Supplemental Materials

*Molecular Biology of the Cell*

Peyro *et al.*
Supplementary Information for

Nucleoporins' exclusive amino acid sequence features regulate their transient interaction with and selectivity of cargo complexes in the nuclear pore

Mohaddeseh Peyro¹, Andrew M. Dickson¹, Mohammad R. K. Mofrad¹,²,*

¹Molecular Cell Biomechanics Laboratory, Departments of Bioengineering and Mechanical Engineering, University of California, Berkeley, CA 94720

²Molecular Biophysics and Integrative Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

*Corresponding Author: mofrad@berkeley.edu

Abbreviations:

lpLCR: largest positive Like charge region
NPC: Nuclear Pore Complex
FG Nup: FG Nucleoporin
TRFDA: Time-resolved force distribution analysis
Further Statistical Analysis on the distribution of force on the FG Nups

Principal component analysis is a mathematical method, which reduces high dimensional data to a representative, but lower dimensional, subspace. For each of our experiments on wildtype, positive control, and negative control proteins, our force data took the form of a list of timestamped vectors, with the vector entries representing punctual stress on protein residues.

PCA attempts to represent all vectors in our dataset as a combination of their average and a weighted sum of a few specifically chosen principal component vectors, which are designed to capture as much of the natural variation as possible. Because combinations of just a few principal components represent data points very effectively, they can be thought of as modes of variance that capture the ways in which data can deviate from the mean. We can produce a low-dimensional visualization of our data by replacing our high-dimensional values for force with the weightings on a handful of ‘modes’, under the premise that these weightings capture most useful information.

PCA Principal Components

PCA results on the individual punctual force datasets generated by wildtype, positive control, and negative control proteins show some interesting results (Figure S8). We plotted the two most significant principal components for each set of force vectors to demonstrate differences within the data.

We see that the principal components typically have a large emphasis on the lpLCR, meaning that residues within the lpLCR frequently experience forces much greater or smaller than the mean, and that the forces on these residues are usually similar. For most components, there is also little weight on the non-lpLCR region, indicating that the forces in this region deviate little from the mean and they are relatively uncorrelated with those in the lpLCR region. This corresponds with the premise that lpLCR residues are generally in closer contact with each other and experience more significant forces from their interactions, and that they are relatively uncoupled from the remainder of the protein.

However, for negative control proteins, we do see that this relationship is disrupted in various ways. Without the positive residues, we see components showing highly disrupted patterns for Nup 49, and much higher variance for non-lpLCR regions for Nup 42.

PCA Embedding Scatterplot

Used on a combination of our three experiments, PCA principal components most strongly weight the residues that dramatically change behaviors between experiments, because these residues account for the most variance in the joint datasets. The 2D embeddings (Figure S9) generated from these components show how the forces on these residues change as the proteins are modified.

Similar to results of Figure S8, the principal components indicate that the most significant forces are within the lpLCR region for all datasets. As shown in the right sides of Figure S9 A, B&C, points from the wildtype simulation generally experience more force along
the first principal component than those from the positive control simulation, while points from the negative control simulation are the furthest to the right, meaning that they experience the highest force. These results demonstrate that positive residues help “buffer” the lpLCR region, and lower both the variability and magnitude of punctual force within it.

Our statistical analyses in Figure S8 and Figure S9 support our interpretations based on visual illustrations in Figure 3 and Figure S7. The residues in the lpLCR domain experience a higher magnitude of force compared to non-lpLCR domain, and positive residues within the lpLCR stabilize force interactions. The reason is that lpLCR domains are more aggregated than non-lpLCR domains and Phe residues interact more strongly in the aggregated domains. On the other hand, the lpLCR domains experience different magnitudes of force in wildtype and mutant cases, with negative control experiencing the highest force and positive control experiencing the lowest. This is directly correlated with the level of aggregation in these molecules. As a result, we can conclude that the number of charged residues in the lpLCR domain regulate the force distribution within FG Nups.”
Figure S 1: Density maps of A) Nups and B) cargo C) FG motifs D) hydrophobic spots on the cargo complex, inside ring of Nup42. The plots show wildtype, negative control and positive control Nup42 simulation results from left to right. Nup42 has a very short charge rich domain (residues 318-382) and a long lpLCR domain (residues 1-317). The lpLCR domain contains 10 positively charged residues. The lpLCR domain forms a collapsed coil conformation and the non-lpLCR domain forms a relaxed coil conformation. The collapsed lpLCR domains interact with their neighboring Nup42 molecules (see part A). In the negative control, the maximum density of Nup42 molecules increases by 100%, while in positive control, this density decreases by 50%. In negative control, the movement of the cargo becomes less even compared to wildtype (see part B). 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.
Figure S2: Density maps of A) Nups and B) cargo C) FG motifs D) hydrophobic spots on the cargo complex, inside ring of Nup145. The plots show wildtype, negative control and positive control in Nup145 simulation results from left to right. Nup145 has an LpLCR domain of 217 residues, containing 7 positively charged residues and a non-LpLCR domain of 216 residues. (see Figure 1). The LpLCR domain forms a collapsed coil conformation and the non-LpLCR domain forms a relaxed coil conformation. The collapsed LpLCR domains interact with their neighboring Nup145 molecules (see part A). In the negative control, the maximum density of Nup145 molecules increases by 30%, while this density decreases by 25% in positive control. Also, in negative control, the movement of cargo becomes less even compared to wildtype, while it becomes more even in positive control (a higher density at the center means more frequent traveling of the cargo complex between FG Nups) (see part B). 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 S 3. Density maps of A) Nups and B) cargo C) FG motifs D) hydrophobic spots on the cargo complex, inside ring of Nup49. The plots show wildtype, negative control and positive control Nup49 simulation results from left to right. The length of the disordered domain of Nup49 is 251 amino acids. This protein contains seven positively charged residue and zero negatively charged residues. Therefore, the entire molecule is an lpcLCR, 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, the maximum density of Nup49 molecules increases by 10%, while it decreases by 20% in positive control. In negative control, 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, density map of cargo shows high density at the center (see part B), which implies that the cargo complex is spending most of its 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.
Figure S4: Density maps of A) Nups and B) cargo C) FG motifs D) hydrophobic spots on the cargo complex, inside ring of Nup57. The plots show wildtype, negative control and positive control Nup49 simulation results from left to right. 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 lpLCR, 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, the maximum density of Nup57 molecules increases by 10% while in positive control, this density decreases by 20%. In negative control, 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, density map of cargo shows high density at the center (see part B), which implies that the cargo complex is spending most of its 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 S5: 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 IpLCR, only the wildtype case is analyzed. This Nup is analyzed as a natural control for other Nups that do contain an IpLCR. The length of the disordered domain of Nup1 is 856 amino acids and its entire length is a charge-rich domain. Therefore, this Nup does not feature an IpLCR. 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. The density map of cargo hydrophobic spots (D) shows a similar pattern to that of cargo.
Since this Nup does not feature an lpLCR, only the wildtype case is analyzed. This Nup is analyzed as a natural control for the other Nups that do contain an lpLCR. The length of the disordered domain of Nup60 is 539 amino acids and its entire length is a charge-rich domain (see Figure 1). Therefore, this Nup does not feature an lpLCR. 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.
**Figure S 7: 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 lpLCR. In wildtype and the two mutants, same residues experience a high punctual stress, but the magnitude of punctual 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 punctual stress and positive control has the lowest punctual stress and wildtype is in the middle of the two. The difference in punctual 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 punctual stress, while positive control has the least aggregated/collapsed conformation and the residues are experiencing the least punctual stress, and wildtype is between the two. As can be seen both in the heatmaps and the time-averaged graphs, the spikes of punctual stress happen at the same residues for wildtype and mutants. This implies that mutating the number of charged residues in the lpLCR domain does not change the network of interacting residues, it only affects the intensity of interaction.
Figure S 8. First and second principal components of absolute force data from several simulations. X-axis represents residue # while y-axis represents weighting of force on residue in principal component. High values for a particular residue indicate that there is significant force variability on that residue. A) Principal components for Nsp 1. The principal components for positive, negative, and wildtype experiments are generally similar, with force variability predominantly in the IpLCR region, and forces being more evenly distributed with more positive residues. B) Components for Nup 42. Again, principal components focus on the IpLCR region (residues 0 through 251). C) Components for Nup 49. The entire protein is an IpLCR, which is reflected by the existence of variable forces throughout the protein. For positive control and wildtype forces are significantly less variable in the center, while they are completely disrupted in the negative control.
Figure S 9: PCA embeddings of wildtype, positive control, and negative control for Nsp 1, Nup 42, and Nup 49. Left column) First and second principal components of combined absolute force data from wildtype, positive control, and negative control simulations. Again, x-axis represents residue while y-axis represents weighting of force on residue, and most variability in force is within the IpLCR region. Right column) Projections of a subset of timesteps onto the first two principal components. A) Principal components are largely the same when positive, negative, and wildtype datasets are combined. Again, we see that the Nsp 1 experiments are harder to differentiate, but that as a general rule forces are lower and more even in the positive control, and highest in the negative control. B&C) Principal components and projections for Nup 42 and Nup 49. Data is clearly separable, and shows a natural differentiation between the experiments. Removal of positive residues for negative control produces widely spread cloud with especially high forces in IpLCR regions, while addition of positive residues lowers variability of forces.
Figure S10. Top view of density map of different layers of FG Nups in whole-NPC model. Density threshold is set to 3% of maximum density. 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.

Table S1: Average radius of gyration of Nsp1 and Nup42 molecules

|                  | Nsp1 (lpLCR domain) | Nup42 (lpLCR domain) |
|------------------|----------------------|-----------------------|
| Wildtype         | 10.60 nm²            | 10.00 nm²             |
| Negative control | 6.98 nm²             | 6.42 nm²              |
| Positive control | 13.10 nm²            | 19.9 nm²              |

Table S2: Force / Inter-distance Correlation Data

|                  | Wildtype | Negative Control | Positive Control |
|------------------|----------|------------------|------------------|
| Nsp1 (lpLCR domain) | -0.31    | -0.362           | -0.286           |

Video S1. Nup42 wildtype ring cross section simulation trajectory

Video S2. Nup42 negative control ring cross section simulation trajectory

Video S3. Nup42 positive control ring cross section simulation trajectory