Title
The Role of FG Nucleoporins Amino Acid Sequence Composition in Nucleocytoplasmic Transport

Author
Peyro, Mohaddeseh

Publication Date
2020

Peer reviewed|Thesis/dissertation

The Role of FG Nucleoporins Amino Acid Sequence Composition in Nucleocytoplasmic Transport

by

Mohaddeseh Peyro

A dissertation submitted in partial satisfaction of the
requirement for the degree of

Doctor of Philosophy

in

Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Mohammad R.K. Mofrad, Chair
Professor Lisa A. Pruitt
Professor Ahmet Yildiz
Professor Rebecca W. Heald

Summer 2020

Abstract

The Role of FG Nucleoporins Amino Acid Sequence Composition in Nucleocytoplasmic Transport

by
Mohaddeseh Peyro

Doctor of Philosophy in Mechanical Engineering
University of California, Berkeley
Professor Mohammad R.K. Mofrad, Chair

Bidirectional transport of molecules through the nuclear envelope (nucleocytoplasmic transport) is a vital, yet very complex process inside eukaryotic cells. This process is facilitated by thousands of nanoscale pores embedded in the nuclear envelope, named nuclear pore complexes (NPCs). The remarkable feature of NPC is that it facilitates transport in a fast, yet selective manner. Despite numerous studies, the underlying mechanism of NPC's function is unknown. The intricate function of NPC cannot be fully captured via experimental techniques due to small scale and rapid movement of the FG Nucleoporins, which are intrinsically disordered proteins rich in phenylalanine-glycine repeats, responsible for facilitating the transport process.

In this dissertation, using a range of computational approaches, the role of FG Nups in the nuclear transport process is explored. FG Nups do not have a well-defined secondary structure and it is believed that their sequence composition and conformational ensemble are critical for their function in the NPC. In this dissertation, for the first time, specific sequence patterns in the charge distribution of FG Nups were identified that were not observed in other intrinsically disordered proteins. These patterns are extended sub-sequences that only contain positively charged residues, have low charge density, and are located toward the N-terminus of FG Nups. We named these evolutionarily conserved patterns *like charge regions (LCRs)*.

Additionally, the role that LCRs play in the conformational ensemble and function of FG Nups was examined, using coarse-grained molecular dynamics simulations. Our simulations in multiple levels (single Nups, ring cross sections of NPC, and whole NPC) show that number of charged residues in the LCR impact the conformational ensemble of FG Nups and movement of the cargo complex. Our simulations also suggest that the number of charged residues in LCR can regulate the interaction of cargo complex with FG Nups. Since conformational ensemble of FG Nups and their interaction with cargo complex are two governing factors of the transport process, we suggest that LCRs, the unique and evolutionarily conserved features of FG Nups, are major regulators of the nucleocytoplasmic transport.

Dedicated to

Maman, Baba, Mohammad,

Mohammad, Borhaan

Acknowledgement

First, I would like to thank my advisor, Professor Mohammad Mofrad, for his continuous support through this journey. It was a great honor to work under his supervision, and a wonderful experience to work with him. I will always be thankful to him for giving me this amazing opportunity, which allowed me to grow academically, and personally.

I would also like to thank my wonderful committee members, Professor Lisa Pruitt, Professor Rebecca Heald, and Professor Ahmet Yildiz, for their support and their valuable comments and feedback, which significantly improved my research.

I am thankful to current and former Molecular Cell Biomechanics members, Mohammad Soheilypour, Brianna Lee, Vikrum Nibber, Andrew Dickson, Ali Ghavami and Andre Montes for their contribution to this dissertation.

I would also like to thank my friends at Berkeley, for being my second family. I was a blessing to have you as friends. I learned a lot from you and enjoyed spending time with you. I would like to specially thank Hengameh, for being always there for me through these years.

I would like to express my deepest gratitude to my parents, Mahnaz and Masih, for always being there for me. Thank you for your unconditional love and support. Thank you for being the best I could ever ask for. I am blessed to have you as my parents, and I could never thank you enough for all you have done for me. Also thanks to my lovely brother, Mohammad, for being there for my parents when I couldn't be with them. Being away from you guys was the hardest thing I had ever done in my life and I look forward to days we never have to leave each other's side for a long time.

I would also like to thank my wonderful husband, Mohammad, for his never-ending support and positivity throughout this journey. Thanks for being my rock, the person I could always count on. Thanks for being my best friend. Thanks for being the one I was able to share my happiness and sadness with. Thanks for your patience, understanding and love. I could not ask for a better partner through this journey, and through the journey of life.

I would also like to thank my amazing son, Borhaan, for inspiring me to be a better human being every day. You are the best thing that has ever happened to me. Thank you for giving me the passion to live, and being the reason to love. I am beyond grateful to have you in my life.

Table of Contents

Table of Figures

Chapter 1 : Introduction

Nuclear Pore Complex

The nucleus and the cytoplasm in eukaryotic cells are separated by a lipid bilayer, called the nuclear envelope, which is perforated by nano-scale pores known as nuclear pore complexes (NPC) (Figure 1-1). NPCs are the sole biochemical gateway for bidirectional transport of cargos between the cytoplasm and the nucleus [1]. Their function is vital to the cell, with cargoes ranging from different functional proteins to RNAs and ribosomes being transported through NPCs [2]. They are the largest molecular machinery in the cell with a mass of 50 MDa in yeast [3] and are made up of nucleoporins (Nups), with each NPC consisting of approximately 30 different types of Nups [4].

The NPC has an hourglass shape and displays octagonal symmetry. Its scaffold features eight flexible filaments that project toward the cytoplasm as well as another set of eight filaments that project toward the nucleoplasm and are attached to a ring to form the nuclear basket[5] (see Figure 1-1). The NPC has an inner diameter that ranges between 35 nm to 50 nm across different species [6], an outer diameter of ~110 nm and a height range of ~35 nm to ~70 nm [5].

Fast yet selective transport is the unique function of the NPC. Each NPC can handle up to 1000 translocations per second via two modes of nucleocytoplasmic transport, namely active and passive transport[2]. While passive transport is simply free diffusion of smaller cargos through the NPC, active transport requires a family of nuclear transport receptors (NTRs) to bind to cargos and transport them across the pore. It is suggested that a molecule's size determines whether it will be transported passively or actively through the NPC. While small molecules (molecular mass ~20–40 kDa, diameter of ~5–9 nm) passively diffuse through the pore, larger molecules (M greater than 40 kDa and up to ~25 MDa, diameter of up to ~40 nm) require transporters [2,4,7].

One of the primary open questions about the NPC function is how it could enable nucleocytoplasmic transport to happen in a fast yet selective manner [8,9]. Despite years of research on NPCs, this feature, known as the transport paradox, is still an active area of research, calling for new approaches to studying NPC function.

FG Nucleoporins

FG nucleoporins (FG Nups) are a subclass of proteins that make up one-third of all Nups. They are defined by intrinsically disordered subsequences rich in Phenylalanine-Glycine (FG) repeats. FG Nups line the central channel of the NPC, and are made of a structured domain that provides anchoring to the NPC and an intrinsically disordered domain that protrudes toward the center of the NPC [10]. These disordered domains possess highly flexible structures and exhibit little or no secondary structure [11–13] (Figure 1-1). FG repeats constitute more than 29% of the average FG Nup sequence[14], indicating the substantial role of these regions in nucleocytoplasmic transport. Weak and transient hydrophobic interactions between NTRs and FG motifs, are the main driving forces that stimulate active translocation of cargos through the NPC [15]. Therefore, FG Nups are the major role players in active transport [1].

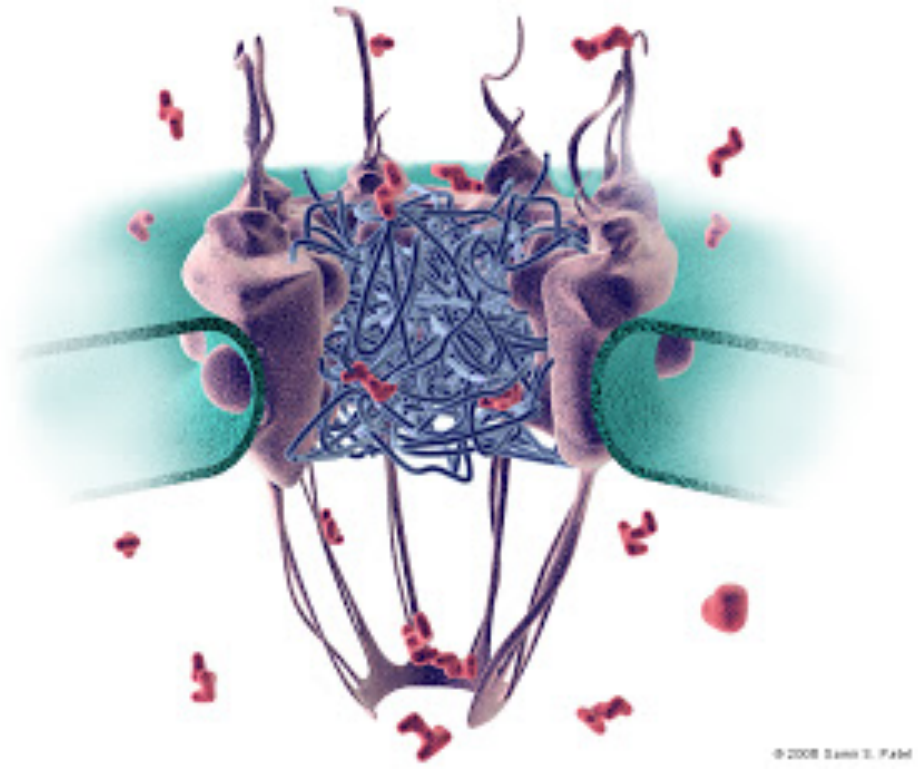

Figure 1-1: (Nuclear transport illustrations by Samir S. Patel, https://sites.google.com/site/sspatel/nuclearporecomplex2)Nuclear Pore Complex (NPC). NPC is embedded in the nuclear envelope. The filaments on top are cytoplasmic filaments. The filaments on the bottom form a basket shape, also known as nuclear basket. FG Nups are intrinsically disordered proteins that cover the lumen of the NPC and are responsible for regulating the transport process.

Nuclear Transport

Large molecules require active transport to successfully travel through the NPC [16]. NTRs provide a mechanism by which they can travel through the NPC in a fast and selective manner. NTRs have high affinity to FG repeats, enabling them to interact with FG repeats to overcome the permeability barrier, while undergoing multiple binding and unbinding events [8]. These interactions must be weak and transient to enable a fast transport of the cargo through the NPC [17,18]. NTRs recognize cargos via specific sequences called nuclear localization signals (NLSs) or nuclear export sequences (NESs).

The direction of active transport is dependent on a RanGTP/GDP gradient that exists between the nucleoplasm and the cytoplasm[19]. RanGTP has a higher concentration in the nucleus, and facilitates unbinding cargo from cargo complexes in the nucleus. On the other hand, when RanGTP binds to NTRs, it helps bind exporting cargos. On the cytoplasmic side, hydrolysis of RanGTP to RanGDP leads to unbinding of the cargo from cargo complex. RanGEF changes GDP to GTP and is located in the nucleus. RanGAP, on the other hand, hydrolyzes RanGTP to RanGDP and is located in the cytoplasm. NTF2, one of the main NTRs, transports RanGDP back to nucleus [8]. Higher abundance of RanGTP in the nucleus and RanGDP in the cytoplasm leads to the RanGTP/GDP gradient that enables transport through the NPC.

Existing models of Nucleocytoplasmic transport

Despite extensive research on the underlying mechanisms of NPC function, there is still no consensus on how cargos are transported through the NPC. So far, a few models have been proposed for nucleocytoplasmic transport. The selective phase model [20–22] suggests that FG Nups inside the NPC form a homogeneous meshwork (hydrogel) via the weak interaction between FG motifs. Due to the high affinity of transporters to FG repeats, active transport of cargo complexes happens through transient breaking of the FG-FG cross-links and breaking of the gel network by transporters. From another point of view, the virtual gate model [23] introduces the brush-like structure formed by FG Nups inside the NPC as an entropic barrier. This FG Nup brush repels nonspecific cargos, i.e. cargos that have not recruited transporters, while cargo-complexes can overcome this barrier due to their weak, transient interactions with FG-repeats. The reversible collapse model [24] justifies the active transport of cargos through the NPC by a conformational change of FG Nups (toward their anchoring point) in the presence of NTRs. The reduction-of-dimensionality model [25] suggests that the transport of cargos through the NPC is facilitated by a two-dimensional random walk of cargo-complexes over a so-called FG-repeat surface formed on the wall of the transport channel. Lastly, the forest model [11,26,27] considers various biophysical characteristics of FG Nups, including Stokes radius and size of individual domains, and suggests that FG Nups show two distinct categories of disordered domains: FG domain and stalk domain. While stalk domains possess a relaxed polymer conformation, FG domains show a more collapsed coil conformation due to the interactions between their residues. These two distinct regions cooperate to form a forest-like landscape inside the NPC and provide two separate pathways.

Different experimental and computational methods have been utilized to study nucleocytoplasmic transport. Due to the complex nature and intricate structure of the NPC, the detailed underlying mechanism of transport through the NPC is usually not tractable via experimental methods. Computational methods, on the other hand, can easily explore the transport mechanism with a higher temporal and spatial resolution. Nevertheless, due to high computational expense, these methods are usually unable to reach timescales high enough to fully describe the transport process. Therefore, the spatiotemporal scale that is necessary for unveiling the transport mechanism is currently unattainable by either experimental or computational techniques [28]. However, integration of experimental and computational techniques can lead to invaluable insights about the function of the NPC.

Why should we focus on FG Nups?

Although nucleocytoplasmic transport has been the focus of many computational and experimental studies in the past few decades [3,11,13,20,29–34], the role of sequence composition of FG Nups has only recently drawn attention amongst researchers of the field [11,26,27,30,31,35]. It has been observed that compositional changes in the sequences of FG Nups significantly influence the distribution of FG Nups inside the pore [30,31]. In addition, specific patterns have recently been discovered in FG Nup sequences[26,27,35] that are believed to play a major role in regulating the distribution of FG Nups inside the pore.

These observations suggest that nucleocytoplasmic transport is strongly dependent upon unique features embedded in the sequences comprising FG Nups. In line with this observation, several diseases including primary biliary cirrhosis, cancer, viral infection, triple A syndrome, and Alzheimers have been directly associated with mutations in the sequences of different FG Nups [36–38]. Therefore, exploring how these sequence-based features determine the NPC function can significantly improve the current understanding of nucleocytoplasmic transport and its relation to different diseases.

Content of this dissertation

In this dissertation, we use a myriad of different computational approaches to study NPC and FG Nups from different perspectives.

In Chapter 2, multiple bioinformatics approaches are adopted with the aim of uncovering unique features in the sequences of FG Nups. Our sequence analysis demonstrates a specific distribution of charged residues in the sequences of FG Nups, which will be referred to as 'like charge regions' or LCRs. Our results on the distribution of FG motifs, polar residues, and LCR sequences demonstrate novel aspects of the formation of the FG network at the center of the NPC.

In Chapter 3, we explore whether LCRs exist in other intrinsically disordered proteins (IDPs) or are specific to FG Nups. Accordingly, we analyzed a large dataset of intrinsically disordered proteins extracted from the DisProt database, which stores experimentally verified disordered proteins. This set of proteins, referred to as DisProt in this dissertation, was analyzed for presence of LCRs. Our results suggest that LCRs are amino acid sequence patterns exclusive to FG Nups, which specifically distinguish them from DisProt. Excluding LCRs from the sequences of FG Nups make them similar to Disprot proteins in terms of charge distribution.

In Chapter 4, we explore the significance of LCRs by examining the biophysical behavior of FG Nups, with the ultimate goal of shedding light on the mechanism of nucleocytoplasmic transport and the role of FG Nups in this selective process. We studied the role of LCRs in FG network formation via a recently developed one-bead-per-amino-acid coarse-grained molecular dynamics model of the NPC. Our results showed that LCRs are important for keeping the FG Nups conformation from collapsing into highly dense aggregates and for facilitating cross-interaction between FG Nups.

In Chapter 5, we extend the studies in Chapter 4 by adding a cargo complex to the analysis to study the effect of LCRs in the conformational ensemble of FG Nups in presence of a cargo complex, movement of cargo complexes in the NPC, and interaction of FG Nups with the cargo complex. Our results suggest that LCRs influence the interaction between FG Nups and the movement of cargo complex. Moreover, the effect of LCRs on the interaction of FG Nups with the cargo complex can be dramatic. Nucleocytoplasmic transport phenomenon is governed by FG Nups' conformational ensembles as well as their interaction with cargo complexes. Accordingly, our multi-method analysis of NPC in multiple levels, suggests a substantial role for LCRs in NPC function and nucleocytoplasmic transport.

Chapter 2 : Evolutionarily Conserved Sequence Features Regulate the Formation of the FG Network at the Center of the Nuclear Pore Complex

The work presented in this Chapter appeared in the following publication:

Peyro, M.*, M. Soheilypour*, B. L. Lee, and M. R. K. Mofrad. "Evolutionarily conserved sequence features regulate the formation of the FG network at the center of the nuclear pore complex." Scientific Reports 5, no. 1 (2015)

*These authors contributed to this paper equally.

Abstract

The nuclear pore complex (NPC) is the portal for bidirectional transportation of cargos between the nucleus and the cytoplasm. While most of the structural elements of the NPC, i.e. nucleoporins (Nups), are well characterized, the exact transport mechanism is still under much debate. Many of the functional Nups are rich in phenylalanine-glycine (FG) repeats and are believed to play the key role in nucleocytoplasmic transport. In this work, a bioinformatics study is conducted on more than a thousand FG Nups across 252 species. Our results reveal the regulatory role of polar residues and specific sequences of charged residues, named 'like charge regions' (LCRs), in the formation of the FG network at the center of the NPC. Positively charged LCRs prepare the environment for negatively charged cargo complexes and regulate the size of the FG network. The low number density of charged residues in these regions prevents FG domains from forming a relaxed coil structure. Our results highlight the significant role of polar interactions in FG network formation at the center of the NPC and demonstrate that the specific localization of LCRs, FG motifs, charged, and polar residues regulate the formation of the FG network at the center of the NPC.

Introduction

Due to the complex and highly dynamic nature of the NPC and its constituents, current imaging and experimental techniques are unable to fully capture a detailed picture of the function and mechanics of FG Nups [29,39]. Computational methods can be effective in predicting the *in vitro* and *in vivo* behavior and function of FG Nups in different spatiotemporal scales. There is a growing body of computational studies each looking through the nuclear pore via a different perspective. Some studies have specifically focused on modeling of the behavior of single FG Nups. Krishnan et al. studied the FG domains of Nup116 through molecular dynamics (MD) simulations [40]. In another comprehensive study on all Nups, Devos et al. combined computational and biochemical methods to assign folds to almost 30 different types of proteins found in the yeast NPC [14]. Other studies examined the interaction between Nups and Kaps. For example, Zhao et al. investigated the interaction of CRM1 with Tpr, a nuclear pore protein located inside the nuclear basket, and identified nine possible binding sites for Tpr on CRM1 [41]. Aggregation and interaction of FG Nups has been the focus of other computational studies. For example, Dolker et al. conducted MD simulations on a few FG repeat sequences and investigated the aggregation of FG Nups and the role of hydrophilic linkers in this phenomenon [42]. Miao and Schulten performed MD analysis on a group of FG Nups grafted on a planar surface [7]. Using a computational framework, Alber et al. determined the molecular architecture of NPC from a set of biophysical and proteomic data [3,43]. Wolf et al. performed an analytical/computational analysis on the significance of eight-fold symmetry of the NPC scaffold [44]. Using a coarse-grained model, Ando et al. [26] performed a comprehensive study on the effect of the sequence on the morphology of single FG Nups and grafted rings of FG Nups. In addition, there are several studies focused on the modeling of the NPC and its constituents as a whole. Ghavami et al. investigated the distribution of FG Nups in the channel by means of a high resolution coarse-grained molecular dynamics model and showed that the distribution is encoded in the sequences of FG Nups[30]. Tagliazucchi et al. [31] also used a computational model to simulate the effect of sequences of FG Nups on translocation through the NPC. Other studies incorporated the transporters in addition to the NPC and its constituents, in an effort to provide a more comprehensive understanding of the NPC. Mincer and Simon used the pairwise agent interaction with rational superpositions (PAIRS) model to predict factors such as the bimodal cargo distribution, the effect of the number of LS tags and RanGTP gradient.[34]. They modeled rings of nups and the NPC as a whole and verified their model using a variety of experimental data. Osmanovic et al. modeled FG Nups as freely jointed chain polymers to explore the effect of NTRs on polymeric structure in the NPC [45]. Moussavi-Baygi et al. studied the functional state of the NPC and measured transportation time versus cargo size via a coarse-grained Brownian dynamics model [29,32]. Zilman et al. developed a mathematical model of the NPC to discuss robustness, efficiency and selectivity of the NPC [46]. Colwell et al. [47] investigated the electrostatic interactions between FG Nups and transporters and found that these interactions play a key role in nucleocytoplasmic transport, as transporters contain negative charge and the NPC's permeability barrier is positively charged. Patel et al. [48] investigated the role of different FG domains in forming the permeability barrier in yeast FG Nups. Using agent-based modeling (ABM), Azimi at al. argued that the presence of an affinity gradient

between Kap and FG Nups across the NPC maximizes the transport rate through the nuclear pore [49]. Azimi et al. further extended their ABM model to study the dynamics of mRNA export through the nuclear pore [50]. Although these studies shed light on the behavior of specific domains of FG Nups, further examinations are needed to capture a comprehensive picture of the overall characteristics, behavior and function of Nups inside the NPC.

Arguably, the well-known correlation between sequence, structure and function in proteins does not hold in the case of IDPs [51–58]. As a result, many studies have incorporated non-classical approaches, e.g. sequence composition and physical property distribution along the sequence, to decipher behavior and function of IDPs [27,59,60]. Therefore, the classical correlation between sequence, structure and function for structured proteins would be translated to a relationship between sequence composition, structural properties and function in IDPs [27,61]. Ando et al. [27] performed a comprehensive bioinformatics analysis on a database of 1138 FG Nups from 252 species. They found important evolutionarily conserved features in FG Nups regarding FG motifs and charged clusters density and location. A slightly modified version of this database was adopted here for a deeper analysis of the distribution of different types of residues along the sequences of FG Nups. We analyzed the number density (the number of a specific type of residue divided by the length of the entire sequence) and location of polar residues and found specific sequences of charged residues common in all FG Nups.

In this study, multiple bioinformatics approaches are adopted with the aim to uncover unique features in the sequences of FG Nups. A large database of FG Nups is used to extract their evolutionarily conserved characteristics. Comparing FG Nups against a database of experimentally approved disordered proteins, i.e. DisProt [62], it is observed that there is a considerable distinction in number density and distribution of charged and polar residues among the two groups of proteins. More in-depth analysis of FG Nup sequences reveals that being rich in FG repeats is not the only evolutionarily conserved property the FG Nups possess. Our sequence analysis demonstrates a specific distribution of charged residues in the sequences of FG Nups, which shall be referred to as ‘like charge regions’ or LCRs. Our results on the distribution of FG motifs, polar residues, and LCR sequences demonstrate novel aspects of the formation of FG network at the center of the NPC. Our approach is verified by comparing the results against available data [11,26,27,30].

Results

FG Nups do not share the same properties with other disordered proteins

The combination of low mean hydrophobicity and relatively high net charge are suggested to be the prerequisites for the lack of structure in IDPs [53,58]. In order to explore these properties in FG Nups, they were compared against proteins of the DisProt database using the charge-hydrophobicity plot [58] (results not shown)— please refer to "Methods" for the definition of database of FG Nups and Disprot. Although FG Nups feature average hydrophobicity patterns similar to those in DisProt proteins (less than %7 difference on average), they exhibit significantly lower absolute net charge. The mean absolute net charge, i.e. absolute net charge divided by length of the sequence, was compared in FG Nups versus that in DisProt proteins (see Figure 2-1). Interestingly, DisProt proteins show a considerably wider range of mean absolute net charge in their sequences compared to FG Nups. Average mean absolute net charge of FG Nups is about one-fifth of that in DisProt proteins. Therefore, it could be deduced that lower net charge is a shared feature of FG Nups across different species.

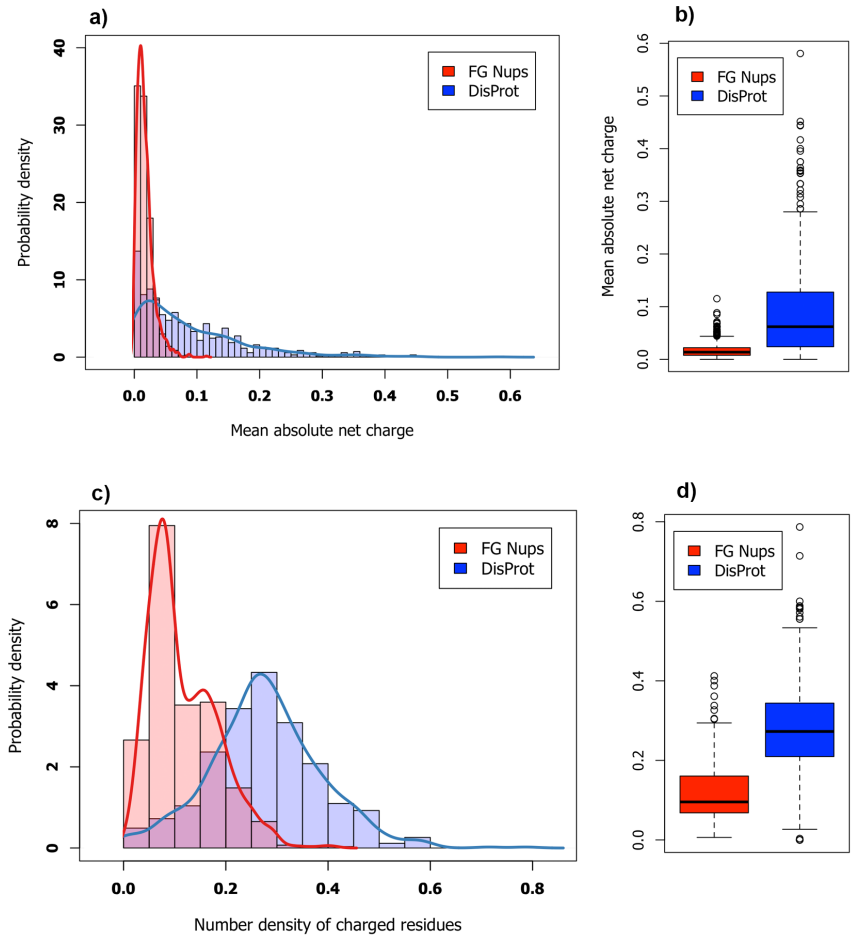

Figure 2-1: Comparison of FG Nups and DisProt proteins in terms of their charge content and distribution. Top Row: Probability density distribution (a) and box plot (b) of mean absolute net charge, defined as absolute net charge divided by the length of the sequence. FG Nups exhibit lower absolute net charge compared to the

DisProt database — please refer to "Methods" for the definition of database of FG Nups and Disprot. Interestingly, DisProt proteins show a much wider range of mean absolute net charge compared to FG Nups, implying that FG Nups share the same charge content (number density and distribution) characteristics across species. Bottom Row: Probability density distribution (c) and box plot (d) of number density of charged residues, defined as the number of charged residues divided by the length of the entire sequence. Sequences of FG Nups feature lower density of charged residues compared to DisProt. The two comparisons imply that charged residues are not only responsible for the disorderedness of FG Nups, but are also carefully distributed to give specific functionalities to FG Nups. Note that the area under each curve equals one, which explains why the values on vertical axes could exceed unity.

Most of the charged amino acids within the sequences of FG Nups are located within the FG linkers, i.e. domains that link FG motifs together. Analysis of these linker regions in FG Nups reveals that about 93% of them possess a net charge of either zero or one (Figure 2-S1,[42]).

Therefore, although net charge is suggested as a key parameter for distinguishing structured and disordered proteins [58], in the case of FG Nups the total content (number density) of charged residues would be a more meaningful parameter [11]. Comparing number density of charged residues, i.e. number of charged residues divided by length of the sequence, between FG Nups and DisProt proteins demonstrates that average charge density of FG Nups is approximately one-half of that of DisProt proteins (see Figure 2-1). Therefore, FG Nups not only have lower net charges but also exhibit significantly lower number density of charged residues as compared to DisProt proteins.

These findings are not completely in line with what is previously stated about FG Nups [63]. Although high net/total charge leads to an increase in the electrostatic repulsion within Nups [58], it seems that, in the context of the NPC, charged residues are not solely responsible for disorderedness of FG Nups. To gain a better insight on the distribution of charges inside the nuclear pore, the distribution of positive and negative charges along the sequences of FG Nups in yeast NPC were examined (see Figure 2-2). Evidently not all regions of FG Nups are rich in charged residues; rather there are regions that include very few charged residues. Consequently, one could speculate that disordered behavior is desired to the extent that it satisfies the function of an FG Nup. In line with this hypothesis, some FG Nups are observed to form local compact collapsed coil configurations to regulate nucleocytoplasmic transport [11,26,64]. For instance, Nsp1, one of the most studied FG Nups from yeast, possesses two distinct regions with one rich in FG motifs, i.e. FG domain, and the other rich in charged residues, i.e. stalk domain [11,26]. As claimed by Ando et al. [26], stalk domains are responsible to extend all the way to the central axis of the pore, locating FG domains in the center of the pore to form the central transporter [11,26]. Central transporter is one of the proposed NPC transport mechanisms in which a network of FG Nups is formed in the center of the NPC. Active transportation needs the cargo complex to interact with this hydrophobic network to be able to pass through the nuclear pore [11,26].

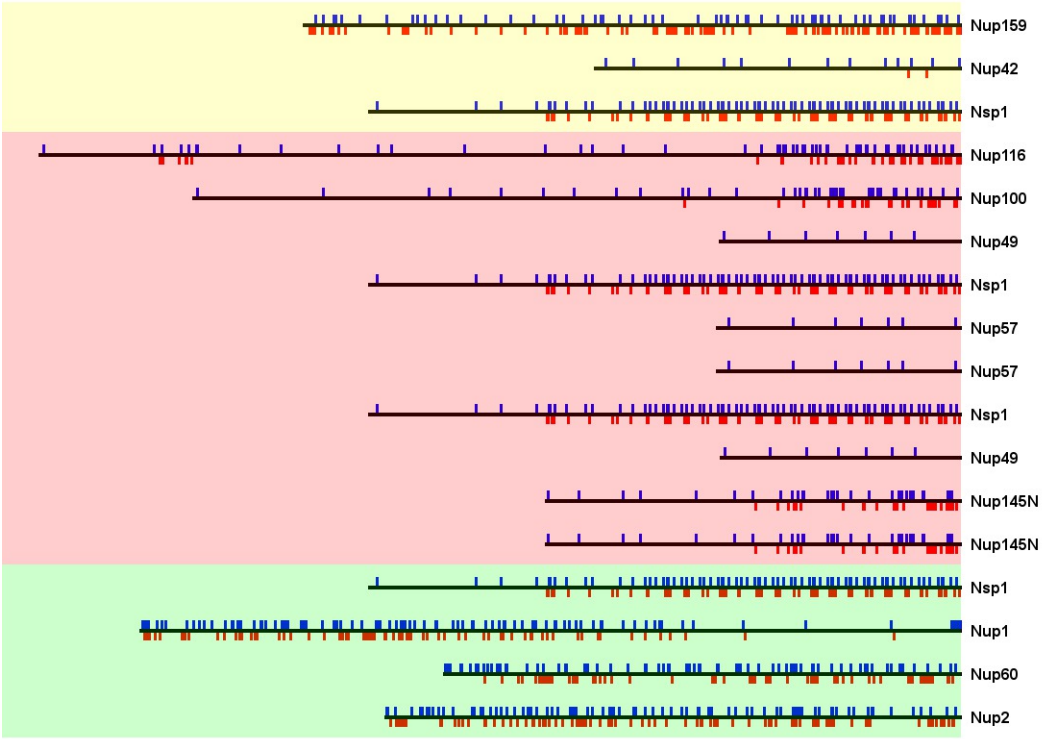

Figure 2-2: Distribution of charged residues along the sequences of yeast FG Nups. Positive charges are shown in blue while negative charges are shown in red. The far right side of the figure represents the NPC wall. FG Nups are drawn in their extended form. It is clearly observable that not all domains of FG Nups are rich in charged residues, which implicitly proposes specific functionalities for different regions of each FG Nup (also discussed in [11] and [27]). Yeast is chosen as the example to represent the distribution of charged residues in the sequence of FG Nups since it is the most well studied species.

Further analysis of amino acid abundance, defined as the number of residues in a certain group normalized by the number of residues in the entire database, was performed using FG Nups and DisProt databases and revealed more significant differences between the two groups. Amino acid abundance of the two groups sorted by disorder-promoting property of each amino acid with Tryptophan as the most order-promoting amino acid and Proline as the most disorder-promoting one [65] is shown in Figure 2-3. FG Nups are no exception to the fact that IDPs have a low content of hydrophobic residues (Ile, Leu, Val, Trp, Tyr, Phe) compared to structured proteins [66,67]. However, FG Nups show a significantly lower content (or number density) of hydrophobic residues as compared to DisProt proteins except for Phenylalanine, since FG Nups are characterized by abundance of Phenylalanine and Glycine residues. One could speculate that FG Nups compensate for the order-promoting property of large number of Phenylalanine by lowering the number of other order-promoting residues. Among order-promoting amino acids, only Phenylalanine and Asparagine have higher abundance in FG Nups. Asparagine is a polar residue and interestingly FG Nups exhibit higher contents (or number densities) of all polar residues (Asn, Ser, Thr, Gln), likely to regulate FG linker interactions forming a network of FG Nups [42]. Furthermore, in line with the aforementioned comparison of charge content between FG Nups and DisProt proteins, the number density of individual charged residues is also lower in FG Nups compared to that in DisProt database.

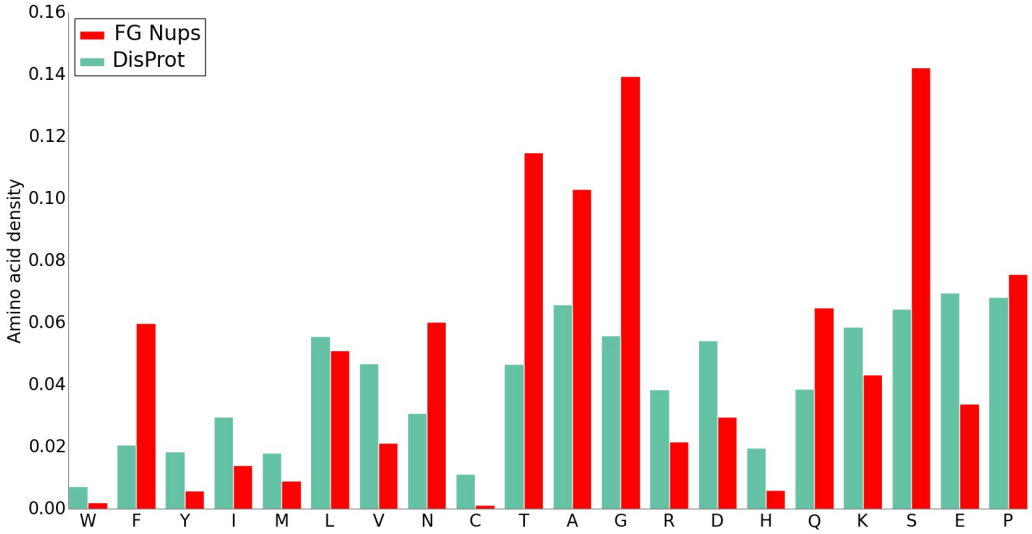

Figure 2-3: Amino acid abundance of FG Nups and DisProt proteins. Amino acids are sorted according to their disorder-promoting property, with Tryptophan as the most order-promoting amino acid and Proline as the most disorder-promoting one [65]. Both groups show low abundance of hydrophobic residues (Ile, Leu, Val, Trp, Tyr, Phe) compared to structured proteins. However, FG Nups have a lower content (number density) of hydrophobic residues compared to those in DisProt proteins, except for Phenylalanine. Furthermore, FG Nups demonstrate higher abundance of polar residues and lower abundance of charged residues compared to DisProt proteins. Abundance is defined as the number of residue normalized by the number of residues in the entire database.

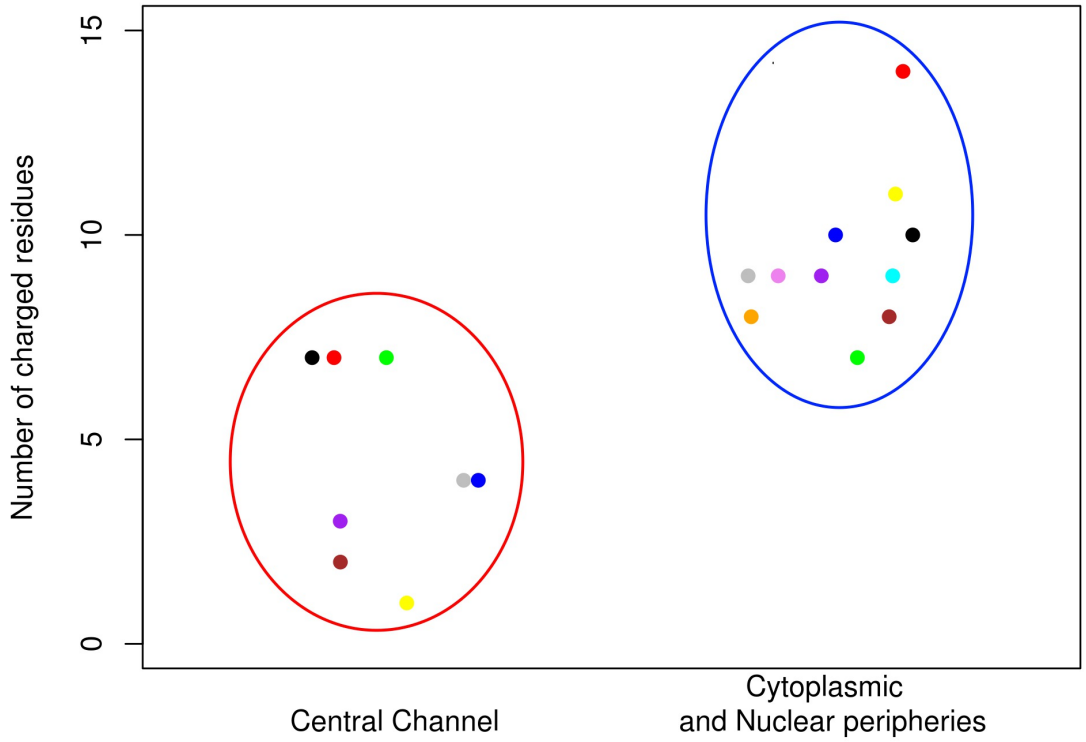

Figure 2-4: Comparison of central channel and periphery FG Nups in terms of number of charged residues embedded in LCR. Data points are jittered to enable the reader to observe all of the data points. FG Nups fall into two distinct groups, with the central channel group possessing LCRs of lower than eight and the periphery group containing LCRs of larger than seven residues. The analysis is conducted on FG Nups from yeast and

vertebrates with known location, i.e. central channel or periphery. Analyzed FG Nups are: [Central Channel] Nsp1 (blue), Nup49 (black), Nup57 (red), Nup145N (green), Nup62 (purple), Nup54 (yellow), Nup45 (gray), Nup58 (brown), [Cytoplasmic and Nuclear peripheries] Nup1 (orange), Nup2 (violet), Nup159 (cyan), Nup42 (blue), Nup100 (black), Nup116 (red), Nup98 (green), Nup358 (purple), Nup214 (yellow), Nup153 (gray), Nup50 (brown). Location information is obtained from [68].

Based on this sequence analysis of FG Nups, it could be concluded that in addition to FG motifs, charged and polar residues might be other important role-players in the behavior of FG Nups inside the pore. Therefore, a more in-depth analysis is conducted to extract unique patterns in the distribution of these types of residues along the sequence of FG Nups across all species.

Like charge regions can partly explain characteristics of FG Nups network

Like charge regions (LCRs), also referred to as uncompensated charged groups [58], are specific domains of protein sequences that contain only one type of charged residues, i.e. either positively or negatively charged. These regions are suggested to enhance the disorder behavior of IDPs [58]. We focus our bioinformatics analysis on largest LCRs, as they exhibit interesting evolutionary conserved characteristics (explained in the following sections). Therefore, to avoid confusion and for the sake of simplicity, hereafter, LCR refers to the largest LCR. The number of charged residues embedded in each LCR was studied among all FG Nups. The histogram of LCRs among all FG Nups is shown in Figure 2-S2. FG Nups indicate a wide range of number of charges in LCRs mostly falling between two to ten charged residues. Investigating FG Nups with known locations, i.e. central channel or cytoplasmic and nuclear peripheries, reveals that FG Nups can be categorized into two groups; those located in peripheries possess LCRs of equal or more than seven charged residues, while this parameter is less than eight for central channel FG Nups. Since the location information of most of the FG Nups is lacking, this inference is based on a partial fraction of FG Nups (i.e. only those with experimentally verified localization data inside the NPC), however it could potentially offer a distinguishing parameter that helps isolate the two groups of FG Nups, i.e. central channel versus peripheral FG Nups. Further computational and experimental studies are indeed necessary to unravel the functional significance of number of charges in LCR.

Another interesting characteristic of LCRs is reflected in the distinction between positive and negative LCRs. The distribution of FG Nups within the length-content space of LCRs is shown in Figure 2-5, where the horizontal axis is showing the length of LCRs and the vertical axis demonstrates the charge content (calculated as number density, i.e. number of charged residues per the length of the LCR) in that region. The length of each LCR is defined as the number of residues between the first and the last charged residue within the LCR. Interestingly, negative LCRs are distributed closely with short lengths and high content (number density) of charged residues. On the other hand, positive LCRs exhibit significantly lower content (number density) of charged residues. Furthermore, the positive LCRs are typically longer than the negative LCRs but show a wide range of lengths among different species. It could also be observed for yeast FG Nups in Figure 2-2, with seven out of ten FG Nups including a long and low-content (low-density) positive LCR. Therefore, it can be concluded that this unique sequential feature of FG Nups is an

evolutionarily conserved characteristic shared across all species. It is worth mentioning that clustering analysis reveals that positive LCRs have almost zero overlap with charged clusters (results not shown), emphasizing the low content (i.e. low number density) of positive LCRs in charged residues. This observation is in line with the results of a previous work on yeast FG Nups [11]. Dividing regions of FG Nups into low and high content charge domains, it was discovered that some yeast FG Nups only include high density charge regions, while some only include low density charge region, and some, e.g. Nsp1 and Nup1, include both high and low density charge regions [11]. Therefore, positive LCRs that have virtually no overlap with charge clusters correspond to low density charge regions, while charge clusters correspond to high density charge regions.

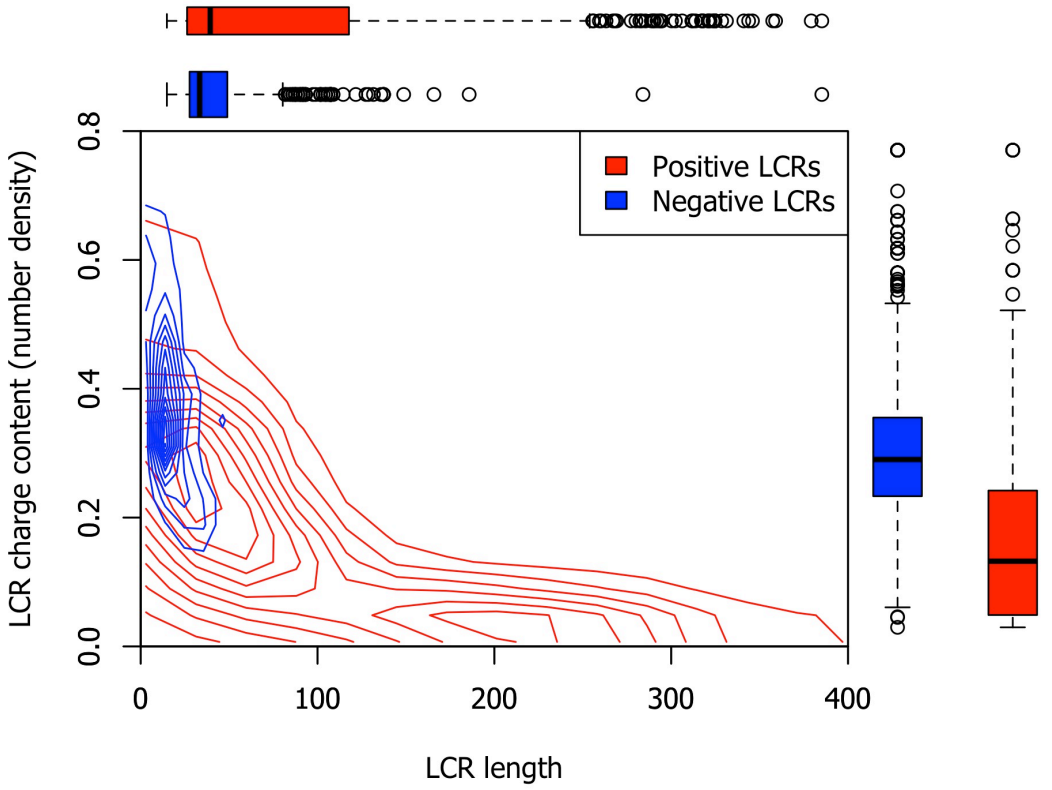

Figure 2-5: Distribution of FG Nups in length-content space of LCRs. Positive LCRs are shown in red while negative LCRs are shown in blue. Charge content here refers to the charge number density defined as number of charged residue normalized by the length of LCR. Negative LCRs are short and contain high charge density, while positive LCRs show very low charge density and longer sequences. In addition, positive LCRs have a wide range of lengths across different species.

Polar and charged residue contents are strongly correlated in FG linkers

The histogram for the polar residue content in linker regions of 10-30 residues in length across all FG Nups is shown in Supplementary Figure 2-S3. A wide range of polar

residue number densities is exhibited in linker regions, varying from zero to almost 100% with a normal distribution around 50%. However, a more in-depth analysis of the distribution of polar residues shows a strong correlation between the number density of charged residues and polar residues in FG linkers. The average density of polar versus charged residues is shown in Figure 2-6. It is clear that the higher number densities of charged residues in FG linkers tightly corresponds to a lower number density of polar residues. In order to show the significance of the evolutionarily conserved trend observed here, the same analysis was performed on a control group. The control group, Disprot proteins, exhibited a similar pattern but the trend was sharper in FG Nups, implying that FG Nups are more likely to have polar residues in their FG linkers when the number of charged residues decreases.

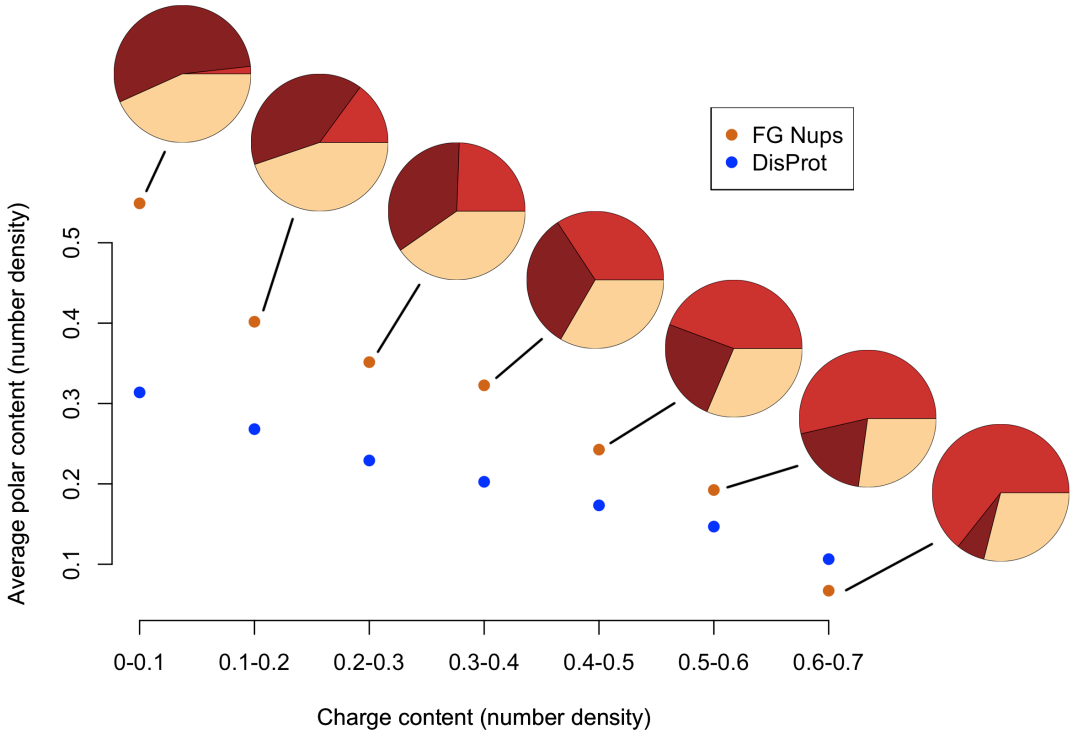

Figure 2-6: Average polar residue content versus charged residue content in FG Nups and Disprot proteins. For FG Nups, only FG linkers with the length of 10 to 30 residues are taken into account. For Disprot proteins, random sequences of 10-30 residues long are selected from Disprot proteins. For the sake of simplicity, these sequences from both groups are called linkers. Each point indicates the average number of polar residues among all linkers falling within that specific range of charged residue densities. Data points with at least 50 samples are shown (FG linkers with charged residue number densities of higher than 0.7 are excluded). Pie charts show average charged residue number density (red), average polar number density (dark brown), and average hydrophobic residue number density (cream) for the corresponding data point. Average number density of polar residues in FG Nups displays a sharp decreasing trend with respect to an increase in the number density of charged residues as compared to Disprot proteins, implying that FG Nups are more likely to have polar residues as the number of charged residues decreases. Disprot proteins are compared to FG Nups as a control group to show the significance of the observed evolutionarily conserved trend.

This conserved feature of FG Nups could be directly related to the results observed in computational studies on Nsp1 [26,42]. Linker regions of Nsp1 are observed to

interact via polar interactions [42]. Furthermore, the stalk domain of Nsp1, which is identified by linkers dense in charged residues with low densities of polar amino acids, forms a relaxed-coil configuration [26]. These observations collectively suggest that highly charged FG linkers with low densities of polar residues form relaxed coil configurations. On the other hand, FG linkers with a high density of polar residues and a low number density of charged residues act as facilitators for efficient interaction of FG motifs and form collapsed domains[26,42].

Clustering analysis reveals relative localization of FG motifs, charged and polar residues

A comprehensive study by Ando et al. [27] on FG clusters and charged clusters demonstrated that in 80% of the cases, there is no overlap between these clusters. In other words, they have a polarity in their location along the sequences of FG Nups. Here, we extend the analysis by taking into account the polar residues as well. Overlap analysis between clusters of charged and polar residues (Figure 2-7) demonstrates a significant polarity between the two types of clusters. The low overlap indicates that charge-rich regions, named stalk domains [27], are low in polar residue content. The same analysis was performed on a control group (Disprot proteins) and the results are shown in Supplementary Figure 2-S4. The relatively high overlap between charged and polar clusters of Disprot proteins shows the significance of the evolutionarily conserved trend observed in Figure 2-7. On the other hand, a similar analysis between clusters of FG motifs and polar residues (Supplementary Figure 2-S5) demonstrates a wide range of overlap between FG and polar clusters. However, further analysis reveals a deeper relation between polar and FG clusters. To gain a better insight on the distribution of polar residues along FG Nups, a spatial analysis was performed on polar clusters by dividing FG Nup sequences into three zones (see Figure 2-8), with zone1 representing the N-terminal tip near the central axis of NPC and zone3 representing the C-terminal end of FG Nups mostly embedded inside the NPC scaffold. The distribution of polar clusters along the sequences of FG Nups, is presented in Figure 2-8 with brighter areas representing a higher number density of polar clusters. FG Nups are rarely rich in polar clusters in their third zone, implying that most FG Nups have a low density of polar residues in their C-terminus zone, which is a mostly structured region. On the other hand, most of the polar clusters are distributed in zones 1 and 2 with more FG Nups being rich in their first zone. While 56% of FG Nups have more than 50% of their polar clusters located in zone1, this value is only 19% for zone2. Based on this observation, we divided FG Nups into two groups according to abundance of polar clusters in the first zone. The FG Nups with more than 50% of their polar clusters located in zone1 are denoted as group A and the rest of the FG Nups as group B. Performing the same overlap analysis between FG and polar clusters of groups A and B (Figures 2-S6 and 2-S7), it is observed that FG Nups in group A have larger overlaps between FG and polar clusters. Group A shows a significantly higher overlap between FG and polar clusters (see Figure 2-9). This observation implies a functional significance of overlapped clusters of FG and polar residues at the tip of FG Nups.

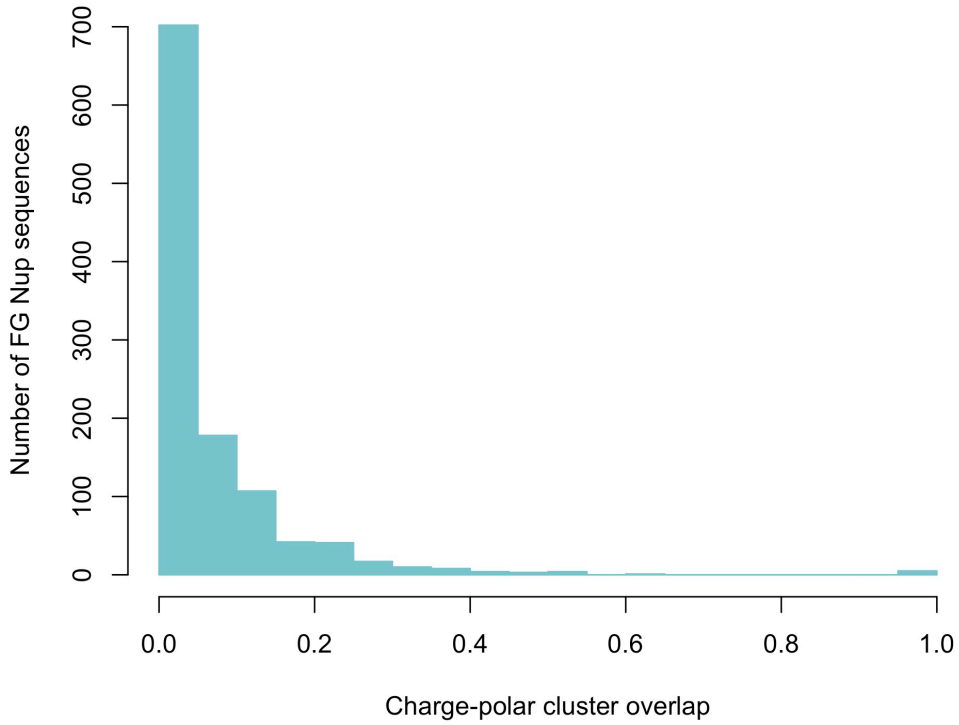

Figure 2-7: Overlap between charged and polar clusters. The relatively low overlap indicates that charge-rich regions, termed stalk domains [26], have a low number density of polar residues. The overlap percentage calculation method is explained in the "Methods", clustering method section. The same analysis was repeated on Disprot proteins as control group (Supplementary Figure S4). Relatively high overlap of polar and charged residues in Disprot shows the significance of the trend observed here.

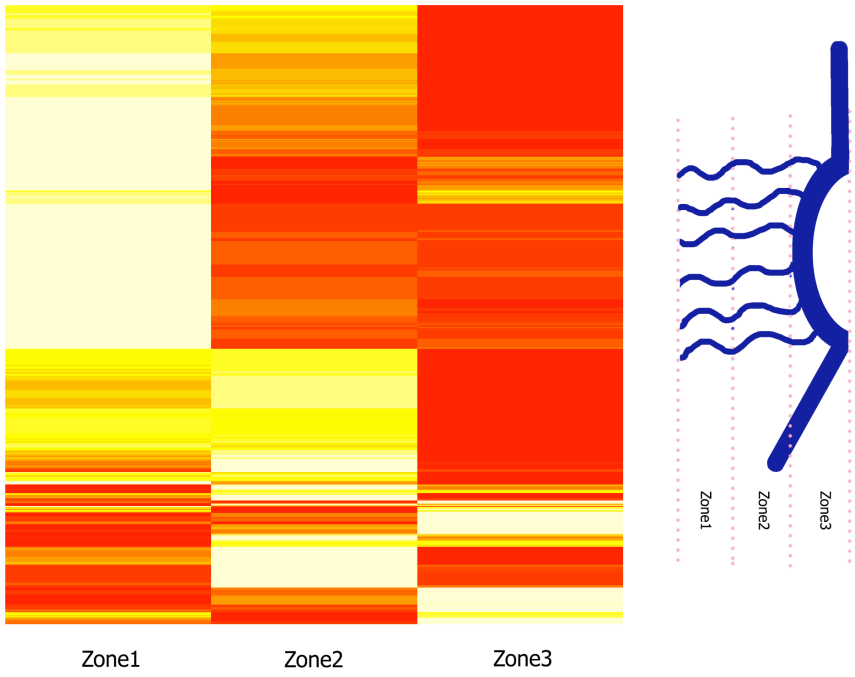

Figure 2-8: Distribution of polar clusters along the sequences of FG Nups. Right) A schematic showing half of the NPC from side-view. FG Nups are divided into three zones with zone1 representing the N-terminus and zone3 representing the C-terminus, which is mostly the structured region of the pore. Left). A heatmap showing abundance of polar clusters in the corresponding region. Brighter areas show higher content of polar clusters. FG Nups can be categorized into three groups. FG Nups of the first group are very rich in polar clusters in their first zone, while FG Nups of the second and third group have more polar clusters in their second and third zones, respectively.

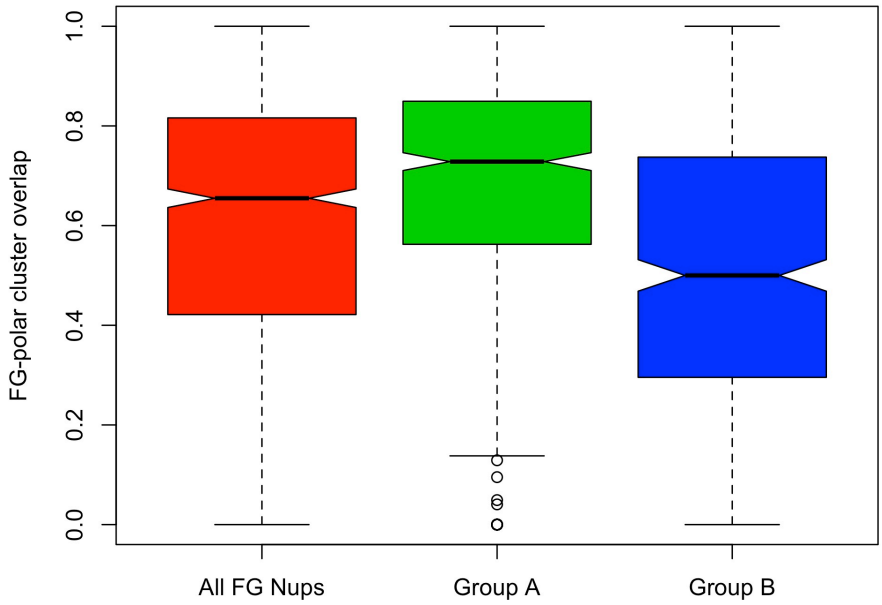

Figure 2-9: Overlap between FG and polar clusters in three different categories of FG Nups. Boxplots of the overlap among all FG Nups (red), FG Nups with more than half of their polar clusters located within the first zone (green), termed Group A, and FG Nups with more than half of their polar clusters located within second and third zones (blue), termed Group B are shown. Higher overlap of FG and polar clusters in group A implies a functional significance of overlapped FG-polar clusters at the tip of FG Nups.

Discussion

The intrinsic bimodality of sequences of some FG Nups has been previously proposed in several studies [11,26,27]. It was observed that FG and charged clusters are well separated within the sequences of FG Nups, forming two distinct regions with different physical behavior and function. However, previous studies were primarily based on charged residues and FG motifs, with some studies focusing only on a very few FG Nups. Here, we considered more than 1000 sequences of FG Nups and extended this hypothesis by extracting other specific features of these two regions along the sequences of FG Nups. We observed that charged clusters and polar clusters have little overlap. We also found specific low-density sequences of charged residues, named LCRs. In line with polar-charge cluster polarity, it was observed that positive LCRs have almost zero overlap with charged clusters. Collectively, one could therefore deduce that sequences of FG Nups are generally partitioned into two domains, one rich in charged residues and one including positive LCRs and a large fraction of polar residues along with FG motifs. This hypothesis is further supported by the overlap analysis between polar and FG clusters. It was previously observed that FG motifs are mainly clustered at the N-terminal tip of FG Nups [27]. We observed that polar clusters have a fair amount of overlap with FG clusters, however, FG Nups that have more than half of their polar clusters located within the first one-third of their sequence show higher overlap of FG and polar clusters. This observation implies that the presence of polar residues among FG motifs is more significant as we move toward the N-terminal tip of FG Nups, i.e. the central axis of NPC.

As mentioned, based on recent observations [11,26,27], sequences of some of the FG Nups can be divided into two functional domains; stalk domain, which is highly charged and relatively straight, and an FG domain that possesses a compact structure and forms the central network of FG Nups, also known as the central transporter [11]. The stalk domain is responsible for bringing the FG domain near the central axis of the pore that is suggested to act as a hydrophobic core, facilitating transportation of cargo complexes via hydrophobic interactions [11]. Our results add to this model by proposing positive LCRs and polar residues as two major regulators of the central transporter. Since nuclear transport signals (NTRs) are negatively charged [47], the presence of positively charged LCRs facilitates translocation of cargo complexes. It was also previously argued that electrostatic interactions are essential for selective transport [47]. In addition, electrostatic repulsion between like charges of LCRs can be considered as a regulatory mechanism that keeps the central network of FG Nups in a certain size. According to the forest model [11], which suggests trafficking zones through the NPC, size of this network will significantly affect nucleocytoplasmic transport. Charged residues embedded in positive LCRs are sparse (Figure 5) to prevent FG Nups from forming a relaxed coil structure while enabling them to maintain a certain radius of the central transporter. Furthermore, it can be postulated that LCRs allow the central transporter to possess sufficient amount of mobility and dynamic behavior via electrostatic repulsion of like charges, hindering hydrophobic forces from turning the network into a more compact structure. On the other hand, the presence of polar residues would improve the formation of the central network of FG motifs via polar interactions and facilitate cargo transport. This is supported by the computational study conducted on a number of segments of FG repeat sequences. It was

observed that presence of hydrophilic linkers and polar residues promote FG repeat meshwork formation [42].

Collectively, our bioinformatics study on sequences of FG Nups enables us to divulge the role of polar residues and LCRs in the formation of the central transporter, which was not previously examined. The sequential features discussed above are cooperating together to enhance nucleocytoplasmic transport. Cooperation of positive LCRs, FG motifs and polar residues is schematically summarized in Figure 2-10. The central transporter is positively charged and is dynamic enough due to the charge repulsion in LCRs. In addition, low charge densities of positive LCRs assists these domains of FG Nups to form a collapsed coil conformation [11]. Furthermore, the central transporter has a high density of polar residues to promote FG repeat hydrophobic interaction and meshwork formation. Finally, the central FG network is placed at the center of the nuclear pore by the other domain of proteins that have a high number density of charged residues and a low density of polar residues. All these sequential features are extracted from more than a thousand FG Nups across 252 species and therefore the results can be considered as the evolutionarily conserved features and function of the NPC.

Although this study cannot predict the doughnut distribution of FG repeats previously seen in experimental studies and coarse-grained simulations [30,69], our results are in general agreement with those studies. According to their observation, while charged residues are mostly located near the wall, FG repeats form a doughnut around the central axis. This is in agreement with our results as well as previous studies [11,26,27] that divide most of the FG Nups into two distinct domains with one richer in FG motifs and the other richer in charged residues. In fact, while our study suggests that FG repeats are placed at the center of the pore, it does not predict the shape they form at the center, whether a central plug or a doughnut. Furthermore, it was observed that replacement of charged residues with neutral ones results in a high density, wide distribution of all FG repeats inside the pore [30]. This is in line with the hypothesis that relaxed coil conformation of charge-rich domains places FG domains at the center of the NPC. In addition, random shuffling of FG Nup sequences completely disturbs formation of the dense FG region [30], which is consistent with our results that the specific distribution of different types of residues is the key parameter in FG network formation.

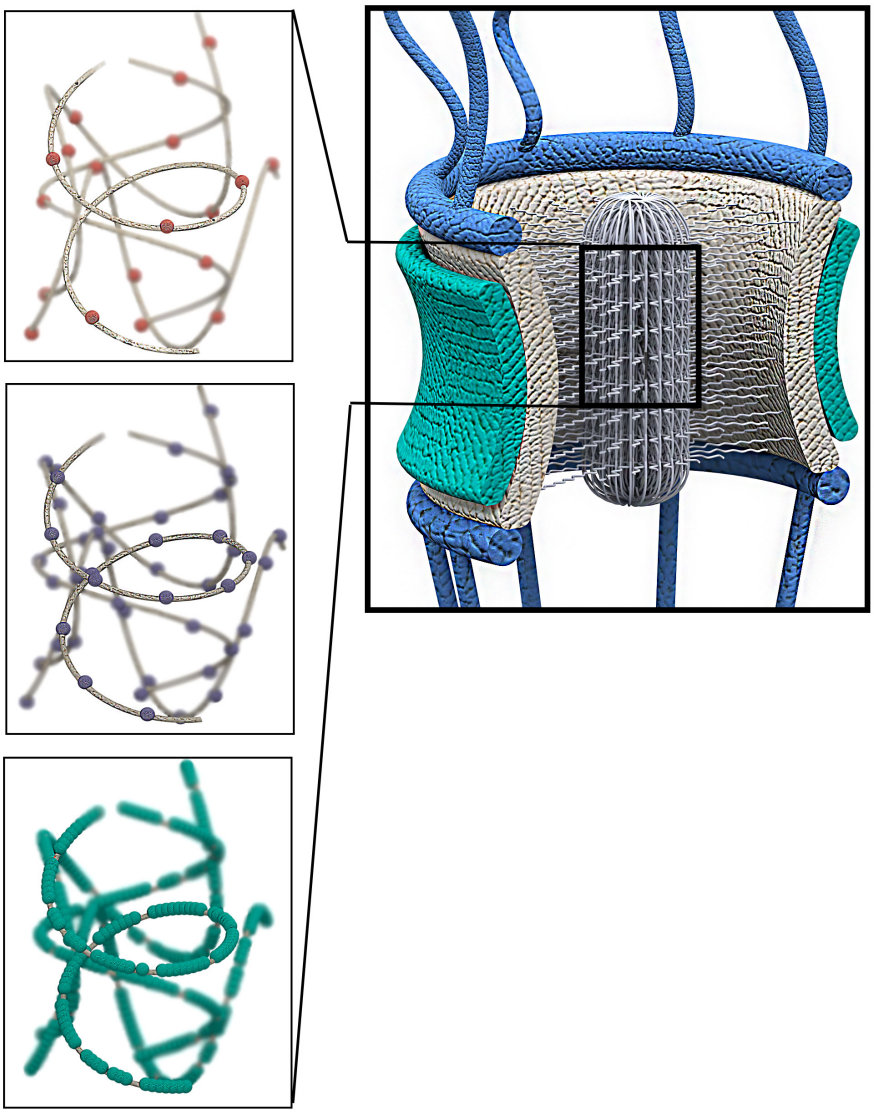

Figure 2-10: Schematic demonstrating cooperation of positive LCRs, FG motifs, and polar residues to form the FG network. Red spheres (top left box) represent positive charges embedded in positive LCRs, purple spheres (middle left box) represent FG motifs, and green spheres (bottom left box) represent polar residues within the central transporter. The highlighted FG Nups domain in the three left boxes represent N-terminal domain of NSP1, i.e. spheres represent location of charged (red) and polar (green) residues and FG motifs (purple).
Due to the presence of positive LCRs, the central transporter is positively charged, facilitating transport of negatively charged NTRs. The dynamics of the central network of FG motifs is also regulated via charge repulsion in LCRs. In addition, low charge content of positive LCRs assists these domains of FG Nups to form a collapsed coil conformation. Furthermore, high densities of polar residues among FG repeats promote the hydrophobic interactions and meshwork formation. On the other hand, highly charged stalk domains of FG Nups, which are depleted of polar residues, place the central transporter in the center of the NPC by forming a relaxed coil conformation.

Interestingly, a small fraction of LCRs are negatively charged, with a short length and high number density of charged residues (Figure 2-5). In yeast, specifically, LCRs of all FG Nups are positive, except for the three outermost FG Nups, namely Nup159, Nup 1 and Nup2. Therefore, it could be postulated that these FG Nups, when compared to central channel FG Nups, play a distinct role in nucleocytoplasmic transport. Being

located at the outermost regions of the pore, these FG Nups are primarily responsible for attracting or rejecting importins and exportins to initiate nucleocytoplasmic transport. As a result, it is not desirable for them to form collapsed coil configurations. Moreover, It can be speculated that the straightened structure of these FG Nups enables them to explore a larger space and attract cargos more easily and deliver them to the central channel where the central transporter has been formed. It is important to note that, as the length of these negative LCRs are short, and the rest of the sequence is full of both positive and negative charges, the electrostatic repulsion between the negative LCRs and negative NTRs should not have a significant negative effect on binding of NTRs and FG Nups.

Our bioinformatics study was conducted on all FG Nups across all different species. It remains an open question to explore how the presence of multiple cargo complexes inside the nuclear pore and their cross-interactions with FG Nups can potentially influence the conformational dynamics of the FG Nups. This calls for further computational and experimental studies to unravel the underlying mechanisms.

Methods

Most analyses presented herein were conducted via in-house Python scripts, except disorder prediction and clustering, which were conducted through ESPRITZ Web server (version 1.3) and ELKI data mining software, respectively [70,71]. Considering the high risk of generating deceiving results, studies were initiated by comprehensive validation and verification steps. All results were verified against previous studies [11,27,53]. Data visualization was performed using Python and R scripts.

Protein datasets

Two groups of proteins were analyzed in this study; the first group, named FG Nups, was initially taken from a recent bioinformatics study [27], including Nups from 252 species. The database was then updated according to the latest changes made to the Uniprot database until March 2014, eventually including 1138 FG Nups. The database was originally created by extracting entries associated with the keyword 'nucleoporin' from the Uniprot database, and selecting those with a high percentage of disorder (>30%) and high percentage of FG motifs (>0.15 FG/AA) as FG Nups [27]. All studies on FG Nups were conducted solely on their disordered regions, unless otherwise specified. Therefore, hereafter, FG Nups refer to disordered domains of these proteins. The second group, termed DisProt, included more than 1500 disordered regions of 694 proteins adapted from DisProt database [62]. DisProt is a database of proteins that contain at least one experimentally determined disordered domain along their sequence. Disordered domains were extracted from the latest version of the database, i.e. version 6.02.

Sequence analysis

Mean absolute net charge was defined as absolute net charge along the sequence of a protein divided by its length, with residues "R" and "K" as positive charges and residues "D" and "E" as negative charges. All charged residues were assumed to have the same amount of charge. Mean hydrophobicity was defined as total hydrophobicity of the sequence of the protein divided by its length. Hydrophobicity scale proposed by Kyte and Doolittle [72] was used, with hydrophobicity of each individual amino acid normalized to a scale of 0 to 1 [53]. FG linkers in FG Nups were characterized by being located between two FG motifs and consisting of 10 to 30 residues. In Figure 2-6, in order to show the significance of the observed trend, the results are compared to those of Disprot proteins. Random sequences of 10-30 residues long were chosen to be compared to FG linkers from FG Nups. Like charge regions (LCRs) were determined by multiple charged residues of the same sign that are not interrupted with charged residues of the opposite sign.

Clustering analysis

For clustering purposes, initially, motifs/residues of interest, i.e. FG motifs, charged residues, and polar residues were located along the sequence of FG Nups. FG clusters were identified by "FG" and "GF" motifs while charged and polar clusters were determined by "D, E, R and K" and "S, N, T, and Q", respectively. Clustering was conducted with a density based method. PreDeCon clustering algorithm was used through ELKI data mining software [71]. Clustering parameters of this algorithm, i.e. minimum points and epsilon, were tuned to maximize Dunn's validity index [73] and they were found to be 35, 8, and 2 for FG Nups, charged residues, and polar residues, respectively. Minimum points were set to two, one motif/residue for each boundary with the exception of polar residues, due to their compact distribution, with four minimum points. We found that on average, FG, charged, and polar clusters contain 128, 25, and 5 amino acids, respectively. The size of charged and FG clusters are consistent with the results reported by [27]. The sizes of the clusters are reversely correlated with the content (number density) of that residue type in the sequence. As polar clusters are very abundant in FG Nups, the average size of the cluster is very small relative to charged residues and FG motifs that are less abundant.

Overlap between a specific type of cluster with other clusters was quantified by the total number of residues shared between them divided by the sum of the sizes of the desired cluster type in each FG Nup. Residue location approximation was performed on whole sequences of FG Nups. The sequence of each FG Nup was divided into three zones, with the first zone located on the N-terminus and the third one on the C-terminus. Therefore, zone3 approximately represents the structured region of the sequence embedded in the scaffold, while zones two and one represent near the wall and near the central axis regions of FG Nups inside the NPC, respectively. For each zone, the proportion of the clusters contained within that zone was calculated. For instance, if 10 residues of one cluster were in the first zone, and there were 50 residues in that cluster type in the sequence of FG Nup, the proportion of that cluster type in the first zone one would be 10/50=0.2.

Appendix

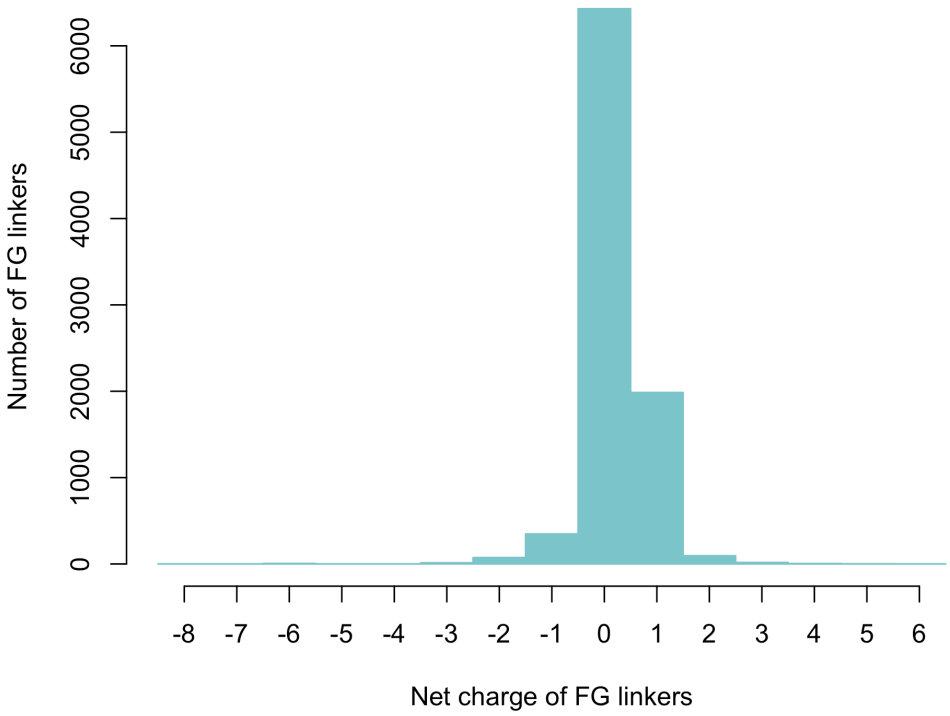

Figure 2-S1: Histogram of net charge of FG linkers. More than 90% of FG linkers possess a net charge of either zero or one. Therefore, although net charge is considered as a primary parameter for distinguishing ordered and disordered proteins [1], in the case of FG Nups, the total number density of charged residues would be a more meaningful parameter [2].

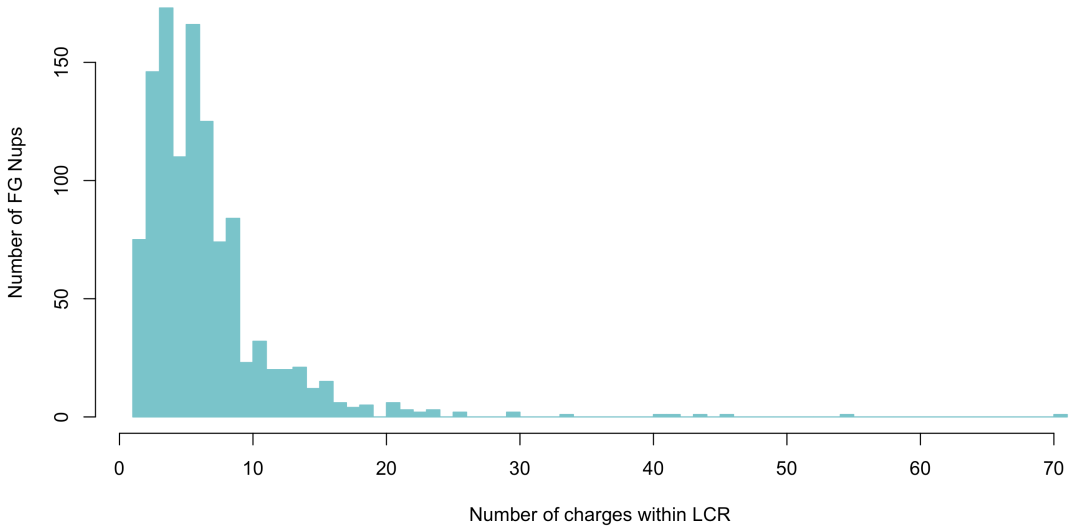

Figure 2-S2: Histogram of number of charged residues within LCRs among all FG Nups. Although FG Nups indicate a wide range of number of charges in LCRs, location analysis of FG Nups with known location, i.e. central channel or cytoplasmic and nuclear peripheries, reveals that FG Nups can be grouped into two categories; those located in the peripheries possess LCRs of equal or more than seven charged residues, while this is less than eight for central channel FG Nups (see Figure 2-4).

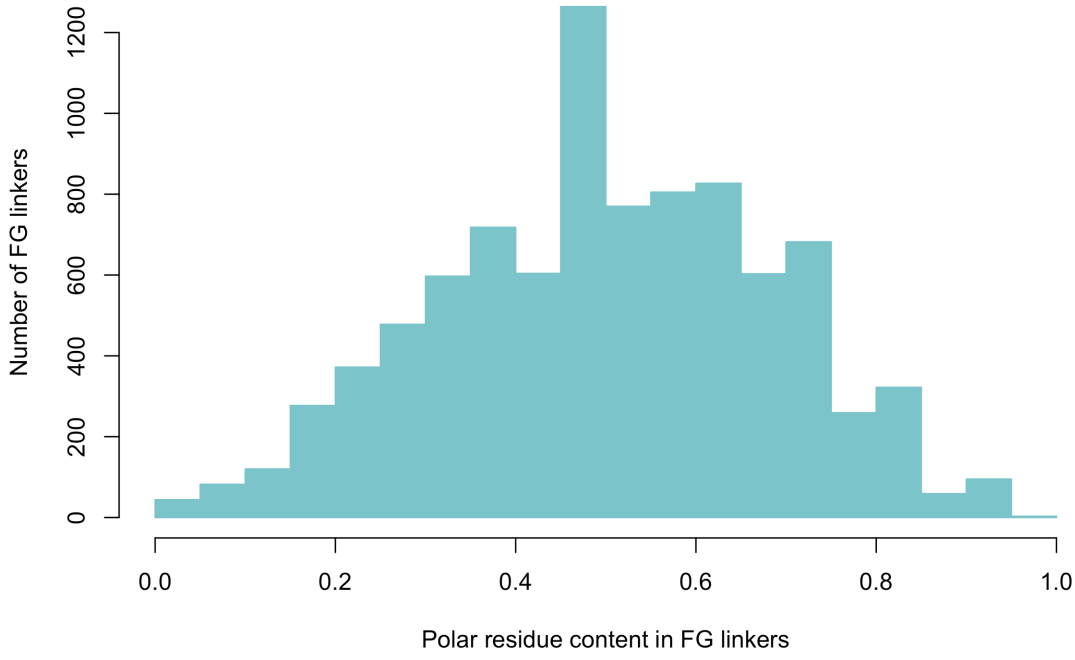

Figure 2-S3: Histogram of polar residue content within FG linkers of 10-30 residues long. FG linkers exhibit a wide range of polar content with approximately a normal distribution about 50%. However, in-depth analysis of FG linkers reveals a strong relation between number density of charged and polar residues within the sequence of FG linkers.

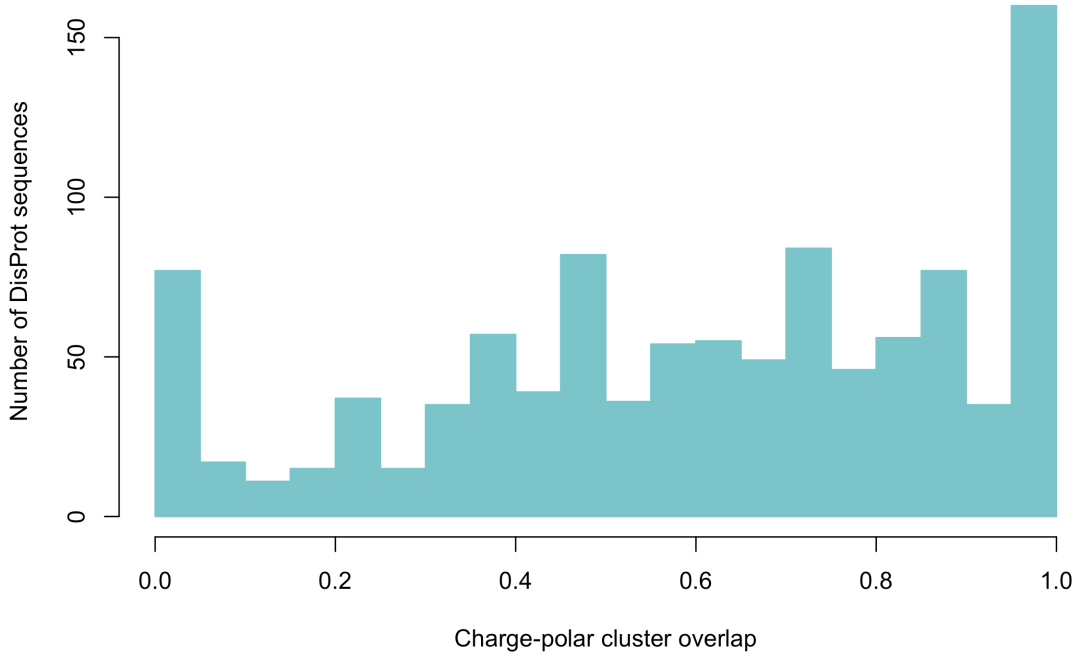

Figure 2-S4: Overlap between charged and polar clusters of Disprot. The relatively high overlap for the control group shows the significance of the trend found in Fig. 7 for FG Nups.

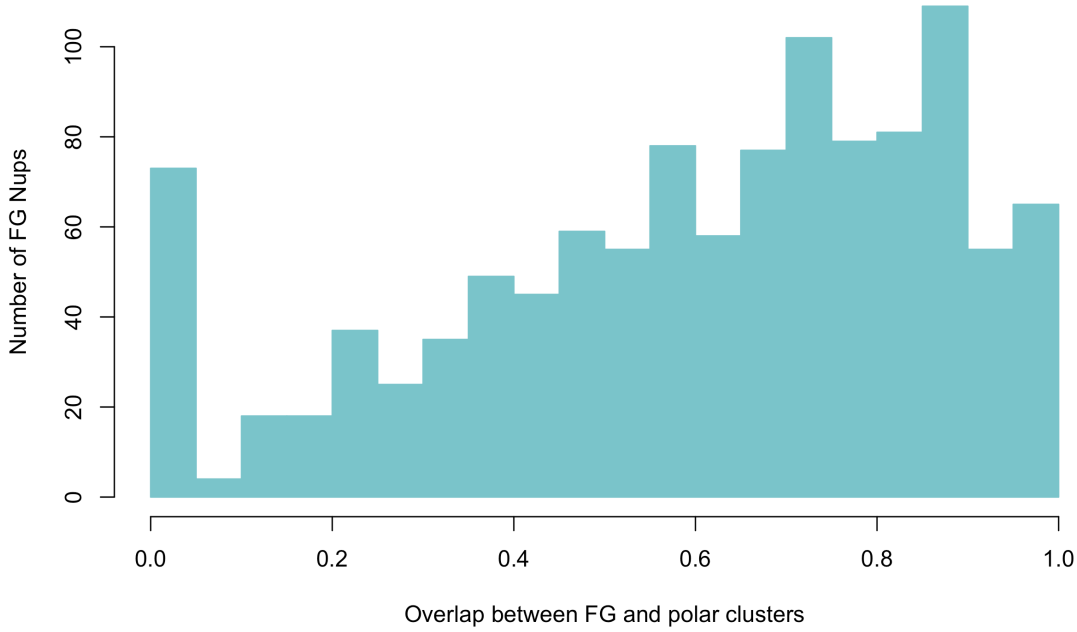

Figure 2-S5: Overlap between FG and polar clusters. Although FG Nups show a wide range of overlap between FG and polar clusters, further analysis reveals that FG Nups with more than half of their polar clusters located in the first one-third of their sequence exhibit a higher overlap between FG motifs and polar residues, which implies a potential functional significance of overlapped FG-polar clusters at the tip of FG Nups.

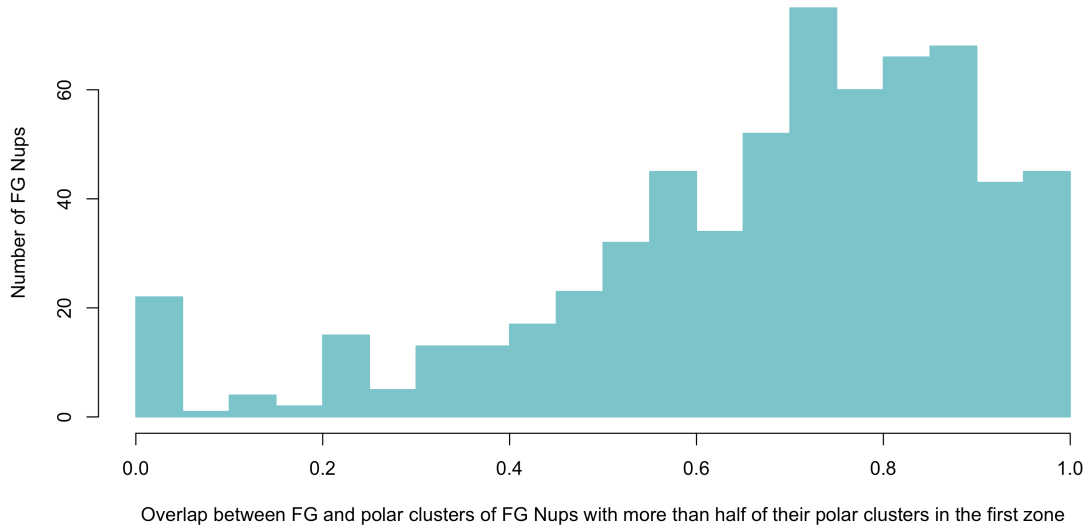

Figure 2-S6: Overlap between FG and polar clusters of FG Nups with more than half of their polar clusters located in the first section of their sequence.

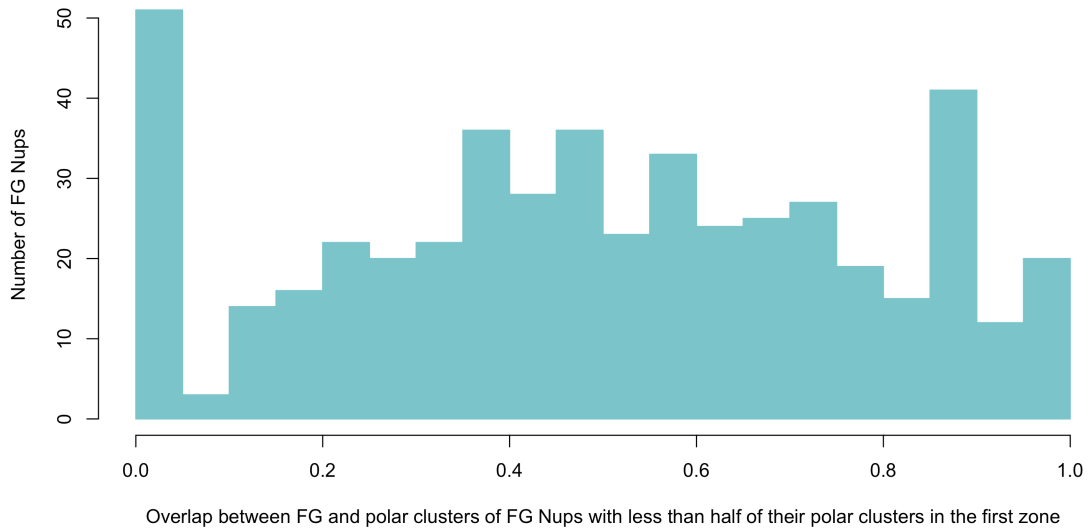

Figure 2-S7: Overlap between FG and polar clusters of FG Nups with less than half of their polar clusters located in the first section of their sequence.

Chapter 3 : FG Nucleoporins feature unique amino acid sequence patterns that distinguish them from other IDPs

The work presented in this Chapter has been submitted for publication:

Peyro, M.*, M. Soheilypour*, V. S. Nibber, A. Dickson, and M. R. K. Mofrad.
"FG Nucleoporins feature unique amino acid sequence patterns that distinguish them from other IDPs", Submitted to Biophysical Journal

*These authors contributed to this paper equally.

*** Special note: Please note that throughout this Chapter, the term "largest positive LCR" is equivalent to the term "LCR" in the rest of this dissertation. The term "LCR" in this Chapter, refers to like charge regions of any length including any type of charge.**

Abstract

FG Nucleoporins (FG Nups) are intrinsically disordered proteins and are the putative regulators of nucleocytoplasmic transport. They allow fast, yet selective transport of molecules through the nuclear pore complex (NPC). The underlying mechanism of nucleocytoplasmic transport is not yet fully discovered. As a result, FG Nups have been the subject of extensive research in the past two decades. While most of the studies have been focused on analyzing the conformation and function of FG Nups from a biophysical standpoint, some recent studies have investigated the sequence-function relationship of FG Nups, with a few investigating amino acid sequences of a large number of FG Nups to understand common characteristics that might enable their function. Previously, we identified an evolutionarily conserved feature in FG Nup sequences, which are extended sub-sequences with low charge density, containing only positive charges, and located toward the N-terminus of FG Nups. We named these patterns largest positive like charge regions (largest positive LCRs). These patterns are specific to positively charged residues and negatively charged residues do not show such a pattern. In this study, we compare FG Nups with other disordered proteins obtained from the DisProt database, in terms of presence of LCRs. Our results show that largest positive LCRs are exclusive to FG Nups and are not observed in DisProt proteins. Also, largest positive LCRs show up once in the amino acid sequence of FG Nups and are what differentiate FG Nups from DisProt proteins, meaning that excluding largest positive LCRs from the sequences of FG Nups make them similar to Disprot proteins in terms of charge distribution. We also previously showed the biophysical effect of LCRs in conformational ensemble of FG Nups. The results of this study are in line with our previous findings and imply that largest positive LCRs are exclusive and functionally significant characteristics of FG Nups and nucleocytoplasmic transport.

Introduction

The transport process through the NPC have been the focus of a large body of research in the past few decades. These studies have investigated this protein complex from different perspectives, including but not limited to structure of the NPC [74,75], transport selectivity [1,76,77], etc. In recent years, however, researchers have identified that in order to better understand the transport process and the specific role of FG Nucleoporins (FG Nups) in the transport process, one should pay a closer attention to the amino acid sequences of FG Nups [11,26,27,30,31,35,78,79]. Similarly, the sequence-conformational ensemble-function relationship in unstructured proteins, in general, have been the subject of substantial research in the past decade [80–88].

Studying disordered proteins and their conformations have become increasingly important with the growing number of intrinsically disordered proteins (IDPs) and IDP regions (IDPRs) discovered [89]. IDPs and IDPRs are known to lack a fixed secondary structure due to their specific amino acid sequence characteristics [53,54,90,91]. In the past decade, many studies have focused on studying the sequences of disordered proteins in detail and investigated how amino acid sequences correlate with conformational properties of IDPs. It is well established that presence of several uncompensated charge groups causes IDPs or IDPRs to have a net charge. This feature along with low hydrophobicity content are the two major reasons that inhibit IDPs and IDPRs from forming a fixed secondary structure [53,92,93]. In addition, IDPs and IDPRs are known to have a high content of disorder-promoting amino acids and a low content of order-promoting ones [90]. IDPs and IDPRs are shown to make two general conformations, namely intrinsic coil and intrinsic premolten globule. The intrinsic coil conformation has a more relaxed coil shape, while the premolten globule shows a more compact conformation. While the premolten globules are closer to structured proteins in terms of conformation, they still fall under the umbrella of IDPs [91].

Among the sequence properties that influence the conformation of IDPs, the effect of charged residues is more widely investigated [80–88]. Fraction of positive and negative charges f_+ and f_- respectively, net charge per residue ($f_+ - f_-$) and fraction of charged residues ($f_+ + f_-$) in the sequence were early metrics employed to explore this effect [81,84,91]. However, since these are averaged metrics, they could only explain specific features of IDPs as a whole and are not detailed enough to relate the effect of charge decoration on conformation of sub-sequences of FG Nups although the combination of these metrics can classify the IDPs into different conformational classes [83,94]. Another metric, κ, was introduced by Das and Pappu that can capture the effect of distribution of opposite charges on global conformation of IDPs effectively [83]. In addition, more studies explored the correlation between charged patterning and conformation of IDPs. Sequences with the same charge content but different charge patterning were shown to demonstrate different conformations [83,95,96]. For instance, Huihui et al. showed that mutation of even one charged residue can induce significant changes in conformation of IDPs [86]. Therefore, recent studies focus more on correlating the *charge decoration* with conformation of IDPs. For example, Firman and Ghosh showed that charge decoration in sequence determined the coil-globule conformation change in IDPs [97].

While the sequence composition-conformational ensemble-function of FG Nups have been studied before [26,30,78,79,98,99], very few studies have investigated the effect of charge decoration on FG Nups, their conformation and their function [35,79,98]. In our previous study, we found that FG Nups contain specific sequence features named largest positive like charge regions (largest positive LCRs) that are evolutionarily conserved across more than 250 species [35]. These patterns are extended sub-sequences that have a long length and a low charge density, which only feature positively charged amino acids (see Figure 3-1). Of note, negatively charged amino acids did not show such a distinct pattern. As a follow up to our previous study, here we explore whether these features are specific to FG Nups or could be found in other disordered proteins as well. Accordingly, we analyzed a large dataset of disordered proteins, i.e. DisProt, for presence of LCRs. Our results suggest that largest positive LCRs are amino acid sequence patterns exclusive to FG Nups and do not appear in DisProt proteins, which further suggests that largest positive LCRs may be important exclusive FG Nups features that dictate their specific function in nucleocytoplsmic transport.

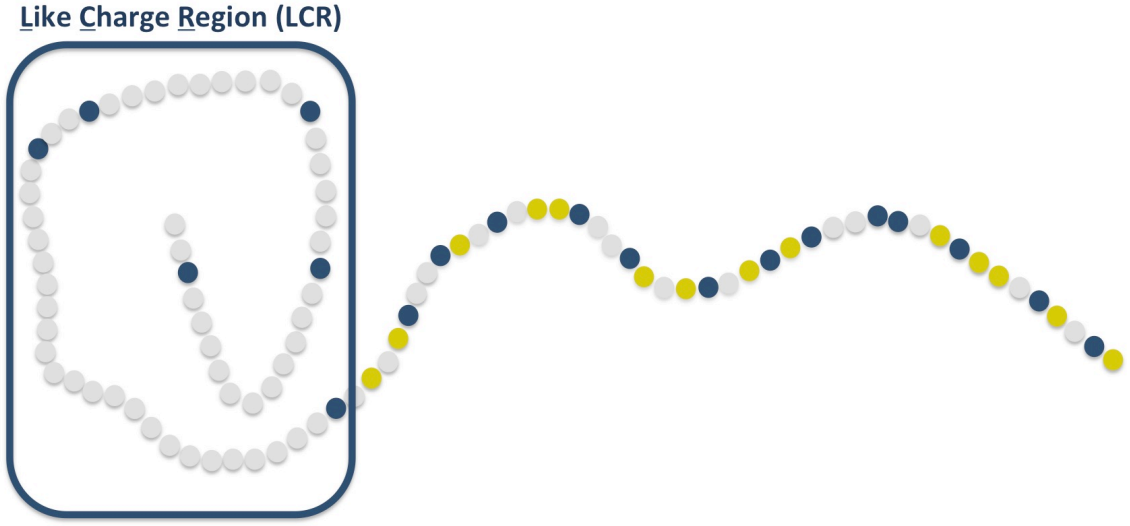

Figure 3-1: Schematic of the disordered domain of an FG Nup and its LCR. Blue residues are positively charged, while yellow residues are negatively charged.

Results

Charged residue abundance and charged residue distribution are different in FG Nups versus DisProt proteins

In this study, we analyzed presence of negative and positive LCRs in the sequences of disordered domains of FG Nups and DisProt proteins. LCRs are defined as any stretch of amino acid sequence that contains only one type of charged amino acids; either positively charged (lysine, arginine) or negatively charged (aspartic acid, glutamic acid). In each dataset only the disordered domains were analyzed and the method of extraction of each of these datasets is explained in the Methods section.

Intrinsically disordered proteins (IDPs) and Intrinsically disordered protein regions (IDPRs) are known to be rich in charged residues and have a low density of hydrophobic residues compared to structured proteins. Therefore, we compared the abundance of these amino acids between the two datasets (Figure 3-2). The amino acids are shown in the order of order-promoting to disorder-promoting property[65]. FG Nups proteins have a lower abundance of all of the hydrophobic amino acids except for Phenylalanine (Phe). Higher abundance of Phe is expected due to naturally high abundance of Phe and Gly in FG Nups. The lower abundance of W, Y, I, L and V in FG Nups is approximately compensated by the higher abundance of F, considering that F is the second order promoting residues. Therefore, the two datasets have almost equal abundance of order promoting amino acids. On the other hand, abundance of all of the charged residues is higher in DisProt proteins compared to FG Nups. Charged residues are the most disorder promoting types of amino acids. This means that while FG Nups have the same abundance of order-promoting amino acids as those in DisProt amino acids, they feature lower density of disorder-promoting amino acids. This difference in amino acid composition implies that, overall, FG Nups are on average or partly less disordered compared to DisProt proteins.

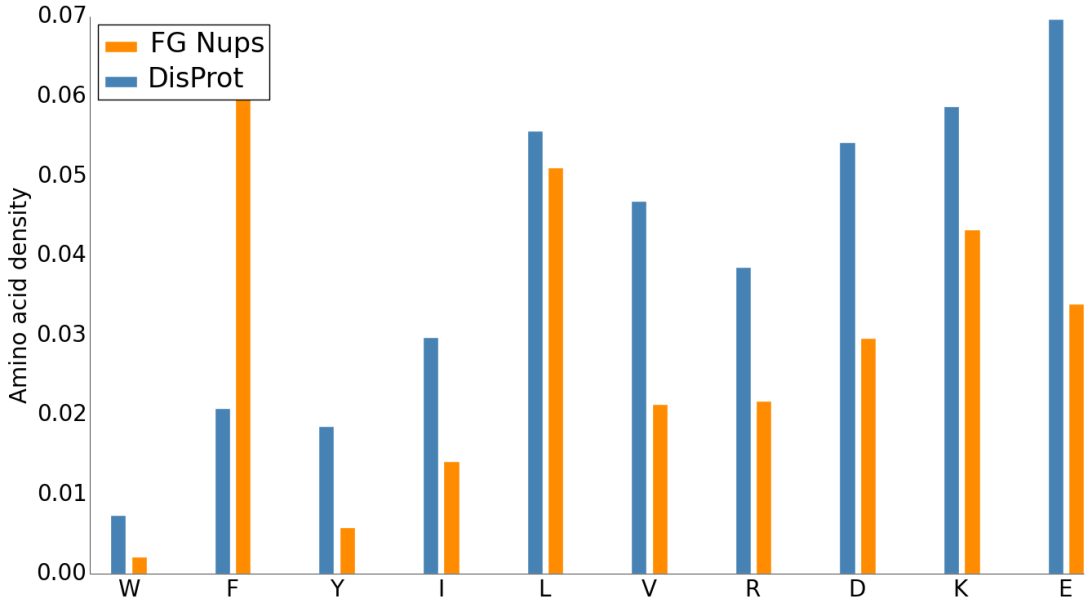

Figure 3-2: Comparison of abundance of the most order-promoting and the most disorder-promoting amino acids in FG Nups vs. DisProt. FG Nups have a lower abundance of all of the order-promoting amino acids except for F. The sum of the abundance of order-promoting amino acids are nearly equal. On the other hand, the abundance of all of the charged residues are lower in FG Nups compared to DisProt. Considering that charged residues are the most disorder-promoting types of amino acids, the results in this figure imply that FG Nups are on average, or partly less disordered than DisProt proteins.

Subsequently, we compared charge distribution between the two datasets. A schematic of charged distribution of all of the proteins in the two datasets were drawn. The entire datasets were too large to be presented here, therefore, a snapshot of part of the schematic for each of the datasets are presented (Figure 3-3). While this is a small fraction of the entire dataset, it represents a meaningful comparison on the difference between the charge distribution in FG Nups and DisProt. The black lines represent the amino acid sequences, the small vertical red lines depict negatively charged residues and the small blue lines depict the positively charged residues. Only charged residues are represented in this schematic for the sake of clarity. The first noticeable difference is that DisProt sequences are more "*colorful*", implying that DisProt proteins have a higher charged residue density, which is supported by Figure 3-2, where abundance of charged residues are compared in the two datasets. As mentioned before, considering that low hydrophobic density and high charge density are characteristics of disordered proteins, lower charge content implies that FG Nups are on average, or partly, less disordered than DisProt proteins. The second difference is that, on average, the length of protein sequences in the DisProt database is shorter than that in FG Nups. The third and main difference is that in the FG Nup sequences, several extensively large sub-sequence exist that have a low charge density and only contain positively charged residues (i.e. largest positive LCRs). Some of these regions are boxed with red. The proteins in FG Nups dataset depict this pattern very frequently. Additionally, many of the FG Nups have two segments in their sequence, one is charge-rich (similar to sequences in the DisProt database) and the other has a low density of charged residues that are oftentimes positively charged (LCR, the boxed pattern). Notably, some of the FG Nups do not contain the charge-rich segment and are only composed of a large positive LCR.

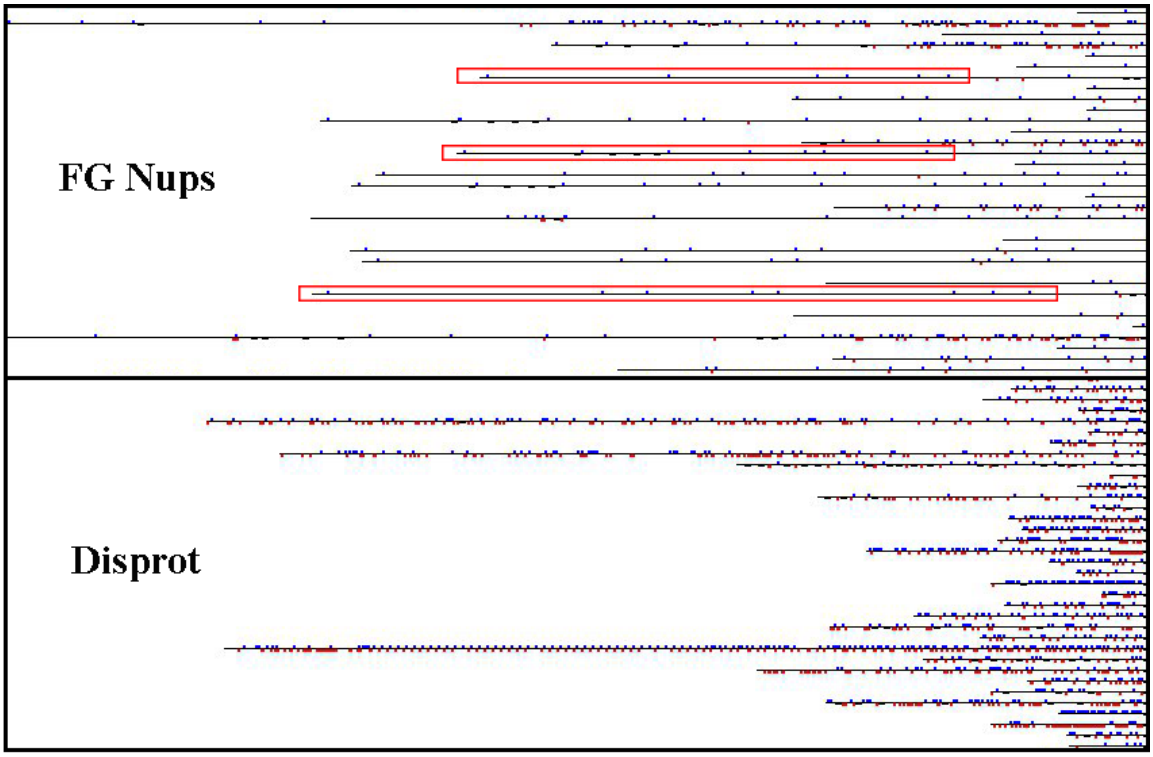

Figure 3-3: A snapshot of a small part of the schematic comparing charged residue distribution in FG Nups vs. DisProt (The entire schematic was too large to be presented). The figure shows that FG Nups have a lower charged residue density compared to DisProt. DisProt have a lower length compared to FG Nups. The most interesting difference is that in the FG Nup sequences, several regions can be observed that are extensively large segments of the protein that have a low charge density only have positively charged residues. Some of these regions are boxed with red. We have named these regions largest positive Like Charge Regions (largest positive LCRs). The proteins in FG Nups dataset depict this pattern very frequently. Many of the FG Nups have two segments in their sequence, one is charge rich (similar to sequences in DisProt) and the other has a low density of charged residues that are almost always, positively charged (LCR, the boxed pattern). Some of the FG Nups also do not contain the charge rich part and are only composed of a large positive LCR.

Long sub-sequences that contain only a few positively charged residues (largest positive LCRs) are exclusive to FG Nups and do not exist in DisProt proteins

To gain a better insight on the characteristics of positively charged and negatively charged residues in each of these two datasets, the amino acid sequence length and charge contents in the largest LCR region were analyzed and compared between the two datasets (Figure 3-4). LCR length is the number of amino acids between the charged residue that the LCR starts with and the charged residue that the LCR ends with (see Figure 3-8). LCR charge content (number density) is the number of charged residues present in the LCR divided by the number of amino acids in the sequence of each protein. Blue datapoints represent largest positive LCRs, while gold datapoints represent largest negative LCRs. A clear distinction can be observed between the two graphs in terms of the distribution of blue datapoints. For the DisProt dataset, positive and negative LCR datapoints cover similar areas of the graph. Distribution of datapoints on the left side of the graph shows that largest positive and negative LCRs of short length with low and

high charge density are present in the DisProt dataset. In FG Nups dataset, a similar pattern can be observed on the left side of the graph, where positive and negative LCRs with short length are present. However, the distinctive feature of the FG Nups dataset is the presence of largest positive LCRs that are long but have a low charge density (boxed in red). It is worth noting that the negative LCRs in FG Nups dataset do not show such a distribution and this pattern is specific to positive LCRs. The average length of largest positive LCRs in FG Nups is 127 amino acids, while the average length of largest negative LCRs is 19 amino acids. These values are 19 and 21 for the DisProt dataset, respectively. Our results clearly demonstrate that the presence of long largest positive LCRs is an exclusive characteristic of FG Nups.

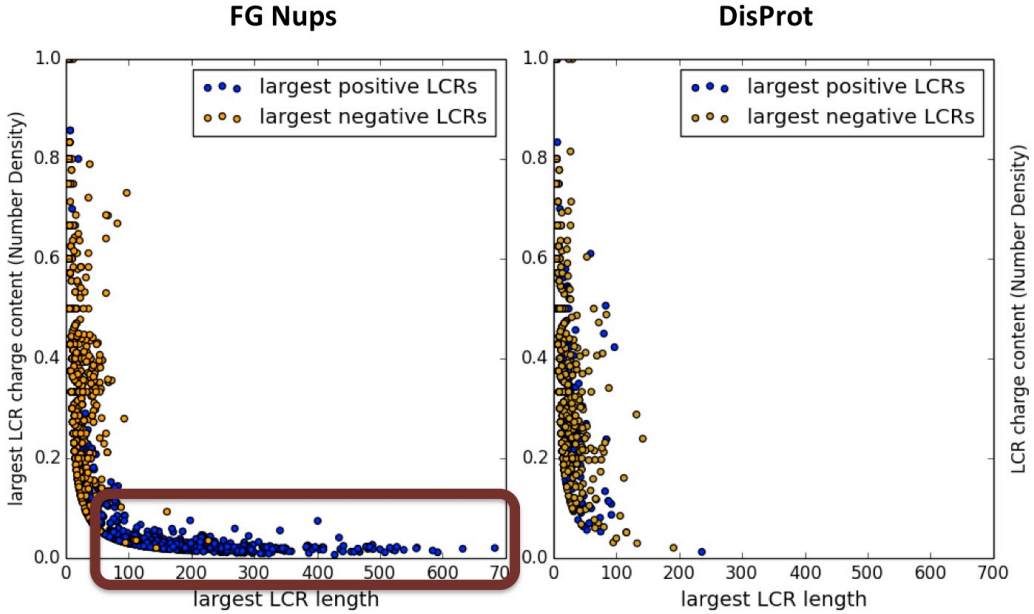

Figure 3-4: Scatterplots showing charge content plotted against LCR length for each dataset. Each gold dot represents the longest negative LCR from one entire protein sequence in the dataset along with the number of charged residues in that LCR; blue dots represent the same information but for positive LCRs. Only the longest positive and negative LCRs from each sequence were plotted, and only if the length was greater than at 2 amino acids. The data shows a clear distinction between FGNups and DisProt, as FG Nups feature largest positive LCRs that are long but have a low charge density (boxed with red). It is important to note that this pattern is not observed in case of largest negative LCRs in FG Nups or in DisProt. Therefore, this data suggests that the specific pattern observed in largest positive LCRs are specific to FG Nups.

As the next step, the positive and negative LCRs (of any length above 2 amino acids) of the two datasets, were compared (Figure 3-5a). It is important to note that this data includes all of the LCRs present in the dataset, and not just the largest LCR in each protein. These results show that the majority of LCRs for all the four categories, i.e. positive LCRs in FG Nups, negative LCRs in FG Nups, positive LCRs in DisProt, and negative LCRs in disport, are short, charge-rich domains. However, the outliers are different in the four groups. Positive LCRs in FG Nups have several outliers with extensively large length. These outliers cause the average length of positive LCRs in FG Nups to be significantly higher than the others (25 versus 8 or 10). The very large standard deviation is also representative of the very large range that the data covers and presence of significant number of outliers with large lengths.

Based on the results from Figure 3-5a, it was shown that a significant difference exists between the distribution of length of positive LCRs in FG Nups and the other three groups. Also, based on the observation from Figure 3-4, FG Nups and DisProt are different in terms of the length and density of largest positive LCRs. We were interested to see if the difference observed in the distribution of lengths of LCRs in Figure 3-5a is caused only by presence of largest positive LCRs boxed in Figure 3-4. Accordingly, we conducted the same comparison in Figure 3-5a, while removing the largest positive LCR of FG Nups from the data (shown in Figure 3-5b). By removing the largest positive LCRs in FG Nups, the results look very similar across the four groups of data in terms of average, standard deviation, and outliers. As a result, it is conceivable to conclude that the only feature that differentiates FG Nups positive LCRs from other groups of data is the largest positive LCR in each sequence; otherwise, the four groups of data look very similar. Collectively, our results demonstrate that largest positive LCRs in FG Nups, that are long stretches of amino acids that have a low charged density and contain only positively charged residues, are exclusive to FG Nups and the only region that differentiates FG Nups from DisProt.

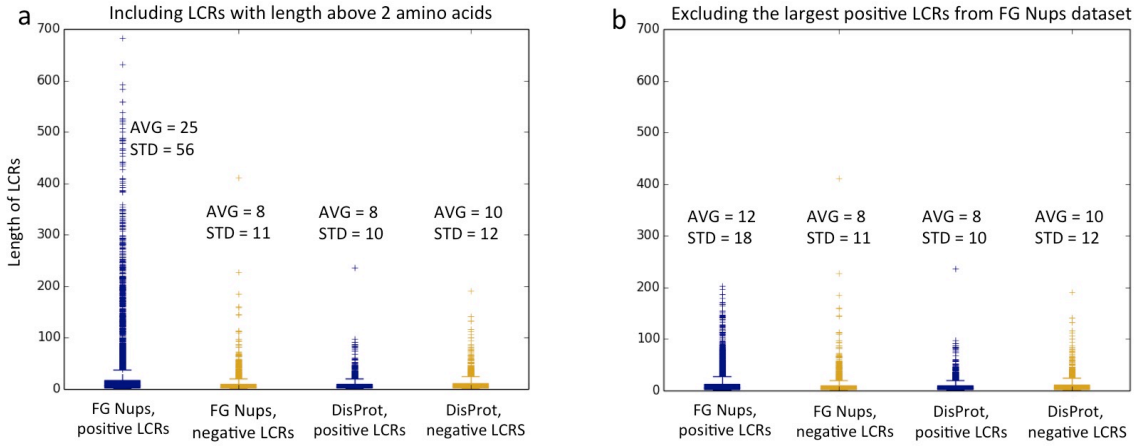

Figure 3-5: Boxplot of length of LCRs for positive LCRs in FG Nups, negative LCRs in FG Nups, positive LCRs in DisProt and negative LCRs in DisProt. All of the LCRs above 2 amino acids length are considered in these datasets. A) posititve LCRs in FG Nups show a significant difference in terms of average and range from negative LCRs in FG Nups and positive and negative LCRs in DisProt b) largest positive LCRs in FG Nups are removed from the data and the rest of the data looks similar to the other three segments of data. This means that the major difference that can be observed in plot (a) between positive LCRs in FG Nups and the other three segments of data is due to presence of the largest LCRs in FG Nups, which mainly form the outlier section of positive LCRs. Other than that, the rest of positive LCRs in FG Nups have similar characteristics to the other three segments of data. This implies that, what differentiates FG Nups form DisProt is presence of largest positive LCRs. These patterns are exclusive to FG Nups and cannot be observed in DisProt.

For a more marked representation of largest positive LCRs in FG Nups, we excluded LCRs shorter than 20 and 40 amino acids respectively (Figure 3-6 b&c). As shorter LCRs are excluded, expectedly, average length is increased for all the four data groups. However, the average for positive LCRs in FG Nups reaches significantly higher values compared to the other three categories (Figure 3-6 b&c). Considering that short LCRs are formed randomly due to high density of charged residues in some regions, removing

them from the data provides a better insight about the length of non-random LCRs. Based on these results, the average length of positive LCRs in FG Nups is 2-3 folds higher than that of negative LCRs in FG Nups as well as positive and negative LCRs in DisProt.

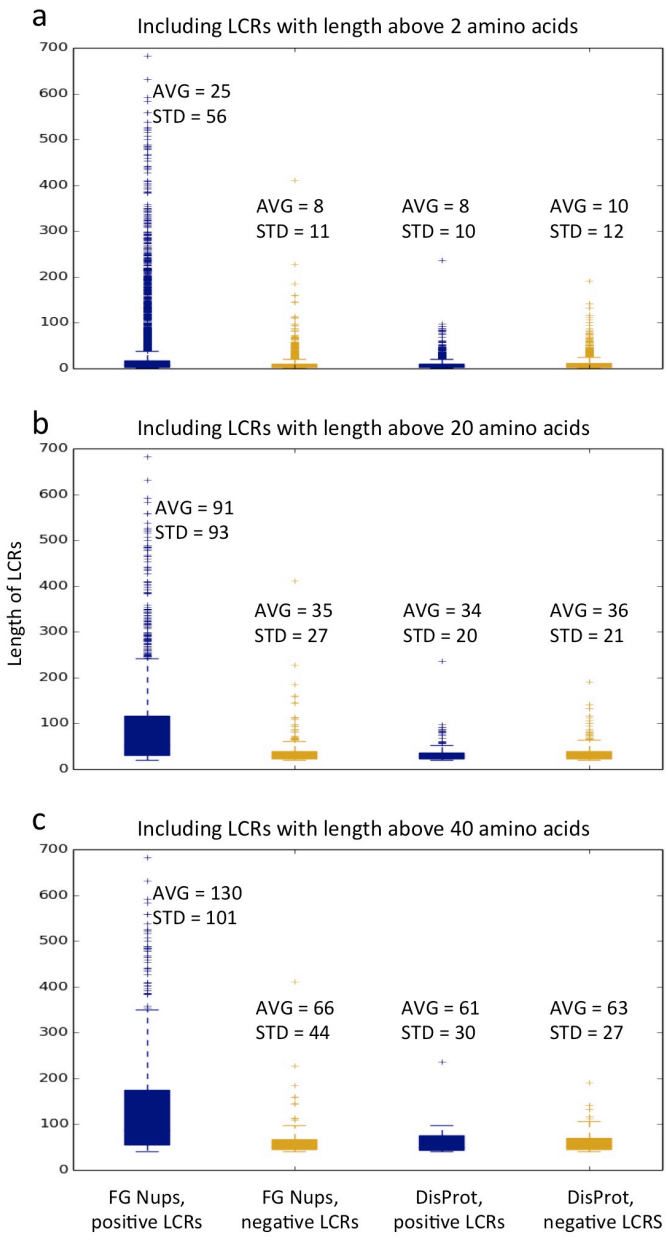

Figure 3-6: Boxplots of length of positive LCRs in FG Nups, negative LCRs in FG Nups, positive LCRs in DisProt and negative LCRs in DisProt a) Considering LCRs of any length (above 2 amino acids), the average length of positive LCRs in FG Nups is 3 times higher than the other three categories. Panels (b) and (c)) Removing shorter LCRs (below 20 in panel b and below 40 in panel c) increases the average length of all the categories, but average length of positive LCRs in FG Nups remains significantly higher than the other three categories, and emphasizes the difference in the length distribution of positive LCRs in FG Nups.

To further explore the differences in charge distribution of the two datasets, we examined the fraction of each amino scid sequence that is covered by LCRs. Lengths of all of the LCRs present in each sequence were added and then normalized by the length of the

sequence. We call this factor *LCR-covered-fraction* (Figure 3-7). This metric was evaluated for the entire proteins in both datasets (Figure 3-7a), as well as those with LCRs with a length of over 20 amino acids (Figure 3-7b), and LCRs with a length of over 40 amino acids (Figure 3-7c). As the LCR threshold increases from 2 to 20 and then 40, all of the histograms shift toward left, meaning that the number of sequences with high LCR-covered-fraction decreases and the number of sequences with low LCR-covered-fraction increases. However, it could be observed that as we exclude more short LCRs, most of the remaining data for the DisProt dataset as well as negative LCRs in FG Nups have a low LCR-covered-fraction for the respective length-exclusions. However, positive LCRs in FG Nups do not demonstrate the same behavior. When considering the LCRs of larger than 40 amino acids, DisProt sequences will have a low positive and negative LCR-covered fraction and FG Nups will have a low negative LCR-covered fraction, but FG Nups will still have a high positive LCR-covered fraction. This implies that the high LCR-covered-fraction for FG Nups positive LCRs are due to presence of largest positive LCRs, but for the other categories, it is due to presence of several short LCRs.

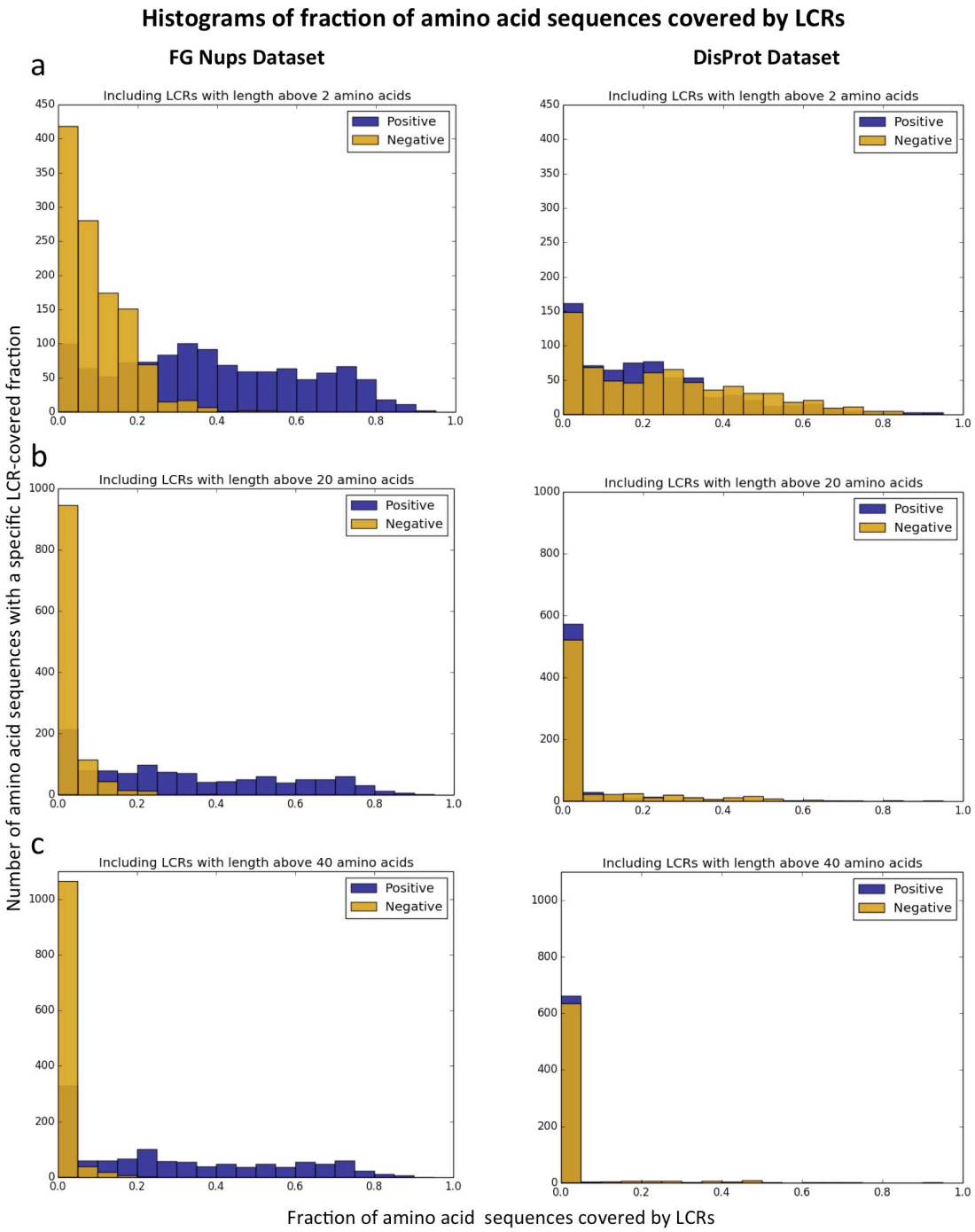

Figure 3-7: Histogram of positive and negative LCR-covered-fraction in FG Nups and DisProt dataset. Blue histograms represent positive LCRs and gold histograms represent negative LCRs. By increasing the threshold of the length of the LCRs considered in the data (from 2(a) to 20(b) to 40(c)), the LCR-covered-fraction decreases for negative LCRs in FG Nups and positive and negative LCRs in DisProt, but will still hold considerably large values for positive LCRs in FG Nups. This indicates that the high positive LCR-covered-fraction in FG Nups is caused by long positive LCRs, while the large values on LCR-covered-fraction for the other three categories were the result of presence of several short LCRs.

Discussion

We previously identified an evolutionarily conserved feature in the sequences of FG Nups [35]. In this study, we further demonstrate that these features are exclusive to FG Nups and are not found in other disordered proteins. Specifically based on our results, it can be concluded that FG Nups are different from DisProt proteins in terms of charged residue content and charged residue distribution. FG Nups have a lower charge content than DisProt (Figure 3-2). We also showed that they contain the same density of order-promoting amino acids, meaning that FG Nups are on average or partly less disordered than other disordered proteins (represented by DisProt proteins). On the other hand, based on the analysis presented in Figure 3-4 and Figure 3-5, the main difference between the two datasets is presence of exceedingly long largest positive LCRs in FG Nups. Also, the data suggests that largest positive LCRs are exclusive to FG Nups.

Disordered proteins like proteins in DisProt dataset are rich in charged residues because these amino acids are known to be disorder promoting [65]. FG Nups are not exempt from this general rule since they fall under the same category of disordered proteins. Therefore, they contain segments in their sequences that are charge rich and look similar to DisProt sequences. These sub-sequences are non-LCR regions of FG Nups. However, interesting patterns are also embedded in FG Nups (largest positive LCRs) that are evolutionarily conserved based on the data in this study and our previous work [35] and, therefore, likely correlate with a function in nucleocytoplsmic transport [79]. Presence of the largest positive LCRs is the reason the charge density of FG Nups is lower than other disordered proteins. DisProt proteins and the charge-rich regions of FG Nups have a high charge density and the large number of charged residues is what forms several small LCRs that cover a large portion of the sequence and make up for a high LCR-covered-fraction (Figure 3-7). However, the origin and function of these small and large LCRs are different. Small LCRs can be considered to be mainly randomly formed, while the largest positive LCRs are embedded in the sequence to facilitate a function. The charge-rich segments of the protein form a rather straight conformation [11,83]. On the other hand, the charge-poor regions (largest positive LCR) are the regions that make FG Nups less disordered because they are less charge-rich and more phenylalanine-dense [35]. Therefore, they are able to form more compact conformations as the hydrophobic effects overcome the electrostatic repulsion [79].

The results of this study suggest that two domains can be observed in the sequences of FG Nups considering charge distribution. One domain is the charge-rich domain that is rich in charged amino acids and therefore forms a relaxed-coil conformation. This domain includes several negative and positive short-length LCRs that are randomly formed due to high density of charged residues. This domain is mainly the only domain that exists in DisProt proteins (in terms of charge distribution). The other domain existing in FG Nups sequences are extended sub-sequences that contain only positively charged residues and a low charge density. This is a unique sequence feature that shows up only once along the amino acid sequences of FG Nups. These patterns (named largest positive LCRs) are evolutionarily conserved, exclusive to FG Nups and what differentiate FG Nups from DisProt. We had also previously shown that these patterns are mostly located toward the N-terminus of FG Nups which is toward the center

of the NPC [35]. All of the above-mentioned facts about largest positive LCRs strongly imply that they are important for the conformation and function of FG Nups.

Our previous coarse-grained molecular dynamics study on the effect of largest positive LCRs had suggested that mutation of charged residues in the LCR make the conformation of FG Nups collapsed and they form an aggregated conformation at the center of the NPC [79]. Our ongoing unpublished studies on the effect of largest positive LCRs on the conformation of FG Nups and their interaction with cargo complexes has shown the regulatory role of largest positive LCRs in this interaction. Our results demonstrate that largest positive LCRs exist because the conformation of FG Nups rely on a low-enough charge density to be able to form compact conformations. Also, the few charged residues should exist for the largest positive LCRs not to fully collapse. Removing the charged residues in LCR make the Nups too sticky for the cargo complex and this may interfere with the transient interactions that are necessary for nucleocytoplasmic transport to happen. Adding charged residues to the LCRs greatly shortens the interactions between FG Nups and the cargo complex which may interfere with the selectivity of the transport process.

Conclusion

This study shows that FG Nups are different from DisProt proteins not inly in terms of fraction of charged residues, but also in terms of charge distribution and charge decoration. The difference is primarily driven by presence of large sub—sequences in FG Nups that contain only positively charged residues, have a low charge density and are located toward the N-terminus of these proteins. These patterns, named largest positive LCRs, are evolutionarily conserved.and occur once in the sequences of FG Nups. Removing these sub-sequences from the dataset, FG Nups will be similar to DisProt proteins in terms of charge distribution. These findings imply that largest positive LCRs are essential for the function of FG Nups. Our previously published data as well as our ongoing research reveals that presence of largest positive LCRs significantly affect the conformation of FG Nups and their interaction with cargo complex.

Methods

Most analyses presented herein were conducted via *in-house* Python scripts, except disorder prediction, which was conducted using ESPRITZ (version 1.3).

Protein Datasets

Two groups of proteins were analyzed in this study; the first group, named FG Nups, was taken from our previous study [35]. This dataset includes 1138 FG Nups from 252 species. This dataset was originally extracted from Uniprot by extracting entries associated with the keyword 'nucleoporin' and choosing proteins with a high percentage of disorder (>30%) and high percentage of FG motifs (>0.15 FG/AA) [27]. The second group, termed DisProt, included more than 1500 disordered regions of 694 proteins adapted from DisProt database [100,101]. DisProt is a database of proteins that contain at least one experimentally determined disordered domain along their sequence. For all of the proteins, only the disordered domain were considered and analyzed.

Sequeqnce Analysis

Abundance of residues is defined as the number of target residues divided by the total length of the disordered region of the protein. A like charge region (LCR) is defined as a region that starts at the first position of a certain charge, and continues until reaching an opposite charge in the amino acid sequence. The residues that were considered negatively charged were aspartic acid (D) and glutamic acid (E), while the positively charged residues included lysine (K) and arginine (R). For example, Figure 3-8 shows a hypothetical positive LCR, with a length of 10 and charge count of 4, followed by a negative LCR, with a length of 9 and charge count of 3. All charged residues were assumed to have the same amount of charge. LCR charge content is defined as the number of charged residues divided by the total length of the LCR. LCR-covered-fraction is defined as the sum of length of LCRs in a protein sequence divided by the total length of its disordered region.

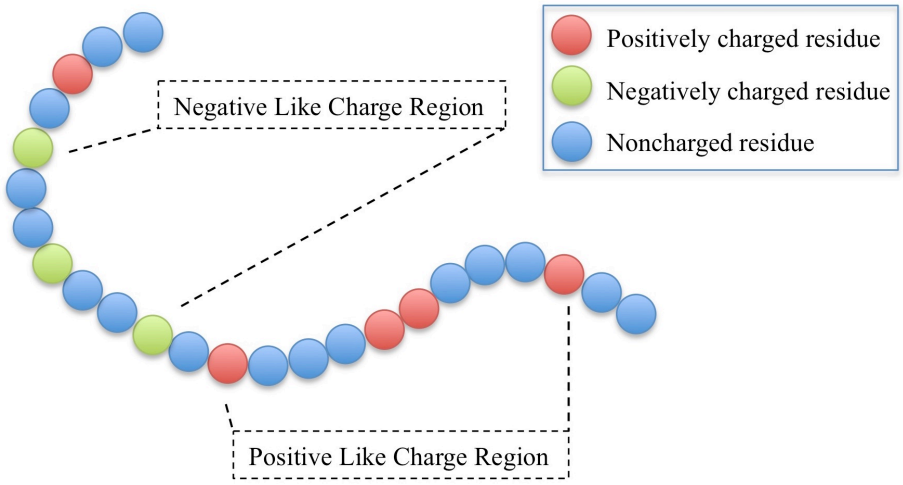

Figure 3-8: Hypothetical disordeded region of a protein containing a positive like charge region (LCR), with a length of 10 and charge count of 4, followed by a negative LCR, with a length of 9 and charge count of 3.

Chapter 4 : Nucleoporin's Like Charge Regions Are Major Regulators of FG Coverage and Dynamics Inside the Nuclear Pore Complex

The work presented in this Chapter has appeared in the following publication:

Peyro, M., M. Soheilypour, A. Ghavami, and M.R.K. Mofrad. "Nucleoporin's like charge regions are major regulators of FG coverage and dynamics inside the nuclear pore complex." PloS One 10, no. 12 (2015): e0143745.

Abstract

Nucleocytoplasmic transport has been the subject of a large body of research in the past few decades. Recently, the focus of investigations in this field has shifted from studies of the overall function of the nuclear pore complex (NPC) to the examination of the role of different domains of phenylalanine-glycine nucleoporin (FG Nup) sequences on the NPC function. In our recent bioinformatics study, we showed that FG Nups have some evolutionarily conserved sequence-based features that might govern their physical behavior inside the NPC. We proposed the 'like charge regions' (LCRs), sequences of charged residues with only one type of charge, as one of the features that play a significant role in the formation of FG network inside the central channel. In this study, we further explore the role of LCRs in the distribution of FG Nups, using a recently developed coarse-grained molecular dynamics model. Our results demonstrate how LCRs affect the formation of two transport pathways. While some FG Nups locate their FG network at the center of the NPC forming a homogeneous meshwork of FG repeats, other FG Nups cover the space adjacent to the NPC wall. LCRs in the former group, i.e. FG Nups that form an FG domain at the center, tend to regulate the size of the highly dense, doughnut-shaped FG meshwork and leave a small low FG density area at the center of the pore for passive diffusion. On the other hand, LCRs in the latter group of FG Nups enable them to maximize their interactions and cover a larger space inside the NPC to increase its capability to transport numerous cargos at the same time. Finally, a new viewpoint is proposed that reconciles different models for the nuclear pore selective barrier function.

Introduction

In our recent bioinformatics study[35], we analyzed more than a thousand sequences of FG Nups across different species and identified evolutionarily conserved specific features within their sequences. One of these features is a region of sequence of FG Nups, in which the charged residues are only either positively or negatively charged. We named these regions as "like charge regions" (LCRs). Our results suggested that largest positively charged LCRs are potentially significant regulators of the FG network inside the central channel of the NPC. We focused our study on largest positively charged LCRs and for the sake of brevity will henceforth call them LCRs. We observed that LCRs co-localize with regions in the sequence that feature a high density of FG repeats and polar residues.

Here, we seek to further explore the significance of these evolutionarily conserved LCRs by examining the biophysical behavior of FG Nups, with the ultimate goal of shedding light on the mechanism of nucleocytoplasmic transport and the role of FG Nups in this selective process. We study the role of LCRs in FG network formation via a recently developed coarse-grained molecular dynamics model of the NPC [30]. The role of LCRs will be explored in three different levels: individual FG Nups, rings of FG Nups including one type of FG Nup, and the whole NPC. In the individual FG Nup simulations we explore the effect of LCRs on the dynamics of isolated FG Nups, whereas simulations featuring the FG Nups rings are performed to understand how Nups within each layer of the NPC interact and form FG networks. In a higher scale, simulations involving the entire NPC model would provide further insight on how the interactions of FG Nups within each layer and across different layers promote FG network formation and facilitate transport. At each scale, the conformational physics presented by the wildtype sequences are compared to that in the mutated sequences with charged-to-alanine scanning in the LCR.

Results

Like Charge Region (LCR) charged residues govern the size and dynamics of FG Nups

We previously showed that the largest positively charged LCRs might play a regulatory role in FG Nup distribution [35]. As the first step, dynamics of yeast (Saccharomyces Cerevisiae) FG Nups, which contain positive LCRs (including Nsp1, Nup42, Nup49, Nup57, Nup100, Nup116, Nup145N) were individually explored in the present study. Hereafter, for the sake of simplicity, the largest positively charged LCRs would be referred to as LCRs. Each FG Nup was modeled both in wildtype and with its charged residues embedded within its LCR mutated to Ala (Figure 4-1). For the sake of accuracy, each simulation was repeated three times. Comparison of the physical behavior of these two configurations demonstrates the role of LCRs in FG Nup behavior. End-to-end distance as well as RMSD of the LCRs and the whole sequence were measured for each FG Nup for the wildtype and mutated sequences (Figure 4-2). It is worth mentioning that LCRs of yeast FG Nups have a significant overlap with FG-rich regions [35].

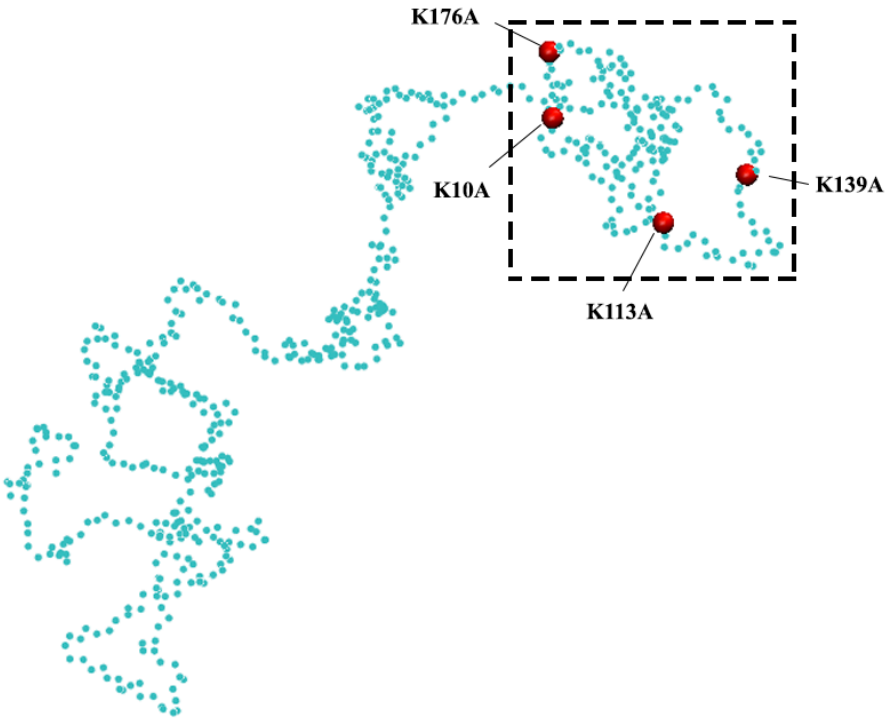

Figure 4-1: A random snapshot of Nsp1 simulation. The "like charge region" (LCR) (see text for definition) is boxed and demonstrates the four Lys residues that form the LCR of Nsp1. Mutated Nsp1 refers to the same sequence with these charged residues mutated to Ala.

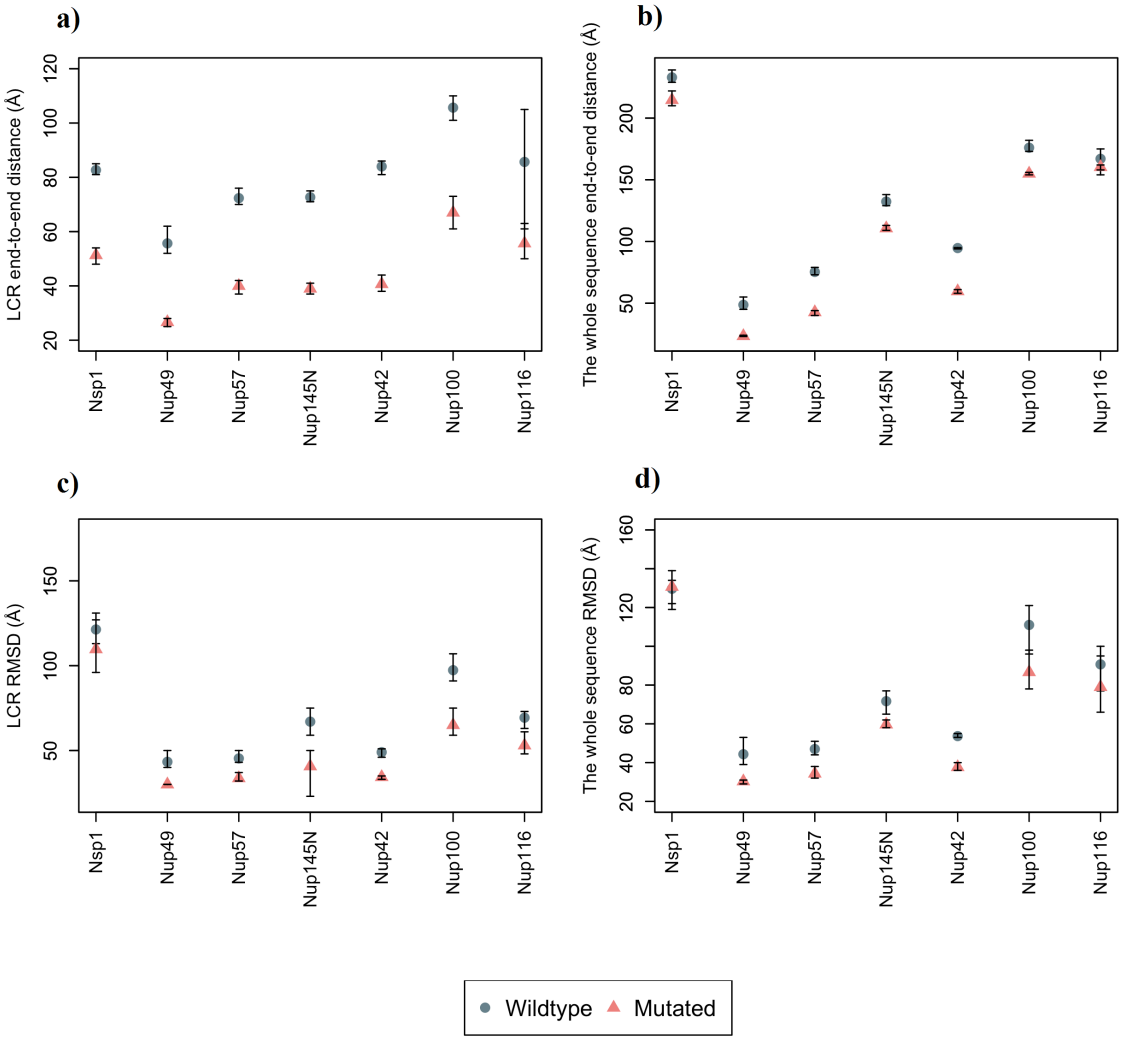

Figure 4-2: End-to-end distance and root mean square deviation (RMSD) measured for wildtype and mutated individual FG Nups. Mutated sequences represent a charged-to-Ala scanning within their like-charge regions (LCRs). Each simulation was repeated three times and the average and the error bars are represented in the graphs. (a) End-to-end distance of the LCR of FG Nups. Mutated sequences demonstrate a significant (35%-52%) reduction in end-to-end distance of their LCR compared to wildtype sequences, which demonstrates the role of LCR charges in maintaining the size of these regions. (b) End-to-end distance of the whole sequence of FG Nups. FG Nups with a long LCR compared to their length (Nup42, Nup49, Nup57) show nearly identical drop in end-to-end distance for LCRs as well as the whole sequence (37%-52%). On the other hand, FG Nups with a shorter LCR compared to their length, demonstrate a significantly lower drop for the end-to-end distance of the whole sequence (4%-17%). (c) RMSD of the LCR drops appreciably (9%-39%) in the mutated sequences, which suggests charges of LCRs as regulators of the highly dynamic nature of these regions. (d) RMSD for the whole sequence drops for up to 32% compared to the wildtype sequence.

Our results demonstrate that charges embedded in LCRs influence the size and conformational dynamics of these regions (Figure 2-2). The end-to-end distance of the LCR drops significantly upon the corresponding charged-to-Ala mutations. FG Nups with a long LCR compared to their length (Nup42, Nup49, Nup57) show nearly identical drop in end-to-end distance for the LCRs as well as the whole sequence. However, Nsp1,

Nup100, Nup116, and Nup145N, which have relatively short LCRs with respect to the length of their entire sequence, exhibit a lower reduction in the end-to-end distance of the whole sequence compared to that of LCRs. This implies that the end-to-end distance of these Nups is primarily governed by the rest of the sequence, which means that regions containing LCRs maintain a more compact conformation as compared to the rest of the sequence. Our results are in agreement with previous studies, which reported that some of the yeast FG Nups, including Nsp1, Nup100, and Nup116, show two distinct domains with completely different physical characteristics [11,26]. One domain forms collapsed coil networks, while the other forms a stalk-like domain. In addition, variations in RMSD in LCRs as well as the whole sequences of mutated FG Nups demonstrate the influence of LCR charges on dynamics of FG Nups. Our previous study showed that LCRs overlap with FG-rich regions[35]. Therefore, here we conclude that LCRs of yeast FG Nups are the main regulators of size and dynamics of FG-rich regions.

Analysis of FG Nup rings highlights the role of LCRs in FG Nup distribution

The NPC features an octagonal structure composed of spokes arranged in an eightfold rotational symmetric manner [44]. We divided the NPC into cross-sectional sections of different FG Nups to explore the mechanics of different regions ("rings") of the NPC. Each ring consists of eight copies of a single FG Nup, arranged in an eightfold symmetric manner. Each ring was simulated for three times, each for 1μs, both in its wildtype as well as mutated forms. Since FG-repeats are suggested to be the main regulator of the active nucleocytoplasmic transport via weak interactions with cargo complexes [2,13,22], FG density plots were used as readouts of the simulations. Density of FG motifs was averaged over the simulation time for both wildtype and mutated rings to compare the two models. The results of the first trial of the simulations for Nup145N, Nup42 and Nsp1 are presented in the main text, while results of their second and third trial as well as simulations for the rest of the FG Nups, including Nup49, Nup57, Nup100 and Nup116, are presented in the supplementary material. It is worth mentioning that in these simulations, the blobs of the NPC scaffold are taken into account [30] which cause the empty space near the scaffold in some of the simulations.

Nup145N: LCRs promote cross-interactions between FG Nups

Nup145N is one of the relatively short FG Nups with a disordered region of 433-residues long. Its LCR (215-residues long) is half of the length of the whole sequence. According to Figure 2-2, mutation of LCR charges in Nup145N cuts its end-to-end distance of the LCR into half, while decreases the end-to-end distance of the whole molecule by just ~16%. Since LCRs are co-localized with FG regions[35], mutation of LCR charges in mutated ring leads to highly-dense FG regions (Figure 4-3). On the other hand, the wildtype ring exhibits a more even distribution of FG-repeats. The relatively short length of Nup145N causes its FG network to form a doughnut-like cloud, close to the NPC wall, rather than a network at the center [30]. Mutating charges in LCR not only causes the FG motifs to aggregate into high-density complexes, but also limits the interaction between different copies of Nup145N. The cross-interactions between Nup145N copies are limited and, consequently, they get trapped (in groups of two or three) into local minima of the energy landscape (S1 and S2 Videos, please see the publication). Since the

repulsion between same charges does not exist anymore, the system does not get enough energy to jump out of these local minima and, hence, FG Nups remain interacting in the same groups to the end of the simulation. Reduction in interconnections between different copies of the FG Nup leads to restricted space covered by FG-repeats. These results are also discernible in the other two trials of the simulation, which confirms the regulatory role of LCRs (Figure 4-S1). As a result, one could postulate that the regulatory effect of LCRs in nucleocytoplasmic transport (in the case of Nup145N) is to increase the communication between different copies of the Nup and, subsequently, maximize the covered area by FG-repeats to facilitate the transport.

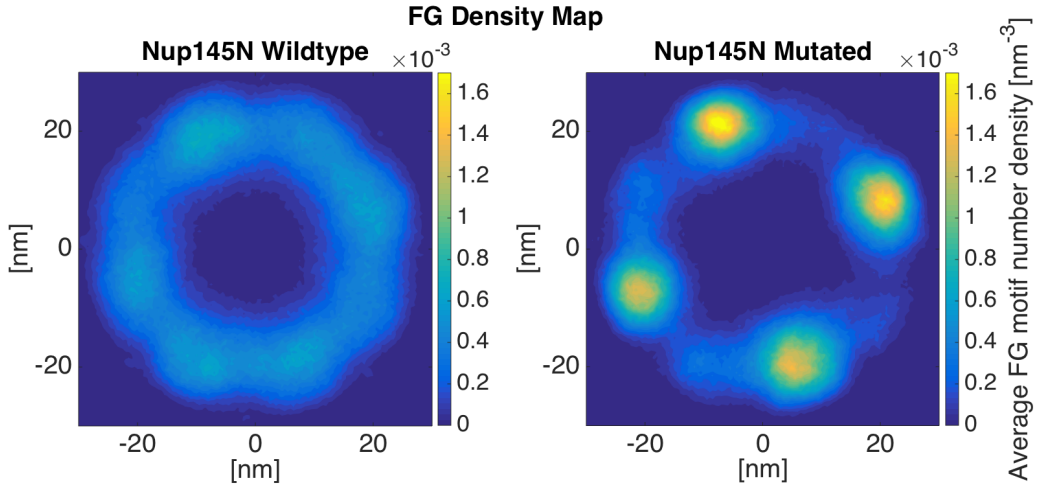

Figure 4-3: FG density plot of Nup145N ring. The wildtype ring shows a more even and inter-connected network of FG-repeats. On the other hand, FG network formed in the mutated ring is more aggregated to high-density regions with more limited interactions between different copies of Nup145N. This highlights the regulatory effect of like charge regions (LCRs) in effective distribution of FG-repeats inside the NPC. The rings are shown from the top view.

Nsp1: LCRs facilitate the formation of a larger FG-covered area

Nsp1 is a relatively long FG Nup, containing 617 residues in its disordered region, while its LCR is only 167 residues long. According to the results from simulating individual wildtype and mutated Nsp1 sequences (Figure 2-2), mutation of LCR charges leads to a 39% decrease in the end-to-end distance of the LCR, while the end-to-end distance of the whole Nup does not change significantly. The same behavior is observable in the case of Nsp1 rings (Fig 4-4). The end-to-end distance of the Nsp1 copies does not change significantly; as the high-FG-density region is formed at the center in both cases [26,35] (wildtype and mutated). However, a considerable reduction in the end-to-end distance of the LCRs results in the formation of a relatively high-density region at the center of the NPC (Fig 4-4-Mutated). The wildtype ring shows a lower peak and a more even distribution of FG motifs compared to the mutated ring. The LCR mutated regions, which no longer contain charged residues, collapse into a more compact configuration and form a high-density region at the center of the ring, with the rest of the ring showing a

relatively lower density of FG motifs. The reason underlying the formation of the density peak at the center of the ring is the long, highly-charged non-LCR region that forms an extended relaxed coil and brings the LCR (or in other words the FG domain) to the center of the NPC [26,35]. As the charges in the LCR are mutated, it collapses into a more compact form and allows the extension of the relaxed coil region (Fig 4-4-Mutated). The same behavior is observable in the case of Nup100 (Figure 4-S2) and Nup116 (Figure 4-S3). The other two trials of this simulation confirm these results as well Figure 4-S4.

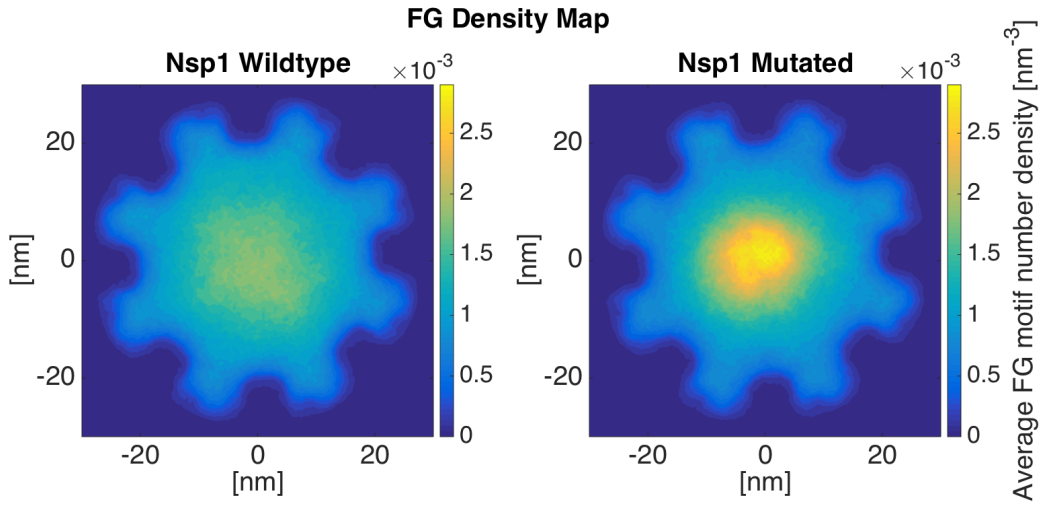

Fig 4-4: FG density plot of Nsp1 ring. FG density is distributed more evenly in the wildtype ring. A significant peak in FG density is observable in the mutant ring due to the collapse of the like charge region (LCR), which has a large overlap with the FG domain. Mutating charged residues in LCR allows the hydrophobic effect of the FG motifs to dominate, resulting in the formation of the high-density regions. These collapsed regions from all copies of Nsp1 stick together and form a relatively low-dynamic, high-density region. For the second and third trials of these simulations, please refer to S4 Fig.

Nup42: LCRs regulate the formation of FG cloud near the NPC wall

Nup42 is one of the relatively short FG Nups with a short disordered region, containing 382 residues. Nup42's LCR is a 305-residue long domain, covering almost the whole disordered region of the Nup. Shorter Nups (Nup42, Nup49, and Nup57) mostly fluctuate near the wall of the NPC rather than covering the center of the pore (Fig 4-5, Figure 4-S5 to Figure 4-S7). Therefore, the wildtype ring shows a doughnut-like high-FG density shape (Fig 4-5-Wildtype). Comparing the wildtype and mutated results, again, the lack of inter-communication between different copies of Nup42 in mutated ring results in local highly-dense regions, while leaving most of the space free of FG motifs. The same behavior is observed for Nup49 and Nup57 (Figure 4-S5 and Figure 4-S6). For the second and third trial of Nup42 simulation, please refer to Figure 4-S7.

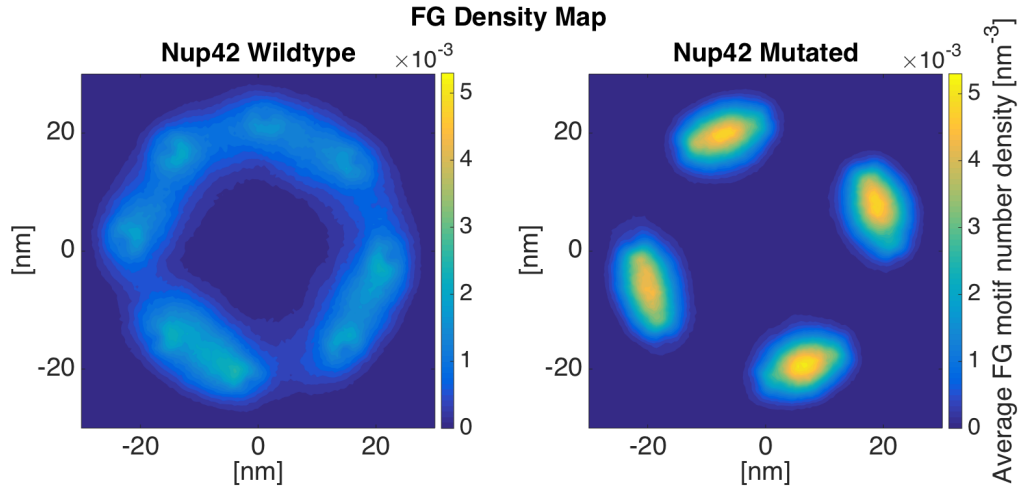

Fig 4-5: FG density plot of Nup42 ring. FG density shows a doughnut-like FG cloud that covers the near-the-wall space through dynamic inter-communication between different copies of Nup42. Mutated ring shows locally trapped groups of two Nups that are not able to escape the conditions due to the lack of dynamicity.

To evaluate the functional role of LCRs and their regulatory effect in the behavior of FG Nups, a control simulation was designed. A new mutation was applied to the Nsp1 sequence, where four Lys residues (the same type and number of charges as in the LCR) were mutated to Ala in a non-LCR region of Nsp1. The mutated residues are located at positions 365, 369, 370 and 376. The results of this control simulation along with wildtype simulation are presented in Figure 4-S8. In contrast to the LCR mutation case, where the maximum density of FG repeats increased by 50% (as compared to that in the wildtype case) (Figure 4-4), the control mutation does not significantly change FG density distribution. The maximum density of FG repeats is altered by less than 10% compared to wildtype simulation. This further highlights the substantial regulatory role of LCRs in FG Nups distribution inside the NPC.

The whole NPC: How LCRs assist FG Nups to dynamically cover the space inside the NPC

In order to see the overall effect of LCRs on the distribution of FG motifs inside the NPC, the whole NPC, including the hourglass-shape wall of the central channel and the disordered domains of all FG Nups was simulated using the coarse-grained molecular dynamics model. In the case of the mutant NPC, LCRs with positive charges were mutated in the corresponding FG Nups, i.e. Nsp1, Nup42, Nup49, Nup57, Nup100, Nup116, Nup145N. For the rest of the FG Nups (Nup1, Nup2, Nup159 and Nup60), which do not contain positive LCRs[35], wildtype sequences were used.

Wildtype NPC shows a doughnut-shaped FG density plot, which was previously observed as well (Figure 4-6)[30]. The only difference is that the peak density is slightly shifted upward compared to the density plots in [30]. The reason is that while we used only the disordered regions of Nup49 and Nup57 [11] in our study, the whole sequences of these Nups were used in [30]. Therefore, it seems that using the whole sequences results in a smoother high-density region, where the longer sequences of Nup49 and Nup57 slightly

pull down the high-density region that forms the doughnut-like region [30]. It was suggested that the doughnut-shaped FG distribution is compatible with the transportation mechanism observed by single molecule tracking experiments [102], where passive transport of small molecules occur through the center of the NPC. Despite modest differences, our results are in line with this proposed mechanism. In addition, the low-density central region of FG Nups is in line with experimental observation of a 10 nm empty hole at the center of the NPC [3]. Interestingly, mutation of charges in LCRs dramatically dysregulates this distribution (Figure S4-6). In the mutated NPC, the size of the high-density region is significantly shrunk and the peak of the density substantially increases. On the other hand, the rest of the space inside the NPC shows a very low density, implying that the dynamics of the system has significantly reduced and the system is almost frozen with very small vibrations. Therefore, the system has fallen into a system of collapsed FG regions and relatively straight stalk domains with significantly low dynamics. As a result, passive transport pathway is occluded by the high-density region and active transport pathway is disrupted due to the low dynamics of FG Nups and significantly smaller FG-covered area. It is noteworthy that all these observed dysregulations are due only to a handful of mutations (4-10) per sequence, demonstrating the regulatory role of LCRs.

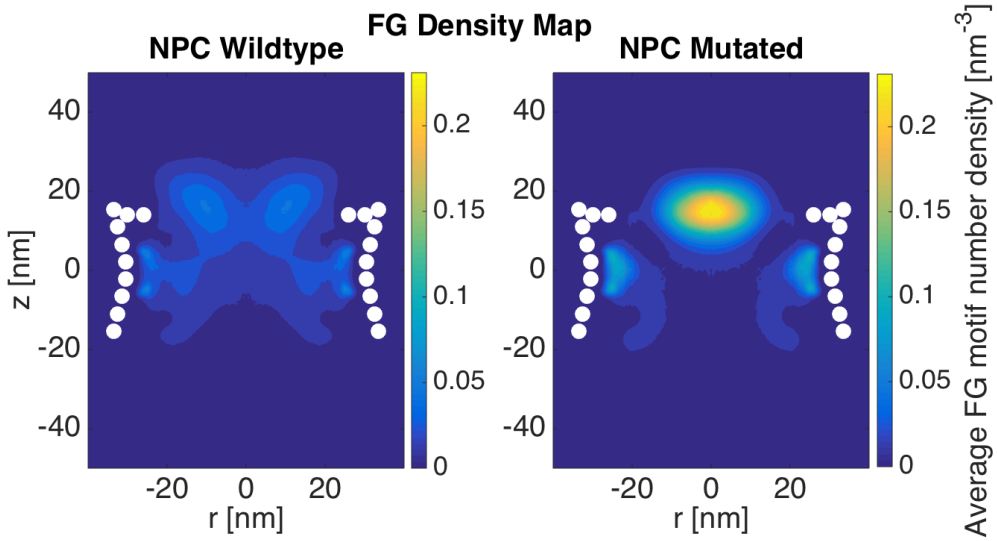

Figure 4-6: FG density plot of the whole NPC. Wildtype FG Nups form a doughnut-shaped high-density region at the entrance of the NPC[30], as well as a region next to the wall in the central channel with a slightly lower FG density. However, mutated like charge regions (LCRs) lead to a highly-dense, collapsed region at the center of the pore as well as near the wall in the central channel, while leaving the rest of the pore free of FG motifs. This comparison demonstrates the role of LCRs in promoting the dynamics of different FG Nups, resulting in a larger covered area.

Discussion

"LCR-regulated Coverage"

Through a bioinformatics study[35], we recently identified some evolutionarily conserved sequences of charged residues within the sequences of FG Nups that possess only positive charges, accordingly called "like charge regions" (LCRs). In this work, we sought to divulge physical significance of these specific domains of FG Nups in the distribution of FG-repeats inside the NPC and nucleocytoplasmic transport. We studied this role on yeast FG Nups and in three different levels, namely individual FG Nups, rings of FG Nups, and the whole NPC.

One of the most interesting aspects of the nucleocytoplasmic transport is the ability to handle more than a thousand cargos in every single second, and transport them with a dwell time of less than 10ms [103–105]. As a result, at every single snapshot, there are multiple cargos in the NPC and, consequently, in each ring of FG Nups. Therefore, each ring is required to accommodate transport of multiple cargos. Our results showed that wildtype rings exhibit a larger area covered by FG motifs, while mutated rings show a more collapsed FG network with the rest of the space left almost free (Figure 4-3- Figure 4-5). Considering that the NPC is a high throughput cellular machine, it should be able to accommodate multiple cargos at each time. Even distributions of FG repeats with a larger coverage inside the pore would facilitate more translocations. Wildtype rings are hence capable of translocating more cargos at each time step. Mutated rings represent a lower FG-covered area with some high FG density spots. This type of FG distribution might be more efficient for traffic of a few cargos per time step but is not necessarily effective for a high throughput selectivity offered by the NPC. This comparison focuses on transportation mechanism in a lower resolution, considering the transport requirements in each layer of the NPC in axial direction. Therefore, it could be speculated that LCRs are primarily co-localized with FG domains to maximize the dynamics of FG Nups and specifically regions of high FG density inside the NPC and facilitate a larger coverage of the central channel by FG-repeats. In other words, without LCRs, FG Nups would not be able to dynamically cover the whole space inside the pore (Figure 4-6) (since their FG domains collapse into compact forms), and cannot handle the numerous cargo-complexes passing through the NPC at each time step. This is in line with our individual FG Nup results as well, where mutation of charges in LCRs resulted in significantly lower RMSD and end-to-end distance (Figure 4-2).

Our simulations demonstrated the categorization of FG Nups into two distinct groups[11,26,35]. While some of them (Nsp1, Nup100, and Nup116) bring their FG domain to the core of the central channel (Fig 4-4 and Figure 4-S2 to Figure 4-S4), others (Nup42, Nup49, Nup57, and Nup145N) cover the area closer to the wall of the NPC (Figure 4-3 and 4-5, Figure 4-S1 and Figure 4-S5 to Figure 4-S7). This observation is in line with the double-zone transport, previously suggested by Yamada et al. [11]. According to our model, Nsp1, Nup100, and Nup116 are among FG Nups that form a highly-dense FG region at Zone1 (Figure 2-7-green arrow) (leaving a low-density region at the center of the NPC (Figure 2-7-yellow arrow)), while others are responsible for transport through Zone2 (Figure 2-7-red arrows).

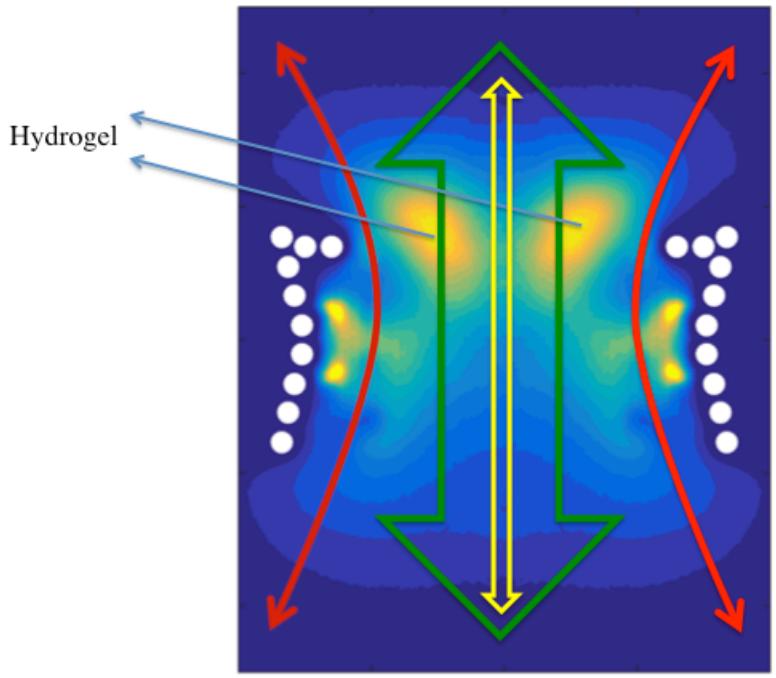

Figure 4-7: A schematic on top of the FG density plot of the wildtype NPC, showing different nucleocytoplasmic models observed in our results. According to the forest model [11], two transport pathways exist (green arrow for Zone1 and red arrows for Zone 2). Hydrogel model suggests that FG Nups form homogeneous FG-networks that transporters need to dissolve in order to pass through the NPC [20,21] (denoted by two blue arrows). Reduction-of-dimensionality, on the other hand, explains the transport as a 2D random-walk of transporters on an FG-rich surface over the wall of the central channel [25](red arrows). Our results demonstrated that like charge regions (LCRs) are the major regulator of the formation of these different regions for an efficient transport.

On the other hand, our model supports the hydrogel model as well. The fact that rings of FG Nups form homogeneous meshwork inside the NPC (Figure 4-3 to 4-6), as opposed to sticking out without merging with each other [11], suggests that they may form a hydrogel (Figure 2-7-blue arrows) [33], where Kaps need to compete with FG-repeats to break FG-FG binds and travel through the NPC [33]. Our results did not specifically demonstrate the reduction-of-dimensionality model, however, this transport mechanism could be implied from our results. As seen in Fig 4-5 and 4-6, Figure 4-S5 to Figure 4-S7, short FG Nups form an FG-rich region next to the wall of the central channel. It could be speculated that, since these regions are much smaller than the high-density region next to the center (Figure 2-6), rather than forming a hydrogel, they form a rich surface of nuclear transport receptors (NTRs) for transport of cargo-complexes. Therefore, rather than dissolving into this region, transporters slide on this surface. Collectively, one can postulate that the nucleocytoplasmic transport could be defined as a combination of three models, namely hydrogel, forest, and reduction-of-dimensionality models. According to the forest model, FG Nups are categorized into two groups, forming two distinct regions of active transport pathways (Figure 2-7-green and red arrows). Zone 1, which is closer

to the center of the pore, is a doughnut-like structure with hydrogel characteristics for active transport (Figure 2-7-green arrow) and a low-density/empty space in between for passive transport [30,69] (Figure 2-7-yellow arrow). Zone 2, on the other hand, is an NTR-rich surface that enables a two-dimensional random walk for transporters (Figure 2-7-red arrows). In this work, we showed that formation of these different regions is highly dependent on LCRs to increase the dynamics of FG Nups for different purposes. While LCRs of some of the FG Nups, including Nsp1, mainly regulate the size of the hydrogel formed in the middle (Fig 4-4 and 4-6, Figure 4-S4), as well as a small opening at the very center of the NPC (Figure 2-6) [3,69], in other FG Nups, including Nup42, they facilitate interactions of different copies of FG Nups to cover a larger space inside the pore (Figure 4-3, 4-5 and 4-6). Accordingly, this model shall be referred to as the "LCR-regulated Coverage" (Figure 2-8), in which LCRs play a major role to enable FG Nups to form two potential pathways for cargo transport. However, further simulations including the cargos are required to lend more support to this model, which would be our future modeling study.

It is worth noting that structural and conformational characteristics of FG Nups are likely to be impacted by the presence of cargos. Nonetheless, the studies of the sequence, conformational dynamics, and biophysical behavior of FG Nups have been traditionally conducted in the absence of cargos[11,26,27,30,35]. Computational limitations have also hindered long-term simulation of these dynamic processes. Although 1μs is a considerable achievement for simulating the entire NPC, it may still be too short to capture a full picture of the FG Nup conformational dynamics. However, it should be noted that due to coarse-graining, time in our simulations is not representative of the physical time and should be multiplied by a speed-up factor to account for that [30]. Altogether, we hope that our study provides valuable insight about the mechanisms of nucleocytoplasmic transport and inspire further experimental and computational studies.

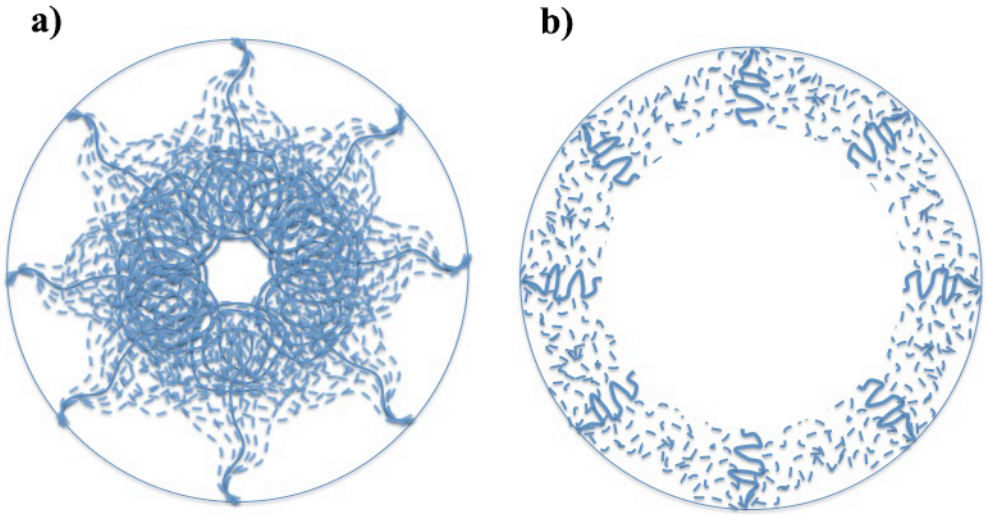

Figure 4-8: A schematic summarizing the role of LCRs in distribution of FG Nups inside the NPC, named "LCR-regulated Coverage", LCR refers to like charge regions. Solid lines show a hypothetical initial

configuration of FG Nups, while dotted lines show different configurations that FG Nups take over time. FG Nups are categorized into two groups. While one group (including Nsp1, Nup100, and Nup116 shown in (a)) bring their FG domain close to the center of the NPC [11,26] and collectively form a doughnut-like FG network, others (including Nup42, Nup49, Nup57, Nup145N shown in (b)) mostly cover the space next to the wall of the pore. In the first group, LCRs are primarily responsible for the size of the hydrogel formed in the middle (Fig 4-4), as well as a small opening at the very center of the NPC (Figure 4-6) [3]. On the other hand, in the second group, LCRs mainly promote interactions of FG Nups in the same layer to cover a larger space inside the pore (Figure 4-3 and 4-5).

Methods

A one-bead-per-amino acid coarse-grained molecular dynamics model was used to model the NPC and explore FG network formation [30]. The model is specifically designed and optimized to study biophysical behavior of intrinsically disordered proteins. Amino acids are represented by single beads with a fixed mass and distance of 120Da and 0.38nm, respectively. In all of the simulations, only disordered domains of FG Nups were considered, which are obtained from [11]. The force-field developed for this model accounts for different biophysical factors including bending and torsion potentials between neighboring beads, an implicit solvent, ion screening effect, and hydrophobic and electrostatic interactions. The interaction between FG Nups is governed by a modified Lennard-Jones potential. The Lennard-Jones potential consists of two parts, the first part takes care of the excluded volume interaction while the second part is related to the hydrophobicity, which changes according to the nature of the amino acids. The model has proven accurate and useful in simulating biopolymeric behavior of intrinsically disordered proteins, specifically FG Nups [30,106].

FG Nups were studied in three different levels. First set of simulations was conducted on individual FG Nups. Each Nup was simulated three times, each for 100ns. Second set of simulations was performed on eightfold ring arrangements of FG Nups. Ring arrangement of FG Nups has been used before for studying nucleocytoplasmic transport in sections of the NPC [26,34]. Each simulation contains eight copies of a single type of FG Nup arranged in a ring configuration to represent various regions of the NPC [26]. Diameter of each ring is equal to the diameter of the NPC at the corresponding tethering point of the FG Nup. Rings are 10nm in height, representing the maximum vertical space that each FG Nup sweeps from *in silico* experiments [30]. Each ring was simulated three times, each for 1 μs. Third set of simulations was done on the whole NPC model. The geometrical model used in [30] was utilized, which is based on the geometry of the yeast NPC scaffold and FG Nup anchoring points [3,43]. The scaffold is modeled as a solid surface with only excluded volume interaction with FG Nups. FG Nups are fixed at their corresponding tethering point with an initial conformation taken from single FG Nup simulations. In the case of ring and whole NPC simulations, each simulation was first energy-minimized in a three-step procedure to remove the possible clashes between the FG Nups. Then, each simulation was initiated with a 100 ns energy minimization run, followed by a 900 ns production run.

Langevin dynamics simulations were performed using Gromacs molecular dynamics simulation software v4.5.3 [107]. The cut-off distances for Van der Waals and Coulombic interactions are set to 2.5 nm and 5.0 nm, respectively. The cut-off length is defined as the maximum radial distance, in which the electrostatic and Van der Waals interactions are calculated for two beads. However, the effective electrostatic force is determined by the Debye screening parameter which is set to κ=1.0 nm^{-1}. The Langevin friction coefficient is set to 50 ps^{-1}, similar to the collision frequency of water molecules [108]. Density plots of FG Nup rings were generated by assuming the ring is centered in a cylinder with a 100 nm diameter and a 10 nm length, which is discretized into 0.5x0.5x0.5 nm unit cells. In the case of the whole NPC, the cylinder has a diameter of 100nm and a length of 140nm. The number of occurrences of FG-repeats in each cell was counted over the total simulation time. Gridcount package was used to generate the

density plots [109]. A time-step of 0.02 ps was used in simulations. All visualizations as well as end-to-end distance and root mean square deviation (RMSD) measurements were performed via VMD 1.9.1 [110].

Appendix

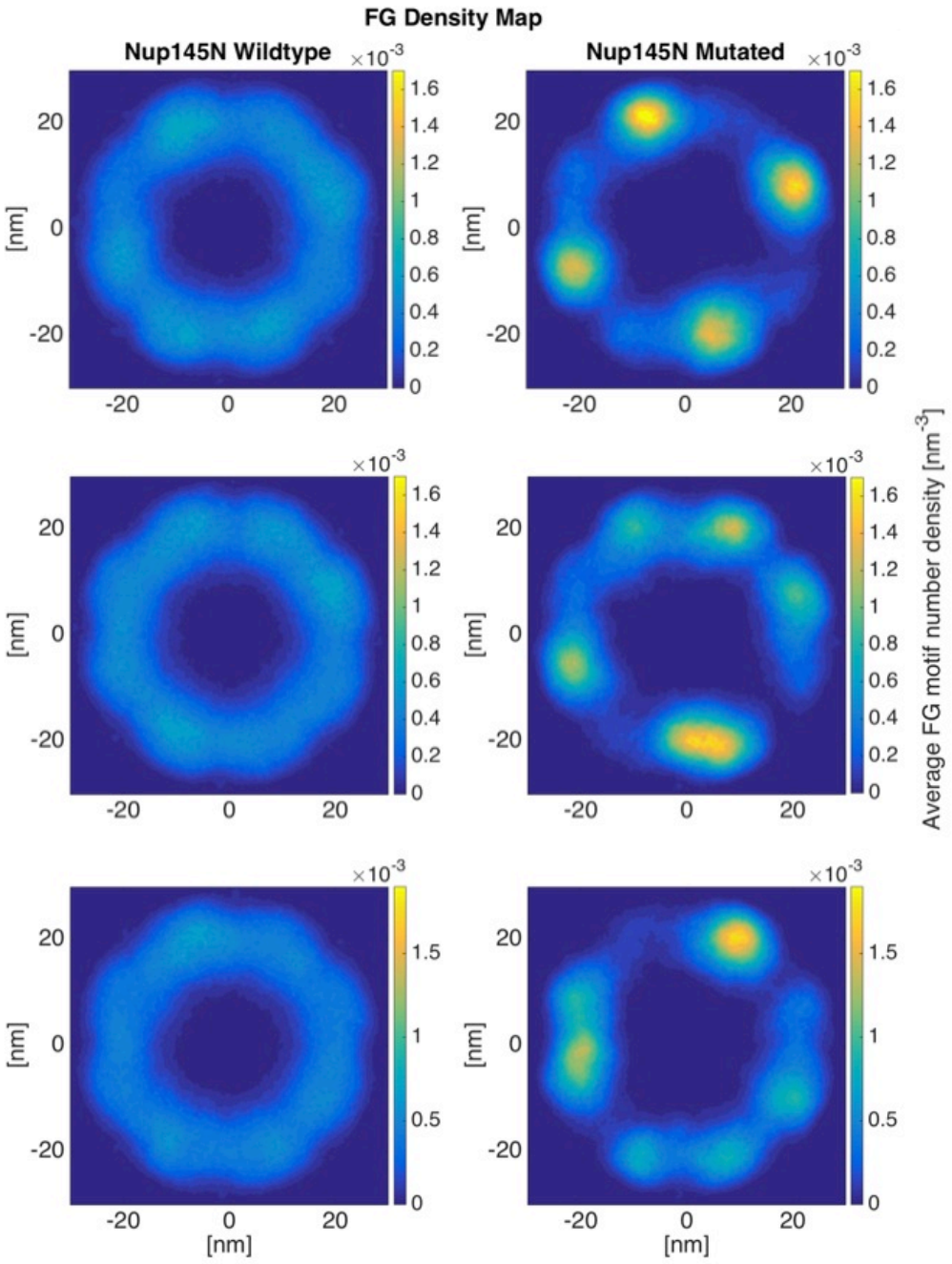

Figure 4-S1: Three simulations of Nup 145N ring. Left figures represent wildtype rings and right figures represent LCR mutated rings. The wildtype ring shows a more even and inter-connected network of FG-repeats. On the other hand, FG network formed in the mutated ring is more aggregated to high-density regions with more limited interactions between different copies of Nup145N. The distribution of FG repeats is different in the mutated rings, which is due to the fact that in each simulation Nups are stuck in a different local energy minimum. However, in all sets of the simulations, the difference between wildtype and mutated rings represent the same regulatory role for LCRs.

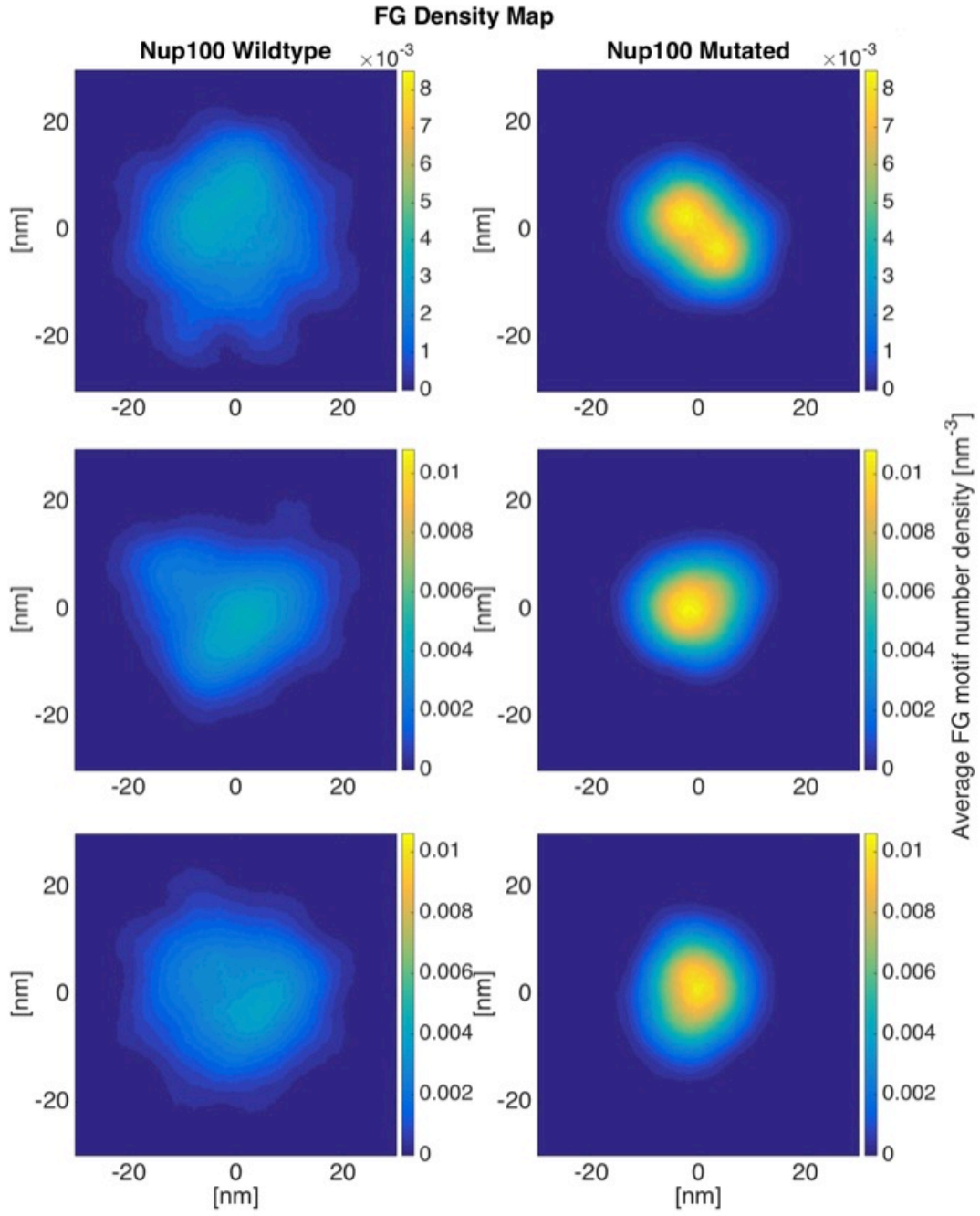

Figure 4-S2: Nup 100 is a relatively long sequence with a disordered domain of 800 residues with the LCR of 506 residues. This Nup behaves similar to Nsp1 and Nup116. The LCR mutation leads to a high FG density region in the center with a less FG covered area. Comparing the three trials of the simulation leads to the same conclusion about regulatory role of LCRs.

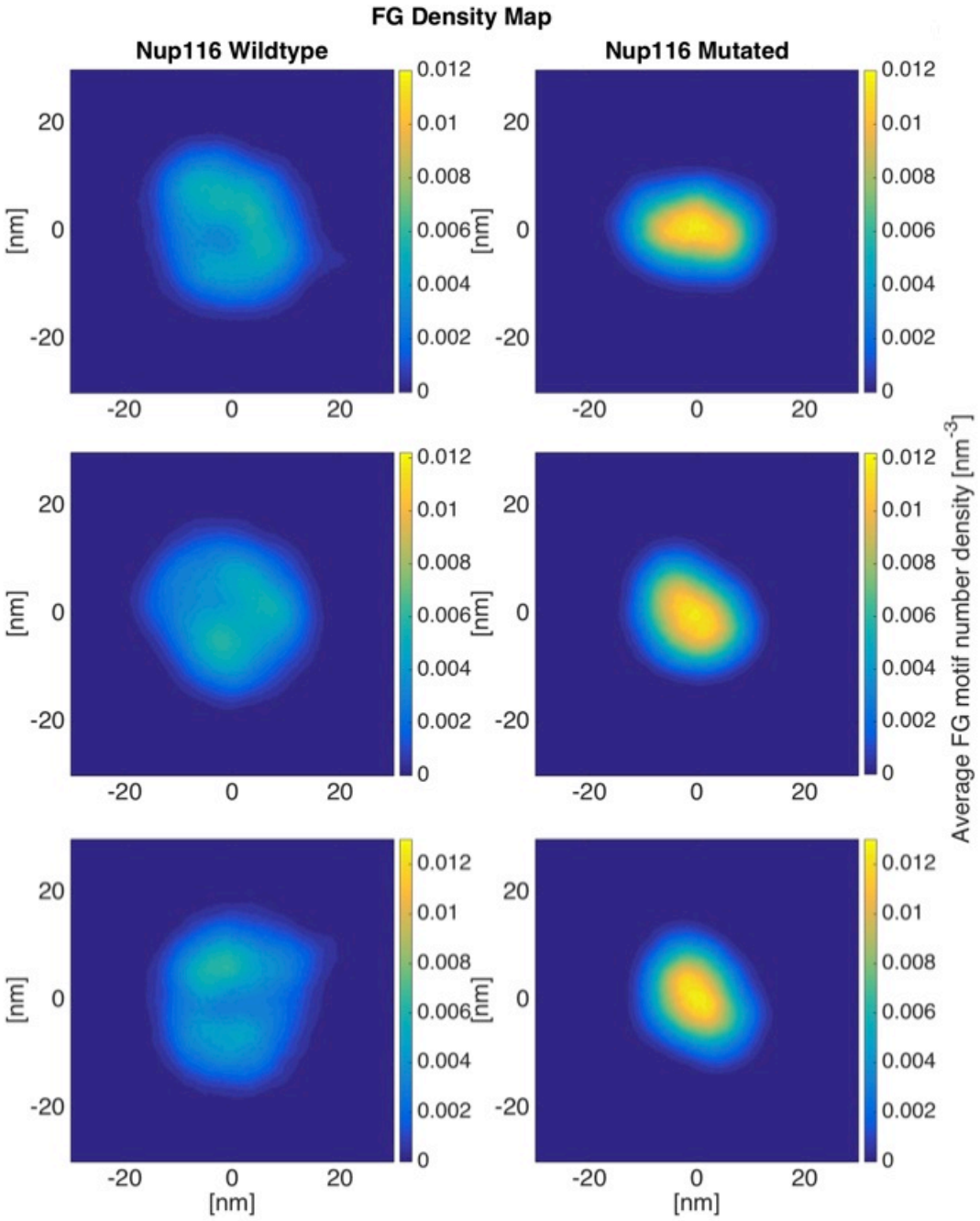

Figure 4-S3: Nup 116 is similar to Nup 100 and Nsp1. It has a relatively long disorder region of 960 residues and contains a 571 residues long LCR. The LCR mutation leads to a high FG density region in the center with a less FG covered area. The three trials represent the same behavior.

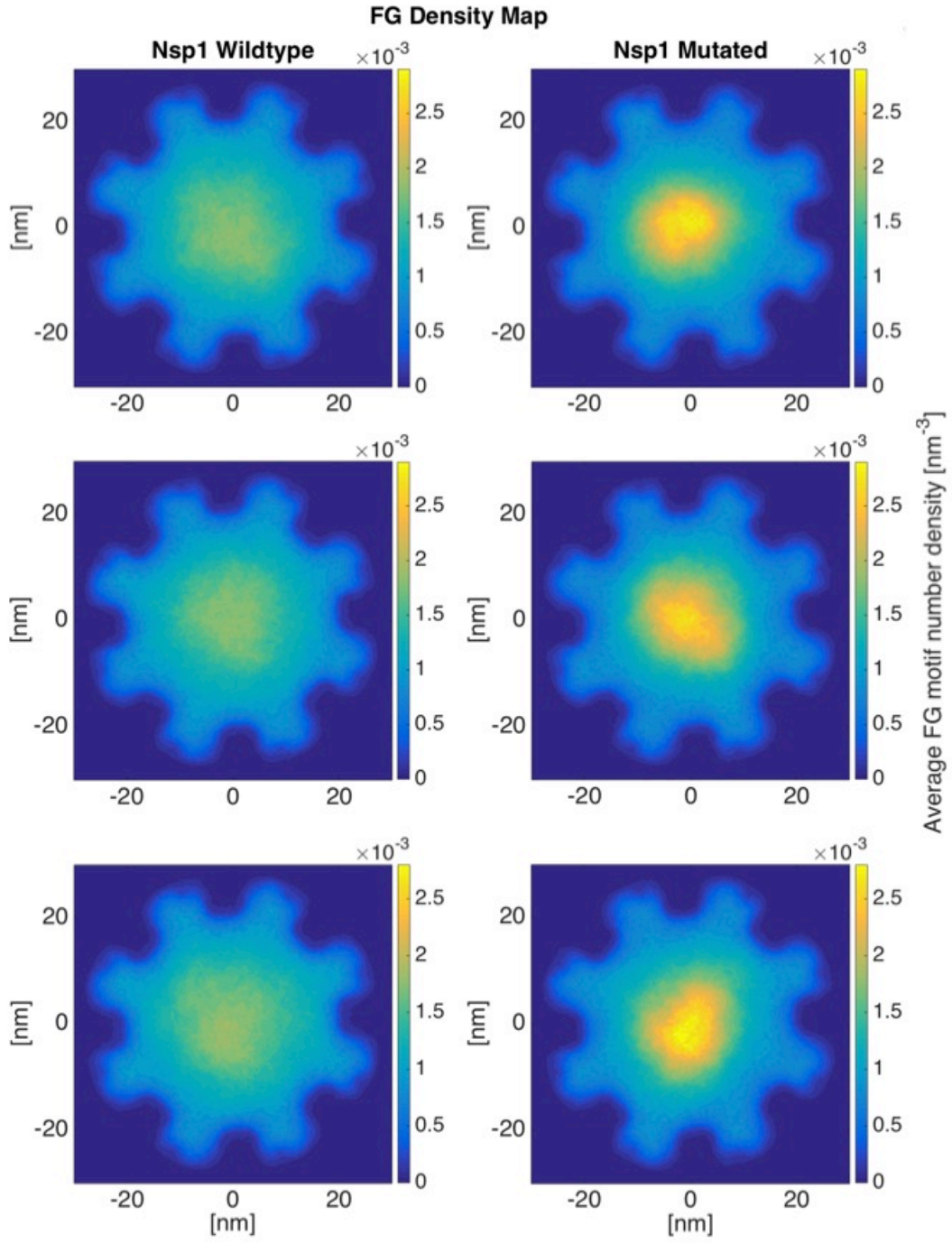

Figure 4-S4: Nsp1 rings. The first trial of the simulations is represented in Figure 4-4 in the main text. This figure shows the results of all the three trials together.

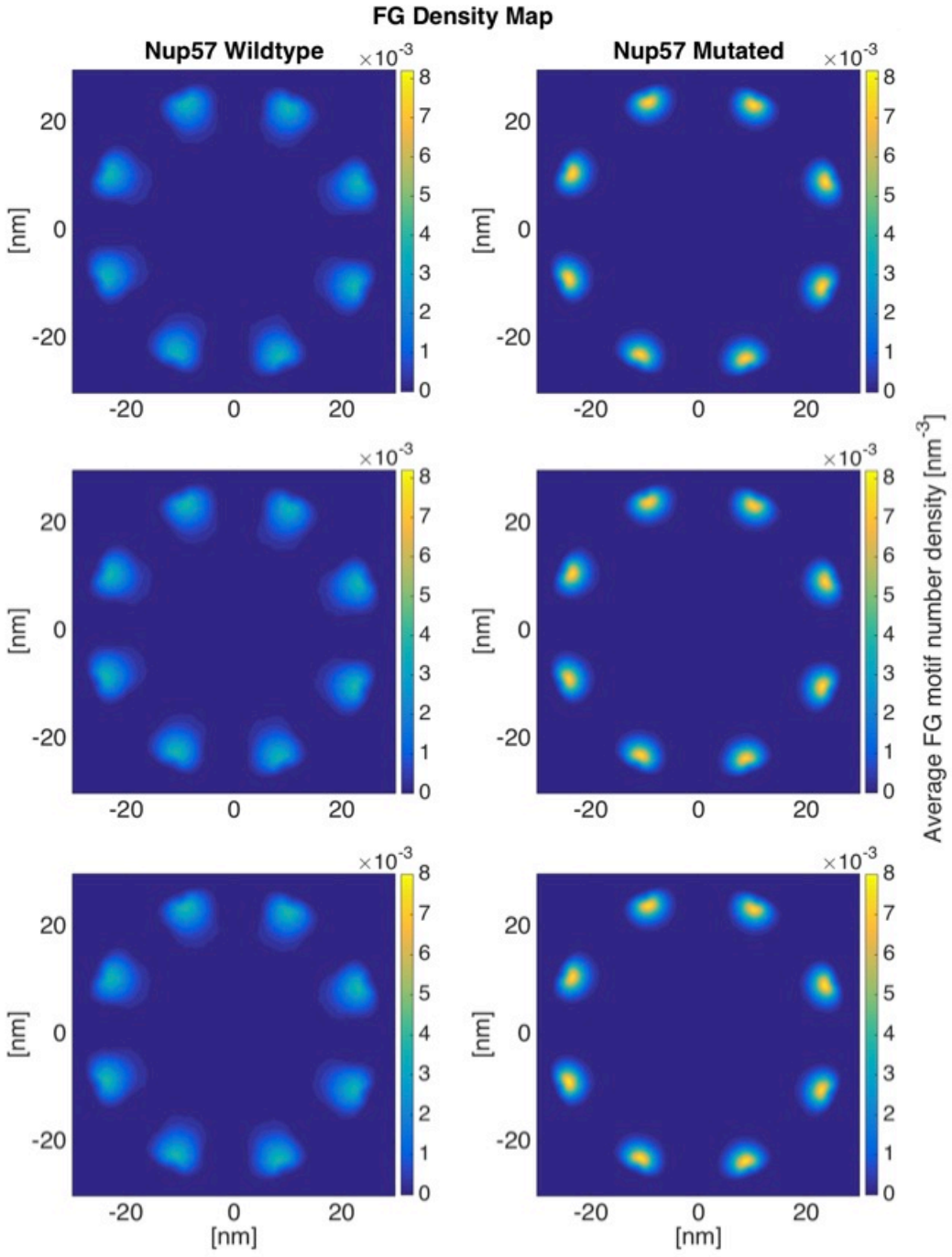

Figure 4-S5: Nup 57 is a relatively short FG Nup, with 255 residues in its disordered region. Its LCR spans the whole disordered region. The wildtype ring shows that, due to the short length of the Nup, an FG-rich region is formed near the wall of the pore. However, in the mutated ring, the size of the FG-rich region considerably decreases, and FG density increases significantly. The three pairs of the figures are the results of the three trials for wildtype and mutated rings.

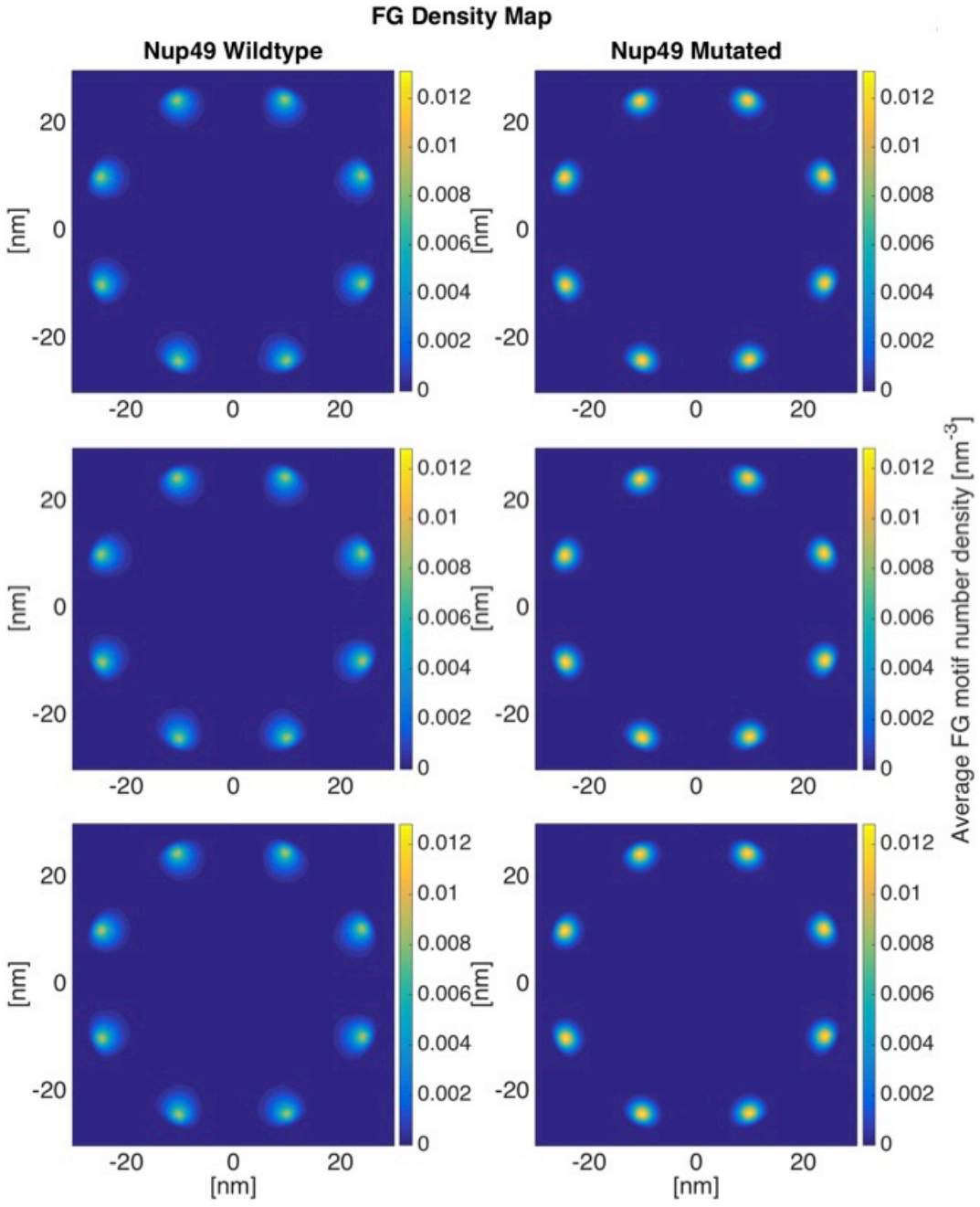

Figure 4-S6: Nup49 has a 251-residues long disordered region, with a 199-residues long LCR. The same behavior as Nup57 is observable here. The three pairs of the figures are the results of the three trials for wildtype and mutated rings.

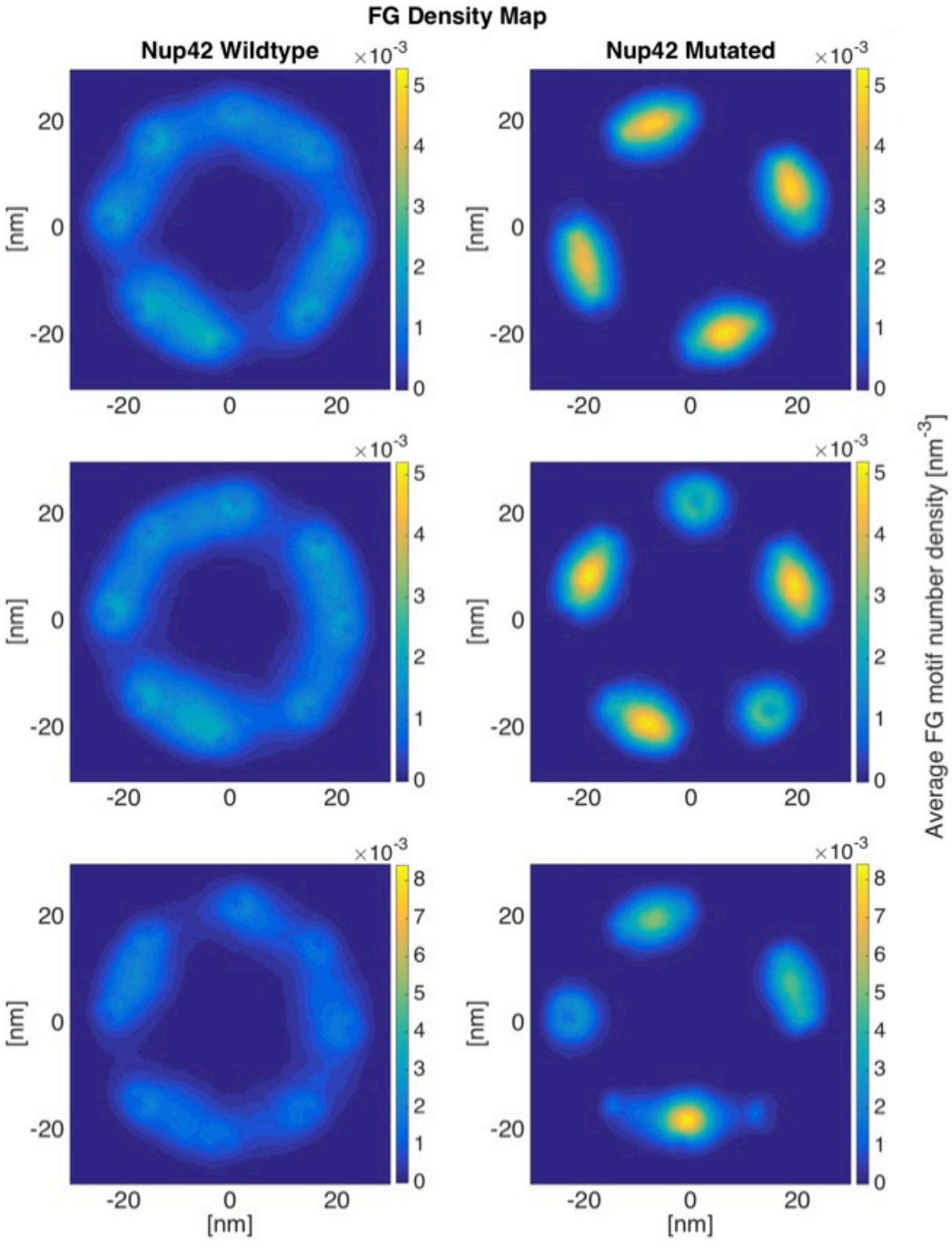

Figure 4-S7: Nup42 rings. The first trial of the simulations is represented in Figure 4-5 in the main text. This figure shows the results of all the three trials together.

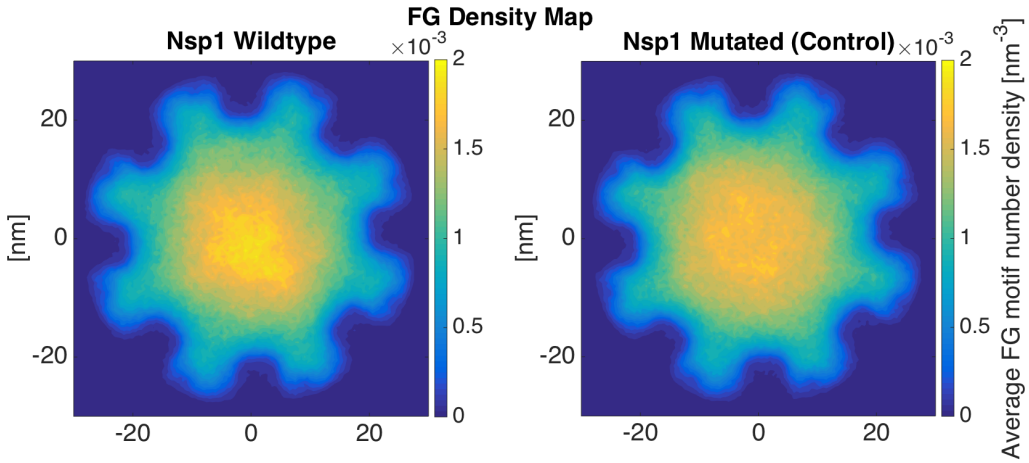

Figure 4-S8: Control Simulation. A new mutation is applied to the Nsp1 sequence, where four Lys residues (the same type and number of charges as in the LCR) are mutated to Ala in a non-LCR region of Nsp1. The mutated residues are located at positions 365, 369, 370 and 376. In contrary to LCR mutation, where the maximum density of FG repeats increased by 50% (compared to wildtype), the control mutation does not significantly change FG density distribution. The maximum density of FG repeats is altered by less than 10% compared to the wildtype simulation. This further highlights the significant functional role of LCRs in FG Nup dynamics.

Chapter 5 : Specific amino acid sequence features keep FG Nups on the border: To collapse or not to collapse

The work presented in this Chapter is in preparation to be submitted for publication:

Peyro, M., A. Dickson, and M. R. K. Mofrad.
" Specific amino acid sequence features keep FG Nups on the border: To collapse or not to collapse ", in preparation

Abstract

Nucleocytoplasmic transport is a vital cellular process that, despite a great body of research, is yet to be fully understood. Specifically, the underlying mechanism of fast but selective transport, known as the transport paradox, is largely unknown. Among several elements that are involved in transport process, FG Nucleoporins (FG Nups) are the major role players. Based on our previous study, specific sequence features (Like Charge Regions (LCRs), i.e. extended sub-sequences that only possess positively charged amino acids) significantly affect the conformation of FG Nups inside the NPC. In this study,.we investigate at multiple levels, how the presence of LCRs affect the interactions between FG Nups and their interactions with cargo complex inside the NPC. Results of our coarse-grained molecular dynamics simulations suggest that the number of charged residues in the LCR domain directly governs the average distance between Phe residues and the intensity of interaction between them. As a result, the number of charged residues within an LCR determines the balance between hydrophobic interactions and electrostatic repulsion and governs how disordered the hydrophobic network of Phenylanalines formed by FG Nups is. Additionally, cargo complex movement is influenced by these mutations. Moreover, changing the number of charged residues in an LCR domain can interfere with fast and transient interactions between FG Nups and cargo complex; Hence, our results imply that LCR domains have a substantial role in nucleocytoplsmic transport process.

Introduction

Our previous work on the distribution of different types of amino acids in sequences of FG Nups showed that, besides previously found patterns such as FG Nups having charge-rich and charge-poor regions [98,111], more specific charge distribution patterns may be exhibited in the FG Nups. By analyzing a large dataset of FG Nups, we found an evolutionarily conserved sequence feature in majority of FG Nups: long sequences of uninterrupted positively charged residues with a low charge density located at the N terminus of FG Nups, towards the center of the NPC. We named these patterns Like Charge Regions (LCRs) [35]. Figure 5-1 shows a schematic of charge distribution and the conformation of the disordered domain of yeast FG Nup with its LCR boxed. As the next step of the study, we used a one-bead-per-amino-acid-coarse-grained model to study the biophysical effect of LCRs on the conformation of FG Nups inside the NPC. Our simulations suggested that LCRs increase the movement of FG Nups inside the NPC and lead to a more even distribution of FG motifs in the NPC as well as more cross-interactions between FG Nups [79].

In the present study, we use a coarse-grained model to further analyze the role of LCRs on nucleocytoplasmic transport by exploring how they affect the conformational ensemble of FG Nups, movement of the cargo complex, and the interaction of the cargo complex with FG Nups. For this purpose, first, the NPC nanochannel is subdivided into 17 ring cross sections that each maintain the eight-fold symmetry of the NPC but contain only one type of FG Nup. A simplified representation of cargo complex is included in the system, considering the cargo as a sphere and the transport protein as 7 hydrophobic elements (phenylalanine) on the sphere. Second, the whole-NPC model including all the FG Nups and a cargo complex initially located at the 2/3 of height of NPC is analyzed. Comparing the overall behavior of the system, specifically conformational ensemble of FG Nups, movement of cargo complex and interaction of FG Nups with cargo complex in wildtype and control simulations, we demonstrate that number of charged residues within LCR regulates the density of FG networks formed by Nups, the movement of the cargo complex, and can also regulate the interaction of the FG Nups with the cargo complex.

Results

It is important to note that throughout this chapter, the term "LCR" refers to the largest positive like charge region existing in the sequence of FG Nups. Ring cross sections of the NPC containing only one type of FG Nup were studied by dividing the axial length of the entire NPC into separate rings that each contain only one type of FG Nup, copied in an eight-fold symmetric manner. The ring complexes were modeled by making coarse-grained models of the section of the NPC scaffold, the anchoring nup copies, and a spherical cargo bound to 7 Phenylalanines that represent transport proteins. This geometry allows for examining the details of FG Nup-FG Nup interactions and FG Nup-cargo complex interactions in each cross section of the NPC. Studying the behavior of Nups and cargo in these isolated rings can provide detailed information about how each type of FG Nup may contribute to the transport process. At the second stage, the whole-NPC model is analyzed comprising of all the FG Nups and a cargo complex located at the 2/3 of the height of NPC. To distinguish and quantify the behavior of FG Nups and cargo complex in each section, the density maps of these entities were studied. Density maps are extracted in wildtype and two mutants (negative and positive control) cases to investigate the role of LCRs in the system. Two types of controls were used for this purpose. In the negative control, the charged residues in the LCR were mutated to Alanine. In the positive control, the same number of charged residues in the LCR was added to the LCR domain of the FG Nup by mutating Alanines to positively charged residues. Each of these mutations helped with analyzing a different aspect of the significance of the presence of LCRs.

The results of our single FG Nup simulations as well as studying their charge distribution suggests that yeast FG Nups can be categorized into 4 groups: 1) Nsp1, Nup100 and Nup160 which are long FG Nups and contain both charge-rich domains and LCRs 2) Nup42 and Nup145, which are medium-length FG Nups and contain both charge-rich domains and LCR 3) Nup49 and Nup57, which are short FG Nups, having a very low charge density and contain only positively charged residues; therefore, the entire sequence is an LCR 4) Nup1 and Nup60 and Nup159, which do not feature LCRs. All yeast FG Nups (except for Nup159 that features a structured domain in its N-terminus which cannot be captured by the model used in this study) and their conformation as well as a representation of initial setup of ring simulations can be seen in Figure 5-1. The ring shown in Figure 5-1 is Nsp1 containing a ring cross-section. The LCR domain of each Nup is boxed in this figure. Based on the categories mentioned earlier, we present the results of one of the FG Nups as a representative of each category.

First, the density maps of important elements of the system, i.e. Nup, cargo, FG motifs and cargo hydrophobic spots are presented for representatives of each category. Second, the force distribution analysis (FDA) of the interaction within and between the Nups in each ring are presented. Third, the density map and the FDA of FG Nups interactions with cargo complex within each ring cross-section will be presented. At last, the whole-NPC model is analyzed by investigating the FG network and movement of the cargo complex in the system.

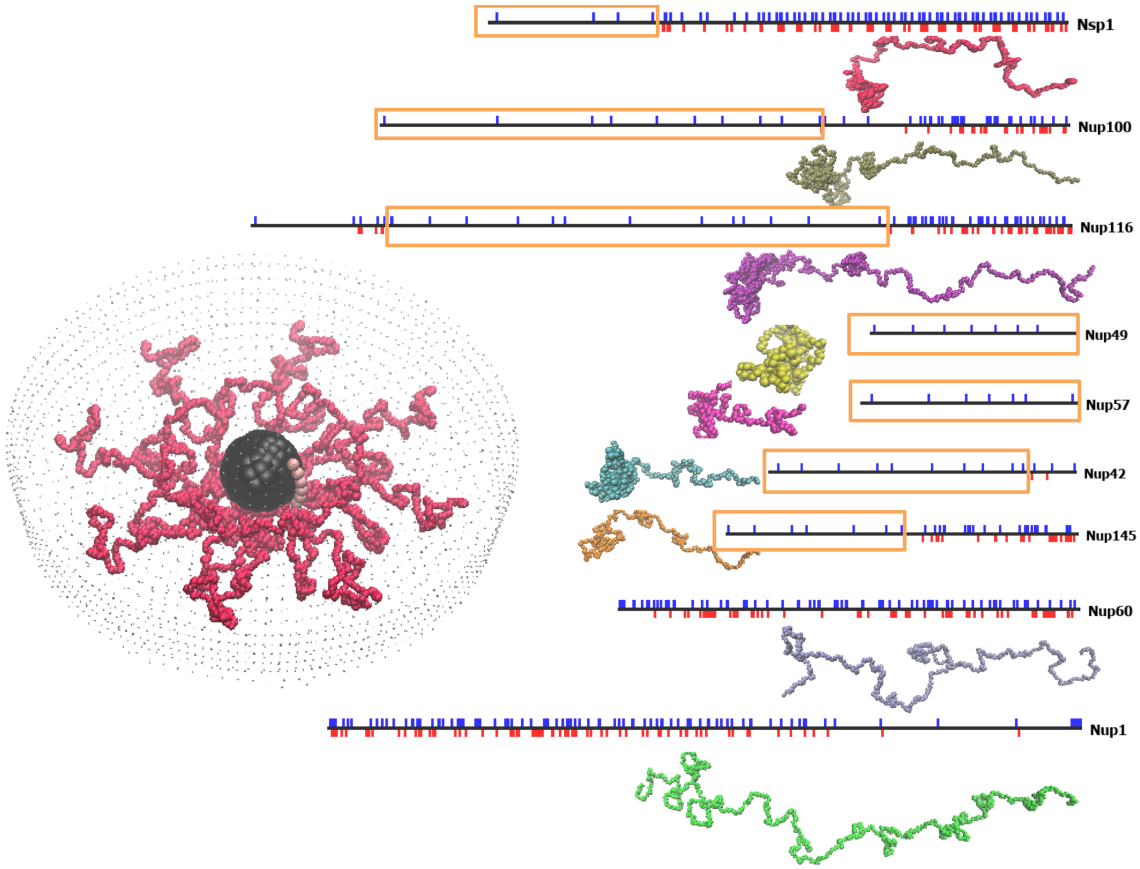

Figure 5-1: Left) Initial conformation of a ring complex Eight copies of a Nup (in this figure, Nsp1) are tethered to the scaffold of the NPC. The scaffold is represented by small non-interacting beads. The cargo complex is made up of a spherical cargo with 7 Phe residues attached to represent Kap [30]. The cargo complex is initially located at the center of the ring. Right) Charge distribution of yeast FG Nups and a snapshot of their conformation in single Nup simulations. In charge distribution schematics, the FG Nup sequence is schematically drawn as a black line with positively charged residues shown with small blue vertical lines pointing upward and negatively charged residues shown with small red vertical lines pointing downwards. The LCR of each Nup is boxed. Only the disordered domain of the Nups are drawn and modeled. First category: Nsp1, Nup100 and Nup116. Nsp1 has 617 residues in its disordered domain. Its LCR domain is 176 residues and contains 4 positive charges. This protein forms a globular head (LCR domain) and a relaxed coil tail (non-LCR, charge-rich domain). Nup100 has 800 residues in its disordered domain with an LCR domain of 511 residues that contains 10 positive charges. Nup116 has 960 residues in its disordered domain with an LCR domain of 572 residues that contains 14 positive charges. The conformations of Nup100 and Nup116 look similar to that of Nsp1. In the next section, it will be shown that FG Nups in category 1 form a high-density conformation at the center of the ring. Second category: Nup42 and Nup145. Nup42 has 382 residues in its disordered domain with an LCR domain of 317 residues that contains 7 positive charges. Nup145 has 433 residues in its disordered domain with an LCR domain of 217 residues that contains 7 positive charges. These proteins form a globular head (LCR domain) and a relaxed coil tail (non-LCR, charge-rich domain). In the next section of Results, it will be shown that FG Nups in category 2 form a donut-shaped high-density conformation near the scaffold of the ring. In the donut shaped conformations, Nup molecules cross-interact with each other. Third category: Nup49 and Nup57. Nup49 has 251 residues in its disordered domain with an LCR domain of 251 residues that contains 7 positive charges. Nup57 has 255 residues in its disordered domain with an LCR domain of 255 residues that contains 7 positive charges. These proteins do not have a charge-rich domain and their conformation only includes a globular head. Fourth category: Nup1, Nup60. Nup1 has 856 residues in its disordered domain. Nup60 has 539 residues in its disordered domain. None of these FG Nups feature an LCR domain. Due to the lack of an LCR domain, the entire length of these proteins form a relaxed coil.

Analysis of trajectory of FG Nups and cargo complex in the ring cross-section of FG Nups

The ring-containing cross-section of FG Nups was analyzed by running the coarse-grained model of the system for 1 μs. In these simulations, eight copies of each type of Nup were tethered to the scaffold and the cargo complex was initially located at the center of the ring (see Figure 5-1). The system was equilibrated for 50 ns and the simulation was run for $1\mu s$. This was long enough to capture the behavior of the system. It is important to note that, due to the disordered nature of the FG Nups, the system does not converge to a certain state, however, the conformational ensemble of the Nups show an average behavior that can be considered equivalent to "convergence" in the case of structured protein simulations. For each Nup, the negative control and positive control simulations were run with the same initial conditions and the results are presented next to the wildtype system. Density maps of key elements of these simulations are drawn for each case, i.e. the Nup, cargo, FG motifs and hydrophobic spots on cargo that represent Kap. Density maps are the best method to analyze the way each of the elements is distributed in the ring. To draw the density maps, the ring is divided into several grids, and the number of times an element is spotted in each grid is counted and a color is assigned proportionate to that number (color range is presented next to density maps).

Nsp1

Density maps of Nsp1-containing ring are presented in Figure 5-2 as a representative of the first category. Density maps show that in wildtype and mutant simulations, Nsp1 occupies most of the space inside the ring and leaves a small free space for the cargo (Figure 5-2a). This behavior is expected, since Nsp1 is a long FG Nup in the yeast NPC (Nsp1 contains 617 residues in it disordered domain) and eight copies of this molecule occupy most of the space inside the ring. As shown in Figure 5-1, Nsp1 sequence is made up of two distinct sections; a highly charged section with a relaxed coil conformation that is located toward the NPC scaffold and, a second section that is located toward the center of NPC and has a lower charge density and a more compact conformation [98]. The second section is the Nsp1 LCR domain, which has four positively charged residues. The highly aggregated LCR domains meet at the center of the NPC and form a high-density region at the center of Nsp1 ring. Comparing the behavior of the system in the case of wildtype and mutants shown in Figure 5-2 demonstrates that removing the charged residues from LCR (negative control) makes the Nups form a higher density regions in the center of the ring. As a result, the density of Nsp1 molecules at the center of the ring increases by 25% (Figure 5-2a). Also, this mutation slightly decreases the movement of cargo complex inside the ring (Figure 5-2b). On the other hand, in the positive mutation where more charged residues are added to the Nups, they form a more even distribution inside the ring, implying the fact that they have more cross-interactions rather than being trapped in compact configurations. The density of Nups at the center of the ring decreases by 25% (Figure 5-2a). Also, the cargo has slightly more freedom of movement compared to wildtype system (Figure 5-2b). FG

motif density maps show the same trend as Nups density maps (Figure 5-2c) since FG motifs

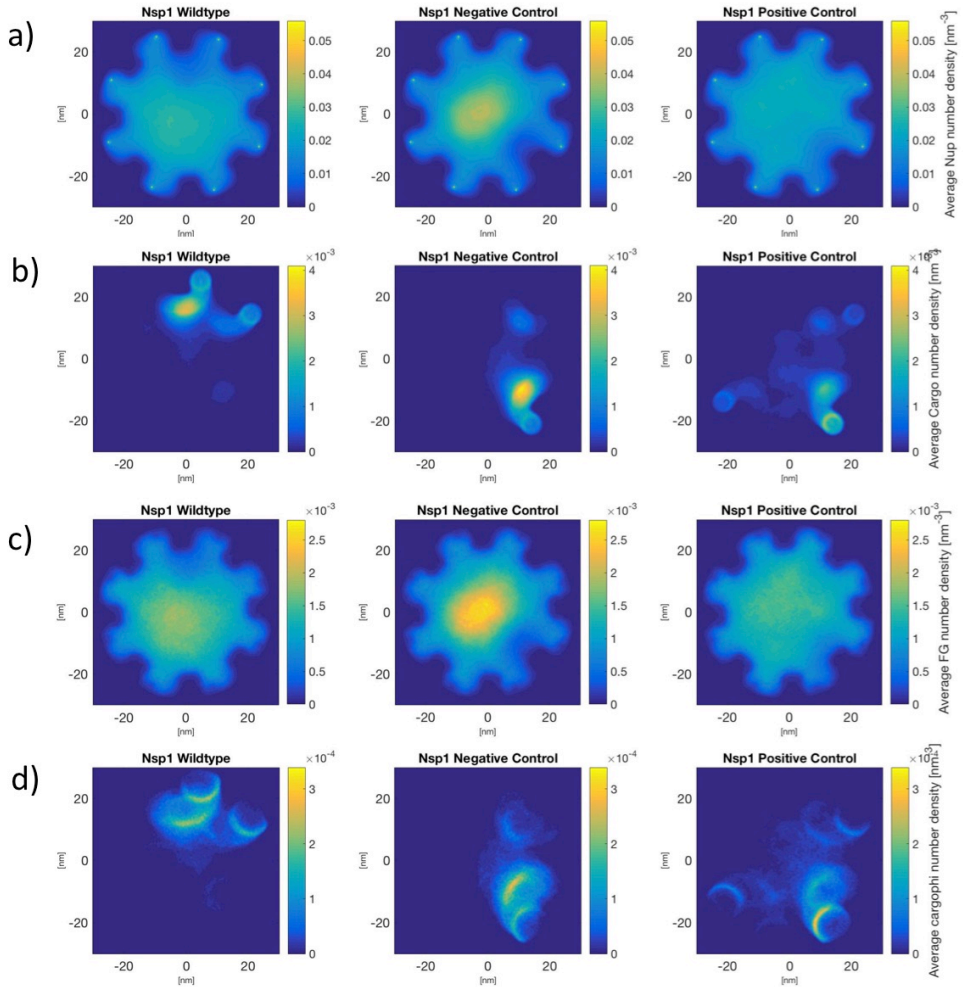

Figure 5-2: Density maps of a) Nups and b) cargo c) FG motifs d) hydrophobic spots on the cargo complex, inside ring of Nsp1. The plots on the left column show wildtype Nsp1 simulation results, plots in the middle column show results of negative control (charged residues in LCR are mutated to Alanine), and the plots on the right show the results of positive control (Charged residues are added to the system by mutating 4 Alanines to Lysine). Nsp1 has a charge rich domain (residues 177-617) and a domain with a low charge density (LCR) (residues 1 to 176). The LCR domain only contains 4 positively charged residues (see Figure 5-1). The LCR domain forms a globular conformation and the non-LCR domain forms a straight conformation. The aggregated LCR domains meet at the center of the ring and form a highly aggregated conformation at the center (see part a). In the negative control, where the 4 charged residues in LCR domain are mutated to Alanine, density of Nsp1 molecules at the center increases by 25% and the Nsp1 molecules form a denser configuration in the center. Also, movement of cargo becomes slightly more limited compared to wildtype (see part b). In the positive control, where 4 charged residues are added to the LCR domain by mutating 4 Alanines to positively charged residues, the density map of Nups becomes more uniform. The density of Nsp1 molecules decrease by 25% which implies the fact that Nsp1 molecules are less aggregated. Also, the cargo complex gains more freedom of movement. Density map of FG motifs (c) show similar pattern to Nups because Nsp1 are abundant in FG motifs along the length of their amino acid sequence. Density map of cargo hydrophobic spots (d) shows a similar pattern to that of cargo, and shows a slight difference in freedom of movement of hydrophobic spots in the ring as a result of mutation.

are abundantly present along the length of Nsp1 Nups. The density map of the cargo hydrophobic spots shows the same trend as the cargo itself, as they are attached to it

(Figure 5-2d). The density maps of FG motifs and cargo hydrophobic spots are presented, as they bind to each other to facilitate transport process.

Furthermore, in order to quantify the distribution of Nsp1 density in the ring, a new representation of the aggregation of Nsp1 inside the ring is shown in Figure 5-S1. To make these figures, the area inside the ring was divided into small grids and each grid had a density assigned to it. A relative density threshold is used to differentiate densities in different grids, i.e. grid cells that have a higher density than the threshold are shown in white. In the wildtype ring, the covered area by Nsp1 is 74%, 43%, and 0.08%, respectively for thresholds of 10%, 30%, and 60%. These numbers are 72%, 36%, and 6% in the negative control and 76%, 48% and 0.08% for positive control. When the threshold is low (10% and 30% of max density), the covered area is the largest in positive control followed by wildtype. However, as the threshold increases, the covered area in the negative control gets larger. This trend implies that in the negative control ring, Nsp1 Nups are more aggregated, while they are less aggregated in the wildtype ring and more evenly distributed in the positive control.

Nup42

Nup42 is the representative of the second category. It is one of the medium-length FG Nups in yeast NPC. This Nup has a short charge-rich domain and a long LCR domain. Over the course of the simulation, the LCR domain of Nup42 forms a globular conformation, and the short charge-rich domain forms a small relaxed coil conformation that keeps the globular LCR domain shortly distant from the scaffold (see Figure 5-1 and Figure 5-3a). Presence of the relaxed coil domains provides more freedom of movement for the LCR domains and allows cross-interaction between the Nup42 molecules (Figure 5-3a). In all three simulations (wildtype and two controls), the Nups form cross-interacting aggregates near the scaffold (Figure 5-3a) and the cargo complex explores the remaining space in the ring complex. In the negative control, the LCR domains collapse and form highly dense conformations. The cross-interaction between the Nups decrease as they get trapped into local interactions with their neighboring molecules and are not able to escape it. The maximum density of Nup molecules increases by a significant 100% value. The cargo's movement is a little less uniform compared to wildtype system (Figure 5-1b). In the positive control, the added charged residues cause the LCR domain to form a more even distribution compared to wildtype. The maximum density of the Nup molecules is decreased by ~50%, which means that the Nups are more evenly distributed. The mutated Nup42 molecules form dynamic cross-interactions in which the Nups easily unbind and bind another molecule. The FG motif density maps follow the same trend as the Nup density maps (Figure 5-3c). The density map of the hydrophobic spots on the cargo show that presence of these elements are uniform across the ring except for a high density in one spot (Figure 5-3d). In the negative control, three high density spots exist that are very close to the location of high density of FG motifs (Figure 5-3c). This implies that the hydrophobic spots interact with FG motifs at these spots. In the positive control, the density map of cargo hydrophobic spots shows a high density at the center. Considering that Nup molecules are located toward the scaffold, a high density of cargo hydrophobic spots at the center means that the cargo and its hydrophobic spots spend

most of the time traveling between Nup42 molecules rather than interacting with few of the Nups for a long time.

The following results provide another perspective of the difference between distribution of Nups in wildtype vs. mutants. In the wildtype ring, the covered area by Nup42 is 33%, 0.19%, and 0.04%, respectively for thresholds of 10%, 30%, and 60%. In the negative control, the respective values are 24%, 8.2% and 1.27%, and in the positive control, the respective values are 27%, 0.28%, and 0.06%, respectively (Figure 5-S2). These numbers confirm the hypothesis that in negative control Nup42 molecules are more compact and in positive control, they are less aggregated.

Nup145 is another Nup in this category that is also of medium length. Nup145's LCR domain is shorter than that of Nup42, but its non-LCR domain is longer. The density maps of Nup, cargo, FG motifs and cargo hydrophobic spots presented in Figure 5-S3 show that the behavior of Nup145 is similar to Nup42.

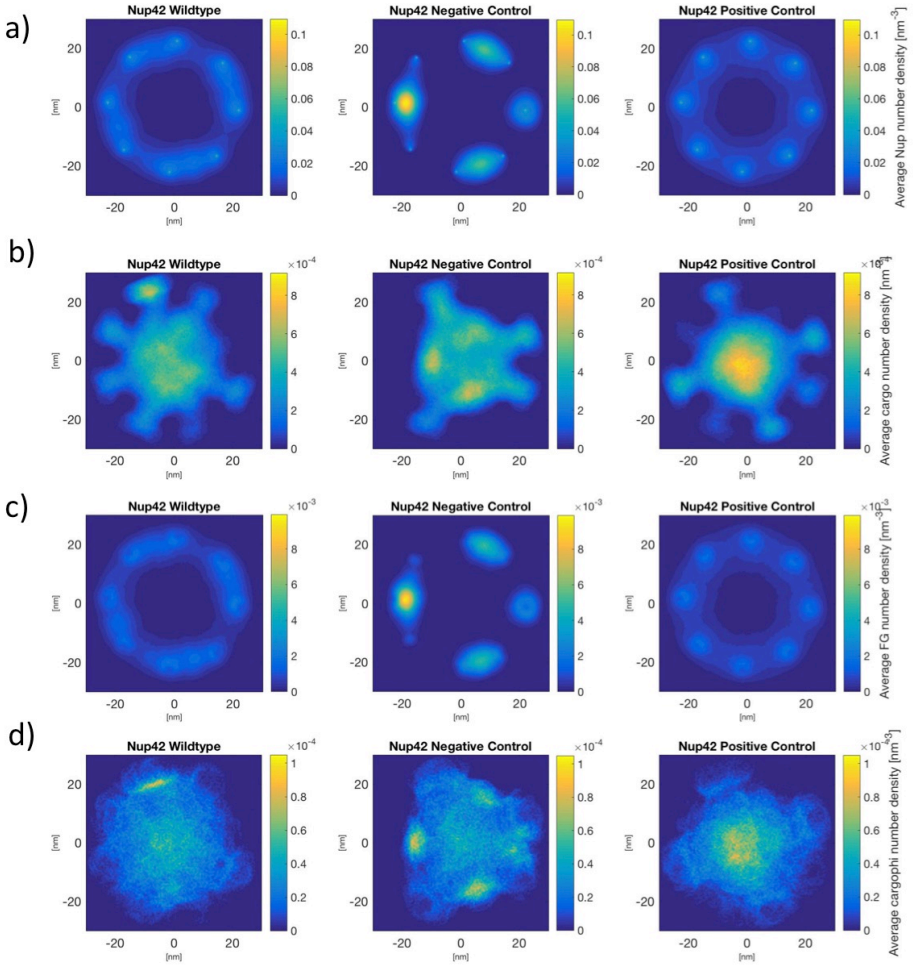

Figure 5-3: Density maps of a) Nups and b) cargo c) FG motifs d) hydrophobic spots on the cargo complex, inside ring of Nup42. The plots on the left column show wildtype Nup42 simulation results, plots in the middle column show results of the negative control, and the plots on the right show the results of the positive control. Nup42 has a very short charge rich domain (residues 318-382) and a long LCR domain (residues 1-317). The LCR domain contains 10 positively charged residues (see Figure 5-1). The LCR domain forms a globular

conformation and the non-LCR domain forms a straight conformation. The aggregated LCR domains interact with their neighboring Nup42 molecules (see part a). In the negative control, where the 10 charged residues in LCR domain are mutated to Alanine, the maximum density of Nup42 molecules increases by 100%. Also, the movement of the cargo becomes less even compared to wildtype (see part b). In the positive control, where charged residues are added to the LCR domain by mutating 10 Alanines to positively charged residues, the density map of Nups becomes more uniform. The density of Nup42 molecules decrease by 50%, which implies the fact that the mutant Nup42 molecules form a less aggregated/more evenly distributed cross-interacting network. Density map of FG motifs (c) show similar pattern to Nups because Nup42 molecules are abundant in FG motifs along the length of their amino acid sequence. The density map of cargo hydrophobic spots (d) shows that in the negative control, three high-density spots exist that are in the same location as high-density spots of FG motifs. The high-density spots decrease to one in the wildtype case and zero in case of the positive control. In the positive control, high density is observed at the center which implies that the cargo and its hydrophobic spots spend most of their time traveling between the Nups rather than interacting with some of the Nups for a long time.

Nup49

Nup49 is the shortest FG Nup in the yeast NPC. Nup49 has seven positively charged residues and no negatively charged residues in its sequence. Therefore, the entire sequence of Nup49 can be considered as LCR (see Figure 5-1). The density maps of Nup, cargo, FG motifs, and cargo hydrophobic spots are presented in Figure 5-4. Nup 49 molecules form highly compact configurations toward the scaffold of ring (Figure 5-4a). Their very collapsed conformation and short length inhibits their cross-interaction with each other. The reason behind this behavior is the low charge density of this molecule combined with high hydrophobicity of all FG Nups. In the negative control, Nup molecules form even more compact conformations compared to wildtype molecules. Mutating the charged residues to Alanine causes the density of the isolated Nup49 aggregates to increase by ~10%. In the positive control simulation, addition of charged residues to the molecule allow it to have a larger end-to-end distance and a less compact conformation. The maximum density of the positive control molecules decreases by almost 20%. Density maps of cargo show that in all three cases (Figure 5-4b), cargo has a large space to explore inside the ring, but the area is larger and less dense in the negative control, then the wildtype and then the positive control. This negatively correlates with the end-to-end distance of Nup49 molecules and their mutants. In the positive control, Nups are less aggregated, so cargo has less space to cover inside the ring. In the wildtype, Nups are more compact, so the cargo has a relatively larger area to cover. In negative control where Nups are most collapsed, the cargo has the largest space to explore, but spends long times close to some of the Nups which are shown as high density points in the density maps. Density maps of FG motifs (Figure 5-4c) show the same pattern as Nups and the density map of hydrophobic spots (Figure 5-4d) show the same trend as the cargo itself.

The difference between aggregation of Nups in wildtype and mutant cases is also represented in Figure 5-S4. In the wildtype ring, the covered area by Nup49 is 10%, 4%, and 0.8%, respectively for thresholds of 10%, 30%, and 60%. However, the numbers are 9%, 4.6%, and 1.8% for negative control, and 11%, 1.8% and 0.04% for positive control. These numbers confirm the hypothesis that removing the charged residues (negative control) leads to a more compact conformation and adding charged residues (positive control) causes the Nups to make a more relaxed coil conformation.

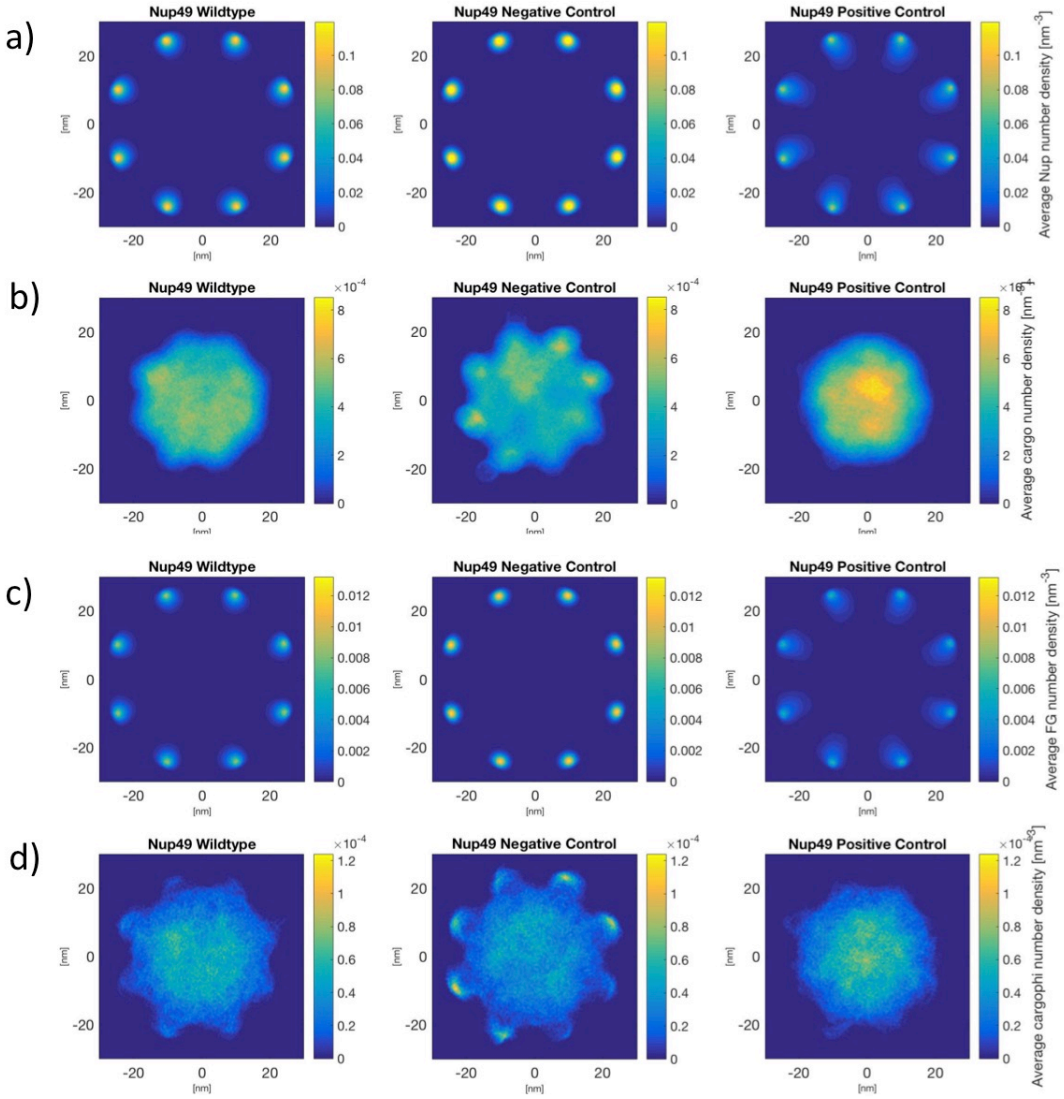

Figure 5-4: Density maps of a) Nups and b) cargo c) FG motifs d) hydrophobic spots on the cargo complex, inside ring of Nup49. The plots on the left column show the wildtype Nup49 simulation results, plots in the middle column show results of the negative control, and the plots on the right show the results of the positive control. The length of the disordered domain of Nup49 is 251 amino acids. This protein contains seven positively charged residue and zero negatively charged residues (see Figure 5-1). Therefore, the entire molecule is an LCR, and due to very low charge density, the molecule collapses into highly aggregated conformation toward the scaffold of the ring. The end-to-end distance of the molecule is not long enough for them to be able to cross-interact with their neighboring molecules (part a). In the negative control, where the seven charged residues in LCR domain are mutated to Alanine, the maximum density of Nup49 molecules increases by 10%. The cargo complex gains more area to move, but the density map shows some high-density spots, which are in the same location as Nups (see part b). This implies that cargo complex and Nup49 molecules are interacting for a long time. In the positive control, where charged residues are added to the LCR domain by mutating seven Alanines to positively charged residues, the Nups become less aggregated by 20% due to presence of more charged residues. Density map of cargo shows high density at the center (see part b), which implies that the cargo complex is spending most of it time traveling between the Nup49 molecules rather than interacting with some of them for a long time. Density map of FG motifs (c) show similar pattern to Nups because Nup49 molecules are abundant in FG motifs along the length of their amino acid sequence. The density map of the cargo's hydrophobic spots (d) shows a similar pattern to that of cargo.

Nup57 simulations show that the ring of Nup57 including the cargo shows similar behavior to the Nup49 ring (Figure 5-S5). Nup57 is the second shortest Nup in yeast NPC with no negatively charged residues and a few positively charged residues in its sequence. In terms of length, charged residue content, density and distribution, Nup49 and Nup57 are very similar and show similar behavior in these simulations. In Nup57 ring, upon mutation, Nups get more aggregated toward the scaffold of the NPC and although the cargo complex has a larger space to explore (due to Nups occupying less space in the ring), it spends a lot of time in certain spots next to a few Nups (Figure 5-S5a and Figure 5-S5b). However, in the wildtype system, the density maps show a more even movement of cargo and hydrophobic spots on the cargo throughout the course of the simulation (Figure 5-S5b and Figure 5-S5d).

Nup1

Nup1 is in the fourth category with Nup60, and Nup159. These Nups do not contain an LCR and are analyzed here as a natural control for the other Nups that do contain LCRs. Notably, these Nups do not feature an LCR, only the wildtype system is analyzed. All of these Nups are long and have long charge rich domains. Nup1 density map (Figure 5-5a) shows that although Nup1 is long and covers the entire area inside the ring, it does not form high-density conformations at the center. The cargo spends some time at the center and some time around the edges (Figure 5-5b). The reason for this is that FG motifs form high-density areas near the tethering point of the Nups (Figure 5-5c) and hydrophobic spots on the cargo complex are also mainly present at the center and around the edges where a high density of FG motifs exists (Figure 5-5d).

Nup60 behaves similarly to Nup1. The Nups do cover the entire area inside the ring, but do not form high-density conformations at the center (Figure 5-S6a). The cargo is mainly present both at the center and around the edges at certain spots (Figure 5-S6b). Its presence around the edges is due to presence of the single FG motif in the sequence of Nup60 that is located close to the scaffold (Figure 5-S6c). Cargo hydrophobic spots density map (Figure 5-S6d) is also similar to the density map of cargo.

The interesting fact is that the FG Nups in the fourth category (Nup1, Nup60 and Nu159) are located at the extremities of the NPC. These Nups are long, but do not feature an LCR. Hence, they do not form high densities and the two ends of the NPC on top and bottom are not blocked by a dense network of hydrophobic residues. However, the cargo will face a dense network formed by Nup42, Nsp1, Nup116 and Nup100 on top and Nsp1 at the bottom.

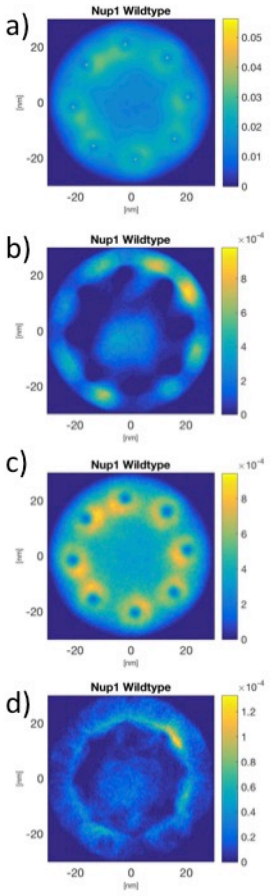

Figure 5-5: Density maps of a) Nups and b) cargo c) FG motifs d) hydrophobic spots on the cargo complex, inside ring of Nup1. Since this Nup does not feature an LCR, only the wildtype case is analyzed. This Nup is analyzed as a natural control for other Nups that do contain an LCR. The length of the disordered domain of Nup1 is 856 amino acids and its entire length is a charge-rich domain (see Figure 5-1). Therefore, this Nup does not feature an LCR. Although this Nup is long and the eight copies of Nup cover the entire area inside the ring, the molecules do not form a high-density conformation at the center. The cargo complex covers some area at the center and some areas near the scaffold (see part b). The presence of cargo near the scaffold is due to a high density of FG motifs (see part c) around the tethering points of Nup1 molecules that is caused by a low density of charges in a small segment of Nup1 sequence near its C terminus (see Figure 5-1). The density map of cargo hydrophobic spots (d) shows a similar pattern to that of cargo.

Force distribution analysis of the interaction between FG Nups

Time-resolved force distribution analysis was performed on the Nups inside the ring with the aim to gain a deeper understanding of effect of presence of LCRs in FG Nups and how mutating LCRs can influence the force distribution within the Nups.

Nsp1

The results of FDA on Nsp1 molecules within Nsp1 containing ring cross-sections are presented in Figure 5-6. Heatmaps show the punctual stress of each residue along the length of the simulation (Figure 5-6a). Time and residue numbers are represented in X and Y axis respectively. Heatmaps of punctual stress are drawn for the wildtype and the negative and positive controls and a schematic of charge distribution in each case is drawn next to it. In the charge distribution schematic, the little blue lines represent positively charged residues and the little red lines represent negatively charged residues. The time-averaged value of the punctual stress for each residue is drawn in Figure 5-6b. The wildtype, negative control, and positive control graphs are drawn in the same plot. As both the heatmaps and the time-averaged plot show, stresses are higher in the LCR domain (residue 1-176). This means that the residues in LCR domain are generally interacting more intensely with each other. LCR domains are the collapsed domain in Nsp1 molecule (see Figure 5-1). Therefore, low charge density in LCR domain allows for forming globular domains in which residues interact strongly with each other. On the other hand, the charge-rich domains form relaxed coil conformations in which residues have less interaction within and between the Nups.

As the time-averaged plot depicts, wildtype and the mutants have the same average punctual stress in the non-LCR domain (residues 176-617) but the values differ in the LCR domain. The average stress values are ordered from higher to lower in the negative control, wildtype, and positive control. The heatmaps show a similar trend, with the negative control having the highest stress, the positive control the least, and the wildtype in the middle. This observation is in line with the behavior observed in Figure 5-2a. Negative control molecules have the most collapsed conformation and experience the highest stress, while the positive control molecules have the least aggregated conformation and the Nups experience the least amount of stress.

An interesting fact about the force distribution in the wildtype and mutants is that in all the three cases, same residues and groups of residues experience a high stress, but the magnitude of stress is different in the three systems (Figure 5-7). The high stress residue and residue groups are mostly FG motifs and Phenylalanines and a few Leucines and one case of a pair of Glutamines located next to each other. This is the case for all the eight copies of the Nsp1 molecules in the ring and the wildtype and the two mutants. This is a significant observation in the mechanism of interaction of Nsp1 molecules with each other, and implies that changing the number of charged residues affects the intensity of interaction of residues in LCR domain, but does not change the network of residues interacting with each other.

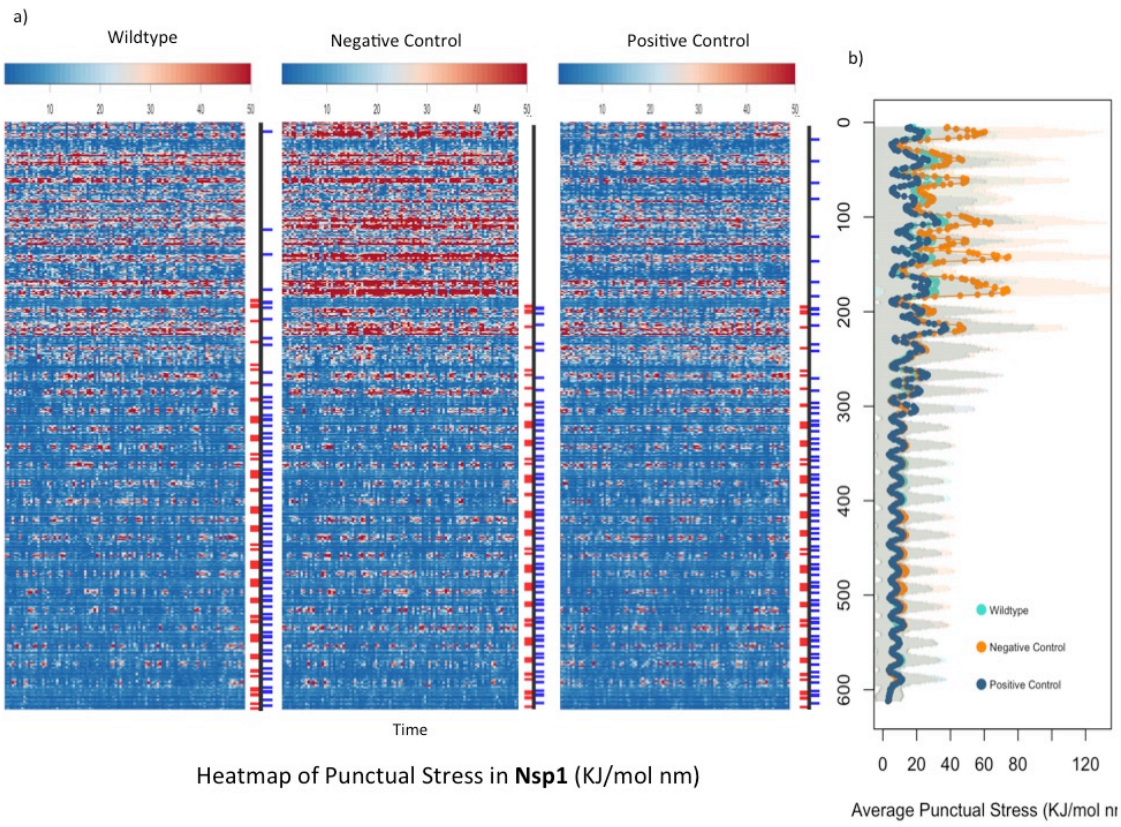

Figure 5-6: Force Distribution Analysis of Nsp1 molecules. a) Heatmaps of punctual stress in Nsp1 molecules caused by interaction of Nsp1 molecules with themselves and with each other. Y-axis represents residue number (1 to 617 from top to bottom) and X-axis represents time. A schematic of charge distribution of each case is drawn next to their heatmap (small blue lines are positively charged and small red lines are negatively charged residues). The residues in the non-LCR domain (177-617) experience on average less stress while the residues in the LCR (1-176) experience on average higher stress. In the wildtype and the two controls, the same residues experience a high stress, but the magnitude of stress is higher to lower in negative control, wildtype and positive control respectively. b) Time-averaged punctual stress for the residues of wildtype, negative control and positive control compared to each other. All the three cases show similar stress values in the non-LCR domain, while the values differ in the LCR domain. The negative control has the highest stress and the positive control has the lowest stress and the wildtype is in the middle of the two. The difference in stress of the three cases is presumably due to the difference in their conformational ensemble. The negative control has the most aggregated/collapsed conformation and the residues are experiencing the highest stress, while the positive control has the least aggregated/collapsed conformation and the residues are experiencing the least stress, and the wildtype is between the two. As can be seen both in the heatmaps and the time-averaged graphs, the spikes of stress happen at the same residues for the wildtype and the mutants. This implies that mutating the number of charged residues in the LCR domain does not change the network of interacting residues; it only affects the intensity of the interaction.

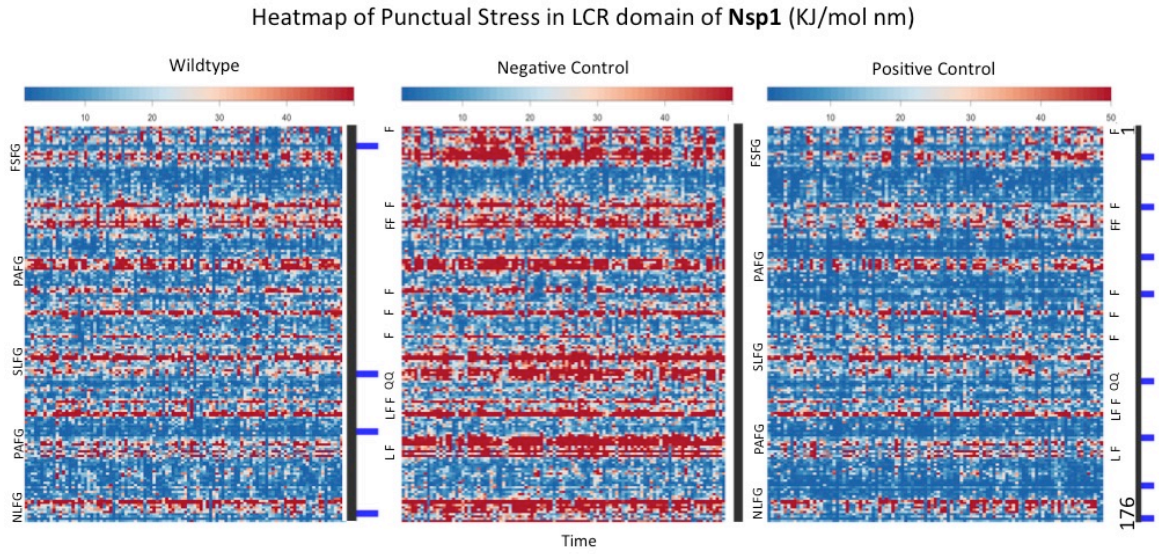

Figure 5-7: Heatmap of punctual stress for the LCR domain of Nsp1 (residues 1-176). The charge distribution of each case is shown next to them (blue lines are positively charged residues). The residues/group of residues experiencing high stresses are shown next to the heatmaps. FG motifs, Phenylalanines, and only few Luecines and two Glutamines are experiencing a high stress. Therefore, interacting residues are mostly Phe that form a hydrophobic network. In negative control, there are no charged residues to disturb the highly dense hydrophobic network. In wildtype, presence of 4 charged residues in LCR domain partially interfered with the hydrophobic interactions and decreases the intensity of interaction. This effect is enhanced in positive control, and while still the same residues are interacting, their stress is decreased as there are more charges to disturb the strong hydrophobic interactions.

Nup42

The results of FDA on Nup42 molecules within Nup42 containing ring cross-sections are presented in Figure 5-8. The charge distribution of each case is schematically shown next to it. Nup42 has a short non-LCR domain (residues 318 to 382) and a long LCR domain (residues 1-317). Similar to the trend seen in Nsp1 (Figure 5-6), LCR domains show a larger stress than non-LCR domain in Nup42 molecules and their two mutants (Figure 5-8). Also, similar to Nsp1, negative control shows the highest stress and positive control shows the lowest stress. This is in line with the conformation of Nup42 molecules shown in Figure 5-3a. Negative control has the most collapsed conformation and allows for stronger interactions to form within and between molecules. Positive control has the least compact conformation and the least stress.

The interesting point is that in case of Nup42, wildtype and negative control case show a closer behavior while positive control is significantly different. The residues experiencing a high stress are similar in wildtype and negative control, with the magnitude of stress being higher in negative control. However, some of those residues experience slightly higher than minimum stress values. The reason might be due to the fact that with adding the charged residues to the LCR domain, the charge distribution in positive control is very similar to the non-LCR domain. Therefore, positive control depicts a significantly different behavior compared to wildtype and negative control. However, the interacting residues (with high stress) are still the same in all the three cases and are mainly

phenylalanines with few other hydrophobic residues. Therefore, similar to Nsp1, the network of interacting residues has not changed, but the magnitude of stress has.

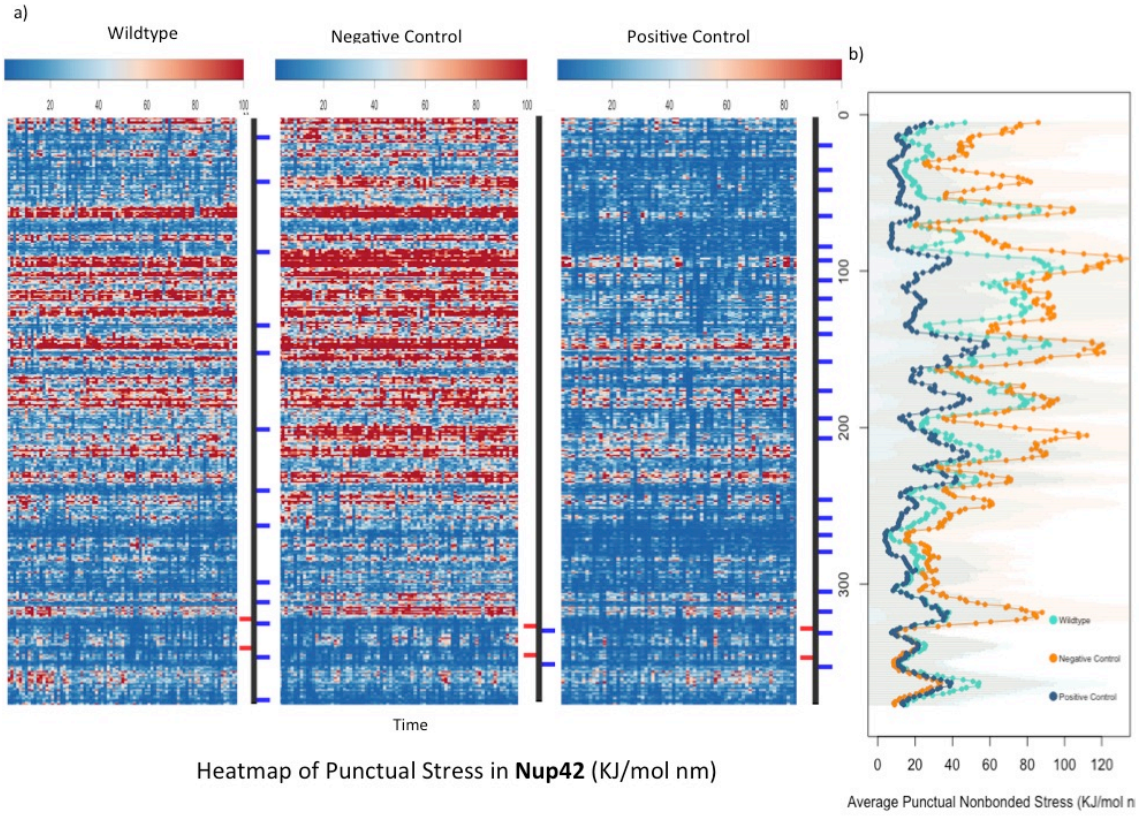

Figure 5-8: Force Distribution Analysis of Nup42 molecules. a) Heatmaps of punctual stress in Nup42 molecules caused by interaction of Nup42 molecules with themselves and with each other. Y-axis represents residue number (1 to 382 from top to bottom) and X-axis represents time. A schematic of charge distribution of each case is drawn next to their heatmap (small blue lines are positively charged and small red lines are negatively charged). The residues in the non-LCR domain (318-382) experience on average less stress while the residues in the LCR (1-317) experience on average higher stress. In wildtype and the two mutants, same residues experience a high stress, but the magnitude of stress is higher in negative control, wildtype and then positive control. b) Time-averaged punctual stress for the residues of wildtype, negative control and positive control compared to each other. All the three cases show similar stress values in the non-LCR domain, while the values differ in LCR. Negative control has the highest stress and positive control has the lowest stress and wildtype is in the middle of the two. The difference in stress of the three cases is presumably due to difference in their conformational ensemble. Negative control has the most aggregated/collapsed conformation and the residues are experiencing the highest stress, while positive control has the least aggregated/collapsed conformation and the residues are experiencing the least stress, and wildtype is between the two. As can be seen both in the heatmaps and the time-averaged graphs, the spikes of stress happen at the same residues for wildtype and mutants. This implies that mutating the number of charged residues in the LCR domain does not change the network of interacting residues, it only affects the intensity of interaction.

Nup49

The results of FDA on Nup49 molecules within Nup49 containing ring cross-sections are presented in Figure 5-9. Nup49 does not have a charge rich domain, so its entire length is an LCR (see Figure 5-1). The trends are similar to those seen in Nsp1 and Nup42, with negative control having the highest stress due to higher aggregation and positive control having the lowest stress due to lower aggregation. Similar to Nsp1 and Nup42, certain

residues/group of residues (FG motifs, Phenylalanines, and few other hydrophobic reisdues) experience a high stress with the stresses being the highest in negative control and lowest in positive control. Therefore, in the case of Nup49 as well, number of charged residues has not changed the network of interacting residues, but has affected the magnitude of stress.

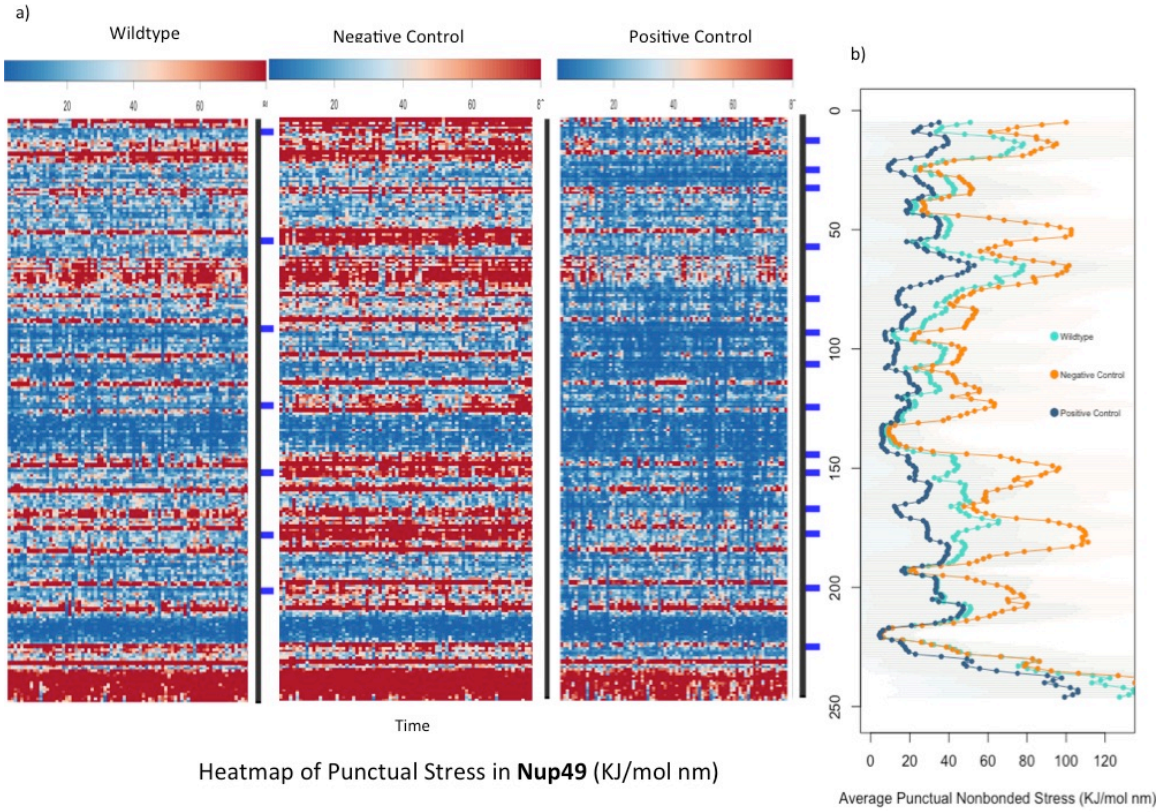

Figure 5-9: Force Distribution Analysis of Nup49 molecules. a) Heatmaps of punctual stress in Nup49 molecules caused by interaction of Nup49 molecules with themselves and with each other. Y-axis represents residue number (1 to 251 from top to bottom) and X-axis represents time. A schematic of charge distribution of each case is drawn next to their heatmap (small blue lines are positively charged and small red lines are negatively charged). The entire length of Nup49 is an LCR. In wildtype and the two mutants, same residues experience a high stress, but the magnitude of stress is higher in negative control, wildtype and then positive control. b) Time-averaged punctual stress for the residues of wildtype, negative control and positive control compared to each other. Negative control has the highest stress and positive control has the lowest stress and wildtype is in the middle of the two. The difference in stress of the three cases is presumably due to difference in their conformational ensemble. Negative control has the most aggregated/collapsed conformation and the residues are experiencing the highest stress, while positive control has the least aggregated/collapsed conformation and the residues are experiencing the least stress, and wildtype is between the two. As can be seen both in the heatmaps and the time-averaged graphs, the spikes of stress happen at the same residues for wildtype and mutants. This implies that mutating the number of charged residues in the LCR domain does not change the network of interacting residues, it only affects the intensity of interaction.

Considering that in all the Nups studied here, Phe residues were the main interacting residues, the average distance of them were compared in wildtype and mutant in each ring cross sections. The results are presented in Table 1 and show that the average distance of Phe residues decrease in negative control and increase in positive control. This implies that the number of charged residues in the LCR regulate the spacing between the Phe residues and influence the force distribution of Nups in this way.

Table 1: Average distance between Phe residues in LCR domain of Nsp1 and Nup42 molecules

	Nsp1 (LCR domain)	Nup42 (LCR domain)
Wildtype	4.08 +- 0.12 nm	3.95 +- 0.24
Negative control	3.25 +- 0.11 nm	3.1 +- 0.31
Positive control	4.55 +- 0.18 nm	5.62 +- 0.29

How does mutating the LCR domain affect the interaction between FG Nups and the cargo complex?

The results presented so far were on the interaction of Nups with each other. In this section, the effect of mutation of LCR on the interaction of FG Nups with cargo complex is discussed. Using FDA, punctual stress of Nsp1 molecules caused by interaction of Nsp1 molecules with cargo complex is calculated. It is worth noting that these stresses are much smaller than the stresses caused by interaction of Nups with each other. The reason is that the number of interactions happening between the residues of the Nups is much larger than the number of interactions between the Nups and cargo complex. The boxplot of the stress values for wildtype and the two mutants are shown in Figure 5-10a. The boxplot depicts that the stress values in the three cases are similar, meaning that mutation of LCR domains has not significantly affected the interaction of Nsp1 molecules with cargo complex. Combined density map of FG motifs and hydrophobic spots on the cargo are shown in Figure 5-10b with blue and white dots respectively. The density threshold was adjusted so that only densities higher than a certain threshold can be observed. Co-localization of white and blue dots can imply long interactions between FG motifs and cargo hydrophobic spots. Density maps in Figure 5-10b are not significantly different as well.

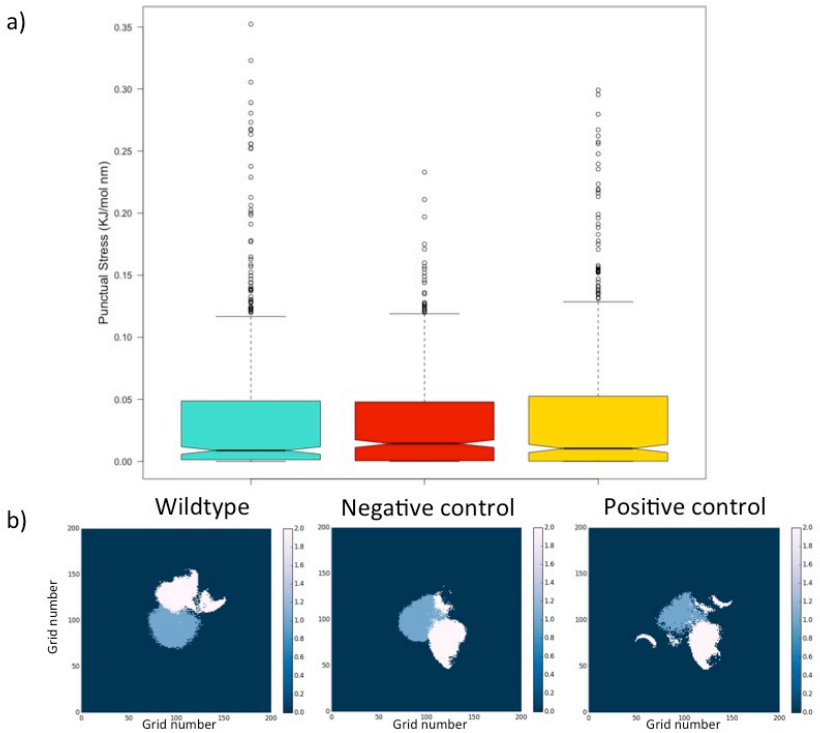

Figure 5-10: Interaction of Nsp1 molecules with cargo complex. a) force distribution analysis was performed to find the punctual stress experienced by Nsp1 molecules caused by interaction of Nsp1 with cargo complex. Boxplot of time-averaged values of punctual stress in all 617 residues of all copies of Nsp1 is drawn for wildtype and two controls. The boxplots of the three cases are not significantly different. b) density maps of FG motifs (shown with blue dots) and cargo hydrophobic spots (shown with white dots) are drawn together. Co-localization of white and blue dots can imply interaction between FG motifs and cargo hydrophobic spots. The results presented in this figure show that interaction of Nsp1 molecules and hydrophobic spots on the cargo are not significantly affected by mutating the LCR domains.

The same analysis is done for interaction of Nup42 molecules with cargo complex (Figure 5-11). The results show that the interaction of cargo complex with wildtype and negative control Nup42 molecules are almost similar, but the positive control case is significantly different. The interactions are much less strong. The density maps (Figure 5-11b) are in line with the boxplots. Mutating the LCR domains by adding charged residues significantly affects the interaction of Nup42 molecules with each other and interaction of Nup42 molecules with cargo complex. The interactions become more transient and less strong.

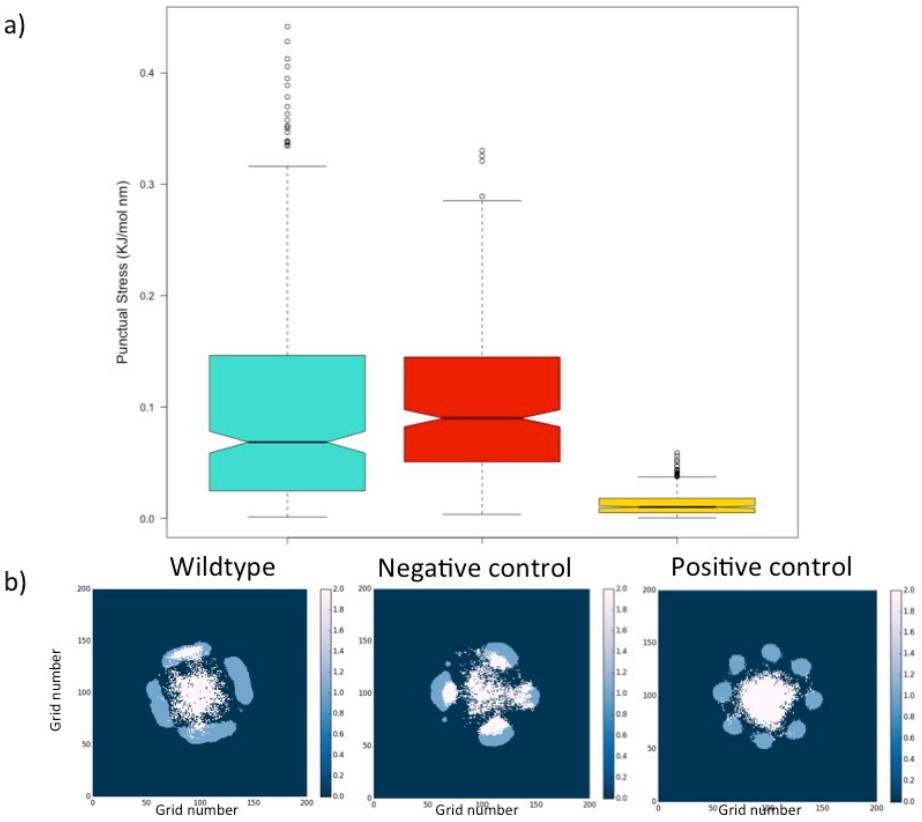

Figure 5-11: Interaction of Nup42 molecules with cargo complex. a) force distribution analysis was performed to find the punctual stress experienced by Nup42 molecules caused by interaction of Nup42 with cargo complex. Boxplot of time-averaged values of punctual stress in all 382 residues of all copies of Nup42 is drawn for wildtype and two controls. The boxplots of wildtype and negative control are similar, but the boxplot of positive control is significantly different. The stresses are significantly lower in positive control. b) density maps of FG motifs (shown with blue dots) and cargo hydrophobic spots (shown with white dots) are drawn together. Co-localization of white and blue dots can imply interaction between FG motifs and cargo hydrophobic spots. Blue and white dots depict co-localization in wildtype and negative control, but do not show any co-localization in positive control. This implies that interactions between Nups and cargo are much weaker and more transient in positive control compared to the other two cases. Results presented in this figure suggest that mutating 10 Alanines in the LCR domain to positively charged residues has remarkably decreases the interaction of mutant Nup42 with cargo complex, while removing the charges (negative control) has not had such an impact.

Nup49 ring complex was also similarly analyzed. Results in Figure 5-12 depict that interaction of Nup49 molecules with cargo complex is largely affected by mutation of LCR domain. The average stress in wildtype is 3 times larger than positive control and the average stress in negative control is 10 times larger than positive control. The combined density maps show how this occurs (Figure 5-12b). The hydrophobic spots on the cargo are co-localizing with FG motifs to a large extent in negative control, but showing no cross-interaction in positive control. The results imply that mutation of LCR domains in Nup49 affects the conformation of molecules (Figure 5-4a) but has a major impact on the interaction of cargo complex and Nup49 molecules.

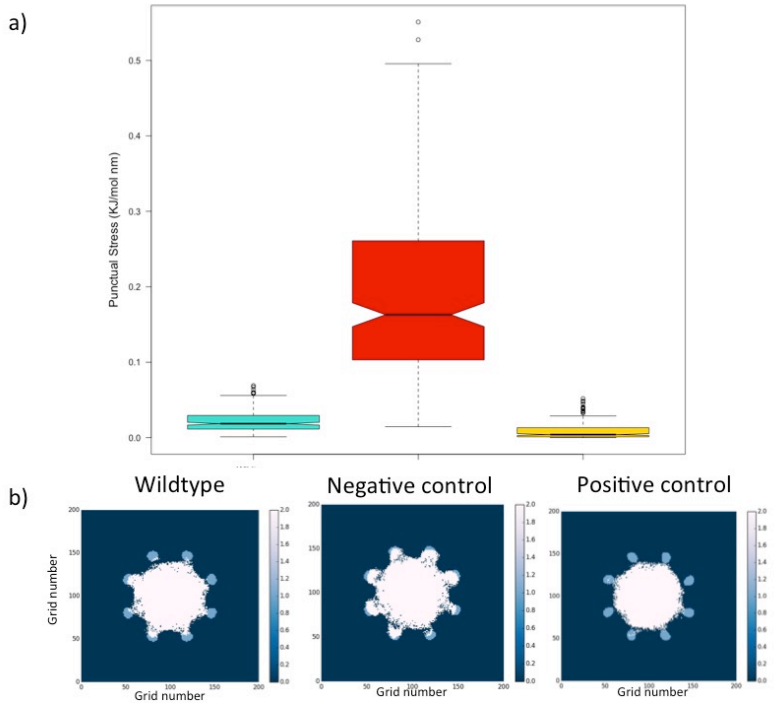

Figure 5-12: Interaction of Nup49 molecules with cargo complex. a) force distribution analysis was performed to find the punctual stress experienced by Nup49 molecules caused by interaction of Nup49 with cargo complex. Boxplot of time-averaged values of punctual stress in all 251 residues of all copies of Nup49 is drawn for wildtype and two controls. The boxplots show that stresses are lowest in positive control, are slightly higher in wildtype, and extremely higher in negative control b) density maps of FG motifs (shown with blue dots) and cargo hydrophobic spots (shown with white dots) are drawn together. Co-localization of white and blue dots can imply interaction between FG motifs and cargo hydrophobic spots. Blue and white dots depict no co-localization in positive control, some co-localization in wildtype, and lots of co-localization in negative control. These are in line with boxplot results. Therefore, interactions between Nups and cargo are the most weak and transient in positive control, and more strong in wildtype. However, in negative control, the stresses are an order of magnitude stronger. Results presented in this figure suggest that mutating 10 Alanines in the LCR domain to positively charged residues decreases the interaction of mutant Nup42 with cargo complex, while removing the charges (negative control) extremely strengthens those interactions, and interferes with the transient quality of interactions known for FG Nups.

A common observation between all the cases of all the rings analyzed is that the conformational ensemble of Nups do not significantly change as a results of presence of cargo complex in the system and its interaction with FG Nups. This can be concluded from comparison of our previous study [79] with density maps of Nups in Figure 5-2, Figure 5-3, and Figure 5-4. The density maps look very similar in presence and absence of cargo complex. This behavior was expected based on comparison of the stress values caused by interaction of Nups with each other and the values caused by interaction of FG Nups with cargo complex. The Nup-Nup are significantly dominant.

It is worth noting that in all of the simulations in this study and our previous study [79], the general behavior of FG Nups is in line with the behavior observed in [28] via high speed AFM. Although the timescale of the observations are different, the average behavior is similar. In line with their observation, our simulations depict a highly dynamic and flexible behavior by FG Nups, which go through transient entanglements yet do not form a cohesive network.

Whole-NPC Results

Distribution of FG motifs in the NPC

A model of whole NPC was created with all the comprising FG Nups, the scaffold of NPC, and a 10nm spherical cargo accompanied by 7 hydrophobic spots that represent Kap. Initially, the cargo complex was located at the cytoplasmic side of the NPC but it was not able to pass through the pore. The cargo complex would penetrate the hydrophobic FG network to a small extent but was repelled back and could not reach the nucleoplasmic side. To solve this problem, we initially positioned the cargo at 2/3 height of the NPC and ran the simulation after minimizing the energy of the system. With this initial condition, the cargo was able to pass through the pore. We repeated the simulation for negative and positive control by replacing the wildtype Nups with negative and positive control Nups respectively and in both of these cases, the cargo complex was able to reach the nucleoplasmic side. FG density maps of wildtype and two control systems are presented in Figure 5-13 a and b for two density thresholds of 1.5% and 3% respectively.

Based on observations in Figure 5-13, whole NPC shows a similar behavior to ring cross-sections as a result of mutating LCR domains. The FG domains become more concentrated in negative control, and less concentrated/more evenly distributed in positive control. The maximum density of FG motifs in the model increases by 26% in negative control, and decreases by 15% in positive control. The interesting observation is that in both wildtype and negative control, a high-density cluster of FG motifs is formed at the top-central region of the NPC, but the 15% decrease in density of FG motifs significantly affected the high-density FG network formed toward the center of the pore and scattered them in the rest of the space inside the NPC.

Results of Figure 5-13b can be compared to Figure 8 in [79] where whole NPC is modeled in absence of cargo complex. In the absence of cargo complex, the wildtype system showed a donut-shaped cluster of FG motifs toward the upper side of NPC, while negative control showed a high-density network at the center of the NPC that blocked the pore. The wildtype system including the cargo complex shows a similar behavior (Figure 5-13b-left figure). However, the negative control system including cargo complex (Figure 5-13b-middle figure) behaves differently from the system in the absence of cargo. The system including the cargo does form high-density FG clusters but they have an opening at the center. The reason is presence of cargo complex. As will be shown in Figure 5-14, in the wildtype and negative control system, the cargo spends most of its time at the center of the NPC. Presence of several FG Nups filling up the space inside the NPC makes the cargo to stay toward the central area, and this prevents FG motifs to form a cluster that blocks the center of the NPC. This observation implies that while FG motif network is more clustered in negative control, presence of even one cargo complex would prevent FG Nups from collapsing into high-density conformations at the center. The difference in positive control seems to be more fundamental though, as FG clusters become very scattered and significantly decrease in density.

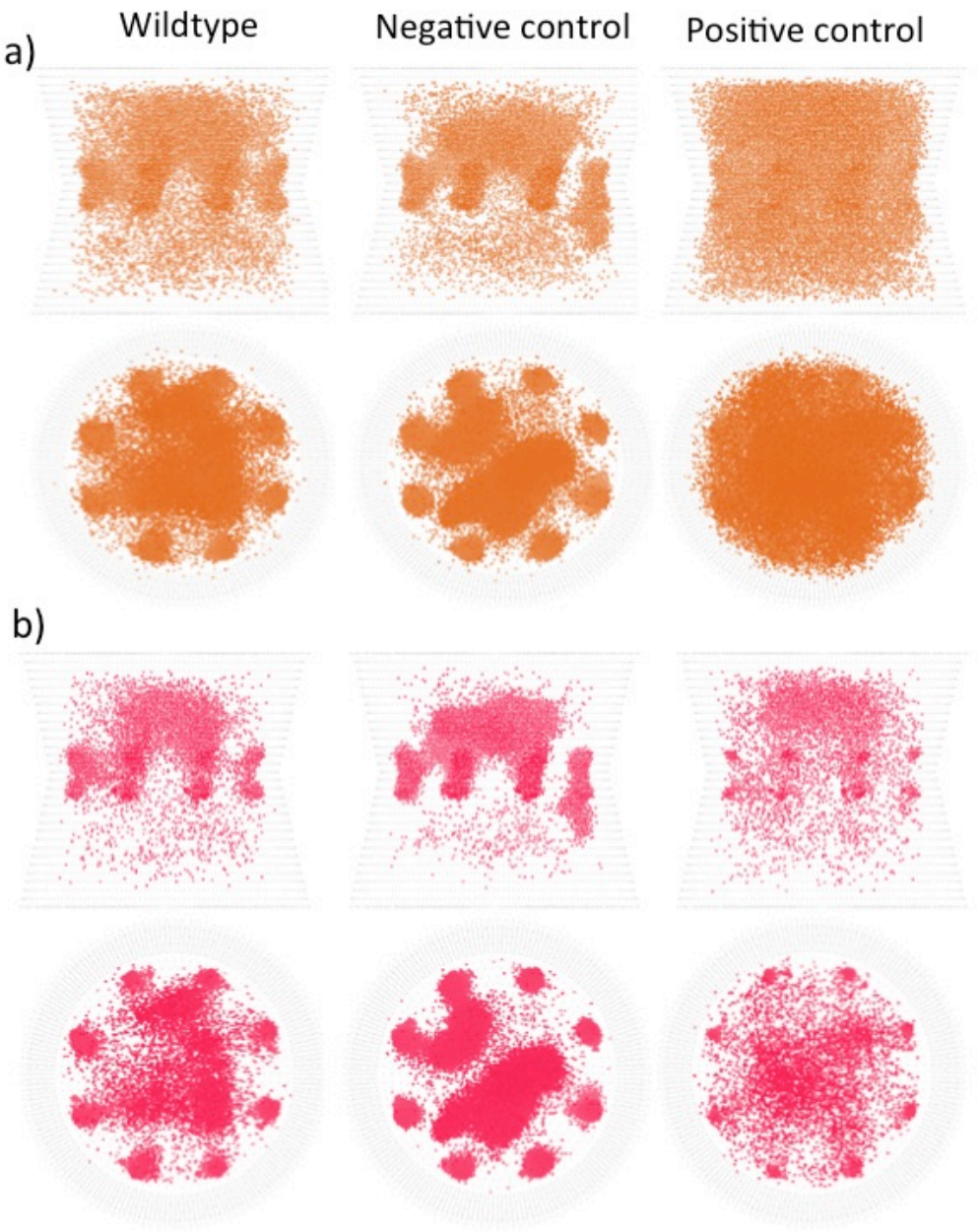

Figure 5-13: FG density maps of the whole-NPC model for wildtype and two controls for density threshold of 1.5% (a) and 3% (b). The side view of the NPC is presented at the top row and top view is presented at the bottom row. Similar to the behavior observed in ring cross-sections, negative control makes the FG motifs more clustered, and positive control makes them more evenly distributed within the space inside the NPC. Maximum density of FG motifs increases by 26% in negative control and decreases by 15% in positive control. Both wildtype system and negative control depict high-density FG clusters in the system, but the 15% decrease in density affect the shape of clusters significantly and scatters them remarkably. Comparing the results with the system excluding the cargo complex [79] shows that the FG network of wildtype system behaves almost similarly in absence or presence of cargo, but this behavior is different in case of negative control. In absence of cargo complex, FG motifs in negative control form a very high-density cluster at the center. In presence of cargo, FG motifs cluster density increases, but presence of cargo at the center of the system prevents FG networks from forming a very dense cluster at the center. Rather, they form high-density clusters with an opening at the center.

Movement of cargo complex inside the NPC

Movement of cargo complex in the whole-NPC model was analyzed and compared between wildtype and the two controls. The density map of center of mass of the cargo in wildtype and the two control cases was drawn in Figure 5-14. Green dots represent center of mass of cargo in wildtype system, red dots in negative control and blue in positive control. It is clearly visible in Figure 5-14 that the cargo has more freedom of movement in positive control (shown with blue) and less freedom of movement (shown with red). These results are in line with our observations in the ring cross-section of NPC.

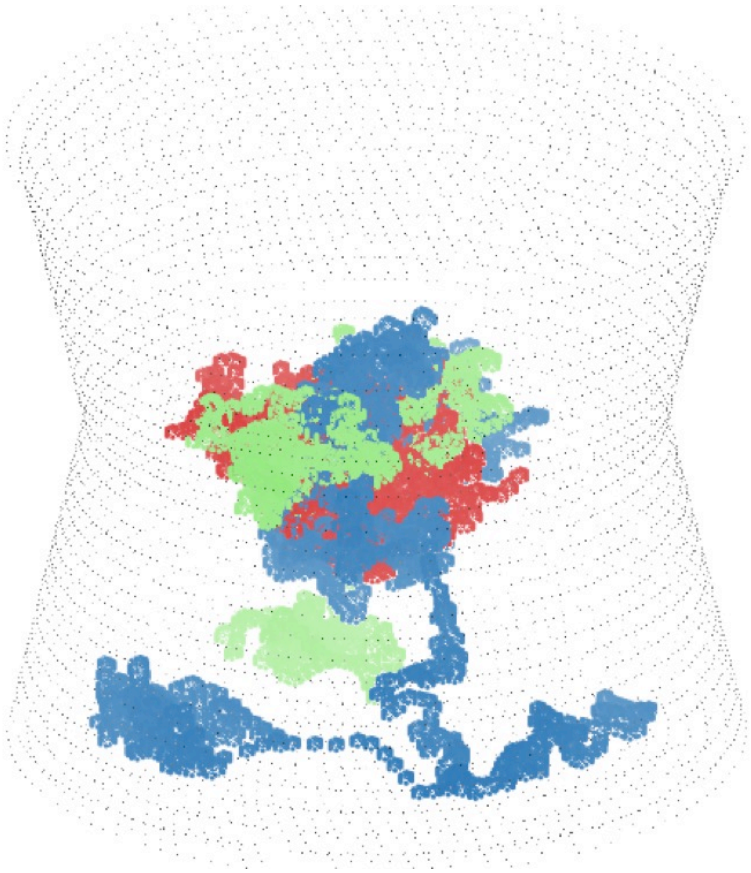

Figure 5-14: Density map of center of mass of cargo. Green dots are for cargo in wildtype system, red dots are for cargo in the negative control system, and blue dots are for cargo in positive control system. Significant difference is observed in the movement of cargo in the 3 different systems, with the cargo in negative control system sweeping a smaller space in the system and the cargo in positive control system sweeping a larger space. This observation correlates with results in Figure 5-13 where in negative control, FG motifs are more clustered and in positive control, they are more scattered.

Conformational ensemble of each of FG Nup layers inside the NPC

Here, the density map of each layer of FG Nups is visualized in the whole-NPC model. The side view and top view of the density map of each layer of FG Nup is presented in Figure 5-15. Density threshold in set to 1.5%. The Nups are ordered in the way they appear in the NPC from cytoplasmic side to nucleoplasmic side [3] (ordered from left to

right and top to bottom in Figure 5-15). Some of the FG Nups like Nsp1 appear at multiple layers in the NPC, so each of the different layers are drawn separately. The side-view results show that Nsp1 copies and Nup116 travel far from their tethering point within the NPC. The other FG Nups stay within their tethering point region. However, density maps of almost all of FG Nups overlap in terms of their location within the NPC.

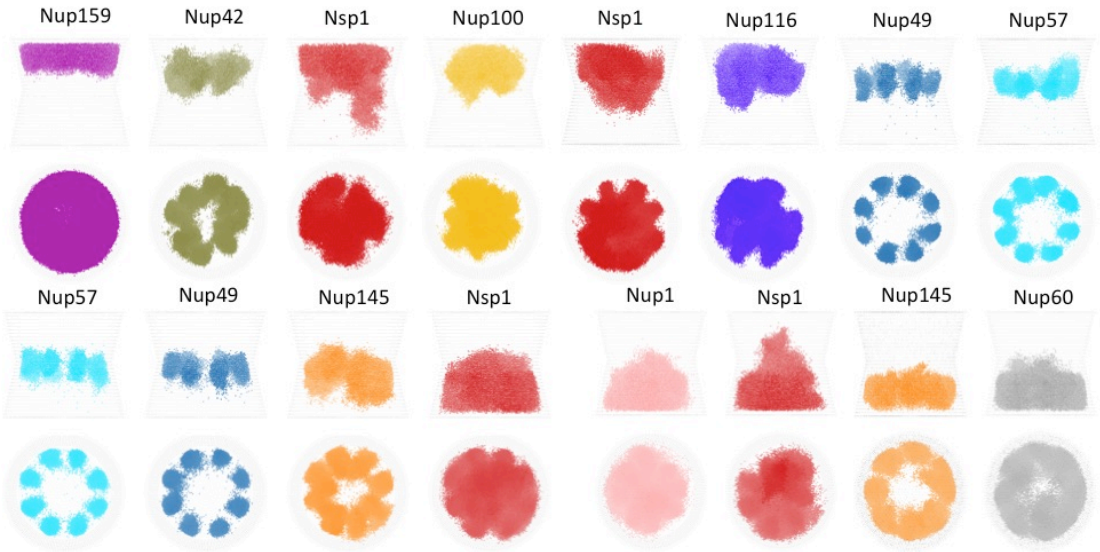

Figure 5-15: Density map of each layer of FG Nups in the whole-NPC model. Density threshold is set to 1.5% for all the cases. The density maps are ordered in the same manner they appear in NPC (Alber et al. 2007)**. Density maps from left to right and from top to bottom correspond to layers of FG Nups in yeast NPC from cytoplasmic side to nucleoplasmic side. Nsp1 copies and Nup116 travel far from their tethering point and explore a large space inside the NPC. The other Nups stay within short distance from their tethering point but the spaces each of the FG Nups sweep overlap with each other.**

Density maps of different layers of FG Nups are drawn in Figure 5-16 with a 3% density threshold to compare with results from isolated ring cross-sections. Slight differences exist in density maps of Nsp1 copies and Nup116 rings in the whole NPC model and ring cross sections. The reason might be because the whole NPC is a crowded system and the movement of cargo complex happens slowly. Therefore, the cargo spends a long time at each layer, and causes Nsp1 and Nup116 molecules not to fully cluster at the center. It is also interesting to note that while the other FG Nups are located close to their neighboring FG Nups and share the space with them, their conformational ensemble is not influenced by presence of neighboring molecules and looks similar to their isolated ring behavior.

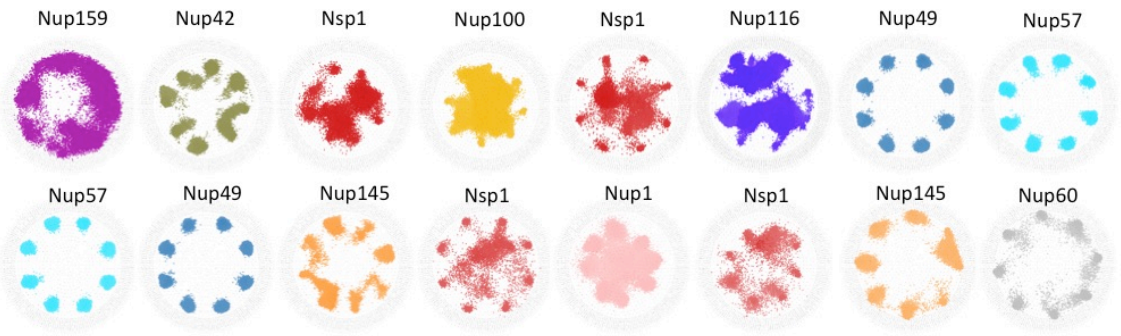

Figure 5-16: Top view of density map of different layers of FG Nups in whole-NPC model. Density threshold is set to 3%. Comparing these results with those obtained from ring cross-section simulations show that Most of the FG Nups maintain their conformational ensemble when positioned next to neighboring FG Nups in the whole-NPC model. Density maps of Nsp1 copies and Nup116 is slightly different, since the whole NPC is a crowded system and the movement of cargo complex happens slowly. Therefore, the cargo spends a long time at each layer, and causes Nsp1 and Nup116 molecules not to fully cluster at the center.

Discussion

The role of FG Nup sequences in NPC function is still not fully understood. In this study, we employed a biophysics approach to analyze the effect of presence of a specific sequence feature in FG Nups, namely Like Charge Regions (LCRs) in nucleocytoplasmic transport. In order to make the analysis tractable, NPC was analyzed at three different levels, i.e. single FG Nups, ring cross sections and whole-NPC model of yeast NPC. Our results suggest that in different ring cross sections of the NPC, FG Nups make different conformational ensembles that depend upon the length and distribution of charged residues in their sequence. We also observed that forces are not evenly distributed between residues of FG Nups. Phe residues experience the largest stress, meaning that Phe residues are interacting with each other and make a hydrophobic network. Our results also suggest that the stresses are not evenly distributed between different domains of the FG Nups. In all the proteins analyzed in this study, the residues in LCR domain experience on average a higher force caused by Nup-Nup interactions compared to the non-LCR domain. LCR domains are more collapsed and aggregated and have a lower average distance between their Phe residues, therefore, they interact more strongly with minimal disruption caused by electrostatic repulsion of charged residues. In other words, in the LCR domains, hydrophobic interactions dominate and the residues form a dense hydrophobic network, while in non-LCR domains, electrostatic repulsion dominates against hydrophobic interactions and the Nups do not collapse into aggregated conformations.

Significance of presence of LCRs was analyzed by comparing the Nup-Nup interactions, Nup-cargo interaction, movement of cargo complex inside the NPC, and characteristics of network of FG motifs in the wildtype case and two controls. In case of Nup-Nup interactions, despite different conformational ensembles of FG Nups due to the variations in their length and sequence structure, a similar pattern could be observed in the difference between the wildtype and mutated rings for all of the yeast FG Nups, Negative control had a similar effect in all the Nups analyzed in this study. By mutating the charged residues in LCR domain to Alanine, the LCR domain of all the Nups become more collapsed, the Phe residues become on average closer to each other, and therefore form a stronger hydrophobic network. Hence, the stress in Phe residues increases as a result of stronger interactions. Stronger interactions are harder to break, therefore in all of the cases, the Nups get trapped into highly aggregated conformations and can hardly overcome the energy barrier to break the bonds and interact with new molecules. Therefore, the system transforms into a less “disordered”, frozen state. On the other hand, positive control also had a similar effect on all the FG Nups presented in this study. By mutating some of the Alanines in the LCR domain to positively charged residues, the LCR domain becomes less collapsed/aggregated. Phe residues become on average more distant and form weaker bonds with each other. Therefore, the stress in Phe residues decreases. Weaker interactions formed in positive control are easier to break; hence the Nups easily bind and unbind and can more easily cross-interact with new neighboring molecules. Therefore, the system transforms to a more “disordered” dynamic network of residues. Based on these results, number of charged residues in the LCR regulates the “orderedness” and “disorderedness” of the network of FG Nups.

Mutating the LCR domains had a similar effect on the whole-NPC model as well, confirming that the results observed in ring cross-sections can be observed at a higher level, which is whole NPC. Interestingly, while wildtype and negative control system form high-density FG network toward the central area of the NPC, the density of FG motifs in positive control is decreased to an extent that the high-density network is significantly influenced. Mutating the LCR domains has also a remarkable effect on the movement of cargo complex inside the NPC. Similar to ring cross-section results, movement of cargo complex is limited in negative control and the cargo has more freedom of movement in positive control.

The number of charged residues in LCR domain can also regulate the interaction of FG Nups with cargo complex. Our results showed that in some of the cases, changing the number of charged residues in LCR did affect the interaction of FG Nups with cargo complex and in some of the cases it did not. For example, in Nsp1, the interaction of LCR domains did not significantly influence the interaction of FG Nups with cargo complex. In Nup42, removing the charged residues in LCR did not affect the interaction of the Nups with cargo complex, but adding the charged residues significantly hindered the interaction of cargo complex with mutant Nup42 molecules. In Nup49 rings, adding charged residues to the system decreased the interaction of mutant Nup49 molecules with cargo complex by order of three, but removing the charged residues in negative control decreased the interaction of mutant Nup49 molecules with cargo complex by a significant order of 10.

Interaction of FG Nups with cargo complex and the characteristics of FG network within the NPC are two of the governing factors of nucleocytoplasmic transport. Results of this study suggest that LCR domains directly regulate the mobility of FG network inside the NPC. Also, mutating LCR domains in some of the Nups hinders the transient interactions that are necessary for nucleocytoplasmic transport.

Fast and selective are two principal features of nucleocytoplasmic transport. The interaction between FG Nups and transport factors can be described as 'kiss-and-run' behavior, which is comprised of super fast on and off rates [75]. However, it is not yet known what enables the fast interaction and prevents transport factors from sticking to FG Nups [112]. According to our results, LCRs could be one of the reasons that fast interaction is enabled between transport factors and FG Nups. LCRs enable the cargo complex to go through binding and unbinding in a transient fashion and removing the charged residues causes the cargo complex to "stick" to FG Nups. Interaction of multiple FG Nups with one transport factor has been correlated to a weak and fast interaction and short residence times on each of the interactions [112] and LCRs facilitates that by enabling cross-interactions between the Nups and keeping the energy barrier low enough for Nups to be able to unbind and form new interactions with other molecules. Also, it has been reported that a highly reactive and dynamic surface is required for a very fast binding [18] and LCR facilitates that by keeping the FG Nups from forming excessively dense clusters. Additionally, since LCRs promote movement of the FG Nups and cargo complex, they probably increase the fluctuation of FG motifs around transport factors, which might facilitate transport factors hopping between different FG Nups [112]. Moreover, LCRs allow FG Nups to maintain their disordered properties and dynamic behavior, and this is in line with the observations from multiple studies that report FG Nups remain highly disordered while transiently interacting with transport factors [18,76,113].

On the other hand, low density of charged residues in LCR domain allows for a dense network to form at the center of NPC. Presence of more charge residues in LCR domain would inhibit formation of those aggregates that could compromise selectivity of the NPC. Presence of dense FG networks is known to give NPC its selective quality.

Based on multiple observations at different levels in this chapter, in negative control the effect of presence of LCR domains is enhanced, meaning that the roles that LCR play is magnified in negative control. However, in positive control, the effect of presence of LCRs are removed to a large extent, since by adding charged residues, the LCR domains become more similar to charge-rich domains. This hinders the characteristics, and the role of LCR in the NPC at multiple levels.

Our unpublished data shows that LCRs are exclusive features of sequences of FG Nups that differentiate FG Nups from DisProt proteins. By removing the LCR domains from sequences of FG Nups, their charge distribution looks like DisProt proteins.

Based on observations of this study and abovementioned unpublished data, charged residues in LCR domain keep these domains on the fine line of being fully disordered (like DisProt proteins) and being vary collapsed and globular. The number of charged residues in LCR is optimized to keep the network of Phe residues dense enough to form a dense hydrophobic barrier inside the NPC, while enabling fast and transient interactions to take place between FG Nups and cargo complex.

The results from this study emphasize and add to the findings from our previous study [79] on the importance of presence of LCRs on NPC function. While, only a few charged residues are mutated in each sequence, a significant change can be observed in the conformational ensemble of FG Nups and interaction of cargo complex. This is more remarkable considering the fact that NPC is a very robust macromolecular machine [13], which is viable even when some of the FG Nups are removed. Although our results cannot directly predict how transport process would be affected by mutation of LCRs, the effect of mutating LCRs on the governing factors of transport process is predicted. The results from this study call for experimental investigations to study the effect of LCRs on transport process. Investigating the effect of presence of this specific sequence feature can be useful in understanding the underlying mechanism of transport as well as designing artificial nanopores.

Methods

A one-bead-per-amino acid coarse-grained molecular dynamics model was used to analyze the behavior of FG Nups and cargo complex in the ring cross sections of the NPC. This coarse-grained model is designed to study intrinsically disordered proteins. Amino acids are modeled as spherical beads with a mass of 120Da and distance of 0.38nm. The force-field developed for this model, originally developed by Ghavami et al. accounts for different biophysical factors including bending and torsion potentials between neighboring beads, an implicit solvent, ion screening effect, and hydrophobic and electrostatic interactions [30,114]. Langevin dynamics simulations were performed using GROMACS molecular dynamics simulation software [107]. All visualizations were performed via VMD 1.9.3 [115].

Density plots were generated by discretizing the ring into 0.5x0.5x0.5 nm unit cells. The number of occurrences of each element was counted over the total simulation time. Gridcount package was used to generate the density plots [109]. A time-step of 0.02 ps was used in simulations. In all of the simulations, only disordered domains of FG Nups were used in the simulation, which are taken from [11]. This model has been previously used in modeling FG Nups [30,79,116]. Density plots of whole-NPC simulations were extracted using GROmaps [117].

A simple representation of cargo complex was used in the model. Cargo was modeled as non-interacting beads that form a spherical shape with a diameter of 10 nm. Transport receptor (specifically Karyopherin β) was modeled as 7 hydrophobic spots (Phenylalanine) on the surface of the cargo [116]. In the initial configuration of the rings, the cargo complex was located at the center of the ring, and FG Nups had a random conformation as the starting point of the simulation.

Time-Resolved Force Distribution Analysis (TRFDA) was performed on the ring complexes using TRFDA GROMACS tool [118]. Punctual stress on each residue based on the non-bonded interaction of Nup-Nup or Nup-cargo complex was calculated through simulation time. This tool calculated punctual stress by summing the absolute values of the pairwise forces acting on each atom [118].

Appendix

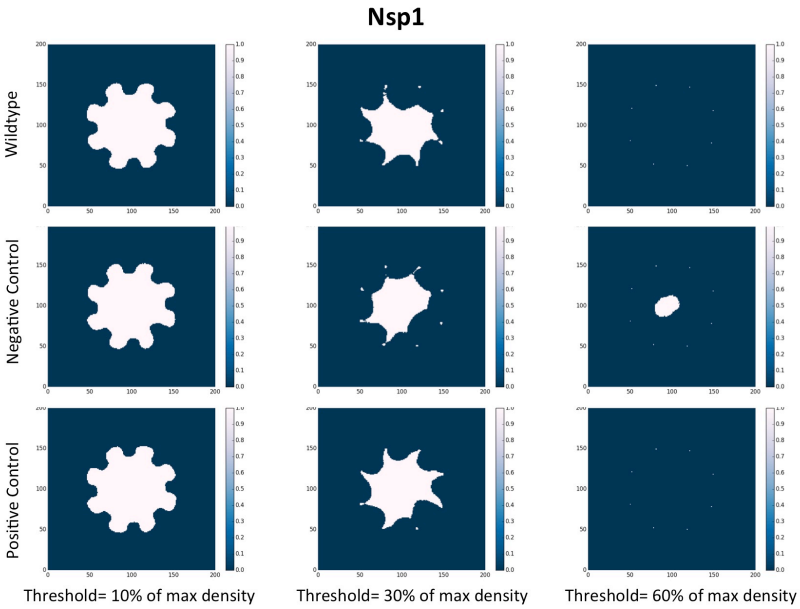

Figure 5-S1: Aggregation of Nsp1 molecules in wildtype versus two mutants. In order to better visualize the difference between the three cases, the relative density threshold is increased and the grid cells with a density higher than the threshold are shown in white. The threshold is increased from left to right. The percentage of the area of the square covered by white dots are 74%, 43%, and 0.08% for wildtype, 72%, 36%, and 6% for negative control, and 76%, 48% and 0.08% for positive control. In the lower thresholds (10% and 30%) the percentages are higher in positive control and then wildtype. In higher threshold (60%), the percentage is almost zero for wildtype and positive control but is 6% for negative control. This clearly demonstrates that adding charged residues to LCR (positive control) decreases the density of the aggregated conformation at the center while removing the charged residues (negative control) increases the density of the Nsp1 at the center. On the other hand, in positive control, the Nups are more uniformly distributed, hence forming a larger covered area at lower thresholds. The Nups become less evenly distributed in wildtype and then negative control.

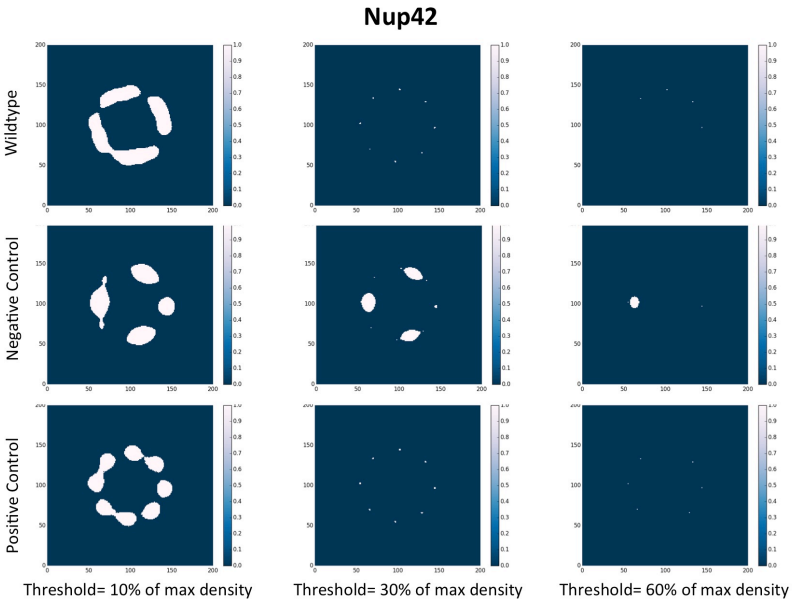

Figure 5-S2: Aggregation of Nup42 molecules in wildtype versus two mutants. In order to better visualize the difference between the three cases, the relative density threshold is increased and the grid cells with a density higher than the threshold are shown in white. The threshold is increased from left to right. The percentage of the area of the square covered by white dots are 33%, 0.19%, and 0.04% for wildtype, 24%, 8.2%, and 1.27% for negative control, and 27%, 0.28% and 0.06% for positive control. In the higher thresholds (60% and 30%) the percentages are higher in negative control and very small or zero in wildtype and positive control. In lower threshold (10%), the percentage is smallest for negative control and largest for wildtype. This clearly demonstrates that adding charged residues to LCR (positive control) decreases the density of the aggregated conformation at the center while removing the charged residues (negative control) increases the density of the Nsp1 at the center. On the other hand, in positive control, the Nups are more uniformly distributed, hence forming a larger covered area at lower thresholds. The Nups become less evenly distributed in wildtype and then negative control.

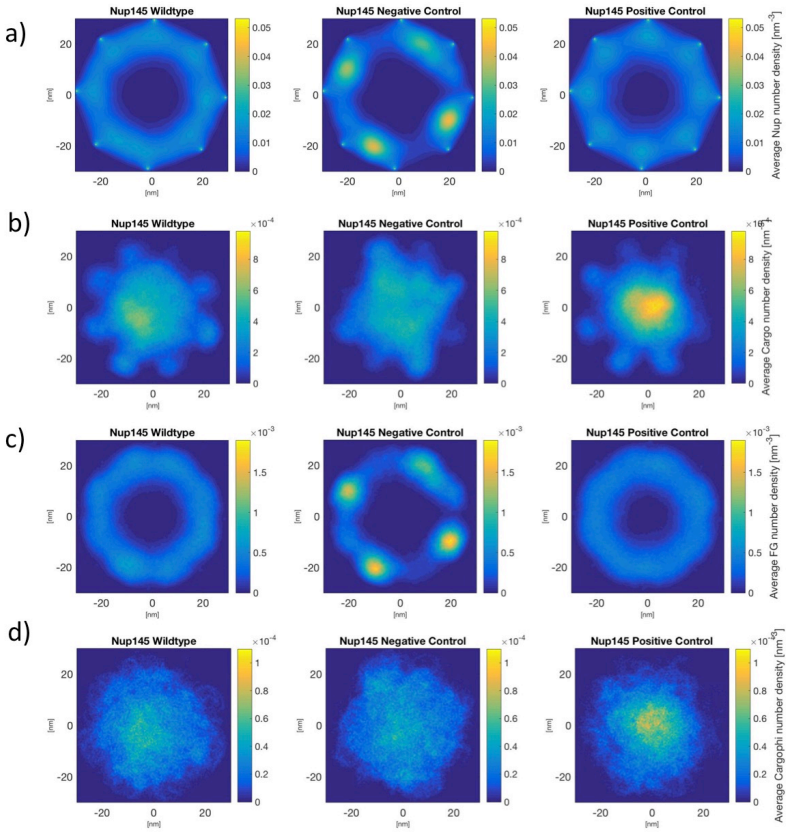

Figure 5-S3: Density maps of a) Nups and b) cargo c) FG motifs d) hydrophobic spots on the cargo complex, inside ring of Nup145. The plots on the left column show wildtype Nup145 simulation results, plots in the middle column show results of negative control, and the plots on the right show the results of positive control. Nup145 has a middle length charge-rich domain and a middle length LCR domain. The LCR domain contains 7 positively charged residues (see Figure 1). The LCR domain forms a globular conformation and the non-LCR domain forms a straight conformation. The aggregated LCR domains interact with their neighboring Nup145 molecules (see part a). In the negative control, where the 7 charged residues in LCR domain are mutated to Alanine, the maximum density of Nup145 molecules increases by 30%. Also, movement of cargo becomes less even compared to wildtype (see part b). In the positive control, where charged residues are added to the LCR domain by mutating 7 Alanines to positively charged residues, the density map of Nups becomes more uniform. The density of Nup145 molecules decrease by 25%, which implies the fact that the mutant Nup145 molecules form a less aggregated/more evenly distributed cross-interacting network. Density map of FG motifs (c) show similar pattern to Nups because Nup145 molecules are abundant in FG motifs along the length of their amino acid sequence. Density map of cargo hydrophobic spots (d) shows similar pattern to cargo.

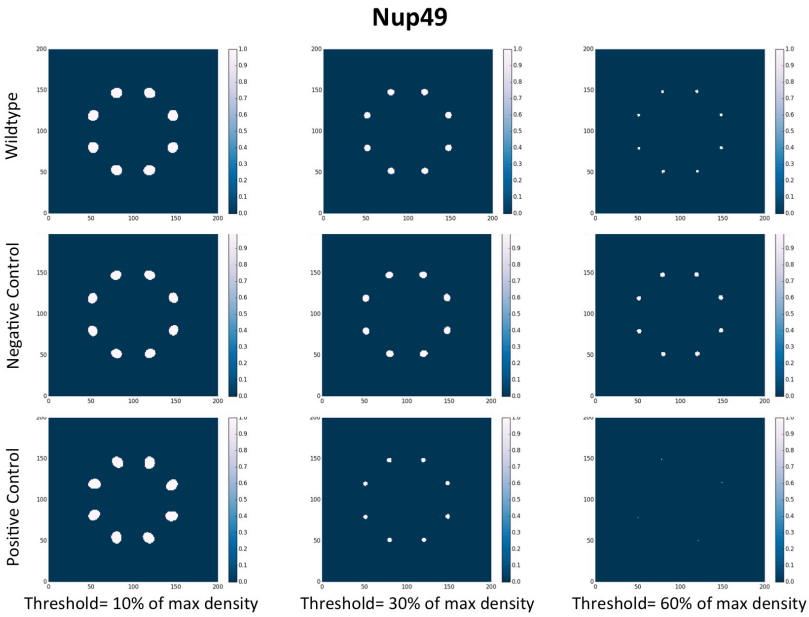

Figure 5-S4: Aggregation of Nup49 in wildtype versus mutants. In order to better visualize the difference between the three cases, the relative density threshold is increased and the grid cells with a density higher than that of the threshold are shown in white. The threshold is increased from left to right, and the area covered by white dots is 10%, 4%, and 0.8% for wildtype, 9%, 4.6%, and 1.8% for negative control and 11%, 1.8% and 0.04% in positive control. In the 10% threshold, the covered area is larger for positive control, and then wildtype, and then negative control. In 30% and 60% threshold, the covered area is larger in negative control, then wildtype and then positive control. This pattern clearly shows that the Nup 49 molecules form a more aggregated conformation in negative control, and more relaxed coil conformation in positive control, compared to wildtype.

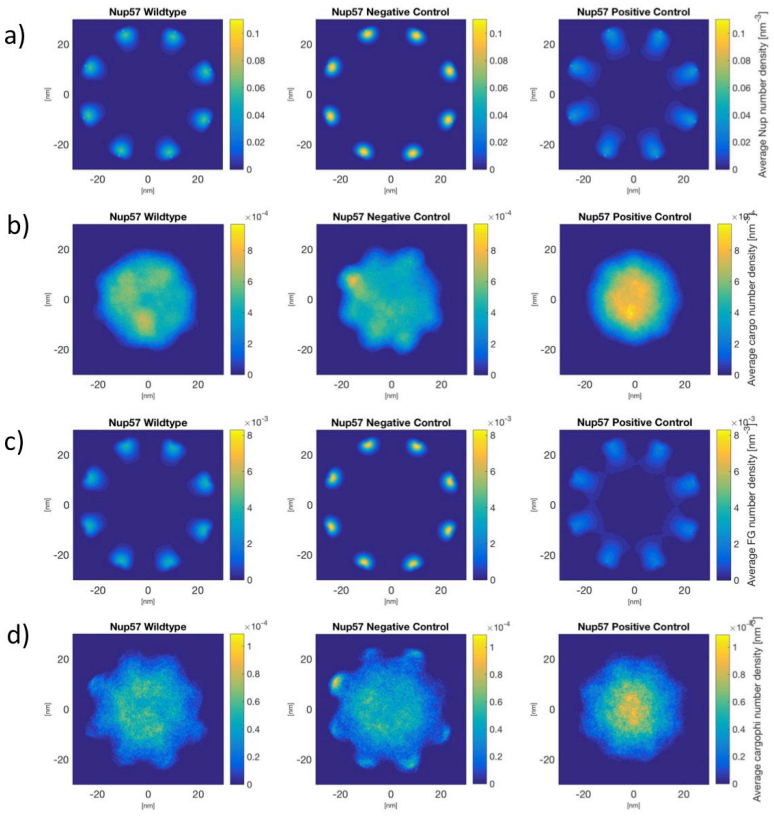

Figure 5-S5: **Density maps of a) Nups and b) cargo c) FG motifs d) hydrophobic spots on the cargo complex, inside ring of Nup57. The plots on the left column show wildtype Nup57 simulation results, plots in the middle column show results of negative control, and the plots on the right show the results of positive control. The length of the disordered domain of Nup57 is 255 amino acids. This protein only contains 7 positively charged residue and zero negatively charged residues (see Figure 1). Therefore, the entire molecule is an LCR, and due to very low charge density, the molecule collapses into highly aggregated conformation toward the scaffold of the ring. The molecules make a very small end-to-end distance that is not long enough for them to be able to cross-interact with their neighboring molecules (part a). In the negative control, where the 7 charged residues in LCR domain are mutated to Alanine, the maximum density of Nup57 molecules increases by 10%. The cargo complex gains more area to move, but the density map shows some high-density spots, which are in the same location as Nups (see part b). This implies that cargo complex and Nup57 molecules are interacting with each other for a long time. In the positive control, where charged residues are added to the LCR domain by mutating 7 Alanines to positively charged residues, the Nups become less aggregated by 20% due to presence of more charged residues. Density map of cargo shows high density at the center (see part b), which implies that the cargo complex is spending most of it time traveling between the Nup57 molecules rather than interacting with some of them for a long time. Density map of FG motifs (c) show similar pattern to Nups because Nup57 molecules are abundant in FG motifs along the length of their amino acid sequence. Density map of cargo hydrophobic spots (d) show a similar pattern to that of cargo.**

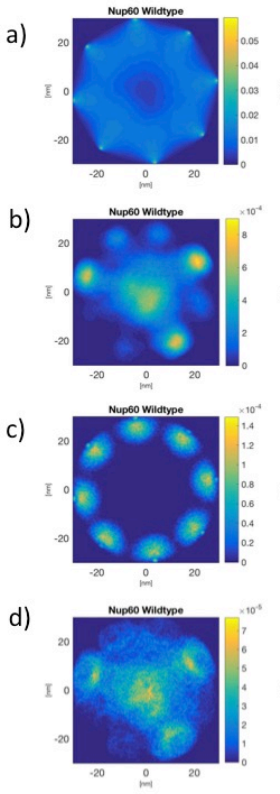

Figure 5-S6: Density maps of a) Nups and b) cargo c) FG motifs d) hydrophobic spots on the cargo complex, inside ring of Nup60. Since this Nup does not feature an LCR, only the wildtype case is analyzed. This Nup is analyzed as a natural control for the other Nups that do contain an LCR. The length of the disordered domain of Nup60 is 539 amino acids and its entire length is a charge-rich domain (see Figure1). Therefore, this Nup does not feature an LCR. Although this Nup is long and the 8 copies of Nup cover the entire area inside the ring, the molecules do not form a high-density conformation at the center. Nup60 has only one FG motif, therefore, density map of FG motifs is different than that of Nups. (see part c) Density map of cargo hydrophobic spots (d) show a similar pattern to that of cargo.

Chapter 6 : Closure

Summary and Conclusion

The main motivation behind the work done in this dissertation was *to identify the unique characteristics of FG Nups that enable them to facilitate fast and selective nucleocytoplasmic transport.* Considering that FG Nups are intrinsically disordered proteins, we started by exploring their sequence composition for any specific features that might enable their function. The main findings of this work are as follows:

- **LCRs are evolutionarily conserved, exclusive amino acid sequence features of FG Nups**

We found a new feature in amino acid sequences of FG Nups that has never been reported before. Like charge regions (LCRs) are evolutionarily conserved extended sub-sequences of FG Nups that contain only positively charged residues, have a low charge density and are located at the N-terminus of FG Nups. We found that these patterns co-localize with dense clusters of FG motifs to form a dense, hydrophobic conformation toward the center of the NPC. We also found that these features are exclusive to FG Nups and are not observed in other disordered proteins. LCRs are what differentiate the amino acid sequence composition of FG Nups from other disordered proteins. If these features are removed from the sequences of FG Nups, they will look similar to other disordered proteins in terms of charge distribution. The fact that LCRs are evolutionarily conserved and exclusive to FG Nups strongly implies that they should be important for the function of FG Nups.

- **LCRs are important role-players in the conformational ensemble of FG Nups, hydrophobic network FG motifs, the movement of the cargo complex, the interaction of FG Nups with the cargo complex, and the distribution of force in the sequences of FG Nups**

In the next phase of this work, we tested our hypothesis in phase I, and explored whether LCRs are important in NPC function. The results of our molecular dynamics simulations at multiple levels of (1) single Nup, (2) ring cross-section of NPC containing one type of Nup, and (3) the whole NPC showed that LCRs are important for the conformational ensemble of FG Nups. Mutating the few charged residues in LCRs to Alanine will result in the collapse of the LCR domain, and will reduce cross-interactions between FG Nups. In this case, the Nups form high-density, aggregated, high-stress conformations that are a result of intense hydrophobic forces that form due to reduced average distance between Phenylalanine residues. On the other hand, when charged residues are added to the sequences (positive control), FG Nups form less aggregated, more cross-interacting and evenly distributed conformations. The average spacing between the hydrophobic residues increases; therefore, the intensity of hydrophobic interactions and the stress in the FG Nups decreases. It is important to note that in all the cases, the stresses in the LCR domain are significantly higher than non-LCR domains because LCR domains are more collapsed and interact more within and between the Nups.

Our results show that LCRs can also have a regulatory role in the interaction of FG Nups and the cargo complex. Some of the mutations studied in chapter 5 significantly affect the interaction between the cargo complex and FG Nups. In some of the cases, adding charged residues drastically decreased the interaction between the FG Nup and cargo complex, while in other cases, removing the charged residues significantly intensified this interaction. Additionally, our whole NPC simulations showed that the number of charged residues in LCR domains directly regulates the freedom of the cargo complex inside the NPC. As the number of charged residues is increased, the cargo gains more freedom of movement and explores a larger space in the NPC.

Nucleocytoplasmic transport phenomenon is influenced by the interaction of FG Nups and cargo complex as well as the conformational ensemble of FG Nups and hydrophobic network of FG motifs. Changing the number of charged residues in LCR domain of FG Nups can significantly affect the conformation of FG Nups, and/or the interaction of FG Nups with the cargo complex and impact nucleocytoplasmic transport. These findings depict how charge decoration in FG Nups can have a regulatory role in FG Nups' function. . Based on our findings , we can conclude that the number of charged residues in LCR is optimized to keep the network of Phe residues dense enough to form a dense hydrophobic barrier inside the NPC, while enabling fast and transient interactions to take place between FG Nups and cargo complex. Increasing the number of charged residue can interfere with selectivity of the transport process, while decreasing the number of charged residues can interfere with fast transport of molecules. As a result, we can conclude that the specific way that charged residues are distributed in FG Nups and presence of LCRs are one of the key features that enable the FG Nups to perform their function.

Future Work

LCRs are one of the features that we discovered to exist in amino acid sequences of FG Nups that were shown to be important for their function. However, there is more to be discovered about sequence composition of FG Nups. This dissertation focused on a pattern in charge distribution but we believe there are other features in the distribution of FG motifs and decoration of different types of residues with respect to each other that enable FG Nups to perform their function. It is worth noting that FG Nups are intrinsically disordered proteins with high mutation rates. Therefore, structure-function relation is more plausible at sequence composition level, rather than at the level of individual residues.

The results presented in Chapter 5 can be expanded in several different directions. This dissertation was the first to study force distribution analysis on intrinsically disordered proteins, and specifically on FG Nups. There is a lot to be discovered from the big data that was generated in this study as a result of simulating a large, dynamic system of intrinsically disordered proteins. The force distribution analysis creates a large data of stresses for each residue at each time step of the simulation. The way the force propagates along the length of FG Nups, and how this force transmission differs between the simulations by mutating only a few charged residues is one of the questions that could be

further explored. Additionally, force distribution data along with the trajectory of residue positions can be used to learn more about other sequence features that facilitate FG Nups' function.

The model of the whole NPC can also be continued in many different directions. How much does each of the FG Nups interact with the cargo complex? Are some of them interacting more than the others? What happens if we have more than one cargo complex in the system?

The results presented in this dissertation could be complemented by experimental validation in different scales, i.e. single Nups, ring cross-section and entire NPC. Our computational modeling suggests that LCRs are major sequence features of FG Nups that significantly affect FG Nups conformation and function, but experimental studies can provide additional insights.

Last but not least, NPC-inspired design of nanopores has attracted attention in recent years. The findings of this dissertation suggest that LCRs are important features that should exist in NPC to enable its function, which could guide the design NPC-inspired artificial nanopores.

Final Remarks

I got introduced to the amazing world of cell biomechanics and mechanobiology when I started my PhD at Molecular Cell Biomechanics Laboratory at UC Berkeley. To me, using the rules of Physics, Math, Statistics and Mechanics to uncover intricate biological systems was beyond interesting and I was astonished how two parts of science, which seemed very distant in my mind, could go hand in hand and answer the most fundamental questions of life like how cells work.

During our initial meetings, Professor Mofrad mentioned Nuclear Pore Complex (NPC), a vital cellular machinery that is significantly under-studied due to limitations of experimental techniques in uncovering its function. Computational techniques on the other hand, could study this system in detail and lend valuable insight into its mechanism. I started with this profound, yet overwhelming question of "Are there any specific features in amino acid sequences of FG Nups that enable their function?". This question never stopped to amaze me throughout my years of PhD studies, and the more I researched, the more I understood how much is yet to be discovered in this field. However, the trend of discoveries and recent advances in experimental and computational techniques shows a promising path toward understanding the function of NPC. Hopefully, understanding the function of this cellular machinery can help with curing diseases that are associated with its malfunction, including cancer and viral infections.

References

1. Jovanovic-Talisman, T. & Zilman, A. Protein Transport by the Nuclear Pore Complex: Simple Biophysics of a Complex Biomachine. *Biophys. J.* **113**, 6–14 (2017).
2. Jamali, T., Jamali, Y., Mehrbod, M. & Mofrad, M. R. K. *Nuclear Pore Complex. Biochemistry and Biophysics of Nucleocytoplasmic Transport in Health and Disease*. *International Review of Cell and Molecular Biology* **287**, (Elsevier Inc., 2011).
3. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701 (2007).
4. Yang, W. & Musser, S. M. Nuclear import time and transport efficiency depend on importin beta concentration. *J. Cell Biol.* **174**, 951–61 (2006).
5. Knockenhauer, K. E. & Schwartz, T. U. The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell* **164**, 1162–1171 (2016).
6. Beck, M. & Hurt, E. The nuclear pore complex: understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* **18**, 73–89 (2017).
7. Miao, L. & Schulten, K. Transport-related structures and processes of the nuclear pore complex studied through molecular dynamics. *Structure* **17**, 449–59 (2009).
8. Aramburu, I. V. & Lemke, E. A. Floppy but not sloppy: Interaction mechanism of FG-nucleoporins and nuclear transport receptors. *Semin. Cell Dev. Biol.* **68**, 34–41 (2017).
9. Mulder, F. A. A. Fuzzy and fast nuclear transport. *J. Biol. Chem.* **293**, 4564–4565 (2018).
10. Wagner, R. S., Kapinos, L. E., Marshall, N. J., Stewart, M. & Lim, R. Y. H. Promiscuous Binding of Karyopherinβ1 Modulates FG Nucleoporin Barrier Function and Expedites NTF2 Transport Kinetics. *Biophys. J.* **108**, 918–927 (2015).
11. Yamada, J. *et al.* A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics* **9**, 2205–2224 (2010).
12. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2450–5 (2003).
13. Strawn, L. A., Shen, T., Shulga, N., Goldfarb, D. S. & Wente, S. R. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat. Cell Biol.* **6**, 197–206 (2004).
14. Devos, D. *et al.* Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2172–7 (2006).
15. Bednenko, J., Cingolani, G. & Gerace, L. Nucleocytoplasmic Transport: Navigating the Channel. *Traffic* **4**, 127–135 (2003).
16. Adams, R. L. & Wente, S. R. Uncovering Nuclear Pore Complexity with Innovation. *Cell* **152**, 1218–1221 (2013).
17. Hough, L. E. *et al.* The molecular mechanism of nuclear transport revealed by atomic-scale measurements. doi:10.7554/eLife.10027.001
18. Milles, S. *et al.* Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors. *Cell* **163**, 734–745 (2015).
19. Kalab, P., Weis, K. & Heald, R. Visualization of a Ran-GTP gradient in interphase

and mitotic Xenopus egg extracts. *Science* **295**, 2452–6 (2002).
20. Frey, S., Richter, R. P. & Görlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–7 (2006).
21. Labokha, A. A. *et al.* Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. *EMBO J.* **32**, 204–18 (2013).
22. Hülsmann, B. B., Labokha, A. A. & Görlich, D. The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell* **150**, 738–51 (2012).
23. Rout, M. Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol.* **13**, 622–628 (2003).
24. Lim, R. Y. H. *et al.* Nanomechanical basis of selective gating by the nuclear pore complex. *Science* **318**, 640–3 (2007).
25. Peters, R. Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. *Traffic* **6**, 421–7 (2005).
26. Ando, D. *et al.* Nuclear pore complex protein sequences determine overall copolymer brush structure and function. *Biophys. J.* **106**, 1997–2007 (2014).
27. Ando, D., Colvin, M., Rexach, M. & Gopinathan, A. Physical motif clustering within intrinsically disordered nucleoporin sequences reveals universal functional features. *PLoS One* **8**, e73831 (2013).
28. Sakiyama, Y., Mazur, A., Kapinos, L. E. & Lim, R. Y. H. Spatiotemporal dynamics of the nuclear pore complex transport barrier resolved by high-speed atomic force microscopy. *Nat. Nanotechnol.* **11**, 719–723 (2016).
29. Moussavi-Baygi, R., Jamali, Y., Karimi, R. & Mofrad, M. R. K. Brownian dynamics simulation of nucleocytoplasmic transport: a coarse-grained model for the functional state of the nuclear pore complex. *PLoS Comput. Biol.* **7**, e1002049 (2011).
30. Ghavami, A., Veenhoff, L. M., van der Giessen, E. & Onck, P. R. Probing the disordered domain of the nuclear pore complex through coarse-grained molecular dynamics simulations. *Biophys. J.* **107**, 1393–402 (2014).
31. Tagliazucchi, M., Peleg, O., Kröger, M., Rabin, Y. & Szleifer, I. Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3363–3368 (2013).
32. Moussavi-Baygi, R., Jamali, Y., Karimi, R. & Mofrad, M. R. K. Biophysical coarse-grained modeling provides insights into transport through the nuclear pore complex. *Biophys. J.* **100**, 1410–9 (2011).
33. Frey, S. & Görlich, D. FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties. *EMBO J.* **28**, 2554–67 (2009).
34. Mincer, J. S. & Simon, S. M. Simulations of nuclear pore transport yield mechanistic insights and quantitative predictions. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E351–E358 (2011).
35. Peyro, M., Soheilypour, M., Lee, B. L. & Mofrad, M. R. K. Evolutionarily Conserved Sequence Features Regulate the Formation of the FG Network at the Center of the Nuclear Pore Complex. *Sci. Rep.* **5**, 15795 (2015).
36. Duarte-Rey, C. *et al.* Primary biliary cirrhosis and the nuclear pore complex.

Autoimmun. Rev. **11**, 898–902 (2012).
37. Cronshaw, J. M. & Matunis, M. J. The nuclear pore complex: disease associations and functional correlations. *Trends Endocrinol. Metab.* **15**, 34–39 (2004).
38. Sheffield, L. G., Miskiewicz, H. B., Tannenbaum, L. B. & Mirra, S. S. Nuclear pore complex proteins in Alzheimer disease. *J. Neuropathol. Exp. Neurol.* **65**, 45–54 (2006).
39. Lim, R. Y. H., Aebi, U. & Fahrenkrog, B. Towards reconciling structure and function in the nuclear pore complex. *Histochem. Cell Biol.* **129**, 105–16 (2008).
40. Krishnan, V. V *et al.* Intramolecular cohesion of coils mediated by phenylalanine--glycine motifs in the natively unfolded domain of a nucleoporin. *PLoS Comput. Biol.* **4**, e1000145 (2008).
41. Zhao, C. L., Mahboobi, S. H., Moussavi-Baygi, R. & Mofrad, M. R. K. The interaction of CRM1 and the nuclear pore protein Tpr. *PLoS One* **9**, e93709 (2014).
42. Dölker, N., Zachariae, U. & Grubmüller, H. Hydrophilic linkers and polar contacts affect aggregation of FG repeat peptides. *Biophys. J.* **98**, 2653–61 (2010).
43. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–94 (2007).
44. Wolf, C. & Mofrad, M. R. K. On the octagonal structure of the nuclear pore complex: insights from coarse-grained models. *Biophys. J.* **95**, 2073–85 (2008).
45. Osmanović, D., Ford, I. J. & Hoogenboom, B. W. Model inspired by nuclear pore complex suggests possible roles for nuclear transport receptors in determining its structure. *Biophys. J.* **105**, 2781–2789 (2013).
46. Zilman, A., Di Talia, S., Chait, B. T., Rout, M. P. & Magnasco, M. O. Efficiency, selectivity, and robustness of nucleocytoplasmic transport. *PLoS Comput. Biol.* **3**, e125 (2007).
47. Colwell, L. J., Brenner, M. P. & Ribbeck, K. Charge as a selection criterion for translocation through the nuclear pore complex. *PLoS Comput. Biol.* **6**, e1000747 (2010).
48. Patel, S. S., Belmont, B. J., Sante, J. M. & Rexach, M. F. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* **129**, 83–96 (2007).
49. Azimi, M. & Mofrad, M. R. K. Higher nucleoporin-Importinβ affinity at the nuclear basket increases nucleocytoplasmic import. *PLoS One* **8**, e81741 (2013).
50. Azimi, M., Bulat, E., Weis, K. & Mofrad, M. R. K. An agent-based model for mRNA export through the nuclear pore complex. *Mol. Biol. Cell* **25**, 3643–53 (2014).
51. Dunker, a K. *et al.* Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput.* 473–484 (1998).
52. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–31 (1999).
53. Uversky, V. N., Gillespie, J. R. & Fink, a L. Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427 (2000).
54. Dunker, A. K. *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59 (2001).
55. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533

(2002).
56. *Protein Folding Handbook*. (Wiley-VCH Verlag GmbH, 2005). doi:10.1002/9783527619498
57. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim. Biophys. Acta* **1804**, 1231–64 (2010).
58. Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **1834**, 932–951 (2013).
59. Vucetic, S., Brown, C. J., Dunker, A. K. & Obradovic, Z. Flavors of protein disorder. *Proteins* **52**, 573–84 (2003).
60. Vuzman, D. & Levy, Y. DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21004–9 (2010).
61. Uversky, V. N. & Dunker, A. K. Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal. Chem.* **84**, 2096–104 (2012).
62. Sickmeier, M. *et al.* DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* **35**, D786-93 (2007).
63. Denning, D. P. & Rexach, M. F. Rapid evolution exposes the boundaries of domain structure and function in natively unfolded FG nucleoporins. *Mol. Cell. Proteomics* **6**, 272–82 (2007).
64. Milles, S. & Lemke, E. A. Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153. *Biophys. J.* **101**, 1710–9 (2011).
65. Campen, A. *et al.* TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* **15**, 956–63 (2008).
66. Hemmings, H. J., Nairn, A., Aswad, D. & Greengard, P. DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Purification and characterization of the phosphoprotein from bovine caudate nucleus. *J. Neurosci.* **4**, 99–110 (1984).
67. Gast, K. *et al.* Prothymosin .alpha.: A Biologically Active Protein with Random Coil Conformation. *Biochemistry* **34**, 13211–13218 (1995).
68. Terry, L. J. & Wente, S. R. Flexible gates: dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryot. Cell* **8**, 1814–27 (2009).
69. Ma, J. & Yang, W. Three-dimensional distribution of transient interactions in the nuclear pore complex obtained from single-molecule snapshots. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7305–10 (2010).
70. Walsh, I., Martin, A. J. M., Di Domenico, T. & Tosatto, S. C. E. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–9 (2012).
71. Achtert, E., Kriegel, H.-P. & Zimek, A. ELKI: A Software System for Evaluation of Subspace Clustering Algorithms. *Sci. Stat. Database Manag.* 580–585 (2008). doi:10.1007/978-3-540-69497-7_41
72. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
73. Dunn, J. C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *J. Cybern.* **3**, 32–57 (1973).
74. Hampoelz, B., Andres-Pons, A., Kastritis, P. & Beck, M. Structure and Assembly

of the Nuclear Pore Complex. *Annu. Rev. Biophys.* **48**, 515–536 (2019).
75. Beck, M. & Hurt, E. The nuclear pore complex: understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* **18**, 73–89 (2017).
76. Raveh, B. *et al.* Slide-and-exchange mechanism for rapid and selective transport through the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2489-97 (2016).
77. Li, C., Goryaynov, A. & Yang, W. The selective permeability barrier in the nuclear pore complex. *Nucleus* **7**, 430–446 (2016).
78. Huang, K., Tagliazucchi, M., Park, S. H., Rabin, Y. & Szleifer, I. Molecular model of the nuclear pore complex reveals a thermoreversible FG-network with distinct territories occupied by different FG motifs. *bioRxiv* 568865 (2019). doi:10.1101/568865
79. Peyro, M., Soheilypour, M., Ghavami, A. & Mofrad, M. R. K. Nucleoporin's Like Charge Regions Are Major Regulators of FG Coverage and Dynamics Inside the Nuclear Pore Complex. *PLoS One* **10**, e0143745 (2015).
80. Müller-Späth, S. *et al.* From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14609–14 (2010).
81. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8183–8 (2010).
82. Hofmann, H. *et al.* Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. **2**, (2012).
83. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13392–7 (2013).
84. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).
85. Schuler, B., Soranno, A., Hofmann, H. & Nettels, D. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annu. Rev. Biophys.* **45**, 207–231 (2016).
86. Huihui, J., Firman, T. & Ghosh, K. Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. *J. Chem. Phys.* **149**, 085101 (2018).
87. Beveridge, R. *et al.* Ion Mobility Mass Spectrometry Uncovers the Impact of the Patterning of Oppositely Charged Residues on the Conformational Distributions of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **141**, 4908–4918 (2019).
88. Martin, E. W. *et al.* Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J. Am. Chem. Soc.* **138**, 15323–15335 (2016).
89. Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **42**, 147–154 (2017).
90. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 1231–1264 (2010).
91. Uversky, V. N. Natively unfolded proteins: A point where biology waits for

physics. *Protein Sci.* **11**, 739–756 (2002).
92. Uversky, V. N. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J. Biol. Chem.* **291**, 6681–6688 (2016).
93. Uversky, V. N. Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. *Front. Phys.* **7**, 10 (2019).
94. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).
95. Srivastava, D. & Muthukumar, M. Sequence Dependence of Conformations of Polyampholytes. *Macromolecules* **29**, 2324–2326 (1996).
96. Sawle, L. & Ghosh, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* **143**, 085101 (2015).
97. Firman, T. & Ghosh, K. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *J. Chem. Phys.* **148**, 123305 (2018).
98. Yamada, J. *et al.* A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics* **9**, 2205–24 (2010).
99. Tagliazucchi, M., Peleg, O., Kröger, M., Rabin, Y. & Szleifer, I. Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3363–8 (2013).
100. Piovesan, D. *et al.* DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* **45**, D219–D227 (2017).
101. Hatos, A. *et al.* DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.* **48**, D269–D276 (2019).
102. Ma, J., Goryaynov, A., Sarma, A. & Yang, W. Self-regulated viscous channel in the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7326–31 (2012).
103. Ribbeck, K. & Görlich, D. Kinetic analysis of translocation through nuclear pore complexes. *EMBO J.* **20**, 1320–30 (2001).
104. Kubitscheck, U. *et al.* Nuclear transport of single molecules: dwell times at the nuclear pore complex. *J. Cell Biol.* **168**, 233–43 (2005).
105. Yang, W., Gelles, J. & Musser, S. M. Imaging of single-molecule translocation through nuclear pore complexes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12887–92 (2004).
106. Ghavami, A., van der Giessen, E. & Onck, P. R. Coarse-Grained Potentials for Local Interactions in Unfolded Proteins. *J. Chem. Theory Comput.* **9**, 432–440 (2013).
107. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–54 (2013).
108. Izaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **114**, 2090 (2001).
109. Beckstein, O. & Sansom, M. S. P. Liquid-vapor oscillations of water in hydrophobic nanopores. *Proc. Natl. Acad. Sci.* **100**, 7063–7068 (2003).
110. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

111. Huang, K., Tagliazucchi, M., Park, S. H., Rabin, Y. & Szleifer, I. Molecular model of the nuclear pore complex reveals a thermoreversible FG-network with distinct territories occupied by different FG motifs. doi:10.1101/568865
112. Hayama, R. *et al.* Thermodynamic characterization of the multivalent interactions underlying rapid and selective translocation through the nuclear pore complex. *J. Biol. Chem.* **293**, 4555–4563 (2018).
113. Hough, L. E. *et al.* The molecular mechanism of nuclear transport revealed by atomic-scale measurements. *Elife* **4**, (2015).
114. Ghavami, A., van der Giessen, E. & Onck, P. R. Coarse-Grained Potentials for Local Interactions in Unfolded Proteins. *J. Chem. Theory Comput.* **9**, 432–440 (2013).
115. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
116. Ghavami, A., van der Giessen, E. & Onck, P. R. Energetics of Transport through the Nuclear Pore Complex. *PLoS One* **11**, e0148876 (2016).
117. Briones, R., Blau, C., Kutzner, C., de Groot, B. L. & Aponte-Santamaría, C. GROmaps: A GROMACS-Based Toolset to Analyze Density Maps Derived from Molecular Dynamics Simulations. *Biophys. J.* **116**, 4–11 (2019).
118. Costescu, B. I. & Gräter, F. Time-resolved force distribution analysis. *BMC Biophys.* **6**, 5 (2013).