Flexible and Extended Linker Domains Support Efficient Targeting of Heh2 to the Inner Nuclear Membrane

Rempel, Irina L; Popken, Petra; Ghavami, Ali; Mishra, Ankur; Hapsari, Rizqiya A; Wolters, Anouk H G; Veldsink, Annemiek C; Klaassens, Marindy; Meinema, Anne C; Poolman, Bert

Published in:
Structure

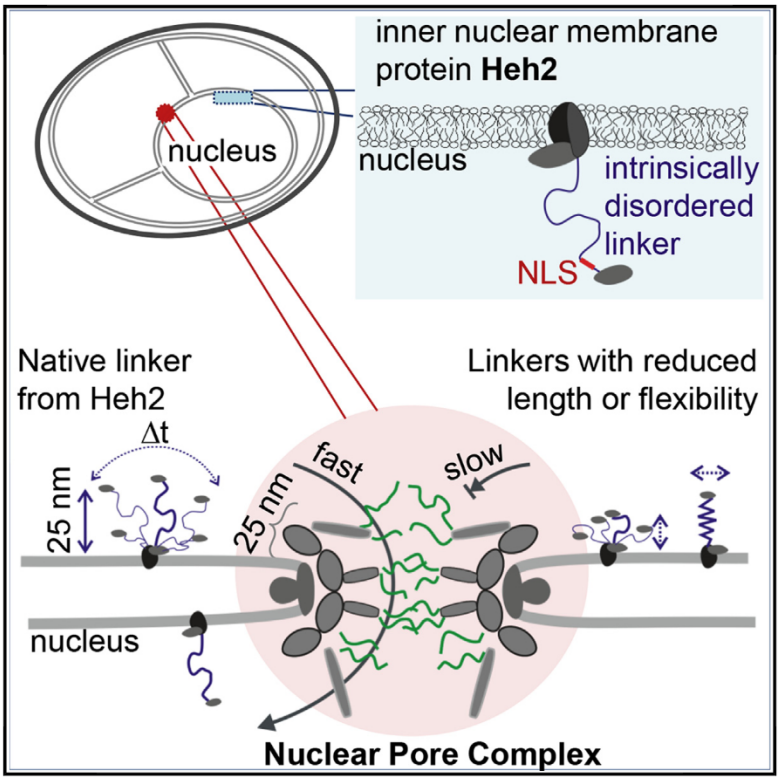
DOI:
10.1016/j.str.2019.11.003

Article

Structure

Flexible and Extended Linker Domains Support Efficient Targeting of Heh2 to the Inner Nuclear Membrane

Graphical Abstract



Authors

Irina L. Rempel, Petra Popken, Ali Ghavami, ..., Patrick R. Onck, Anton Steen, Liesbeth M. Veenhoff

Correspondence

p.r.onck@rug.nl (P.R.O.), a.steen@umcg.nl (A.S.), l.m.veenhoff@rug.nl (L.M.V.)

In Brief

Membrane proteins pass the nuclear pore complex (NPC) en-route to the inner nuclear membrane. Heh1/Heh2 have sorting signals: an NLS and intrinsically disordered linker. Rempel et al. show that linkers that best support import are flexible and match the dimensions of the NPC scaffold, consistent with transport through the NPC center.

Highlights

- Inner nuclear membrane proteins Heh1 and Heh2 have unique sorting signals
- Their sorting signals are composed of a long intrinsically disordered linker and NLS
- Linkers that best support import are flexible and match dimensions of NPC scaffold

Rempel et al., 2020, Structure *28*, 1–11
February 4, 2020
https://doi.org/10.1016/j.str.2019.11.003

CellPress

Please cite this article in press as: Rempel et al., Flexible and Extended Linker Domains Support Efficient Targeting of Heh2 to the Inner Nuclear Membrane, Structure (2019), https://doi.org/10.1016/j.str.2019.11.003

Structure
Article

Flexible and Extended Linker Domains Support Efficient Targeting of Heh2 to the Inner Nuclear Membrane

Irina L. Rempel,[1,6] Petra Popken,[1,2,3,6] Ali Ghavami,[2] Ankur Mishra,[2] Rizqiya A. Hapsari,[1] Anouk H.G. Wolters,[4] Annemiek C. Veldsink,[1] Marindy Klaassens,[1] Anne C. Meinema,[3,5] Bert Poolman,[2,3] Ben N.G. Giepmans,[4] Patrick R. Onck,[2,*] Anton Steen,[1,*] and Liesbeth M. Veenhoff[1,7,*]

[1]European Research Institute for the Biology of Ageing (ERIBA), University of Groningen, University Medical Center Groningen, Antonius Deusinglaan 1, Groningen 9713 AV, Netherlands

[2]Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, Groningen 9747 AG, The Netherlands

[3]Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, Groningen 9747 AG, The Netherlands

[4]Department of Biomedical Sciences of Cells and Systems, University of Groningen, University Medical Center Groningen, Antonius Deusinglaan 1, Groningen 9713 AV, Netherlands

[5]Present address: Institute of Biochemistry, Department of Biology, ETH Zürich, Zürich, Switzerland

[6]These authors contributed equally

[7]Lead Contact

*Correspondence: p.r.onck@rug.nl (P.R.O.), a.steen@umcg.nl (A.S.), l.m.veenhoff@rug.nl (L.M.V.)

https://doi.org/10.1016/j.str.2019.11.003

SUMMARY

The nuclear pore complex (NPC) is embedded in the nuclear envelope and forms the main gateway to the nuclear interior including the inner nuclear membrane (INM). Two INM proteins in yeast are selectively imported. Their sorting signals consist of a nuclear localization signal, separated from the transmembrane domain by a long intrinsically disordered (ID) linker. We used computational models to predict the dynamic conformations of ID linkers and analyzed the INM targeting efficiency of proteins with linker regions with altered Stokes radii and decreased flexibilities. We find that flexibility, Stokes radius, and the frequency at which the linkers are at an extended end-to-end distance larger than 25 nm are good predictors for the targeting of the proteins. The data are consistent with a transport mechanism in which INM targeting of Heh2 is dependent on an ID linker that facilitates the crossing of the approximately 25-nm thick NPC scaffold.

INTRODUCTION

The nuclear pore complex (NPC) is embedded in the nuclear envelope (NE) and forms the main gateway to the nucleus. NPCs are roughly built of a scaffold structure stabilizing a cylindrical opening in the NE, and attached to this scaffold is a set of proteins that are intrinsically disordered (ID) and extend into the channel of the NPC. The full structure of the *Saccharomyces cerevisiae* NPC was recently solved at a resolution of 28 Å (Kim et al., 2018), giving detailed insight into the dimensions of the different substructures of this large, 52-MDa complex. The structure reveals that the distance between the pore membrane and the inner surface of the NPC central channel is approximately 25 nm (Figure 1A) (Kim et al., 2018).

NPCs mediate rapid and energy-dependent transport between the nucleus and the cytoplasm. This is best understood for soluble proteins (Aitchison and Rout, 2012) and less so for integral membrane proteins of the NE. Several mechanisms have been proposed for the sorting of membrane proteins to the inner membrane of the NE (reviewed in Antonin et al., 2011; Blenski et al., 2019; Boni et al., 2015; Burns and Wente, 2012; Goodchild et al., 2015; Katta et al., 2014; Laba et al., 2014; Lusk et al., 2007; Meinema et al., 2012; Pawar et al., 2017; Zuleger et al., 2012). The biogenesis and sorting of membrane proteins starts in the cytosol with co-translational or post-translational insertion of proteins into the endoplasmic reticulum (ER) membrane. The membrane proteins then potentially roam the entire NE-ER network, as the membranes of the peripheral ER and the outer and inner nuclear membrane (ONM and INM) of the NE form a continuous system. To reach the INM, membrane proteins pass the NPCs where the INM and ONM are fused. The route through the *S. cerevisiae* NPC is spacious enough for passage of membrane proteins with extralumenal domain sizes of up to 90 kDa and thus the majority of monomeric proteins may enter the INM through the NPC even in the absence of specific sorting signals (Popken et al., 2015).

Many proteins indeed reach the INM by diffusion through the NPC and their accumulation at the INM is through retention mechanisms (Holmer and Worman, 2001). However, there are two *S. cerevisiae* membrane proteins, Src1/Heh1 and Heh2, and likely a third human membrane protein, whose localization in the nucleus are depending on an active transport mechanism involving transport factors, FG-nups and the RanGDP/RanGTP gradient (King et al., 2006; Kralt et al., 2015; Meinema et al., 2011). The sorting signal of Src1/Heh1 and Heh2 consists of a very potent nuclear localization signal (NLS) (King et al., 2006; Lokareddy et al., 2015; Meinema et al., 2011, 2013), and

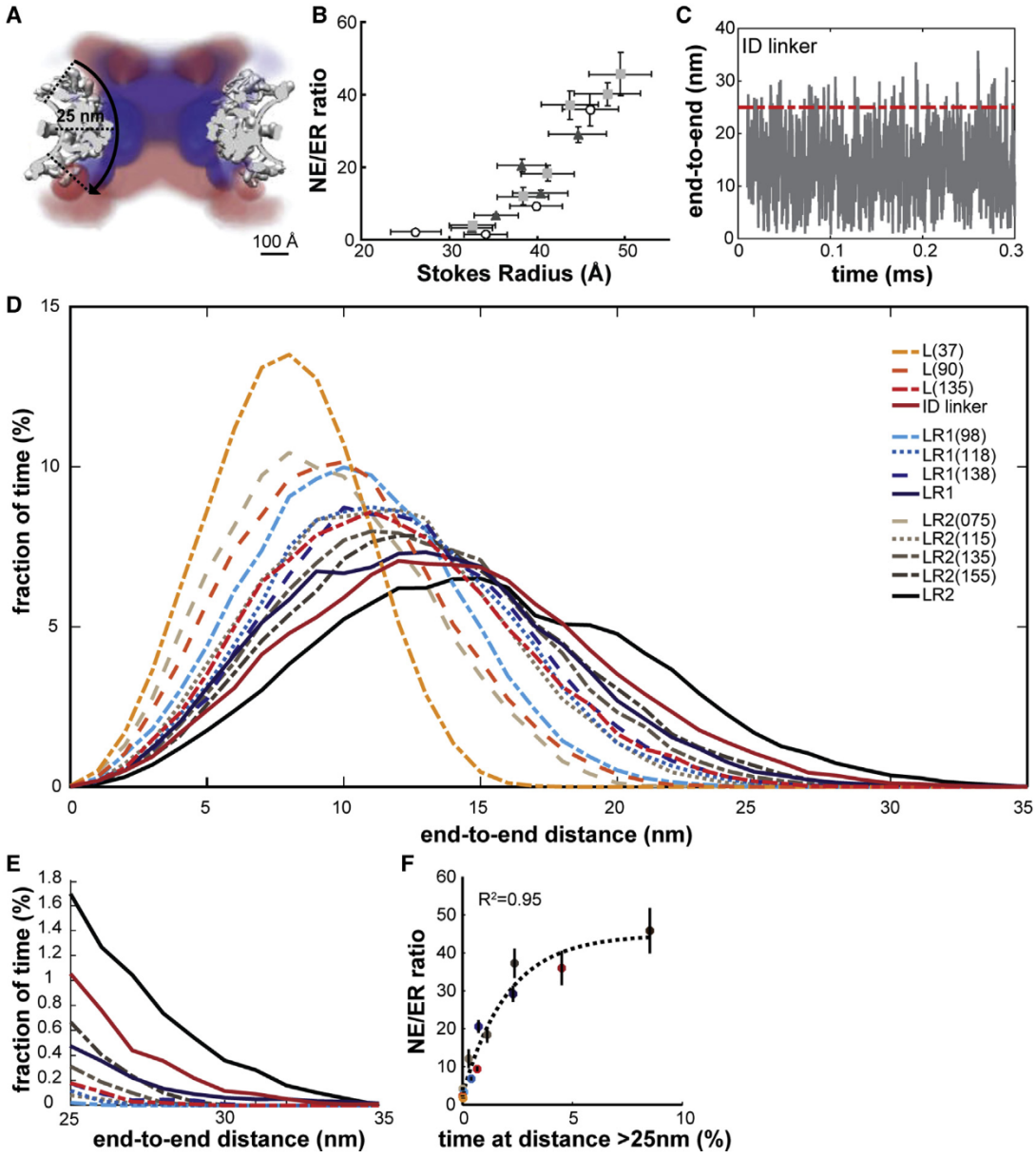


Figure 1. The Time that ID Linkers Spend in an Extended Conformation at an End-to-End Distance Larger than 25 nm Correlates with the Accumulation of Reporter Proteins at the INM

(A) Based on the structure of the NPC (Kim et al., 2018), an end-to-end distance of 25 nm is needed to span the distance between the pore membrane and the central channel.

(B) Correlation between the NE/ER ratio of membrane proteins and the Stokes radius of the linker. NE/ER ratio of GFP-h2NLS-linker-TM reporters with different linkers from (data from Meinema et al., 2011) plotted against Stokes radius predicted with the coarse-grained MD model described in (Ghavami et al., 2013, 2014). Truncations of native linker (L, white circles), randomized linker 1 (LR1, dark gray triangles), and randomized linker 2 (LR2, light gray squares). Error bars are SEM for NE/ER ratio and SD for Stokes radius.

(C) Example of end-to-end distance simulation of the Heh2 native ID linker. The red line marks an end-to-end distance of 25 nm.

(D) Plot showing the frequency at which the linkers extend to the different end-to-end distances.

(E) Like (D), but showing only the frequency at which ID linkers reach end-to-end distances larger than 25 nm.

(F) The percentage of time that ID linkers spend in their most extended conformations, larger than 25 nm, correlates with NE accumulation. Error bars are SEM

additionally a long, ID linker (Meinema et al., 2011). Already in the ER, the NLS ensures recruitment of the transport factor Kap95 via the adapter protein Kap60 (Meinema et al., 2013). The model proposed is that the complex then shuttles through the NPC by binding to the FG-nups in the NPC; a transport mode for which energy input is provided from the gradient Ran proteins across the NE (King et al., 2006; Meinema et al., 2011).

The ID linkers in Heh1 and Heh2 are approximately 200 amino acids long (Meinema et al., 2011). Shortening the linker

decreases the sorting of the proteins to the INM, while randomization of the amino acid sequence, which keeps the linkers ID nature, maintains full functionality (Meinema et al., 2011). Thus, the length but not the sequence of the linker is important for efficient import to the INM. Unique to ID structures is that they are highly dynamic, readily changing between very extended and more collapsed conformations, while the energy input that is required to stabilize a specific conformation is small (Fuxreiter, 2019; Galea et al., 2008; van der Lee et al., 2014; Necci et al., 2016; Uversky, 2017). We proposed two prospective roles for the linker in INM targeting (Meinema et al., 2011, 2012, 2013). In the first, the linker dodges into the scaffold of the NPC, allowing the transmembrane domain to remain in the membrane, while the NLS reaches out into the central channel of the NPC, where the linker enables dynamic interactions of the NLS-bound Kap95 with the FG-Nups. If the linker indeed functions to enter and cross the NPC, we expect that the fraction of time spent in an extended conformation, as well as flexibility, may be important parameters. In addition, or alternatively, we proposed that the linker may increase the efficiency of recruitment of Kap60 and Kap95 by positioning the NLS into the cytosol (Meinema et al., 2011, 2012, 2013). For this, the linker is effectively a spacer and we expect that also less-flexible structures, such as α-helical segments, may suffice.

In this study we used Heh2 linker mutants and reporter proteins with various artificial linker domains, to study the role of length, charge content, and flexibility of the linker in INM targeting. We altered the linker properties by replacing the full, or part of the linker sequence with polyproline sequences, or α helices. Shorter stretches of polyproline peptides and α helices have been used as molecular rulers in FRET, because they form rigid helices (Arai et al., 2001; Schuler et al., 2005). Multiple consecutive prolines form structured helical regions, of which the left-handed polyproline II helix (PPII) with residues in the *trans* conformation is energetically most favorable (Moradi et al., 2009). However, in longer polyproline peptides (24 residues) some residues are in the *cis* conformation (Hanson et al., 2012). Thus, most probably long stretches of prolines are not a rigid all-*trans* PPII helix, but even so such proline-rich sequences are much less flexible than the native ID linker domain. α Helices are right-handed helices formed by hydrogen bonding between the N-H group and the backbone C=O group amino acid located three or four residues apart in the protein sequence. Alanine shows particularly high tendencies to form α-helical structure. α-Helical structures can be further stabilized by the formation of salt bridges, such as between glutamate and lysine/arginine residues; this increases the rigidity of the helix (Marqusee and Baldwin, 1987; Olson et al., 2001). Single charged α-helical segments of up to 60 amino acids have been suggested to act as relatively rigid spacers between protein domains (Süveges et al., 2009), while shorter artificial α helices in the form of up to five EAAAK repeats have been reported to form stable α helices that show an approximately 80% helicity (Arai et al., 2001).

In this study we aimed to describe structural features of the ID linker that are important for targeting of Heh2 to the INM. We studied the nuclear import of membrane proteins that have the ID linker replaced by α-helical segments of defined lengths, and we modulate the conformational freedom of the ID linker by introducing polyproline stretches into the ID sequence. A coarse-grained computational model (Ghavami et al., 2013) was used to gain insight in the structure and dynamics of different linker structures. We find that while also helical and polyproline linkers support transport of the membrane proteins, a flexible and extended linker structure supports the most efficient translocation to the INM. We discuss our findings in the light of a potential transport path through the NPC, keeping in mind that the linkers have to extend 25 nm from the membrane to reach across the scaffold of the NPC.

RESULTS

The ID Linkers that Best Support Import Match the Dimensions of the NPC Scaffold

We set out to answer how the length and dynamics of ID linkers relate to the structure of the yeast NPC as reported in Kim et al. (2018) (Figure 1A). To span the distance from the membrane across the NPC the linker would need to reach 25 nm. Previously a series of reporter proteins with truncated and randomized linkers was made (Meinema et al., 2011). The reporter proteins consist of amino acids 93 to 378 from Heh2, corresponding to the NLS, the linker domain, and the first transmembrane segment of Heh2. Mutant versions of this reporter encoded ID linkers with randomized sequences (LR1 and LR2) and three to four truncated versions of the native and randomized linkers. The amino acid composition in the randomized sequences (LR1 and LR2) and the native linker (L) were the same, but the sequences were randomized. The localization of this set of reporters was studied and quantified as the fluorescence intensity at the NE relative to the ER (the NE/ER ratio) (Meinema et al., 2011). Now, we used a coarse-grained one-bead-per-amino-acid molecular dynamics model, developed to predict the behavior of ID proteins (Ghavami et al., 2013, 2014), to simulate the dynamic structure of each of the linker domains (methods and reagents are detailed in the STAR Methods). The Stokes radius is a measure of how extended the linker is. In Figure 1B we plotted the localization of the different proteins at the NE as a function of the Stokes radius of their linker domain. Indeed, the calculated Stokes radius and the measured NE/ER ratio show a striking correlation (Figure 1B).

The simulations were also analyzed in terms of the end-to-end distances that the linkers visit in time. We see, that the ID linkers are highly dynamic, they frequently change conformations (Figure 1C) and cover a large range of end-to-end distances (Figures 1C and 1D). The ID linkers only very infrequently reach end-to-end distances larger than 25 nm (Figures 1D and 1E). Notably, ID linkers that never reach this end-to-end distance do not support accumulation of the membrane proteins at the NE. We found that the NE/ER ratio of the different ID linkers was closely correlated to the percentage of time that the linkers spend in this most extended conformations of end-to-end distances larger than 25 nm (Figure 1F), while no such correlation was found with the average end-to-end distance of the ID linkers.

The correlation between NE/ER ratio and end-to-end distance follows a negative exponential decay function, meaning the function plateaus at an NE/ER just below 50, when the linker spends more than 10% of its time at an end-to-end distance

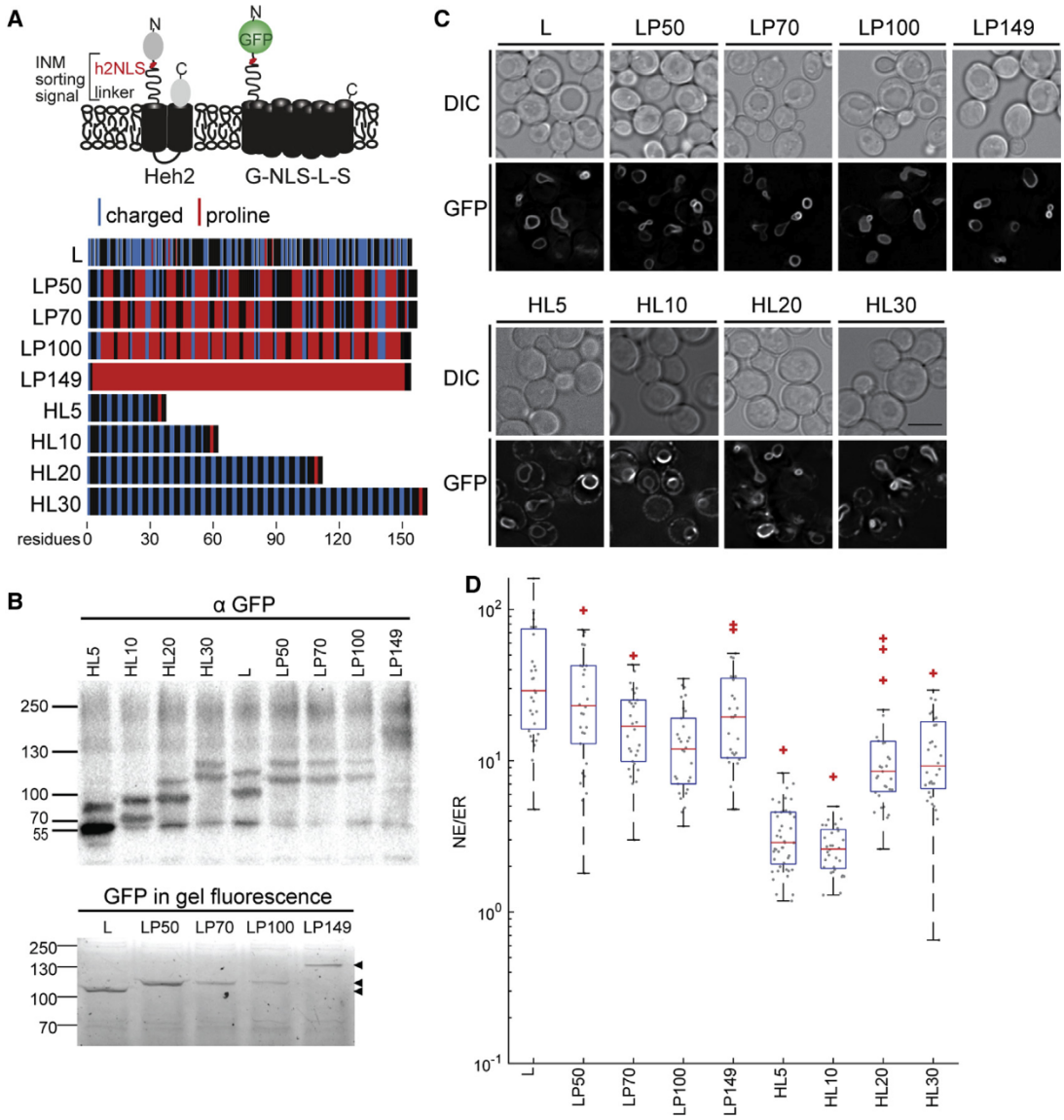


Figure 2. Reporters with Long Linkers Express and Accumulate at the NE

(A) Graphical representation of Heh2 and reporter proteins, where G stands for GFP and S stands for the transmembrane domain of Sec61. Schematic overview of the composition of different artificial linker domains used in this study. Red bars are proline residues, blue bars are charged residues, other amino acids are represented by black bars.

(B) Upper panel: western blot from lysates of cells expressing reporter proteins, showing full-length expression of all reporter proteins, except the LP149. Lower panel: in gel fluorescence of indicated reporter proteins, confirming full-length expression of reporter proteins (indicated by the arrowheads). Although, the LP149 was not detected on the western blot, a full-length green fluorescent signal was clearly detected on the SDS-PAA gel. On a western blot, the reporter proteins run with for GFP-fusion proteins characteristic double bands, one representing the fusion protein with unfolded GFP, while the GFP in the second band is still folded (Geertsma et al., 2008), consequently only one of the two bands is fluorescent *in gel*.

(C) Deconvolved images of cells expressing reporter proteins with indicated linker. Scale bar, 5 μm.

(D) Quantification of NE/ER ratio. Boxplot shows the median ratios of the fluorescence at the NE over that at the peripheral ER (NE/ER). The box shows 25th and 75th percentiles of the data. The whiskers extend to the data points, which are closest to 1.5 times above or below the interquartile range; data points above or below this region are considered outliers and plotted individually. Individual measurements are shown as gray dots, n = 30.

larger than 25 nm. We observe that NE/ER ratios above the threshold cannot be quantified, because the ER is not visible in those samples. However, we also consider that the NE targeting efficiency of the reporter proteins will be limited at some point, when other factors than the linker domain become rate-limiting for the INM targeting (i.e., the availability of Kaps to bind to the NLS). We conclude that INM targeting is effective with a linker domain that can extend to end-to-end distances larger than 25 nm, which coincides with the approximate distance between the pore membrane and the central channel.

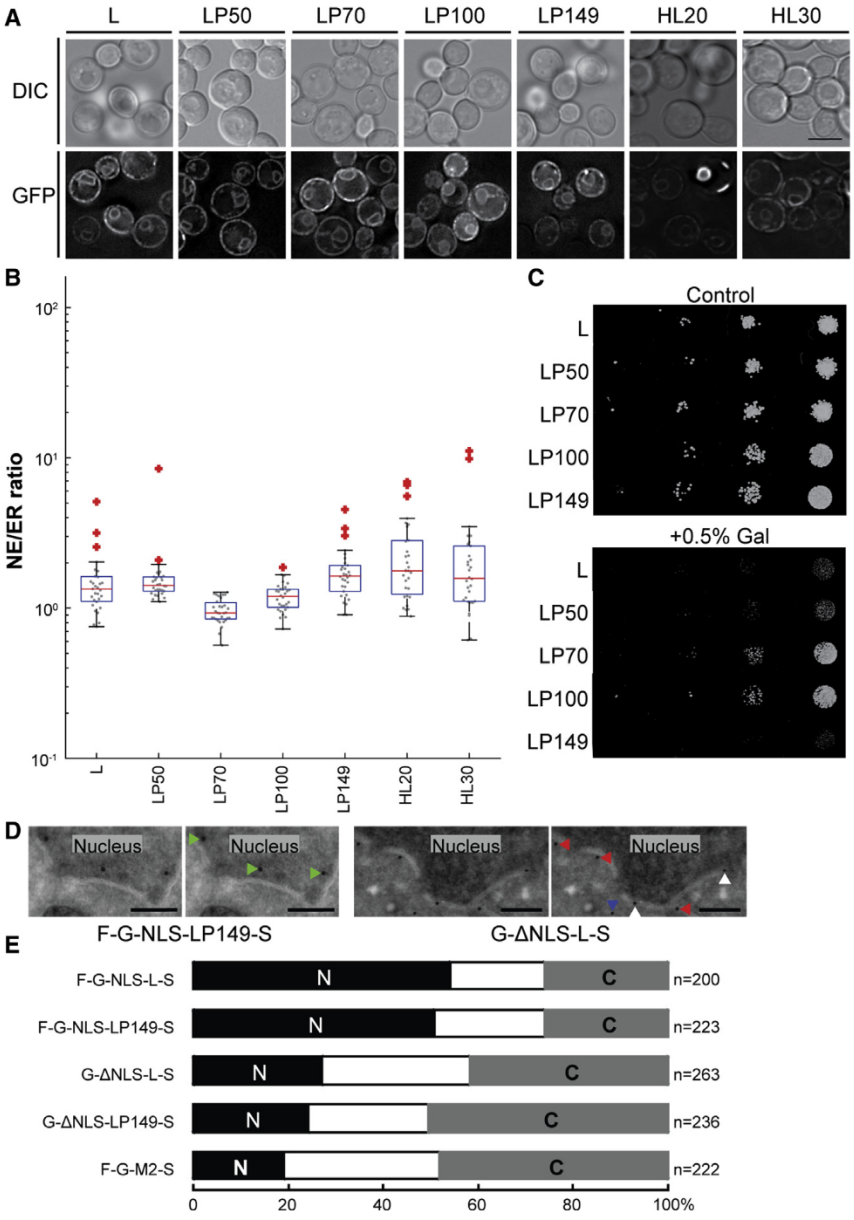


Figure 3. Reporters with Polyproline Linkers Accumulate at the INM

(A) Deconvolved images of cells expressing reporter proteins with indicated linker in the presence of rapamycin, which blocks active import. Scale bar, 5 μm.

(B) Quantification of localization of reporter proteins, in the presence of rapamycin, plotted as in 2D, n = 30.

(C) Spot assay with cells expressing reporter proteins with indicated linker. Top panel: control plate; bottom panel: plate containing 0.5% D-galactose which induces expression.

(D) Two selected examples showing reporter proteins as indicated with mainly INM or mainly ONM localization. Nucleus is indicated. Arrowheads show assignment of gold particles. Green rightward: nuclear; red leftward: cytoplasmic; white upward: ambiguous; blue downward: not assigned (too far from membrane). Scale bars, 0.25 μm.

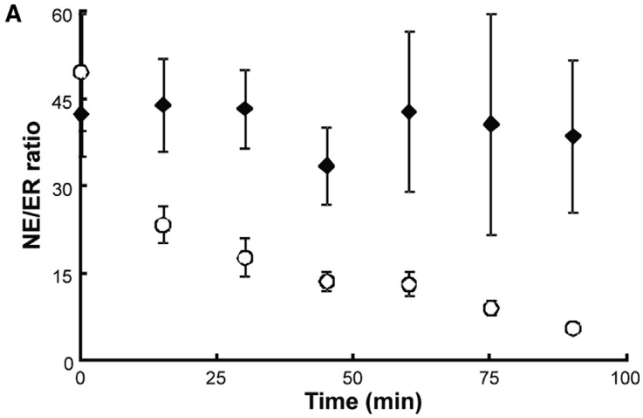
(E) ImmunoEM analysis to determine INM localization of h2NLS-LP149-Sec61. Percentage of gold particles counted at nuclear side (black), cytoplasmic side (gray) of the membrane, or ambiguous (lumen or on membrane, white) are shown, with number of gold particles counted listed. F-G-M2-S was included as a reference sample as this protein is excluded from the INM (Popken et al., 2015). Analysis was done blinded.

Reporter Proteins with Inflexible Linker Regions Show Decreased Accumulation at the NE

The correlations found in Figure 1 are consistent with a role for the ID linker in bridging a certain distance from the membrane and it fits with the distance required to bridge the scaffold of the NPC. If the linker is indeed dodging into the NPC, we expect that flexibility of the linker region might also be relevant to facilitate INM targeting. To alter the flexibility of the linker region, we designed eight different artificial linker domains: four linker domains with the same length, but increasing proline content and four linker domains with α-helical linkers of defined length. We increased the number of proline residues in the linker region by replacing regions of the ID linker with stretches of five to ten prolines, which decreases the charge content and causes a more collapsed structure because of reduced repulsive energy. The resulting linkers contained a total of 57 (LP50), 75 (LP70), 100 (LP100), and 149 prolines (LP149), the last having a continuous stretch of 149 proline residues. The α-helical linkers consisted of 5 (HL5), 10 (HL10), 20 (HL20), and 30 (HL30) EAAAK repeats, respectively (Figure 2A). α Helices in the form of EAAAK repeats (Arai et al., 2001) form a stable α helix that give the linkers an extended, yet less-flexible structure, while keeping the charge content within the linker region high. Each α-helical linker has two proline residues at the C terminus, to set the α helix apart from the transmembrane domain. The reporter proteins consisted of GFP, NLS, an (artificial) linker, and the transmembrane domain of Sec61 (Figure 2A).

First, we confirmed that all reporter proteins were expressed, although at reduced levels with increasing polyproline content (Figure 2B), and that they are correctly inserted into the membrane (Figure 2C). We quantified the efficiency of NE targeting as the ratio of fluorescence intensity in the NE over intensity in the peripheral ER (NE/ER ratio). As expected, based on their length in amino acids, HL5 and HL10 linkers were too short to support the accumulation of the reporter proteins at the NE. The longer α-helical linkers (HL20 and HL30) showed similar accumulation at the NE (median NE/ER of 8.5 and 9.2, respectively) (Figures 2C and 2D). We conclude that proteins with helical linkers also enrich at the NE albeit with lower levels compared with the native linker. The NE/ER ratio decreased

A



B

Reporter	Mobile fraction	D (μm²/s)	n
G-NLS-L-S	75 ± 3 %	1.4 ± 0.3	19
G-NLS-LP149-S	87 ± 3 %	1.3 ± 0.4	12

Figure 4. Mobility of Reporters with L or LP149

(A) Time traces of leak of reporter proteins from INM after abolishing import by adding rapamycin, which depletes Kap95, and stopping synthesis by adding glucose. White, G-NLS-L-S; black, G-NLS-LP149-S. Each time point is average of at least 30 cells; error bar: SEM.

(B) Mobility measurements of reporters at INM using fluorescence recovery after photo bleaching. D is lateral diffusion coefficient averaged over n cells. SEM is indicated.

with increasing proline content, with an exception for the reporter with the all-proline linker LP149, which showed an intermediate accumulation (NE/ER = 19.4 compared with 29 for the native Heh2-linker L) (Figures 2C and 2D). In the cells expressing the reporter with the LP100 linker a small soluble fraction can be observed in the nucleus (Figure 2C). This may be a degradation product of the reporter, containing at least the NLS and GFP, but it does not interfere with the analysis since the ratio of fluorescence intensity in different membrane compartments is used. Thus, having many proline residues in the linker, which increases the amount of structure between the NLS and the transmembrane domain, does not prevent accumulation of the reporter proteins at the NE, while it does affect the efficiency of transport as judged by the NE/ER ratio. Taken together, while all variants (helical and proline-rich) with sufficient linker length accumulated at the NE leaving little protein at the peripheral ER, the native linker is superior. We conclude, replacement of amino acids in the linker region reduces the ability of reporter proteins to accumulate at the NE, probably due to reduced flexibility in the linker domain.

Membrane Proteins with Polyproline Linkers Are Targeted to the INM

Having established that the reporter proteins with helical or polyproline linkers are expressed and localize to the NE, we addressed whether the proteins are actually localized at the INM and not just at the ONM. A first indication was obtained using the inducible Kap95-anchor away system. When reporters were expressed in the presence of rapamycin, Kap95-FRB traps at Pma1-FKBP in the plasma membrane, hereby disabling active import of Kap60/Kap95-dependent cargo through the NPC (Haruki et al., 2008; Meinema et al., 2011). The expression under these conditions caused those reporters that were previously shown to be NE-accumulated reporters, to mislocalize to the peripheral ER resulting in NE/ER ratios around 1.5 (Figures 3A and 3B), showing that the accumulation at the NE is Kap95 dependent. This is consistent with the hypothesis that the enrichment at the NE is the result of nuclear import and that the proteins are localized in the INM.

A second indication that the proteins are INM localized comes from assessing the viability of cells expressing the different polyproline reporter proteins (Figure 3C). Previously we have seen that overexpression of reporter proteins that accumulate at the INM causes a reduction in cell viability, while overexpression of reporters without the NLS or linker does not (Laba et al., 2015). We do not understand what causes the toxicity of excess INM-localized membrane proteins, but we use it to predict if the proteins localize at the INM. Indeed, the reporters with polyproline linkers also show lethality upon overexpression, and there is a clear correlation between the viability of the strains and the NE/ER ratios: the proteins with highest accumulation at the NE show the lowest viability (compare, e.g., LP50 and LP100 in Figures 2C and 2D). Thus, the viability analysis predicts that the reporter proteins with polyproline linkers are INM resident.

Finally, for a subset of the reporters we used quantitative immunoelectron microscopy (iEM) to determine their localization in relation to the INM (Figure 3D). Primary anti-GFP antibodies where visualized with colloidal gold. To allow quantitative and unbiased analysis, hundreds of yeast cells where imaged using a semi-automated large-scale EM at high resolution (also known as nanotomy [www.nanotomy.org; Sokol et al., 2015]). The gold particles were assigned, in a blindfolded fashion, as nuclear (N) or cytoplasmic (C) when within 75 nm of the NE inside or outside the nucleus, respectively, or as ambiguous when they were in the lumen or at the membrane (indicated by the arrowheads in Figure 3D). The LP149 reporters, with and without NLS, were compared with the corresponding reporters with the native linker L. The large FKBP-GFP-2xMBP-Sec61 (F-G-M2-S) was used as a control for a protein that is excluded from the INM (Popken et al., 2015). Because expression of strongly NE-accumulated reporters led to deformed nuclei, which made identification of the NE and assignment of the gold particles as INM or ONM difficult, expression levels were decreased by fusing an FKBP domain to the N terminus of the INM reporters (F-G-NLS-L-S and F-G-NLS-LP149-S). For each reporter at least 200 gold particles were counted, and percentages of N, C, or ambiguous are shown in Figure 3E. The distributions of the LP149 reporters are similar to their native linker counterparts, which shows that the different linker does not greatly affect the localization of the reporters, and that the reporter with LP149 and NLS is indeed accumulated in the INM.

Disentangling Import and Efflux Efficiency

To interpret the NE/ER ratios or the iEM data in terms of nuclear import efficiencies, it is critical to consider all parameters that determine the steady-state localization of the reporters. Most relevant are synthesis rates, retention mechanisms, and the rate of efflux and import through the NPC. We first compared the rates of efflux between the membrane proteins with different

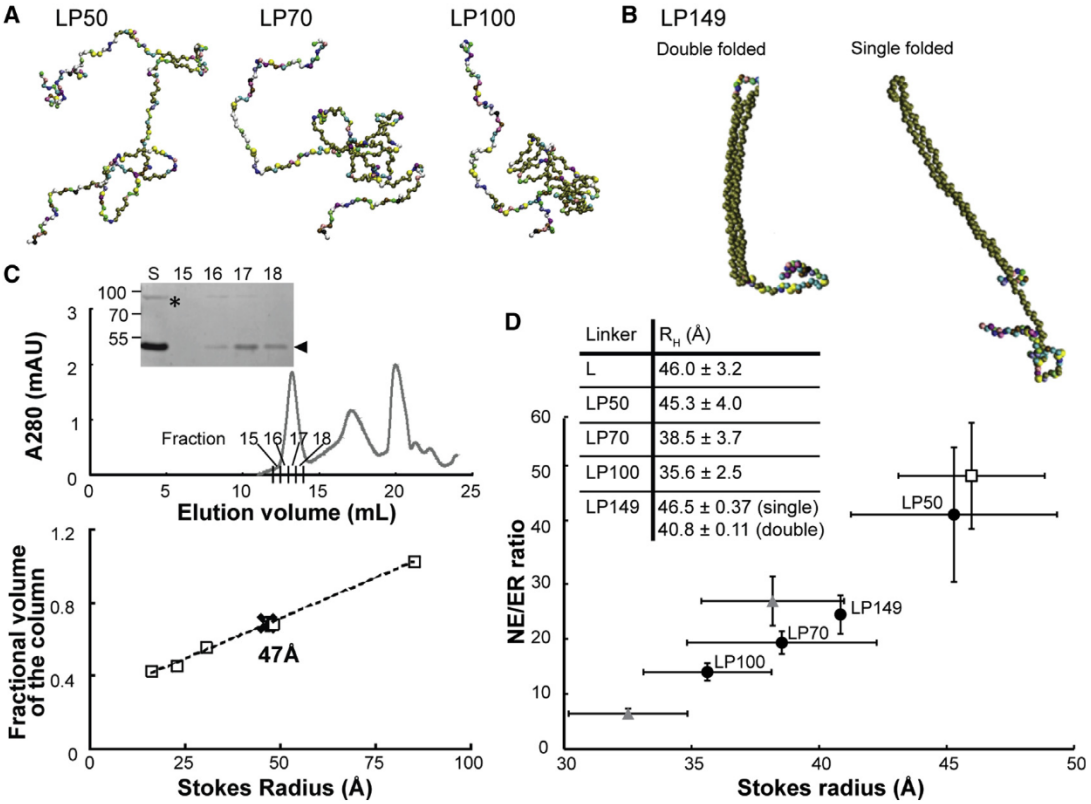


Figure 5. Stokes Radii of Polyproline Linkers and the Correlation with Import

(A) Snapshots from modeling of indicated linker domains.

(B) Snapshots from modeling showing the two distinct conformations assumed by LP149 during the simulations differing in the folding of protein. On the left panel we show the double-folded structure corresponding to lower Stokes radius and larger end-to-end distance. The right panel depicts a single-folded conformation corresponding to a larger Stokes radius and lower end-to-end distance.

(C) Stokes radius determination of LP50. Top graph shows elution profile from gel filtration of LP50 after GFP was removed by TEV-cleavage. Fractions indicated were analyzed together with the loaded sample (S) on SDS-PAGE (inset). Arrowhead indicates LP50, asterisk labels uncleaved LP50-GFP. Bottom graph shows fractional volume of reference proteins (squares) and LP50 (cross) plotted against Stokes radius.

(D) NE/ER ratio of reporters (G-NLS-L-S) with different linkers plotted against the Stokes radius from the computational model (shown in the table, along with the indicated SD). White square, L; black circles, proline linkers; gray triangles, LR1 truncations LR1(138), LR1(78). The Stokes radius for the LP149 represents the double-folded conformation. Error bars are SEM for NE/ER ratio and SD for Stokes radius. See also Figures S1–S3 and Tables S1 and S2.

linker regions. The Kap95-AA system was used to study the efflux of reporters from the INM (Meinema et al., 2011). As expected, when rapamycin and glucose were added to cells expressing G-NLS-L-S, a gradual decrease in NE/ER ratio was observed over time (Figure 4A) as proteins diffuse from the INM to the ONM and ER. The simultaneous addition of glucose halted expression, reducing the impact of newly synthesized proteins interfering with the measurements. In contrast, the NE/ER ratio for the reporter with LP149 did not decrease after blocking its import and stopping synthesis but stayed accumulated at the NE (Figure 4A). This shows that the reporter with an LP149 linker was either retained in the INM at a nuclear component, or it cannot passively go through the NPC to the ONM and ER.

The lack of diffusion to the ER upon depletion of Kap95 prompted us to validate if the protein with LP149 is mobile at the INM. Fluorescence recovery after photo bleaching measurements showed the mobility of G-NLS-LP149-S to be comparable with G-NLS-L-S (Figure 4B), indicating that the lack of re-localization to the ER is not due to retention but rather to an inability to efflux through the NPC. Considering that accumulation is the result of import and efflux (assuming synthesis and degradation rates to be much slower than import), a low efflux rate would result in a higher NE/ER ratio if import rates are the same. The lack of leak for the reporter with LP149 thus shows that also the import is less efficient with this linker when compared with the reporters with similar NE/ER ratios. We conclude that the LP149 reporter is actively imported to the INM at reduced rates compared with other linkers, and its efflux through the NPC is severely hindered by the all-proline linker.

Stokes Radii and Flexibility of Polyproline Linkers

Next, we addressed which properties of the polyproline linkers, Stokes radius, or flexibility, are responsible for the decreased import to the INM. The polyproline linkers are less flexible than ID linkers, but also the charge content is decreased, which might cause a more collapsed structure. The simulation predicts that the Stokes radius decreases, with more prolines present, up to

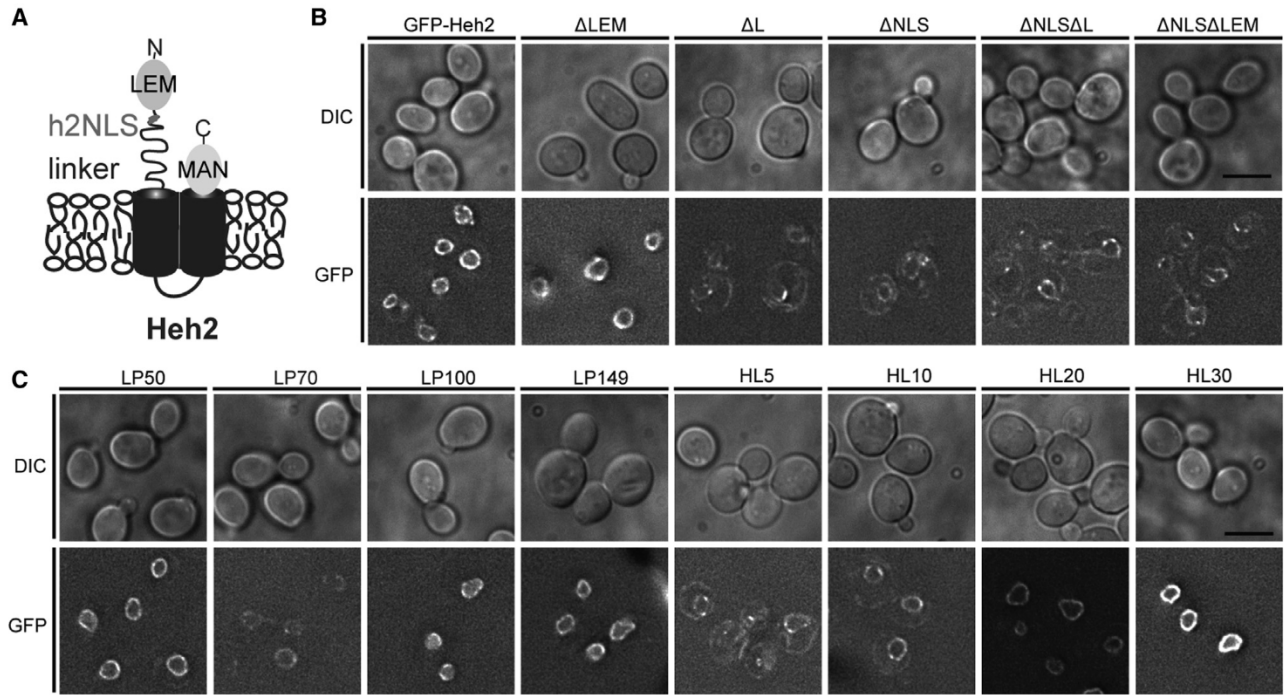


Figure 6. The NLS-L Sorting Signal Is Required for Accumulation of Full Heh2 at the INM

(A) Graphical representation of the different Heh2 domains.

(B) Microscopy images of full-length Heh2-GFP and various Heh2-GFP truncation mutants expressed from the endogenous locus. Full-length Heh2 and Heh2ΔIem show clear residence at the NE and very low amounts of peripheral ER in comparison with the other truncation mutants. Scale bar, 5 μm.

(C) Microscopy images of Heh2-GFP, with artificial linkers. With the exception of HL5 and HL10, all Heh2-GFP linker mutants show clear residence at the NE and very low amounts of peripheral ER. Scale bar, 5 μm.

LP100 (Figure 5A). Snapshots show that the short proline stretches, which can form PPII helices folded back on top of each other. Combined with the lower charge content this makes a more collapsed structure more favorable. Simulations for the LP149 revealed two distinct states: a single-folded conformation and a double-folded conformation (Figure 5B). The double-folded conformation has lower Stokes radius than the single-folded conformation and has higher stability than the single-folded conformation (Figure S1; Table S1). While the single-folded conformation never reaches the 25 nm end-to-end distance the double-folded conformation does (Figure S2).

Complementing the simulations, we experimentally determined the Stokes radius of the LP50-linker. LP50 was expressed in *Lactococcus lactis* as a fusion protein with a C-terminal TEV-cleavage site, GFP, and His-tag. After purification the GFP and His-tag were cleaved off and the Stokes radius was determined by

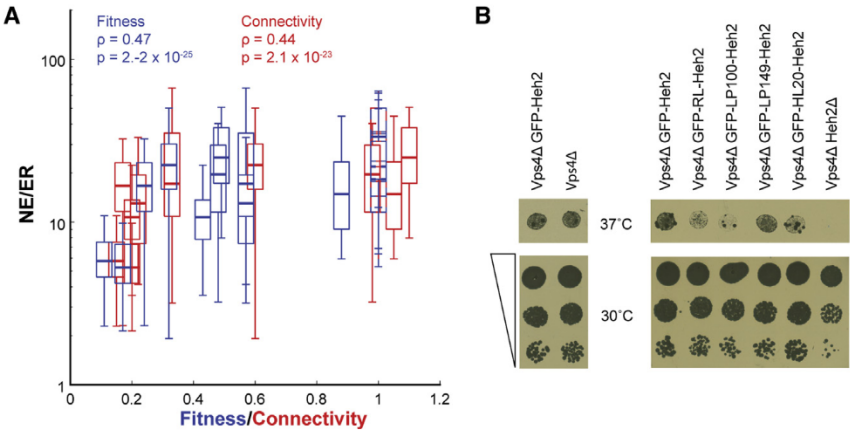


Figure 7. On the Relevance of the Linker for Heh2 Function

(A) The accumulation of G-NLS-L-TM measured in NPC mutants that lack domains of the components of the Nup84 complex (see STAR Methods for specific truncations) and plotted as a function of the connectivity of the NPC and the fitness of the cells as determined in Fernandez-Martinez et al. (2012). Boxplot shows the median ratios of the fluorescence at the NE over that at the peripheral ER (NE/ER). The box shows 25th and 75th percentiles of the data. The whiskers extend to the data points, which are closest to 1.5 times above or below the interquartile range. For each strain minimally 25 cells were analyzed. For both datasets, Spearman's rank-order correlations (ρ) and the according p values show that there is a positive correlation between connectivity/fitness and the accumulation of G-NLS-L-TM at the NE.

(B) Heh2 mutants that have linker variants are synthetic sick with Vps4Δ at 37°C. GFP-Heh2 and linker mutants derived are expressed from the endogenous promotor and location on the chromosome in W303 Vps4Δ, grown to mid-exponential phase, spotted on YPD, and grown at 30°C or 37°C.

size-exclusion chromatography (Figure 5C). The experimental values of 47 Å are in good correspondence with the Stokes radius obtained in the simulations (45Å ± 4). The expression of the other polyproline linkers in *L. lactis* was too low to yield a sufficient amount of protein. LP149-GFP could, however, be expressed and purified from yeast, and a comparison with the experimental Stokes radii of L-GFP, LP50-GFP, L, and LP50 suggest that also the modeled Stokes radius of LP149 in the double-folded conformation, aligning the experimental data (Figure S3; Table S2).

The NE/ER ratio measured for reporters with polyproline linkers, the native Heh2-linker, and two truncated ID linkers (Meinema et al., 2011), was plotted against their simulated Stokes radius (Figure 5D). The plot shows a clear correlation between the hydrodynamic radius of the linker with the accumulation at the NE for the random linkers and the polyproline linkers, which indicates that a more extended structure is favorable for efficient import.

The NLS-Linker Sorting Signal Is Required for INM Localization of Native Heh2

We next asked whether the artificial linker domains tested in the context of reporter proteins would have a similar effect on the localization of natively expressed full-length Heh2. Since previous research suggested that the LEM domain could be a nuclear retention signal (Grund et al., 2008; Taddei and Gasser, 2012), we first investigated whether the NLS-linker sorting signal was actually sufficient for Heh2 accumulation at the INM. The GFP-Heh2 fusions were expressed under the native heh2 promoter from the endogenous genomic location. The deletion of the ID linker domains, as well as the deletion of the NLS, resulted in loss of INM accumulation of Heh2, which is visible as reduced fluorescent intensity at the NE and the appearance of a visible ER signal at the cellular periphery (Figures 6A and 6B). Deletion of the LEM domain of Heh2 did not interfere with the NE localization of Heh2 (compare GFP-Heh2 with ΔLEM) and the LEM domain was insufficient to promote INM accumulation of Heh2ΔNLSΔL. The signals from the natively expressed GFP-Heh2 and GFP-Heh2 mutants are low, especially at the ER, and do not allow for a quantitative analysis of the fluorescence levels at the NE and ER, as we could do for the reporters. Nevertheless, the qualitative data clearly show that the NLS-linker sorting signal is required and sufficient for INM localization of natively expressed Heh2.

Next, we exchanged the native Heh2 linker for the polyproline and α-helical artificial linkers and integrated the different Heh2 linker mutants into the genome in the *heh2* locus, where they are expressed from the native heh2 promoter. In line with previous data on the importance of the length of the linker region, we confirm that the α-helical linkers of five and ten EAAAK repeats (HL5 and HL10) are too short to support NE accumulation; these mutants behave like a Heh2ΔL. All other linker mutants show high accumulation of Heh2 at the NE, indistinguishable from the native ID linker (Figure 6C). We conclude that sufficiently long linker regions, including those with polyproline stretches or helical linkers, support localization of Heh2 at the NE.

Safe-guarding the integrity of the diffusion barrier across the NE is essential to ensure proper nuclear compartmentalization. Because of the fact that the yeast integral INM protein Heh2 is not a fundamental component of the NPC, but has been found in physical and genetic interaction networks with NPC components such as Nup170 or Nup84 and the ESCRT-III/Vps4 surveillance system, it is conceivable that Heh2 functions as a sensor that is able to distinguish between a properly assembled pore and defective intermediates (Thaller et al., 2019; Webster et al., 2014, 2016; Yewdell et al., 2011). Considering the unique energy-dependent nuclear import mechanism and the ID nature of the linker sequence of Heh2, we speculate that these properties are exploited to support NPC surveillance mechanisms. Previously, NPC mutant strains that lack domains of the proteins of the Nup84 complex were characterized in terms of the connectivity of the NPC and the fitness of the strains (Fernandez-Martinez et al., 2012). We used these to assess if transport of Heh2-derived proteins is sensitive to perturbations in NPC connectivity. We indeed find that the nuclear accumulation of GFP-NLS-L-TM reporter proteins in NPC mutants lacking specific domains of the proteins of the Nup84 complex, correlates with the connectivity of these NPCs and also with the fitness of the cells (Figure 7A). Additional preliminary support comes from the observation that Heh2 mutants that have altered linker properties are synthetic sick in strains that also carry a deletion of VPS4 gene (Figure 7B). Altogether, these data preliminarily support a hypothesis that fast transport is what signals "this NPC is intact," and that the disordered linker of Heh2 plays a role in this process.

DISCUSSION

Heh1 and Heh2 are targeted to the INM by a mechanism that requires metabolic energy and specific transport factors (King et al., 2006). The targeting sequence of these proteins is comprised of an NLS and an ID linker domain (Meinema et al., 2011). We set out to study the ID linker as an essential part of the targeting sequence of INM proteins and asked if effective sorting to the NE inner membrane requires a linker that can span the 25-nm distance from the pore membrane to the central channel of the NPC (Kim et al., 2018). We simulated the highly dynamic end-to-end distances of the linker domain (Figure 1). Our simulated data show that the ID linkers spend a small fraction of time at this very extended conformation. Strikingly, the fraction of time that they extend at an end-to-end distance larger than 25 nm is a good predictor for targeting efficiency of the respected reporter proteins. The data thus support the hypothesis that the linker is needed to bridge the distance between the membrane and the central channel of the NPC (Meinema et al., 2011).

We found that the membrane proteins with mutant linker regions of different flexibility vary in their import and efflux dynamics. With the shorter α-helical linker regions (HL5, HL10), we confirm that a minimal length of the linker sequence is required to support INM targeting (Meinema et al., 2011). The polyproline linkers also support the proposed mechanism. From the most extreme linkers, HL30 and LP149, we learn that, in addition to an extended structure, also other properties are important. Based on the absence of leak for the LP149, the accumulation of this linker should be higher than what would be expected based on its Stokes radius. Also, the helical linkers

never support transport as effectively as the ID linkers (not even the long HL30). The explanation may be that the lack of flexibility, as well as the diameter (thickness) of the helix, limits INM targeting of membrane proteins. Our data are consistent with a narrow or highly curved passage way for the linker through the NPC scaffold.

It remains elusive, why Heh1 and Heh2 are transported by this mechanism rather than by a diffusion retention mechanism. Both proteins are part of an ESCRT-III/Vps4-dependent mechanism that senses misassembled NPCs and enables them to be cleared via the proteasome (Webster et al., 2014). Heh1 specifically is needed to initiate the recognition of misassembled NPCs. Heh2 is required for the clearance of misassembled NPCs, as is indicated by the fact that the deletion of Heh2 causes the accumulation of misassembled NPCs (Webster et al., 2014, 2016). One aspect of the mechanism to detect misassembled NPCs is related to the compartmentalization of Chm7 (Thaller et al., 2019). Future studies should resolve if the highly efficient INM transport of Heh1 and Heh2 is itself is a mechanism for sensing the structural integrity of the NPC. Our data showing that transport of Heh2-derived reporters is sensitive to the integrity of the NPC scaffold, and the genetic interactions of linker mutants with Vps4, does suggest this may be the case. We speculate that the N-terminal NLS and LEM domains can probe the integrity of the central regions of the NPC while the transmembrane domains and the C-terminal Man1 domain probe the lateral channels. Jointly these domains could thus probe a large region of the NPC, much larger than common soluble and transmembrane proteins do, and this could potentially provide for a good sensing mechanism where fast transport signals "this NPC is intact." It is, however, early days and more work is needed. The selection of mutants created here can provide a tool for future studies aiming to resolve how the special active transport mechanism may be important in the context of NPC quality control.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Western Blot and in Gel Fluorescence
  - Fluorescence Microscopy
  - Protein Purification and Stokes Radius Determination
  - Molecular Dynamics Simulations
  - ImmunoElectron Microscopy
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2019.11.003.

ACKNOWLEDGMENTS

This work was supported by the University of Groningen, the European Research Institute for the biology of Ageing, the Zernike Institute for Advanced Materials, the Ubbo Emmius Fund and the Netherlands Organisation for Scientific Research (ALWOP.2015.053). Part of the work has been performed at the UMCG Imaging and Microscopy Center (UMIC), which is sponsored by NWO-grants 175-010-2009-023, ZonMw 91111.006 and STW 12718. We thank Michael Rout and his team for sharing their Nup84 complex deletion strains with us. We thank Michael Chang for critical input into this project.

AUTHOR CONTRIBUTIONS

Investigation and Formal Analysis, I.L.R., P.P., and M.K. (Figures 1, 2, 3A, 3B, 3C, 4, and 5); A.C.M. (Figure 1B); A.M. and A.G. (Figures 1 and 5); P.P. and A.H.G.W. (Figure 3D); A.S. (Figure 6); R.A.H. (Figure 7A); A.S. and A.C.V. (Figure 7B). Writing – Original Draft, I.L.R., P.P., R.A.H., A.C.V., B.P., P.R.O., A.S., and L.M.V. Writing – Review & Editing, I.L.R., P.P., A.S., L.M.V. Supervision, B.N.G.G., P.R.O., A.S., and L.M.V. Funding Acquisition, B.P., P.R.O., and L.M.V.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 8, 2019
Revised: August 9, 2019
Accepted: November 8, 2019
Published: December 2, 2019

REFERENCES

Aitchison, J.D., and Rout, M.P. (2012). The yeast nuclear pore complex and transport through it. Genetics *190*, 855–883.

Antonin, W., Ungricht, R., and Kutay, U. (2011). Traversing the NPC along the pore membrane. Nucleus *2*, 87–91.

Arai, R., Ueda, H., Kitayama, A., Kamiya, N., and Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng. *14*, 529–532.

Blenski, M., Kehlenbach, R., Blenski, M., and Kehlenbach, R.H. (2019). Targeting of LRRC59 to the endoplasmic reticulum and the inner nuclear membrane. Int. J. Mol. Sci. *20*, 334.

Boni, A., Politi, A.Z., Strnad, P., Xiang, W., Hossain, M.J., and Ellenberg, J. (2015). Live imaging and modeling of inner nuclear membrane targeting reveals its molecular requirements in mammalian cells. J. Cell Biol. *209*, 705–720.

Burns, L.T., and Wente, S.R. (2012). Trafficking to uncharted territory of the nuclear envelope. Curr. Opin. Cell Biol. *24*, 341–349.

Ellenberg, J., Siggia, E.D., Moreira, J.E., Smith, C.L., Presley, J.F., Worman, H.J., and Lippincott-Schwartz, J. (1997). Nuclear membrane dynamics and reassembly in living cells: targeting of an inner nuclear membrane protein in interphase and mitosis. J. Cell Biol. *138*, 1193–1206.

Fernandez-Martinez, J., Phillips, J., Sekedat, M.D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J.D., Williams, R., Stokes, D.L., Chait, B.T., Sali, A., et al. (2012). Structure-function mapping of a heptameric module in the nuclear pore complex. J. Cell Biol. *196*, 419–434.

Fuxreiter, M. (2019). Fold or not to fold upon binding—does it really matter? Curr. Opin. Struct. Biol. *54*, 19–25.

Galea, C.A., Wang, Y., Sivakolundu, S.G., and Kriwacki, R.W. (2008). Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. Biochemistry *47*, 7598–7609.

Garcia de la Torre, J., Navarro, S., Lopez Martinez, M.C., Diaz, F.G., and Lopez Cascales, J.J. (1994). HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. Biophys J *67*, 530–531.

Geertsma, E.R., Groeneveld, M., Slotboom, D.-J., and Poolman, B. (2008). Quality control of overexpressed membrane proteins. Proc. Natl. Acad. Sci. U S A *105*, 5722–5727.

Ghavami, A., van der Giessen, E., and Onck, P.R. (2013). Coarse-grained potentials for local interactions in unfolded proteins. J. Chem. Theory Comput. *9*, 432–440.

Ghavami, A., Veenhoff, L.M., Van Der Giessen, E., and Onck, P.R. (2014). Probing the disordered domain of the nuclear pore complex through coarse-grained molecular dynamics simulations. Biophys. J. *107*, 1393–1402.

Goodchild, R.E., Buchwalter, A.L., Naismith, T.V., Holbrook, K., Billion, K., Dauer, W.T., Liang, C.-C., Dear, M.L., and Hanson, P.I. (2015). Access of torsinA to the inner nuclear membrane is activity dependent and regulated in the endoplasmic reticulum. J. Cell Sci. *128*, 2854–2865.

Grund, S.E., Fischer, T., Cabal, G.G., Antúnez, O., Pérez-Ortín, J.E., and Hurt, E. (2008). The inner nuclear membrane protein Src1 associates with subtelomeric genes and alters their regulated gene expression. J. Cell Biol. *182*, 897–910.

Hanson, J.A., Brokaw, J., Hayden, C.C., Chu, J.-W., and Yang, H. (2012). Structural distributions from single-molecule measurements as a tool for molecular mechanics. Chem. Phys. *396*, 61–71.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell. *31*, 925–932.

Holmer, L., and Worman, H.J. (2001). Inner nuclear membrane proteins: functions and targeting. Cell. Mol. Life Sci. *58*, 1741–1747.

Katta, S.S., Smoyer, C.J., and Jaspersen, S.L. (2014). Destination: inner nuclear membrane. Trends Cell Biol. *24*, 221–229.

Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., et al. (2018). Integrative structure and functional anatomy of a nuclear pore complex. Nature *555*, 475–482.

King, M.C., Lusk, C., and Blobel, G. (2006). Karyopherin-mediated import of integral inner nuclear membrane proteins. Nature *442*, 1003–1007.

Kralt, A., Jagalur, N.B., van den Boom, V., Lokareddy, R.K., Steen, A., Cingolani, G., Fornerod, M., and Veenhoff, L.M. (2015). Conservation of inner nuclear membrane targeting sequences in mammalian Pom121 and yeast Heh2 membrane proteins. Mol. Biol. Cell *26*, 3301–3312.

Kuipers, O.P., de Ruyter, P.G.G., Kleerebezem, M., and de Vos, W.M. (1998). Quorum sensing-controlled gene expression in lactic acid bacteria. J. Biotechnol. *64*, 15–21.

Laba, J., Steen, A., Popken, P., Chernova, A., Poolman, B., and Veenhoff, L. (2015). Active nuclear import of membrane proteins revisited. Cells *4*, 653–673.

Laba, J.K., Steen, A., and Veenhoff, L.M. (2014). Traffic to the inner membrane of the nuclear envelope. Curr. Opin. Cell Biol. *28*, 36–45.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of intrinsically disordered regions and proteins. Chem. Rev. *114*, 6589–6631.

Lokareddy, R.K., Hapsari, R.A., Van rheenen, M., Pumroy, R.A., Bhardwaj, A., Steen, A., Veenhoff, L.M., and Cingolani, G. (2015). Distinctive properties of the nuclear localization signals of inner nuclear membrane proteins Heh1 and Heh2. Structure *23*, 1305–1316.

Lusk, C.P., Blobel, G., and King, M.C. (2007). Highway to the inner nuclear membrane: rules for the road. Nat. Rev. Mol. Cell Biol. *8*, 414–420.

Marqusee, S., and Baldwin, R.L. (1987). Helix stabilization by Glu-…Lys+ salt bridges in short peptides of de novo design. Proc. Natl. Acad. Sci. U S A *84*, 8898–8902.

Meinema, A.C., Laba, J.K., Hapsari, R.A., Otten, R., Mulder, F.A.A., Kralt, A., van den Bogaart, G., Lusk, C.P., Poolman, B., and Veenhoff, L.M. (2011). Long unfolded linkers facilitate membrane protein import through the nuclear pore complex. Science *333*, 90–93.

Meinema, A.C., Poolman, B., and Veenhoff, L.M. (2012). The transport of integral membrane proteins across the nuclear pore complex. Nucleus *3*, 322–329.

Meinema, A.C., Poolman, B., and Veenhoff, L.M. (2013). Quantitative analysis of membrane protein transport across the nuclear pore complex. Traffic *14*, 487–501.

Moradi, M., Babin, V., Roland, C., Darden, T.A., and Sagui, C. (2009). Conformations and free energy landscapes of polyproline peptides. Proc. Natl. Acad. Sci. U S A *106*, 20746–20751.

Necci, M., Piovesan, D., and Tosatto, S.C.E. (2016). Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. Protein Sci. *25*, 2164–2174.

Olson, C.A., Spek, E.J., Shi, Z., Vologodskii, A., and Kallenbach, N.R. (2001). Cooperative helix stabilization by complex Arg-Glu salt bridges. Proteins *44*, 123–132.

Pawar, S., Ungricht, R., Tiefenboeck, P., Leroux, J.-C., and Kutay, U. (2017). Efficient protein targeting to the inner nuclear membrane requires Atlastin-dependent maintenance of ER topology. Elife *6*, https://doi.org/10.7554/eLife.28202.

Popken, P., Ghavami, A., Onck, P.R., Poolman, B., and Veenhoff, L.M. (2015). Size-dependent leak of soluble and membrane proteins through the yeast nuclear pore complex. Mol. Biol. Cell *26*, 1386–1394.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schuler, B., Lipman, E.A., Steinbach, P.J., Kumke, M., and Eaton, W.A. (2005). Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. Proc. Natl. Acad. Sci. U S A *102*, 2754–2759.

Sokol, E., Kramer, D., Diercks, G.F.H., Kuipers, J., Jonkman, M.F., Pas, H.H., and Giepmans, B.N.G. (2015). Large-scale electron microscopy maps of patient skin and mucosa provide insight into pathogenesis of blistering diseases. J. Invest. Dermatol. *135*, 1763–1770.

Süveges, D., Gáspári, Z., Tóth, G., and Nyitray, L. (2009). Charged single α-helix: a versatile protein structural motif. Proteins *74*, 905–916.

Taddei, A., and Gasser, S.M. (2012). Structure and function in the budding yeast nucleus. Genetics *192*, 107–129.

Thaller, D.J., Allegretti, M., Borah, S., Ronchi, P., Beck, M., and Lusk, C.P. (2019). An ESCRT-LEM protein surveillance system is poised to directly monitor the nuclear envelope and nuclear transport system. Elife *8*, https://doi.org/10.7554/eLife.45284.

Uversky, V.N. (2017). Intrinsically disordered proteins in overcrowded milieu: membrane-less organelles, phase separation, and intrinsic disorder. Curr. Opin. Struct. Biol. *44*, 18–30.

Webster, B.M., Colombi, P., Jäger, J., and Patrick Lusk, C. (2014). Surveillance of nuclear pore complex assembly by ESCRT-III/Vps4. Cell *159*, 388–401.

Webster, B.M., Thaller, D.J., Jäger, J., Ochmann, S.E., Borah, S., and Lusk, C.P. (2016). Chm7 and Heh1 collaborate to link nuclear pore complex quality control with nuclear envelope sealing. EMBO J. *205*, 695–701.

Yewdell, W.T., Colombi, P., Makhnevych, T., and Lusk, C.P. (2011). Lumenal interactions in nuclear pore complex assembly and stability. Mol. Biol. Cell *22*, 1375–1388.

Zuleger, N., Kerr, A.R.W., and Schirmer, E.C. (2012). Many mechanisms, one entrance: membrane protein translocation into the nucleus. Cell. Mol. Life Sci. *69*, 2205–2216.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Lactococcus lactis NZ9000	Kuipers et al., 1998	N/A
Chemicals, Peptides, and Recombinant Proteins		
L VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSKSDGKATSADLTSELETVEELHKKDSSDDKPRVKELPKP ELPNLKVSNEFLAQLNKELASAATENYDHSIKSTDLSSIRIETEEPVG PSTGAETRNESEVMENINLEVQPEVKEAKEELTKISETFDNQDEED TSRLSSKKNIRSPKGRTRHFIGARKVPYNQK	This study	N/A
LP50 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVASEESPPPPPLQFDEVNSKNPPPPPEREDGKDTEL PPPPPVSYPKTLEDPPPPPPEALFEPSRIEPPPPPNIITSVSRVDPP PPPPNRVLGITSAPPPPPRELDAEEPTLPPPPPETNDNESLSKPP PPPSNTHEPEKKDPPPPPKMVIPKLVIGARKVPYNQK	This study	N/A
LP70 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVASEESPPPPPLQFPPPPSKNPPPPPEREPPPPTEL PPPPPVSYPPPPEDPPPPPPPEALPPPPRIEPPPPPPNIIPPPPPRVDPP PPPPNRVLGLTSAPPPPPRELDAEEPTLPPPPPETNDNESLSKPP PPPSNTHEPEKKDPPPPPKMVIPKLVIGARKVPYNQK	This study	N/A
LP100 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVASEEPPPPPPLQPPPPPSKPPPPPPERPPPPPTELP PPPPVSPPPPPEDPPPPPPPEAPPPPPRIPPPPPPNIPPPPPRVDPP PPPPNRPPPPTSAPPPPPRELPPPPPTLPPPPPETNPPPPPSKPP PPPSNPPPPEKKDPPPPPPPVIVNGARKVPYNQK	This study	N/A
LP149 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP PPPPPPPPPPPPPPPPPPPPPPVNGARKVPYNQK	This study	N/A
HL5 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVLAEAAAKEAAAKEAAAKEAAAKEAAAKAAAPPVNG ARKVPYNQK	This study	N/A
HL10 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVLAEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKAAAPPVNGARKVPYNQK	This study	N/A
HL20 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVLAEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKEAAAKAAAPPVNGARKVPYNQK	This study	N/A
HL30 VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDVLAEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKEAA AKEAAAKEAAAKEAAAKEAAAKAAAPPVNGARKVPYNQK	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LR1(138) **LR1(078)** VKDENVETNKRKREQISTDNEAKMQIQEEKSPKKKRKKRSSKANK PPESPPQSDV>VSYPKTLEDPDANPLEALFEPSRIESKTDENIITSV SRVDKRGGSPNRVLGITSAKIVTL<RELDAEEPTLQATATETNDNE SLSKSKLKESNTHEPEKKDKLSSKKMVIGARKVPYNQK	Meinema et al., 2011	N/A
Deposited Data		
EM data related to Figures 3D and 3E on www.nanotomy.org	This study	http://www.nanotomy.org
Experimental Models: Organisms/Strains		
S. cerevisiae: KAP95-AA (W303, MATα tor1-1 fpr1::NAT PMA1-2×FKBP12::TRP1 Kap95-FRB::KanMX)	Haruki et al., 2008	N/A
S. cerevisiae: Htb2-FRB (W303 MATα; tor1-1; fpr1::NAT; Htb2-FRB::kanMX6)	Popken et al., 2015	N/A
S. cerevisiae: heh2Δ (MATa ura3Δ0 leu2Δ0 his3Δ1 met15Δ0 YDR458CΔ::kanMX4)	Thermo Fisher	N/A
S. cerevisiae: GFP-Heh2 (BY4742 heh2::GFP-HEH2-NAT)	Kralt et al., 2015	N/A
S. cerevisiae: GFP-Heh2ΔNLS (BY4742 heh2::GFP-HEH2(Δh2NLS)-NAT)	Kralt et al., 2015	N/A
S. cerevisiae: GFP-Heh2ΔL (BY4742 heh2::GFP-HEH2 (Δ146-299)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2ΔLEM (BY4742 heh2::GFP-HEH2 (Δ1-92)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2ΔNLSΔL (BY4742 heh2::GFP-HEH2 (Δ93-299)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2ΔNLSΔLEM (BY4742 heh2::GFP-HEH2 (Δ1-138)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-LP50 (BY4742 heh2::GFP-HEH2 (h2L::LP50)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-LP70 (BY4742 heh2::GFP-HEH2 (h2L::LP70)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-LP100 (BY4742 heh2::GFP-HEH2 (h2L::LP100)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-LP149 (BY4742 heh2::GFP-HEH2 (h2L::LP149)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-HL5 (BY4742 heh2::GFP-HEH2 (h2L::HL5)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-HL10 (BY4742 heh2::GFP-HEH2 (h2L::HL10)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-HL20 (BY4742 heh2::GFP-HEH2 (h2L::HL20)-NAT)	This study	N/A
S. cerevisiae: GFP-Heh2-HL30 (BY4742 heh2::GFP-HEH2 (h2L::HL30)-NAT)	This study	N/A
S. cerevisiae: W303 Vps4Δ (Vps4::HgrNT2)	This study	N/A
S. cerevisiae: W303 Vps4Δ Heh2Δ (Vps4::HgrNT2 YDR458CΔ::kanMX4)	This study	N/A
S. cerevisiae: W303 Vps4Δ GFP-Heh2 (Vps4::HgrNT2 heh2::GFP-HEH2-NAT)	This study	N/A
S. cerevisiae: W303 Vps4Δ GFP-Heh2-LP100 (Vps4::HgrNT2 heh2::GFP-HEH2(h2L::LP100)-NAT)	This study	N/A
S. cerevisiae: W303 Vps4Δ GFP-Heh2-LP149 (Vps4::HgrNT2 heh2::GFP-HEH2(h2L::LP149)-NAT)	This study	N/A
S. cerevisiae: W303 Vps4Δ GFP-Heh2-HL20 (Vps4::HgrNT2 heh2::GFP-HEH2(h2L::HL20)-NAT)	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae: W303 Vps4Δ GFP-Heh2-RL (Vps4::HgrNT2 heh2::GFP-HEH2(h2L::RL2)-NAT)	This study	N/A
S. cerevisiae: W303 and DF5 with Nup84 complex truncations, namely Nup84-(1-573) NUP85(233-744) Nup85(1-438) Nup120 (397-1037) Nup133(1-1098) Nup133 (301-1157) Nup145c(1-8)-(228-712) Nup145c(1-670) Nup145c(1-468) Nup145c(1-316)-(327-712)	Fernandez-Martinez et al., 2012	N/A
Recombinant DNA		
pACM023: G-NLS-L-TM	Meinema et al., 2011	N/A
pSI6: G-NLS-L-S (pACM023, where the TM is replaced for Sec61 (Residues 25-471))	This study	N/A
pSI7: G-NLS-LP50-S (pSI6, where L is replaced for LP50)	This study	N/A
pSI8: G-NLS-LP70-S (pSI6, where L is replaced for LP70)	This study	N/A
pSI9: G-NLS-LP100-S (pSI6, where L is replaced for LP100)	This study	N/A
pSI11: G-NLS-LP149-S (pSI6, where L is replaced for LP149)	This study	N/A
pSI34: G-NLS-HL5-S	This study	N/A
pSI35: G-NLS-HL10-S	This study	N/A
pSI36: G-NLS-HL20-S	This study	N/A
pSI37: G-NLS-HL30-S	This study	N/A
pSI20: G-ΔNLS-L-S (pSI6, where h2NLS is removed)	This study	N/A
pSI26: **G-ΔNLS-LP149-S** (pSI20, where L is replaced by LP149)	This study	N/A
pAS004: F-G-NLS-L-S (pSI6 with N-terminal 2×FKBP12)	This study	N/A
pPP037: F-G-NLS-LP149-S (pSI11 with N-terminal 2×FKBP12)	This study	N/A
pPP011: F-G-M2-S	Popken et al., 2015	N/A
pSI15: LP50-GFP (Vector based on pNZ for expression of LP50-GFP in *L. lactis*)	This study	N/A
pSI33: LP149-GFP (vector based on pACM23, for the expression of LP149-GFP in yeast)	This study	N/A
pRAH38 pACM023, HIS3 replaced by URA3, *GAL1* promoter replaced by *CUP1-1,* h2NLS replaced by NP NLS (AVKRPAATKKAGQKKKKLD)	This study	N/A
pACM044 Adapted from pACM023, where the L is replaced by a randomized and truncated version of 138 residues and TM is replaced for full-length Sec61 (Residues 25-471)	Meinema et al., 2011	N/A
pACM046 pACM044 with linker truncated to 78 residues	Meinema et al., 2011	N/A
Software and Algorithms		
ImageJ	Schindelin et al., 2012	https://imagej.nih.gov/ij/
Matlab 2016b		https://mathworks.com/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Liesbeth M. Veenhoff (L.M.Veenhoff@rug.nl).

All unique/stable reagents generated in this study are available from the Lead Contact without restriction.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All strains used in this study are listed in the Key Resources Table. All experiments with reporter proteins were done in the KAP95-AA strain (Haruki et al., 2008; Meinema et al., 2011), with the exception of the iEM control sample F-G-M2-S which was expressed in strain Htb2-FRB (Popken et al., 2015). Plasmids (Key Resources Table) were generated by standard molecular biology techniques and validated by sequencing; details are available upon request. Amino acid sequences of the linker used in this study are listed in the Key Resources Table (with NLS sequences underlined). Cells were grown at 30°C in selective drop-out medium, supplemented with 2% D-raffinose. Reporter proteins were expressed under control of the GAL1 promoter by 2 h induction with 0.5% D-galactose. Rapamycin (LC laboratories, Woburn MA) was used at a final concentration of 5 μg/mL where appropriate. Incubation with 1% D-glucose was used to stop expression. The strains carrying domain deletions of the Nup84 complex described in Fernandez-Martinez *et al, 201* are derived from DF5 or W303 strain backgrounds. W303 and DF5 strains were grown at 25 degrees in selective drop-out medium, supplemented with respectively 2% D-raffinose or 2% glucose. Expression of GFP-NP-L-TM in DF5 from the *CUP1-1* promoter (150 μM CuSO4, 2 hours) and in W303 strains from the *GAL1* promoter (0.1% D-galactose, 1.5 hours) provided comparable expression conditions.

Spot assays shown in Figure 3 were done using exponentially growing cells, diluted to $1x10^6$ - $1x10^3$ with drops of 4 μL on SD-His plates supplemented with 2% D-raffinose and 0.5% D-galactose or only 2% D-raffinose as a control.

Spot assays shown in Figure 7 were done using exponentially growing cells, diluted to $1x10^6$ - $1x10^4$ with drops of 5 μL on YPD plates supplemented with appropriate antibiotics. Cells were spotted on 2 separate plates, one grown at 30 degrees for 2 days, the other at 37 degrees for 3 days.

METHOD DETAILS

Western Blot and in Gel Fluorescence

Native whole cell protein extracts were used for western blots and in gel fluorescence. Approximately 2.5 OD units from an exponentially growing culture were harvested by centrifugation. The cells were resuspended in 450 μL 50 mM Tris-Cl pH 7.4, 0.33 M sucrose, 1 mM EDTA, 1 mM DTT, 1:100 solP. Up to 500 μl of resuspended cells were added to the fast prep cups together with ~ 450 μl glass beads (BioSpec, 0.5 mm diameter). Cells were lysed using the FastPrep-24 classic grinder (mpbio) on a standard yeast setting. The cell lysate was separated from the glass beads by pinching the fast prep tubes with hot needles (top and bottom) and draining the liquid into a fresh Eppendorf cup by centrifugation. 55 μl 10x SDS sample buffer was added to the samples, which were left at RT for a few minutes, before they were centrifuged at max speed for 30s to get rid of aggregates. Samples were separated on SDS-PAGE. In gel fluorescence was detected, using the Fluorescein filter of a ChemiDocXRS (Bio-Rad).

For western blots, the SDS-PAGE was transferred to a PVDF membrane (semi dry transfer). An anti-GFP antibody (Merck, Darmstadt, Germany) at a dilution of 1:2,000 (v/v) and a secondary with anti-rabbit-alkaline phosphatase conjugate at 1:10,000 (v/v) (Sigma-Aldrich, St. Louis, USA), was used for detection of the reporter proteins.

Fluorescence Microscopy

Imaging was done on a DeltaVision Deconvolution Microscope (Applied Precision) at 30°C, using InsightSSITM Solid State Illumination of 488 nm and an Olympus UPLS Apo 100x oil objective with 1.4NA. Detection was done with a CoolSNAP HQ2 camera. SoftWoRx software was used, and image-stacks were deconvolved using standard settings. Data was analyzed with open source software Fiji (Schindelin et al., 2012). The data plotted in Figure 1 and originating from Meinema et al., 2011 were measured on a confocal microscope.

FRAP was done on a LSM780 NLO confocal microscope (Carl Zeiss MicroImaging, Jena, Germany) at 30°C, using a "PlanNeofluar" 63x/1.3NA CorrDIC water or glycerine immersion objective, 34channelQuasar detector and ZEN acquisition software. Measurements were done essentially as described in (Meinema et al., 2013). Data were analyzed using the ZEN2010B software package (Carl Zeiss) and fitted to diffusion equation for membrane proteins (Ellenberg et al., 1997).

Protein Purification and Stokes Radius Determination

L. lactis NZ9000 (Kuipers et al., 1998) was used as expression host for LP50. Liquid cultures were grown standing at 30°C in M17 medium (Oxoid, Hampshire UK) supplemented with 0.5% glucose and 5 μg/mL Chloramphenicol. Expression of LP50 was induced at OD 0.5 with 0.01% nisin supernatant for 4 h. Purification and Stokes radius determination of LP50 was done as was described for h2NLS-L in (Meinema et al., 2011).

Overexpression of h2NLS-LP149-GFP was performed in *S. cerevisiae* (BY4741). The cells were grown under standard conditions to late exponential phase, at which point expression of h2NLS-LP149-GFP was induced with 0.5 % Galactose for a total of 2 hours, before the cells were harvested. The subsequent handling of the cell pellet and purification procedures were identical to those described for h2NLS-L-GFP and h2NLS-LP50-GFP.

Molecular Dynamics Simulations

Simulations are performed using a one-bead-per-amino-acid coarse-grained molecular dynamics model (Ghavami et al., 2013, 2014). The model accounts for hydrophobic and electrostatic interactions and can differentiate between all 20 naturally existing

amino acids. The force field parameters of proline are specifically calibrated against the experimentally measured end to end distance of polyproline segments (Schuler et al., 2005). The proteins are set at the fully extended conformation at the beginning of the simulations and are allowed to freely move and rotate through the medium during simulation. The Stokes radii are calculated according to the method explained in (Ghavami et al., 2013), in which each protein is simulated for 10^7 steps and the average Stokes radius for the generated conformations (excluding the first 10^6 steps) is calculated using the HYDRO computer program (García de la Torre et al., 1994). The standard deviation of the Stokes radii (Figure 5) is calculated based on the generated conformations after the first 10^6 steps. LP149 showed 2 different stable conformations (see Figure S1). Thus, in order to acquire relevant statistics, we performed MD simulations for 6 different starting configurations where each simulation was run for 5×10^7 steps long. The Stokes radius and the end-to-end distance data for these 6 realizations are shown in Figures S1 and S2 and the time-averaged Stokes radius and end-to-end distance data is discussed in Tables S1 and S2.

ImmunoElectron Microscopy

Sample preparation essentially the same as published before (Yewdell et al., 2011). Cells for immuno-EM (iEM) were fixed in 2% paraformaldehyde (Merck, Germany) and 0.2% glutaraldehyde (Polysciences, Germany) in PHEM buffer (60 mM PIPES, 25 mM HEPES, 2 mM MgCl2, 10 mM EGTA, pH 6.9), washed with PHEM buffer and incubated for 1h with 1% periodic acid in PHEM buffer. After washing the cell pellet was infiltrated with 12% gelatin (Sigma, Germany) in PHEM buffer. The sample was cut in blocks and incubated in 2.3 M sucrose in PHEM buffer overnight. The blocks were mounted on aluminum pins and put in liquid nitrogen. 70 nm thin sections were cut with a cryo-ultramicrotome (Leica UC7, Austria) and picked up with 2.3 M sucrose (J.T. Baker, the Netherlands) in 2% methylcellulose (Sigma, USA) solution. The sections were placed on copper grids (Veco, the Netherlands) coated with formvar (Sigma, USA) support film sputtercoated (Leica, Austria) with carbon. Grids were placed on a solid gelatin plate at RT (Sigma, Germany; 2% in 0.1 M phosphate buffer), which was placed in an 37°C incubator (30 min). Grids were washed with PBS/glycine (Sigma, Japan) and subsequently blocked with PBS/1% BSA (Sanquin, Netherlands). Next, GFP was probed (Abcam, rabbit anti GFP; 1:200; 2 hr), washed with PBS/0.1% BSA and incubated with protein A gold (purchased from G. Posthuma, UMC-Utrecht; 1:50 for 30 min). After washing with PBS, grids were post-fixed with 1% glutaraldehyde in PBS. After washing with MilliQ water the samples were contrasted with uranyloxalate-acetate (4% uranylacetate in water, 3.8% oxalic acid, pH 7) for 5 min and methylcellulose/uranylacetate (9 mL 2% methylcellulose, 1 mL 4% uranylacetate) for 5 min on ice.

Data was recorded using the Supra55 scanning electron microscope at 29 kV (SEM; Zeiss, Oberkochen), at 2.5 nm pixel size using the transmission detector essentially the same as described before (Sokol et al., 2015). Large-area scans were generated using the external scan generator ATLAS (Fibics, Canada) and tiles were stitched using VEviewer (Fibics) and exported both as a single TIF and a high-resolution html file, which contains the raw data and is open-access available via www.nanotomy.org.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters including the definitions and minimum values of cells analyzed per sample (N) are reported in the figures and corresponding figure legends. Further details about the statistical analysis of the samples, as well as computer programs used can be found in the accompanying materials and methods section of each experiment.

DATA AND CODE AVAILABILITY

The EM dataset generated during this study (Figures 3D and 3E) is available at www.nanotomy.org. Other datasets from fluorescence imaging, biochemical analysis and modeling, supporting the current study have not been deposited in a public repository, but are available from the corresponding author on request.