes (see further below).

Pores in the NE were already proposed and discussed long before their undeniable visualisation (Hertwig, 1876; Flemming, 1882). NPCs were first visualised in the 1950s using electron microscopy and were referred to as “pores” or “annuli” in the initial descriptions of the ultrastructure of the NE (e.g., Callan & Tomlin, 1950; Watson, 1954; Bahr & Beermann, 1954; Gall, 1954; Watson, 1955; Afzelius, 1955; Rebhun, 1956; Wischnitzer, 1958). The term “pore complex” was introduced together with a first description of a larger compartment that extends from the NE-perforating pore into the cytoplasm and also appears to form intranuclear channels (Watson, 1959).

These pore complexes are usually arranged in an octagonal symmetry (Franke, 1966; Gall, 1967; see also, e.g., Löschberger *et al*., 2012; Göttfert *et al*., 2013) with copy numbers that in

most cases represent multiples of eight and could sum up to over 1000 individual proteins with an approximate mass of 120 MDa per NPC in vertebrates (Krohne *et al*., 1978; Reichelt *et al*., 1990) and about 500 proteins per NPC with a mass of roughly 66 MDa in *Saccharomyces cerevisiae* (Rout & Blobel, 1993; Yang *et al*., 1998)

The individual NUPs can be classified, based on their localisation at the NPC and their contribution to the NPC function, into transmembrane NUPs, scaffold NUPs, and phenylalanine-glycine (FG)-rich NUPs (see also Figure 1). The nucleoporins of the first group are embedded in the NE via their transmembrane domains (TMEM) and are involved in anchoring the NPC and maintaining membrane curvature (e.g., Gerace *et al*., 1982; Hallberg *et al*., 1993; Mansfeld *et al*., 2006; Stavru *et al*., 2006a, 2006b; Fujitomo *et al*., 2012; Tang *et al*., 2020).

The second group of nucleoporins constitutes the core scaffold of the NPC, which consists of three major concentric assemblies, namely the cytoplasmic ring (CR), the inner ring (IR), and the nuclear ring (NR; see also Figure 1). The IR is constituted by multiple copies of the NUP93 complex (e.g., Kosinski *et al*., 2016) and NUP155 provides the linkage to the two outer rings (e.g., von Appen *et al*., 2015). The outer rings are in a double ring arrangement, slightly offset relative to each other, with each ring consisting of eight NUP107 complexes, also known as Y-complexes because of their shape. This means that each double ring contains 16 copies of the nine distinct nucleoporins that form the Y-complex (Ori *et al*., 2013; Bui *et al*., 2013). The NPC is not only one of the largest molecular protein assemblies, but parts of its scaffold are also among the longest lasting proteins *in vivo* (e.g., D'Angelo *et al*., 2009; Savas *et al*., 2012; Toyama *et al*., 2013; Hakhverdyan *et al*., 2021).

The perforated NE is a prerequisite for material exchange between the nucleus and the cytoplasm (e.g., Anderson & Beams, 1956; reviewed, for example, by Görlich & Mattaj, 1996), which is necessary as a consequence of compartmentalisation and thus the spatial separation

of nuclear and cytoplasmic processes such as transcription and translation, respectively. The central part of the NPC thus represents a channel, which is equipped with the third group of nucleoporins that have characteristic FG-repeats (e.g., Davis & Fink, 1990; Cordes *et al*., 1991; Carmo-Fonseca *et al*., 1991). These FG-repeat-containing NUPs are bound to the NPC core structures and project primarily into the NPC channel and to some extent also into the cytoplasm and nucleoplasm, respectively. The assigned main function of the FG-NUPs is the selective bidirectional karyopherin-mediated transport with a capacity of about 100 MDa per second and NPC (Ribbeck & Görlich, 2001). A certain degree of permeability by passive diffusion allows the unhindered passage of smaller molecules up to a molecular weight of approximately 40 kDa or a diameter of up to 5 nm (e.g., Harding & Feldherr, 1958; Paine *et al*., 1975; Bonner, 1975; Mohr *et al*., 2009). Larger molecules require the assistance of nuclear transport receptors (NTRs) to traverse the NPC (e.g., Moore & Blobel, 1992; Görlich *et al*., 1994; Adam & Adam, 1994; Imamoto *et al*., 1995; Görlich *et al*., 1995; Fornerod *et al*., 1997; reviewed by Görlich & Kutay, 1999). The interplay of NTRs with the FG-repeat domain meshwork of this group of nucleoporins, i.e. how nucleocytoplasmic transport through the NPC takes place, is still a matter of debate, with different models proposed. This comprises the selective phase/hydrogel model (e.g., Ribbeck & Görlich, 2001, 2002; Frey *et al*., 2006), the virtual gate/polymer brush model (e.g., Rout *et al*., 2003; Patel *et al*., 2007), the forest model (Yamada *et al*., 2010), and the reduction-of-dimensionality model (e.g., Peters, 2005; Chowdhury *et al*., 2022). The upper limit for the transport of large objects was for a long time assumed to be about 40 nm in diameter (e.g., Panté & Kann, 2002), which was consistent with the cryo-EM structures of the NPC available at the time (Bui *et al*., 2013). Recently, it was reported that the diameter of the NPC channel *in vivo* is approximately 60 nm, and thus the upper limit is correspondingly higher (e.g., Mahamid *et al*., 2016; Schuller *et al*., 2021; see

also Matsuda & Mofrad, 2022), which was also impressively visualized by the import of intact HIV capsids into the nucleus (e.g., Zila *et al*., 2021).

The filamentous extension on the cytoplasmic side of the NPC could be categorised as a fourth group of nucleoporins. The large protein NUP358/RanBP2 is the major component of the elongated and flexible cytoplasmic filaments (e.g., Wu *et al*., 1995; Yokoyama *et al*., 1995; Delphin *et al*., 1997; Bley *et al*., 2022), which appear to contribute to the formation of a ribosome exclusion zone in the proximity of the NPCs (e.g., Watson, 1959; Krull *et al*., 2010).

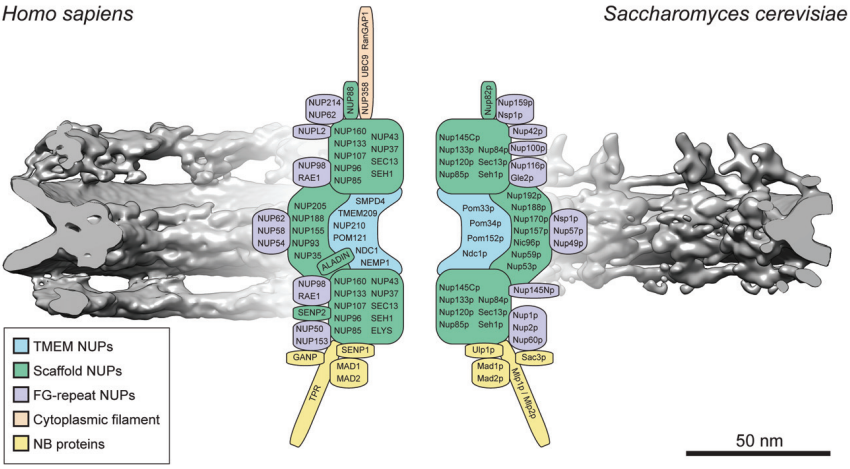

Figure 1. The composition of the NPC and NB in vertebrates and yeast.

The cryo-EM maps of the *H. sapiens* NPC (EMD-12814; Schuller *et al*., 2021) and of the *S. cerevisiae* NPC (EMD-10198; Allegretti *et al*., 2020) were processed with UCSF Chimera, with the surface bisected along the central axis. The cryo-electron tomographic reconstructions of tomograms of intact human and yeast cells are on average at 35 Å and 25 Å resolution, respectively. The five groups of the different nucleoporins are colour-coded with blue for TMEM NUPs, green for Scaffold NUPs, purple for FG-repeat NUPs, orange for the cytoplasmic filament proteins, and yellow for NB proteins. Note that not all associated proteins on the cytoplasmic side and also not all NB proteins are depicted. The relative position and grouping of these nucleoporins into subcomplexes are shown schematically (not to scale). Note that the outer parts, i.e. the cytoplasmic filaments and the NB, are not present in the tomograms, as they are probably lost in the averaging process due to their flexibility or heterogeneity.

The filamentous counterpart on the nucleoplasmic side of the NPC is the so-called nuclear basket (NB). It is morphologically distinguishable from the flexible cytoplasmic appendices by

the rod-like appearance of its fibrils that are laterally interconnected at their ends to form an annular platform, also referred to as the distal or terminal ring (TR). The NB entity could be categorised as the fifth group of nucleoporins in agreement with the first definition of the NPC, which included intranuclear channels (Watson, 1959). It was first visualised by transmission electron microscopy (TEM) as NPC-attached nucleoplasmic filaments (e.g., Franke & Scheer, 1970a, 1970b; Franke, 1970; Maul, 1971). A study with SV40 virus particles revealed an exclusion zone that was correlated with a fish trap-like fibrillar structure (Maul, 1976). Later, these filaments were more commonly referred to as "cage-like" or "basket-like" structures in various species, including frogs, birds, yeasts, slime molds and plants (e.g., Ris, 1989, 1991; Jarnik & Aebi, 1991; Ris & Malecki, 1993; Goldberg & Allen, 1992; Goldberg *et al*., 1997; Kiseleva *et al*., 2004; Beck *et al*., 2004; Fiserova *et al*., 2009).

The basic composition of NPCs and NBs is evolutionarily conserved (e.g., Rothballer & Kutay, 2013) despite the low sequence conservation of several nucleoporins to their homologues in other species (e.g., Kuznetsov *et al*., 2002; Mans *et al*., 2004; Neumann *et al*., 2010; Holden *et al*., 2014; Field *et al*., 2014; Padilla-Mejia *et al*., 2021). The core structure of the NPC has been elucidated by cryo-electron tomography (see also Figure 1). By now, this includes the species *Homo sapiens* (e.g., Maimon *et al*., 2012; Bui *et al*., 2013; von Appen *et al*., 2015; Kosinski *et al*., 2016; Lin *et al*., 2016; Schuller *et al*., 2021; Mosalaganti *et al*., 2022), the African clawed frog *Xenopus laevis* (e.g., Eibauer *et al*., 2015; Huang *et al*., 2020; Zhang *et al*., 2020; Huang *et al*., 2022; Zhu *et al*., 2022), the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomcyes pombe* (Kim *et al*., 2018; Allegretti *et al*., 2020; Zimmerli *et al*., 2021; Akey *et al*., 2022), the algae *Chlamydomonas reinhardtii* (Mosalaganti *et al*., 2018), and the slime mold *Dictyostelium discoideum* (Beck *et al*., 2004). Except for the latter, it had been not possible to visualise the NB in any of these studies. A certain flexibility or asymmetry of the NB, perhaps due to its inherent heterogeneity, does not allow for the averaging required for

this technique and results in the NB structure being blanked out. It remains to be seen if and when new cryo-EM studies with higher resolution NB structures will become available. In the meantime, new advances in super-resolution microscopy will probably help to improve our understanding of the NB ultrastructure.

1.2 NB scaffold protein TPR and its known interactors

The human protein TPR (translocated promoter region) has a molecular weight of approximately 267 kDa. Studies in which fusion events of the *MET* proto-oncogene and the *TPR* gene were discovered, leading to the formation of an active oncogene, gave rise to its name (e.g., Park *et al*., 1986; Tempest *et al*., 1986; Rodrigues & Park, 1993). The protein has a heptad repeat dominated amino-terminus (NT), also called "rod", which forms predominantly coiled-coil structures and can homodimerise. In contrast, the carboxy-terminus (CT), also called "tail", is intrinsically disordered and contains a high proportion of prolines and acidic amino acids (Mitchell & Cooper, 1992; Byrd *et al*., 1994; Cordes *et al*., 1997; Hase *et al*., 2001; Frosst *et al*., 2002). The C-terminus harbours the only known nuclear localisation signal (NLS) to facilitate the translocation of the long rod-shaped protein into the cell nucleus (Bangs *et al*., 1998; Cordes *et al*., 1998), where it is predominantly anchored on the nuclear side of the NPC (Cordes *et al*., 1997). The region around amino acids (aa) 400 to 600 is the so-called NPC binding domain (NBD), which can interact with NUP153, among others (e.g., Cordes *et al*., 1998; Bangs *et al*., 1998; Hase *et al*., 2001; Hase & Cordes, 2003; see further below). The NBD surrounding rod elements form long fibrils (Hase *et al*., 2001), corresponding to the presumed height of the prototypic NB of about 40–60 nm (e.g., Krull *et al*., 2004; Beck *et al*., 2004). The remaining rod parts eventually bifurcate (Hase *et al*., 2001) and are likely to interlock in an antiparallel arrangement, forming what is commonly known as the TR (e.g., Krull *et al*., 2004).

Reports on NB-like structures (NBLS) of cylindrical appearance and heights of up to about 400 nm are discussed further below (see Chapter 5).

In addition to TPR in mammals and amphibians (Cordes *et al*., 1997), homologues of TPR have been shown to occur in many other species throughout the eukaryotic realm, including *S. cervesiae*, where they are named Myosin-like protein 1 and 2 (Mlp1p and Mlp2p; Kölling *et al*., 1993; Strambio-de-Castillia *et al*., 1999; Kosova *et al*., 2000), and also in insects, plants, and other fungi (e.g., Zimowska *et al*., 1997; Kuznetsov *et al*., 2002; Xu *et al*., 2007; Jacob *et al*., 2007; Bae *et al*., 2009; De Souza *et al*., 2009). TPR was predicted to be the major structural element of the NB (Krull *et al*., 2004), and depletion or deletion of TPR or its homologues has been found to result in loss of the NB structure, as shown in both human cells and yeast (e.g., Krull *et al*., 2010; Funasaka *et al*., 2012; Niepel *et al*., 2013; Duheron *et al*., 2014). Some species are even able to tolerate the complete deletion of the TPR homologues, i.e. double knockout (KO) of Mlp1p and Mlp2p in *S. cerevisiae* (e.g., Strambio-de-Castillia *et al*., 1999; Kosova *et al*., 2000) or the KO of the *TPR* gene in *Arabidopsis* (e.g., Xu *et al*., 2007; Jacob *et al*., 2007), whereas in insect and higher vertebrates the deletion of TPR is lethal (e.g., Qi *et al*., 2004; Hart *et al*., 2015; Aksenova *et al*., 2020; our unpublished data).

Even though the proteins of the NB are often considered to be known, a complete protein inventory of the NB, which would also hold for different cell types and species, is not yet available. However, a variety of proteins have been identified as binding partners of TPR or its homologues in other species. Some of these proteins are NB-resident, but have no effect on NB assembly or maintenance and likely use TPR only as an operational platform or as a site for transient interactions with the NB. For simplicity, the following text does not claim to represent a complete list, but focuses in particular on vertebrate TPR interactors that have been shown to be located at the NB, and additionally mentions the corresponding *S. cerevisiae* homologues that are known to be Mlp-interacting proteins.

NUP153 is often referred to as NB protein because initial immuno-localisation data in TEM showed a distribution of the protein in areas at the NE that correspond to all parts of the NB (e.g., Sukegawa & Blobel, 1993; Cordes *et al*., 1993; Panté & Aebi, 1994; Frosst *et al*., 2002; Fahrenkrog *et al*., 2002). NUP153 mediates TPR anchorage at the NPC (Walther *et al*., 2001; Hase & Cordes, 2003; Aksenova *et al*., 2020) but, on the other hand, is not further necessary to subsequently keep TPR bound to the NPC (Lussi *et al*., 2010; Duheron *et al*., 2014; Aksenova *et al*., 2020; Gunkel *et al*., 2021; our unpublished data). In turn, the localisation of NUP153 and its binding partner NUP50 (Fan *et al*., 1997; Guan *et al*., 2000), which is also often referred to as an NB protein, is not TPR-dependent (e.g., Hase & Cordes, 2003). In the yeast *S. cerevisiae*, two possible homologues exist for NUP153, namely Nup1p and Nup60p (Davis & Fink, 1990; Rout *et al*., 2000). The latter, to which the NUP50 homologue Nup2p also binds, plays a role in anchoring Mlp proteins (e.g., Dilworth *et al*., 2001; Denning *et al*., 2001; Feuerbach *et al*., 2002; Zhao *et al*., 2004).

The NB-resident proteins MAD1 and MAD2 (mitotic arrest deficient 1 and 2) form a complex and localise jointly to the NPC during interphase, where they bind directly to TPR via MAD1 (Campbell *et al*., 2001; Lee *et al*., 2008; Lussi *et al*., 2010; Rodriguez-Bravo *et al*., 2014). Furthermore, two of the nine small ubiquitin-like modifier (SUMO)-specific proteases known in mammals, namely SENP1 and SENP2 (sentrin-specific proteases 1 and 2), have been localised to the NPC (Hang & Dasso, 2002; Zhang *et al*., 2002; Bailey & O'Hare, 2004; Goeres *et al*., 2011; Chow *et al*., 2012; Cubeñas-Potts *et al*., 2013). Of these, SENP1, but not SENP2, has been identified as a binding partner of TPR, on which its localisation at the NPC largely depends (Schweizer *et al*., 2013, and our unpublished data). Another NB-resident and TPR-interacting protein is the scaffold protein for the components of the transcription export complex 2 (TREX-2), namely GANP (germinal center associated nuclear protein; Wickramasinghe *et al*., 2010; Umlauf *et al*., 2013; Wickramasinghe *et al*., 2014; Aksenova *et*

al., 2020, and our unpublished data). GANP forms the platform for the associated proteins ENY2 (enhancer of yellow 2), CETN2 (centrin-2), which can also be replaced by CETN3 (centrin-3), and PCID2 (PCI domain-containing protein 2), which is in complex with SEM1 (suppressor of exocyst mutants 1; e.g., Jani *et al*., 2012). In *S. cerevisiae*, Mad1p/Mad2p, Ulp1p and the TREX-2 scaffold protein Sac3p are also localised to the NPC and interact with the Mlp proteins (Takahashi *et al*., 2000; Iouk *et al*., 2002; Fischer *et al*., 2002; Li & Hochstrasser, 2003; Fischer *et al*., 2004; Rodríguez-Navarro *et al*., 2004; Zhao *et al*., 2004; Scott *et al*., 2005; Lewis *et al*., 2007; Wilmes *et al*., 2008; Faza *et al*., 2009).

In addition, various other potential interaction partners have been identified in one or the other species, without the evolutionary conservation of the respective interaction having been demonstrated to date. These include the E3 ligase COP1/RFWD2 (Yi *et al*., 2006; Ouyang *et al*., 2020) or the initially seemingly kingdom-specific protein named Pml39p (pre-mRNA leakage 39) in budding yeast (Palancade *et al*., 2005). Furthermore, CRM1 and importin a/β, which are part of the bulk protein transport machinery, had been reported as interactors, which would have a major impact on the potential functional repertoire of the NB (e.g., Shah *et al*., 1998; Ben-Efraim *et al*., 2009; Funasaka *et al*., 2012; Zhao *et al*., 2014).

This raises the basic question of which functions are performed by the NB? A selection of the proposed NB functions, among others, covers perinuclear chromatin organisation and transcription control (e.g., Galy *et al*., 2000; Skaggs *et al*., 2007; Tan-Wong *et al*., 2009; Krull *et al*., 2010; Bermejo *et al*., 2011; Boumendil *et al*., 2019), protection from DNA damage sensitivity (e.g., Zhao *et al*., 2004; Kosar *et al*., 2021), cell cycle progression and proliferation (e.g., Niepel *et al*., 2005; Lee *et al*., 2008; David-Watine, 2011; Funasaka *et al*., 2012; Rodriguez-Bravo *et al*., 2014), as well as the export of various RNAs and the surveillance of such export (e.g., Shibata *et al*., 2002; Galy *et al*., 2004; Palancade *et al*., 2005; Coyle *et al*., 2011; Rajanala & Nandicoori, 2012; Saroufim *et al*., 2015; Lee *et al*., 2020; Aksenova *et al*.,

2020; Li *et al*., 2021). Some of these functions may not be mutually exclusive, but it remains to be clarified which of them are exclusively attributable to TPR and which functions are carried out via TPR-associated proteins.

However, the NB appears to provide a physical barrier that not only defines the export pathways for different types of mRNPs to the NPC (e.g., Mehlin *et al*., 1991; Kiseleva *et al*., 1996; Iborra *et al*., 2000; Soop *et al*., 2005; Li *et al*., 2021), but also keeps the nuclear forecourt of the NPC free of nuclear materials, like heterochromatin (e.g., Krull *et al*., 2010). One of the open questions in this context is: how far can this forecourt reach into the interior of the cell nucleus and which other proteins are involved in its formation?

1.3 Protein ZC3HC1 and its reported functions

The zinc finger C3HC-type protein 1 (ZC3HC1) was initially named HSPC216 because its cDNA had been isolated from hematopoietic stem/progentior cells (HSPCs; Zhang *et al*., 2000). It later received its current name from the Pfam motif zf-C3HC (Finn *et al*., 2008). In addition, three other names have been published so far. In a study on interaction partners of anaplastic lymphoma kinase (ALK), it was identified and named nuclear interacting partner of ALK (NIPA; Ouyang *et al*., 2003). Another study searched for novel inhibitor of apoptosis proteins (IAPs) based on sequence similarity to already known baculovirus IAP repeat (BIR) domains. Such sequence database mining led to the identification of two ZC3HC1 homologues in *Arabidopsis* and further identification of homologues in other phyla, including the human homologue, which was then named IAP-like protein 1 (ILP1; Higashi *et al*., 2005). More recently, a screening identified a protein whose loss leads to increased global DNA methylation, and which was named methylation elevated mutant 1 (MEM1; Lu *et al*., 2020;