Nuclear Pore Complex Protein Sequences Determine Overall Copolymer Brush Structure and Function

David Ando,[†] Roya Zandi,[‡] Yong Woon Kim,[§] Michael Colvin,[¶] Michael Rexach,[||] and Ajay Gopinathan[†*]
[†]Department of Physics, University of California at Merced, Merced, California; [‡]Department of Physics, University of California at Riverside, Riverside, California; [§]Graduate School of Nanoscience and Technology, Korea Advanced Institute of Science and Technology, Daejeon, Korea; [¶]Department of Chemistry and Chemical Biology, University of California at Merced, Merced, California; and [||]Department of Molecular, Cell, and Developmental Biology, University of California at Santa Cruz, Santa Cruz, California

ABSTRACT The transport of cargo across the nuclear membrane is highly selective and accomplished by a poorly understood mechanism involving hundreds of nucleoporins lining the inside of the nuclear pore complex (NPC). Currently, there is no clear picture of the overall structure formed by this collection of proteins within the pore, primarily due to their disordered nature. We perform coarse-grained simulations of both individual nucleoporins and grafted rings of nups mimicking the in vivo geometry of the NPC and supplement this with polymer brush modeling. Our results indicate that different regions or blocks of an individual NPC protein can have distinctly different forms of disorder and that this property appears to be a conserved functional feature. Furthermore, this block structure at the individual protein level is critical to the formation of a unique higher-order polymer brush architecture that can exist in distinct morphologies depending on the effective interaction energy between the phenylalanine glycine (FG) domains of different nups. Because the interactions between FG domains may be modulated by certain forms of transport factors, our results indicate that transitions between brush morphologies could play an important role in regulating transport across the NPC, suggesting novel forms of gated transport across membrane pores with wide biomimetic applicability.

INTRODUCTION

A particular membrane that sees a lot of cross-membrane traffic is the nuclear envelope in eukaryotes. The nuclear pore complex (NPC) is an important macromolecular structure that spans the nuclear envelope and regulates all nucleocytoplasmic traffic, including important processes such as the import of regulatory proteins from the cytoplasm and the export of RNA from the nucleus (1–5). It is a supramolecular structure composed of ~30 different types of NPC proteins (nups) (2). A subset of the nups forms a structural ring or cylindrical-like structure that is embedded in the nuclear envelope. This structure forms an open aqueous channel, ~50 nm in diameter, that connects the nucleoplasm and cytoplasm (Fig. 1). A second subset of nups, those which form the focus of this article, are grafted along the interior wall of the aqueous channel and are responsible for forming a selective diffusion barrier. Containing many phenylalanine-glycine (FG) repeats, these nups are called FG nups (2) and are structurally unique due to their large FG repeat domains, which are highly flexible and behave natively as polymerlike unfolded proteins. Hence, these FG nups are referred to as being intrinsically disordered or natively unfolded proteins (6). FG repeats further separate into different types, with commonly studied GLFG and FxFG repeat motifs known to have different binding affinities (7). In common Baker's Yeast *Saccharomyces cerevisiae*, there are ~150 copies of FG nups in each NPC and it has been hypothesized that the NPC pore is occupied by dozens of FG nups that interact with each other weakly via hydrophobic attractions to form a network that functions as a semipermeable diffusion barrier (see Fig. 1 *A*). This barrier maintains a tight seal against cytoplasmic particles larger than ~4 nm while allowing smaller particles to passively diffuse through. However, amazingly, it also allows the facilitated transport of specially tagged particles up to 40-nm diameter (8), at rapid speeds of 5–20 ms per transported cargo for even large mRNP complexes (9). Additionally, the NPC does not exert any forces or expend any energy, with the only known free-energy consumption being limited to maintaining a gradient associated with the transport factors that generate the directionality of transport.

Despite the NPC's key role in biology and numerous studies on the structure and properties of individual nucleoporins (1–3,6,10–13), the actual structure of the polymer complex within the nuclear pore and its mechanism of operation are still under debate. For example, prominent models of nuclear pore transport, such as the selective phase/hydrogel (14) and virtual gate/polymer brush (15) models, assume very different morphologies for the polymer complex filling the nuclear pore (Fig. 1). The hydrogel model predicts that FG domains interact via hydrophobic amino acids to form a filamentous meshwork that physically blocks protein diffusion whereas the polymer brush model predicts that FG domains have limited hydrophobic interactions and instead behave as bristles that form an entropic gate at the NPC that blocks protein diffusion.

Most theoretical and polymer physics approaches (16–19) to this problem typically tend to assume a homogenous structure for individual FG nups, resulting in a

Submitted December 9, 2013, and accepted for publication March 11, 2014.
*Correspondence: agopinathan@ucmerced.edu
Editor: David Sept.

0006-3495/14/05/1997/11 $2.00
http://dx.doi.org/10.1016/j.bpj.2014.03.021

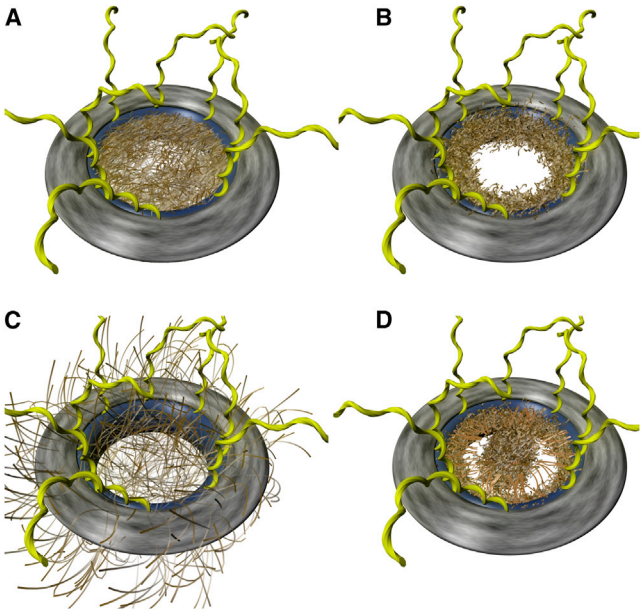

FIGURE 1 Illustration of NPC models. Each NPC depicted spans the double lipid bilayer nuclear envelope, oriented such that the tops of the pores face the cytoplasm, while the bottoms of the pores face the nucleus. (*A*) FG nups form a hydrogel, which transport-factors and their cargo complexes transit by binding with FG motifs and temporarily disrupting cross-links in the FG nup meshwork (14). (*B*) FG Nups are collapsed and lie along the wall of the NPC, creating a surface to which transport-factors bind and diffuse along in a dimensionally reduced manner (36). (*C*) FG nups form a polymer brush that binds transport factors (possibly collapsing in their presence) but excludes unwanted molecules that do not interact with the FG nups (15). (*D*) Individual FG nups can have collapsed coil gel-like regions and extended coil brushlike domains, resulting in a microphase separation of these domains within the NPC. A central pluglike structure and another dense shrub region, which lies along the wall of the NPC, are separated by a polymer brush of extended disordered regions (stalks) of FG nups (26). To see this figure in color, go online.

relatively homogenous NPC architecture. There have also been recent attempts to perform coarse-grained simulations of NPC proteins and the entire transport process through the NPC (20–23), but these too suffer from either the intrinsic drawback of assuming that the NPC proteins are homogeneous, or of overreliance on simple amino-acid characterizations and assumptions, which could have left the ultimate underlying molecular architecture of the NPC unresolved. In this article, we attempt to overcome these issues by using a more reductionist approach starting from FG nups sequences combined with a detailed coarse-grain (CG) model that can reproduce secondary and tertiary structure of proteins. We show that different regions or blocks of an individual NPC protein can have distinct, quantifiably different forms of disorder and that this property appears to be a conserved functional feature. The unique diblock polymer structure of the longer FG nups, critical for transport, under certain conditions result in a mesoscale polymer brush structure with a dense FG domain core surrounded by brush of disordered proteins with a high concentration of charged amino acids. To understand the physics behind the emergent structure we developed a simple polymer brush model based on the properties of individual nups revealed by our CG model. Given that the interactions holding the dense FG domain core together are likely be modulated by certain transport factors, our polymer brush modeling suggests that the approach of certain forms of cargo could alter the local brush structure, to allow insertion and regulated transport.

METHODS

The CG model of Hills et al. (24) was run using the large-scale atomic/molecular massively parallel simulator (LAMMPS) (25) software suite. Using the three separate calibrated CG models with α, β, and γ equal to 3.35, 4.30, and 5.26, respectively, at 300 K, simulated polypeptide chains representing the disordered regions of various FG nups from *S. cerevisiae* and humans were equilibrated for 1 μs, using starting conditions that consisted of fully extended protein configurations for the simulations without the cylindrical boundary conditions. Disordered regions of FG nups that were simulated used the nup-specific definitions defined in Yamada et al. (26) and the Supporting Material. For the porelike simulations with cylindrical boundary conditions, initial nup configurations (identical sequences as in the individual nup simulations) were as depicted in the Supporting Material. During the 1-μs equilibration, all individual FG nups collapsed from their initially fully extended configuration in <300 ns. After equilibration, data were then taken from a 4-μm production run. Postsimulation analysis was done to determine the average radius of gyration of different domains as defined in Yamada et al. (26). Contact proximity between two given side-chain beads was calculated by counting the number of times side chains were within 16 Ångstroms of each other in the trajectory snapshots (saved every 100 ps during the 4 μs of production simulation), then normalizing by the number of snapshots. Initial starting conditions for the ring simulations can be seen in the Supporting Material. Heat maps of the density of FG motifs and charged amino acids in the FG nup with the most FG repeats across 10 different species were only applied to sequence regions that were predicted to be disordered via a PONDR-FIT score >0.5 (27) and to have >40 amino acids. The resulting binary densities of FG motif/amino acids (aa) and charged aa/aa were smoothed by a running average over a 20-aa window.

RESULTS

Coarse-grained simulations

Although all atom molecular-dynamics models can accurately reproduce the local secondary structure and nature of tertiary contacts for proteins, it is not computationally feasible to study even a single FG nup's properties due to their large size (~1000 aa) and the long timescales involved (in μs). Characterizing the overall structure formed by a hundred-odd FG nups becomes next to impossible. To overcome these issues we use a CG model (24) that is detailed enough so that it is able to reasonably reproduce secondary and tertiary structure and even fold small proteins ab initio starting from extended states, but simple enough such that multiprotein systems over microsecond timescales are computationally feasible. This CG model attempts to reproduce the energy landscape of real proteins by representing each amino acid by multiple CG interaction sites and by using specially developed interactions between these sites.

The physical potentials between these CG interaction centers is derived from measuring the forces present in highly accurate all-atom simulations of diverse protein sequences, using the multiscale coarse-graining method (28). This method makes no assumptions about the type of interactions between CG sites, and additionally implicitly includes multibody correlations (29) in the resulting implicit solvent model with an effective CG potential. In this CG model, secondary and tertiary structure emerges naturally through the comprehensiveness of the physical model, and it is general enough to simulate proteins of arbitrary amino-acid sequence.

Typically, CG potentials for a specific protein simulation are scaled such that model results match well with experiment (24,30). This was achieved by multiplying the entire CG potential by a constant scaling factor. Experimentally measured properties of disordered FG nups have yet to converge, therefore we developed a number of CG models with different scaling factors (α, β, and γ), with each scaling factor corresponding to a different class of experimental results. FRET measurements of human nup153 by Milles and Lemke (31) have shown the disordered regions of this protein to consist of nearly entirely compact collapsed-coil configurations. The CG model α was therefore developed to mimic these results, scaled such that it produces a collapsed coil (molten globule) (32) for the disordered region of nup153. (Table S1 compares model and experimental radii in the Supporting Material). Other experiments have shown nup153 to take on more extended configurations, forming relatively extended brush configurations that undergo reversible collapse in the presence of transport factors (33). CG model γ was therefore developed to mimic these results, scaled such that it produces a relaxed coil (32) for the disordered region of nup153 (see Table S1). Bead-halo experiments by Yamada et al. (26) indicate nups are biphasic, with stalk domains noninteracting in general whereas, on the other hand, FG domains are sticky to other FG domains. FG nups were also found to be biphasic in terms of polymer structure, with stalk domains having expanded polymer conformations, whereas FG domains had more-compact polymer conformations. A scaling factor β precisely in the middle of the α- and γ-scalings (Methods) was found to produce such biphasicness, with this scaled CG model labeled β.

The FG nups we simulated were taken from evolutionary distant *S. cerevisiae* (Nsp1, Nup1, Nup100, and Nup116) and *H. sapiens* (Nup98 and Nup153), focusing on long FG nups with large disordered regions which have a biphasic sequence structure that can be separated into an FG domain and stalk domain. An FG domain corresponds to a sequence region with a low concentration of charged amino acids and a high concentration of FG motifs, whereas the stalk domain corresponds to a disordered sequence region with a high concentration of charged amino acids (nup specific amino-acid domain definitions and disordered domains used are defined in the Supporting Material and Yamada et al. (26)).

Structure and dynamics of single nups

Polypeptide chains representing the disordered regions of various FG nups from *S. cerevisiae* and humans were started from an initially fully extended configuration and equilibrated before the start of simulation. Fig. 2 shows a typical snapshot of one such nup, Nsp1, after collapse and equilibration. The snapshot seems to indicate that these proteins can exhibit distinctly different kinds of polymeric structure along a single chain with regions that look like swollen chains to other parts that resemble collapsed coils. In general, the proteins seemed to adopt a block polymer structure to varying degrees, dependent on the form of CG model employed. Amino-acid regions of the FG nups simulated and domain definitions used are in Table S2.

To quantitatively characterize FG nup structure, we define the contact probability between a pair of amino acids as the percentage of time the corresponding CG backbone beads are within 1.6 nm of each other. The diblock polymer structure of these FG nups is particularly clear from the contact probability maps showing the contact probability over all possible amino-acid contact pairs (Fig. 3 and see Fig. S12 in the Supporting Material). In many cases these show a universal block diagonal structure with a part that is almost diagonal—representing an extended stalk and a square block that is uniformly lit up, indicating a collapsed disordered structure—the tip. The lack of contacts between the blocks

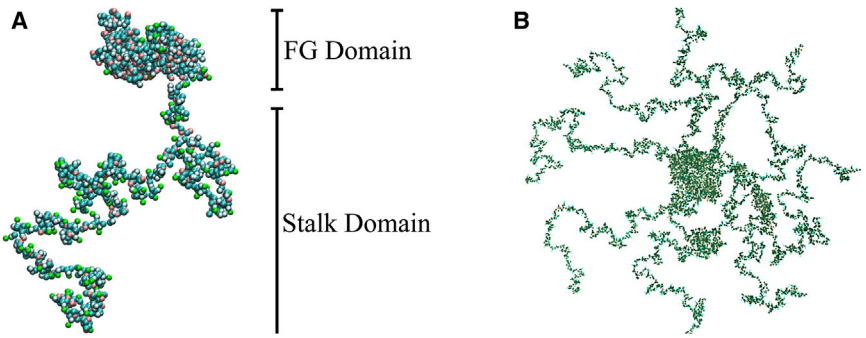

FIGURE 2 (*A*) Simulation snapshot of full length Nsp1 showing the biphasic FG domain and stalk domain structures in the α-model. Snapshots for the other models β and γ are shown in Fig. S7 in the Supporting Material. (*B*) Snapshot of a ring of eight Nsp1s in the α-model. To see this figure in color, go online.

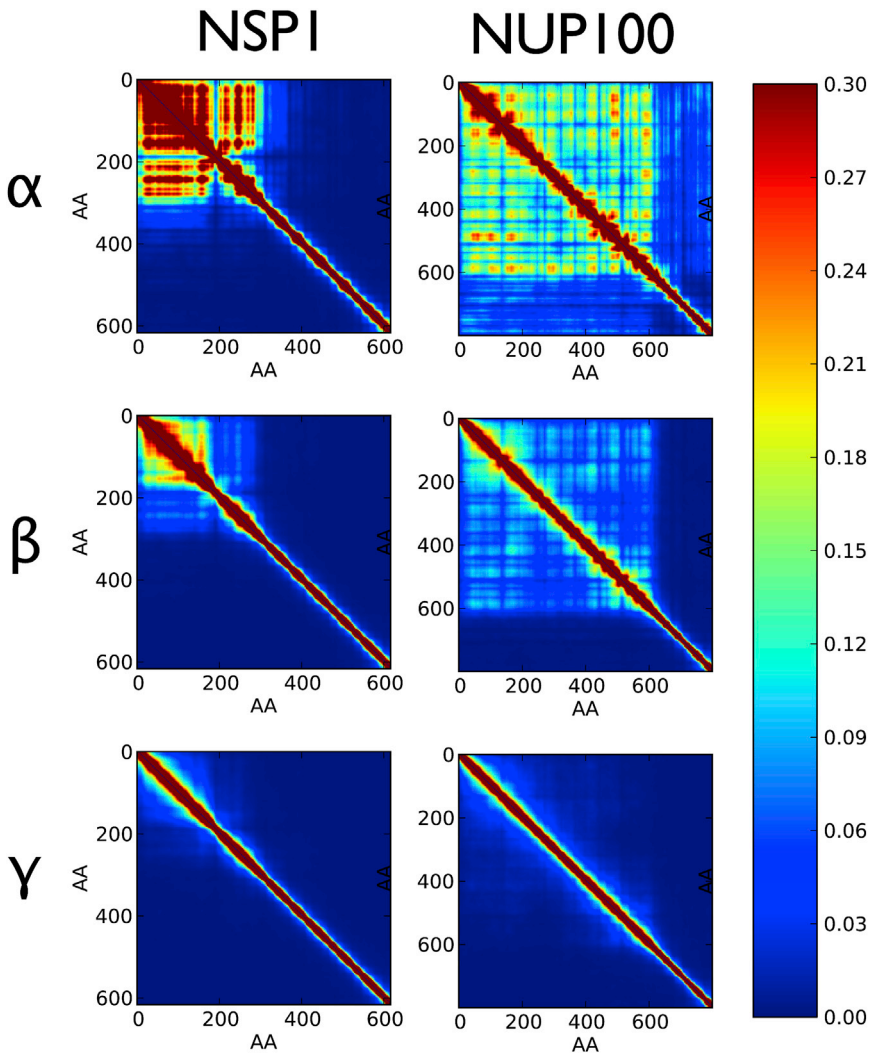

FIGURE 3 Contact maps for the different CG models α, β, and γ for Nsp1 and Nup100. Maps for Nup1, Nup98, Nup116, and Nup153 can be seen in Fig. S12 in the Supporting Material. Contact probability maps show the time-averaged contacts between all pairs of amino acids. A block diagonal structure is noticeable in numerous contact maps, with one block for FG domains and diagonal contacts for stalk domains. A diblock structure can be most strongly seen in FG nups simulated under scenario β. Model α shows some block structure but often produces significant contacts between FG and stalk domains, because the two domains often interact. In striking contrast to models α and β, model γ produced nups with little difference in contact probability between FG and stalk domains with the FG nups representing unstructured extended homogenous polymers. Amino-acid residues shown are with respect to the disordered domains of the simulated FG nups, whereas full protein amino-acid indexes can be determined by domain definitions in Table S2. To see this figure in color, go online.

indicates the blocks maintain their distinct identities over time. The diblock structure was most strongly seen in model β. Model α often produced significant contacts between FG and stalk domains that destroy diblock structure because the two domains often interact, whereas model γ produced nups with little difference between FG and stalk domains with the entire FG nups representing unstructured extended homogenous polymers.

Under all model assumptions, individual FG nups were highly dynamic and disordered. FG domains across all simulated nups maintained a relatively smooth block contact structure, with contacts between amino acids rapidly exchanging and no folded constant contacts visible. The consistency in the diblock structure that is exhibited across the different nups indicates that there must be some underlying general features in the sequence structure that is maintained. In particular, this would suggest that disruption or randomization of the sequence would result in a loss of such structure. To test this, we repeated the simulations with a shuffled version of the amino-acid sequence of Nsp1, which resulted in the loss of diblock structure (see Fig. S1). This clearly suggests that assuming nup sequences have an effectively homogeneous character could be problematic and that the biphasic sequence structure is responsible for producing two distinct types of dynamic domains via weak but specific interactions among portions of FG nups. The sequence features of the FG and stalk domains delineating these domains are discussed in the Bioinformatics subsection later in this article and in Ando et al. (34).

An earlier hypothesis that FG nups can form a treelike structure was based on the experimentally measured properties of fragments of FG nups from *S. cerevisiae* (26). Here, our simulation results on full-length FG nups support this hypothesis by demonstrating that it is clearly possible that individual fragment properties combine to form true diblock copolymers. Additionally, the CG model used here is, to our knowledge, the first such model capable of reproducing the observed interactions (26) between different fragments from different regions of the FG nups.

To quantitatively characterize the levels of disorder within distinct regions of the contact probability maps, we derived looping probabilities from the maps for different domains of the FG nups. We define the looping probability $P_l(s)$ to be the probability that two monomers of size b, separated by a contour length of sb along the amino-acid chain, are within a spatial distance $r < R_{cut}$. In our analysis, $R_{cut} =$ 1.6 nm. For two monomers of an ideal chain, the looping probability for a given monomer separation distance, sb, is given by the integral of the probability distribution of end-to-end distances up to R_{cut},

$$P_l(s) = \int_{r<R_{cut}} dV \left(\frac{3}{2\pi sb^2}\right)^p \exp\left(\frac{-3r^2}{2sb^2}\right), \qquad (1)$$

with exponent $p = 3/2$ (35). An extended chain would have a power of $p = 9/5$ and a collapsed coil would have a $p = 1$, with appropriately scaled constant prefactors. Additionally, we quantitatively determined the statistical significance of the block structure of contact maps in the γ-model as they seem to disappear in this model. At a 40-amino-acid separation distance, we measured the average contact probability and standard deviation (SD) between all possible amino-acid combinations for the shuffled Nsp1 sequence that is shown in the Supporting Material. Measuring the same 40-amino-acid separation contact probabilities for the γ-model contact maps, we determined that the average of this contact probability over all contact maps in this model was only 1.02 SDs away from the corresponding value for the unstructured shuffled Nsp1 sequence. In contrast, the same contact probability averaged over the β-model contact maps was 3.58 SDs away from the shuffled Nsp1 sequence value.

Fig. 2 shows a log-log plot of the average measured looping probability as a function of s for the all FG and stalk domains simulated, with a numerically fitted theoretical looping probability (for the power p) over the amino-acid separations >5 and <65 monomers. Qualitatively, a high looping exponent power p for a particular domain would imply that that part of the polymer is relatively extended spatially, because the contact probability falls off more rapidly as s is increased within that domain.

As can be seen in Fig. 2, FG domains have looping exponents significantly lower than those of stalk domains, indicating the constantly large difference in ensemble structure between these two types of domains. For the α, β, and γ CG models, the average fitted scaling exponents for the FG domains were 1.16, 1.38, and 1.81, respectively. For the stalk domains, the average fitted scaling exponents were 2.08, 2.58, and 2.82 for the different models, respectively. We also note an important caveat that these simulations of individual FG nups by themselves were done ignoring the effects of the crowded pore environment (36) and other FG nups. Although we analyze aggregates of interacting FG nups in the next section, our simulations do not include the crowding effects from transport factors, cargo, and other molecules in the cellular milieu that can significantly modulate FG nup behavior (37).

Structure and dynamics of nup rings

To look at the structures formed by multiple interacting nups in vivo, we performed simulations of rings of Nsp1 and Nup100 (identical sequences as in the individual nup simulations) grafted in an eightfold symmetric manner to the inner surface of a neutral, nonreactive cylinder, with the grafted-end locations chosen in a manner consistent with the grafting orientation and FG nup anchor domain positions within the nuclear pore. For a single ring of FG nup trees, we chose the dimensions of the enclosing cylindrical boundary to be (5 nm axial length, 50 nm diameter), representative of the geometry of the NPC as shown in Fig. S4 (26).

Fig. 4 shows the total mass density of the simulated rings of Nsp1 FG nups. The mass density clearly reveals a dense central plug connected by peripheral cables to the pore walls in the α- and β-models, with a plug mass density which, in all scenarios, is much smaller than the mass density of folded proteins of ~1400 mg/cm^3 (38). The mass density can be tied directly to the microphase separation of the different blocks of the natively unfolded proteins. One block phase (the FG domains) separated along the center of the channel whereas the other block (the stalk regions) aggregated along the periphery (see Fig. S13). This microphase separation could explain the observation of a pluglike structure in cryo-electron microscopic examination of NPCs (39). This pluglike structure is currently not recognized as a structural element of the nuclear pore complex and is instead rationalized as an artifact of cargo in transit (40). Interestingly, in the γ-model, which represents an extended coil scenario, there is no microphase separation present—reminiscent of the virtual gate model.

Spatial net charge density of the ring of Nsp1 nups mirrors closely the mass density of the FG domain, because FG domains are in general positively charged (34). This produces a striking effect in the α- and β-models, with a strongly positively charged plug region forming that has a net charge density much higher than considered earlier by Colwell et al. (41) and Tagliazucchi et al. (23), with the same possible functionality but much greater attraction to negatively charged transport factors due to radial concentration toward the center of the NPC channel.

In contrast to nups like Nsp1 and Nup1 that have a long stalk region and comparatively small sticky tip, other FG nups, like Nup100 have large sticky tips and comparatively short stalk regions. We simulated rings of Nup100 in an identical manner to Nsp1, resulting in similar results except for a key difference that Nup100 rings in isolation are unable to close in the α- and β-models, unlike Nsp1

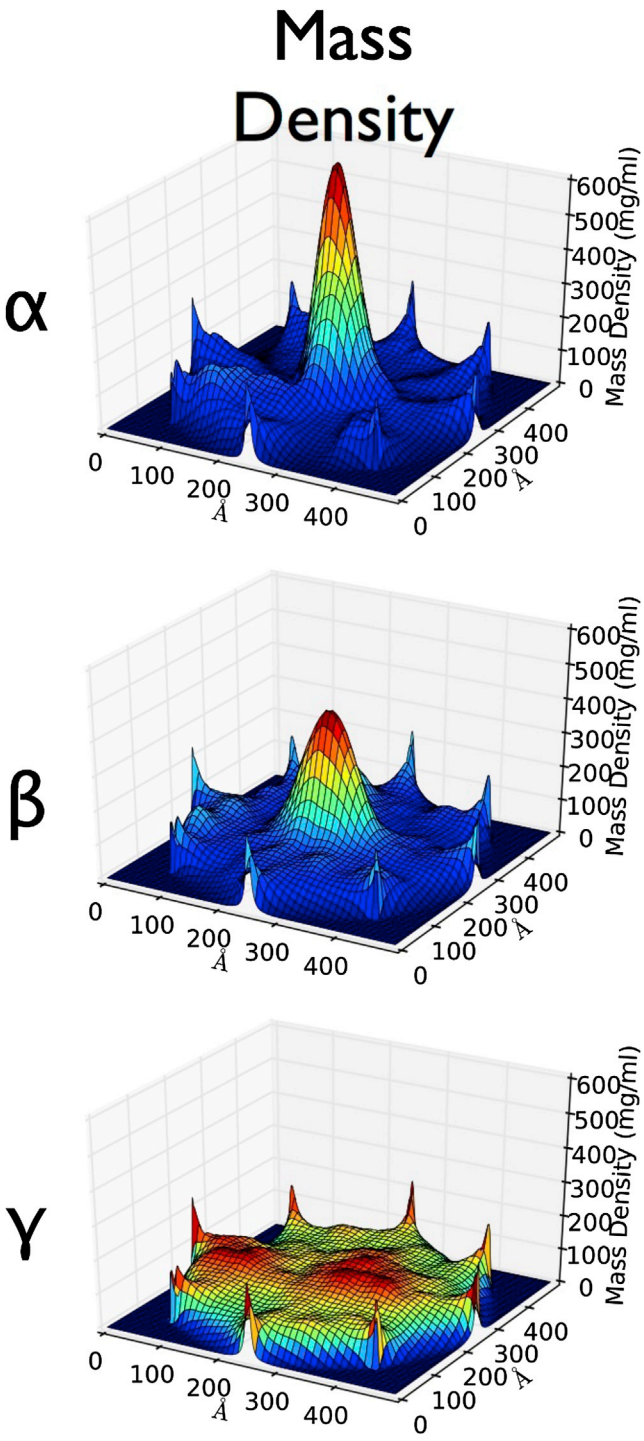

FIGURE 4 Density maps of a simulated pore containing eight Nsp1 FG nups. Plotted for all three models is the total mass density (mg/mL) along the pore. The total mass density clearly reveals a dense central plug connected by peripheral cables to the pore walls in the α- and β-models, whereas the γ-model produces a homogenous extended brush. Plots of the mass density of FG domains, the mass density of stalk domains, and the net charge density can be seen in Fig. S13 in the Supporting Material. To see this figure in color, go online.

which can close the pore in these scenarios (see Fig. S5). In the next section, we use polymer brush modeling to understand the structure of Nup100 rings in the presence of closed Nsp1 rings, and find a cooperative effect where Nup100 rings will close in the presence of closed Nsp1 rings.

We did not observe any rigid contacts from the ring simulations that lasted for a time in excess of 1 μs, so we rule out the existence of amyloid or rigid hydrogel structures that have contacts on these timescales in the β-model used. This can also be seen by the high similarity of the contact map for an individual free Nsp1 (Fig. 3) and the contact map for a grafted Nsp1 (see Fig. S2) in the ring conformation. Even in the compact α-model, the dense FG domain central plug contains individual FG domains that maintain their disordered state and rapidly exchange contacts to form block diagonal contact maps for the full Nsp1 nup.

Polymer brush morphology and dynamics

To understand the underlying physics that governs the formation of different types of mesoscale structures and possible transitions between them, we turn to polymer physics modeling. Here we introduce a simple polymer brush model that treats an idealized polymer brush constructed of diblock polymers representing the FG nups. We assume that the nups can be described by neutral excluded volume polymers (representing the stalk domains) with an appropriate length and stiffness, and grafted to the inside of a cylinder of diameter equal to that of the nuclear pore. We then incorporate the biphasic structure of the nups in the α- and β-models by modeling the ends of the polymers as structureless cohesive blobs that correspond to the cohesive FG domains at the nup tips. The resulting configuration is a polymer brush as shown schematically in Fig. 5 *A*.

The question now becomes: What is the height H of this brush? Do the nups extend all the way into the center of the pore and fill the space, or are they collapsed along the sidewalls or do they form a brush with an intermediate height that still leaves an open transport conduit along the middle of the pore? The equilibrium height of this brush can be derived from a minimization of the free energy of the brush, which has contributions coming from the entropic stretching of the stalks, the excluded volume interactions (steric hindrance) between the stalks, and the cohesive energy interactions between the tip blobs. Even within this simple model, the curvature of the grafting surface and the cohesive tips result in the brush height being given by the solution to a nonlinear differential equation. We can, however, make another mean-field approximation to arrive at an algebraic form for the effective free energy per chain (see the Supporting Material)

$$\frac{F}{k_BT} \sim (H/a)^{5/2}N^{-3/2} + \frac{a^{5/2}N^{3/2}RH^{-1/2}}{d^2(2R-H)} - \frac{\epsilon\delta^2 R}{d^2(R-H)}, \quad (2)$$

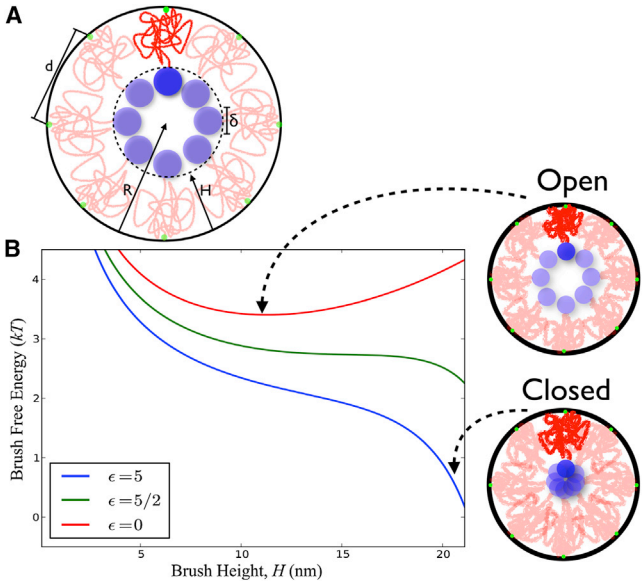

FIGURE 5 (*A*) Schematic of a polymer brush structure formed by diblock FG nups. Parameters H, height of the brush; R, radius of the pore; δ, diameter of the sticky tips; and d, the average distance between anchor points. (*Green circles*) Locations at which FG nups are grafted to the pore. (*B*) Free energy of the Nsp1 brush as a function of brush height for various values of the blob cohesive energy (ϵk_BT). Brush height can extend to a maximum of ~22 nm, which is the radius R of the modeled pore minus the size of the sticky tips. (*Right*) Schematic diagram of the proposed diblock copolymer brush gate model at various minima of the brush free energy. When particular transport factors are present that are able to outcompete the inter-FG domain sticky-tips interactions, the brush is able to open up to a new free energy minimum that can accommodate the cargo. When interactions between sticky tips are able to recover into the several kT range, the pore is able to close with a free energy minimum at $H{\sim}R{-}\delta$. We estimated the self-interaction energy level of the Nsp1 sticky tip to be 4.7 kT (see the Supporting Material), which also sets the energy scale for blob-blob interactions of ϵkT. To see this figure in color, go online.

where ϵk_BT is the effective cohesive energy between a pair of blobs. The first two terms of the free energy follow directly from prior work on polymer brushes in cylinders (42,43).

To begin, we specifically choose one of the Nups to model: Nsp1. The extended stalk domains are modeled by excluded volume polymer chains with sizes as measured from the β-simulation, which determines the Kuhn length a and effective polymerization number N (see the Supporting Material). The cohesive domains at the tip are expected to be in a small, compact, collapsed state. We model these as featureless sticky blobs of size (diameter) $\delta = 3.6$ nm, twice the radius of gyration as measured in the β-CG model simulations. We assume that the average spacing between grafting points of Nsp1 along the pore wall is $d \sim 10$ nm, which we estimate by assuming that there are 32 copies (3) of Nsp1 anchored along the inner wall in a symmetric fashion along the geometry used in Yamada et al. (26) (see Fig. S4). The pore radius is taken to be $R \sim 25$ nm.

Fig. 5 *B* shows a plot of the free energy as a function of the brush height for different values of the cohesive energy with all the other parameters being set by the values for Nsp1 that were detailed above. The minima in these curves correspond to equilibrium values of the brush height. For no cohesion ($\epsilon = 0$) we see that the equilibrium actually corresponds to a brush height that is approximately one-half the pore radius, implying that if cohesion between the tips is lost, there will be a wide-open channel down the center of the pore. We call this the open state of the gate. As the cohesive energy is increased, we see a new minimum developing at $H{\sim}R{-}\delta$, which is the state where the stalks are extended so that the cohesive blobs are at the center, thus completely closing the conduit. This we will refer to as the closed state. This indicates that there will be a transition between the closed and open states as the cohesion energy is varied. Thus, the biphasic functional arrangement of domains with an extended stalk and cohesive tip leads to a brush structure that can be switched between two states of the brush: open and closed, as shown in Fig. 5.

With reference to gated transport, it is known that transport factors, such as karyopherin/importin, bind to cohesive FG motifs (44,45). This provides a mechanism to disrupt the intermolecular cohesiveness between nup tips by the competitive binding of karyopherin. We propose that the binding of certain particular transport factors in vivo results in such a switch from the closed to open state forming the basis for our model. In other words, specific types of transport-mediated cargo complexes could have the ability to dissolve into the central plug that keeps the gate closed, thus effectively opening it up to the desired diameter. These types of cargo can then undergo one-dimensional diffusion down the center of the channel. Enzymes on the nuclear side can cleave the transport factors, reducing the propensity of cargo to remain in the channel leading to escape to the nearest side—the nucleus. This picture works in the other direction with export signals as well and forms the basis of our diblock copolymer brush gate model. It is important to note that small cargo can also have alternate routes through the sides, and this aspect will be fully explored in future work.

We also used polymer brush modeling to understand the structure of Nup100 rings in the presence of closed Nsp1 rings. The free energy per chain is the same as in Eq. 2, but with an additional Nup100-Nsp1 cohesive cross-interaction term. Polymer sizes used were as measured in the Nup100 β-CG model simulations (see the Supporting Material). The per-chain cross-interaction free energy term between the two different types of FG nups can be simply modeled as the overlap fraction of an individual Nup100 sticky tip to the Nsp1 sticky tips at the center of the pore, multiplied by the blob interaction energy of the Nup100 sticky tips, ϵk_BT. The total Nup100 free energy is shown in Fig. S6 for different values of the blob-blob interaction energy, with ϵk_BT on the scale of the estimated self-interaction energy of Nup100 sticky tips (see the Supporting Material).

Rings of Nup100 are unable to close the pore, except in the presence of an adjacent closed ring of Nsp1 nups.

Given how our brush modeling and CG simulation show that rings of Nup100 are unlikely to close the nuclear pore except in the presence of other closed rings of FG nups, cooperativity between FG nups could play an important role in nucleocytoplasmic gating. This idealized analysis where only Nup100 and Nsp1 are present ignores the excluded volume that shrub FG nups like Nup49 and Nup57 along the pore wall could have in crowding out stalk domains and similar effects induced by cargo and transport factors, although we tentatively conclude that Nsp1 or a FG nup with similar properties could be principal to the closing or sealing mechanism of the pore. Consistent with this hypothesis, combinatorial deletions of FG nups of the type performed by Strawn et al. (46), have shown perturbations in transport in mutants when Nsp1 or nups like Nsp1 are functionally impaired. Our idealized analysis also ignores possible compactification, extension, crowding, and cross-linking of FG nups via interactions with cargo and their transport factors, although the existence and/or nature of these effects in vivo remain unclear.

Bioinformatics

Our modeling connects the biphasic sequence structure across FG nups in *S. cerevisiae* to their diblock polymer properties, which, in turn, results in a bistable polymer brush morphology. If this unique brush structure is critical for NPC gating it would imply that the biphasic sequence structure should be conserved across eukarya. To examine this assertion, we analyzed FG motif density versus charged amino-acid density across the disordered regions of the FG nup with the most FG repeats across 10 diverse eukaryotic species; see Fig. S11. We found that these nups all showed a common feature where, within each individual FG nup, there are two domains—one high in FG motifs with another high in charged amino acids, with both domains disjoint from one another along the amino-acid sequence. One domain, labeled the FG domain (26), is low in charged amino acids and high in FG motif repeats, whereas the other domain, the stalk domain (26), has a high charged aa density and typically a low FG repeat density with an anchor domain that attaches the FG nup to the inner pore surface. A quantitative bioinformatic analysis of the distribution of motifs along nup sequences by Ando et al. (34) revealed a broad conservation across species of these bimodal sequence structures in FG nups. Our modeling work here suggests a possible explanation for the functional advantage gained by utilizing the bimodal sequence of FG nups in NPC transport regulation.

DISCUSSION

Many experimental results regarding individual FG nup properties such as structure, location, and cohesive properties of different domains of nups are in conflict. What happens when these FG nups are put together within the confines of the nuclear pore channel is even more uncertain. Here we attempted to understand how the overall architecture of the assembly of disordered FG nups in the NPC channel is governed by the properties of individual FG-nups that make up the assembly using large scale physically accurate coarse-grained simulations. We calibrated our CG model against three main classes of experimental data on individual FG nup properties. We showed that, under certain conditions, a dynamic hydrogel-like structure does form, but only locally across spatial domains that have a high concentration of FG domains. We do not observe any scenario where FG nups form a dense layer located along the pore wall with a relatively open center as seen in some simulations (21–23), highlighting the importance of protein model structure on results. Using a calibrated model where the FG nups are relatively compact, our results indicate that a Forest type model structure (26) emerges, with the sticky tips of FG nups coalescing in the center to form a hydrophobic plug flanked by an extended coil brush zone.

Some recent experimental studies have measured that certain forms of facilitated transport take place near the walls of the NPC (47–51). Facilitated transport through peripheral routes does not appear to be a natural outcome from previous simulation efforts of the NPC (20–23). Such transport, however, appears to be supported by the α- and β-models where there exists a dense pluglike structure along the center of the channel that can block some forms of transport, forcing smaller actively transported cargoes to take the peripheral route along the channel walls. Other cargoes, larger in size and which have yet to be experimentally studied, may contain the proper transport factors that can open up the central plug as our polymer modeling suggests.

Our polymer brush modeling indicates that for the critical treelike nups to functionally open and close the pore in the Forest model, their sequence must follow a specific arrangement. The motifs and charges should form a diblock pattern and have a fixed polarity with the extended stalk region coming after the anchor region followed by the cohesive FG regions. Our analysis of 10 selected nups (with the largest number of FG motifs) across different species as well as earlier bioinformatic analysis of thousands of FG nups (34) clearly show that a biphasic sequence pattern is repeated across widely varying species, indicating that a diblock copolymer brush gating mode of transport may be a generic feature of certain types of nuclear import and export. Our modeling, therefore, sheds light on the underlying physics and biological function responsible for the presence of this biphasic pattern. Because the models α, β, and γ can also be viewed as roughly simulating FG nup properties under differing solvent conditions, other models of transport that do not rely upon individual FG nup biphasic sequence structure may also emerge as limiting cases of our model in different effective solvent conditions. We imagine that the

γ-model loosely corresponds to poor solvent conditions because hydrophobic-hydrophobic interactions are weakened in poor solvent conditions, model γ represents a weakening of hydrophobic interactions, and that the FG nups are dominated by hydrophobic amino acids due to their compact size relative to neutral polymers of the same length. In the γ-model, the FG nups form a homogenous brush within the pore similar to the virtual gate polymer brush model, a gating mechanism the NPC may share with the diblock copolymer brush gate model under different types of solvent conditions.

Throughout this article, we have focused on the larger tree-type FG nups that have been shown to be critical for many forms of transport (46). Some details of NPC structure such as cytoplasmic filaments, nuclear basket structure, and short shrub FG nups that presumably lie along the pore wall have not been considered in our analysis. Because the cytoplasmic filaments and the basket are spatially well separated from the interior of the pore, we anticipate that they would not interfere with the copolymer brush structure. Given that the properties of the shrub FG nups are similar to the sticky-tip domains of tree FG nups in terms of bimodal adhesion (26), we anticipate that the shrub-sticky attractive tip interactions could help stabilize the wide open configurations of the copolymer brush in the presence of very large cargo in the middle of the pore. This sort of cooperativity could be relevant for the export of large ribonucleoprotein particles whose transport is facilitated by surface-bound transport factors. In the work of Strawn et al. (46), shrublike nups Nup49 and Nup57 were required for the viability of yeast cells, which indicates that the effects of shrubs are important and possibly necessary for the transport of large cargoes. It is interesting to note that we have shown that another form of cooperativity exists between different types of FG nups. Our modeling shows that the closure of the Nup100 ring requires the presence of a closed ring of Nsp1 or Nsp1-like nups. These and possibly other forms of cooperativity could play an important role in heightening the selectivity over a wider range of cargo sizes. Therefore, as is the case with numerous other biomolecular assemblies (52), cooperativity may be significant for NPC function.

Indirect evidence for the existence of a diblock copolymer brush gate mechanism in vivo include the fact that in mammals and yeast, weak alcohols such as hexanediol can loosen the permeability barrier of the NPC at low concentrations (5%) (53,54). Also, exposed electrical charges in cargos bound to karyopherins can slow down their passage across the NPC (55). Both these seem to indicate a hydrophobic plug but are also consistent with the hydrogel model. One way to truly differentiate between these models is to shorten/extend the stalk regions or to rearrange the cohesive regions within the same nups. We propose a plausible experiment where genetically engineered organisms have the anchor domains of their core FG nups moved away from their stalk domains and toward the end of their FG domains, a genetic rearrangement that we would predict to be lethal, whereas most other models, which assume homogenous FG nups, would predict identical function and viable organisms. Similarly, we also propose that genetically engineered organisms that have had their disordered FG nup sequence regions randomly shuffled should be lethal, whereas again most other models would predict identical function. Physically, we predict lethality in these genetic engineering experiments because they would destroy the specific microphase separation of the cylindrical brush that we predict would be key to the transport of certain types of important cargoes.

Our results not only increase the number of physical mechanisms by which we can understand nucleocytoplasmic transport, but also allow us to extend our model to designing and optimizing novel forms of biomimetic transport. Biomimetic membranes (56) have a wide variety of applications ranging from chemical and biological separation and purification (57), as a platform for analytical detection of antioxidants, antibiotics, antiviral, psychotropic and other substances, drug delivery (58), and self-contained reactors and mock cells (59,60). The regulated cross-membrane transport of material in such systems is of vital importance and our investigations of a well-defined biocompatible mechanism should be very useful. In general, any diblock polymers that follow the sequence-determined polymer properties of FG nups (excluded volume polymer stalk, with collapsed sticky tips) could be used to design biomimetic pores that regulate traffic using our diblock copolymer brush gate mechanism. In addition, amino-acid sequences of disordered proteins can be modified to change their overall polymer properties to be similar to FG nups in our diblock copolymer pore model, allowing similar regulatory pores to be created in many different contexts. Indeed, even though the NPC transport is poorly understood, there have already been efforts to make biomimetic gates inspired by the NPC (61,62).

SUPPORTING MATERIAL

Thirteen figures, two tables and supplemental information are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00321-X.

This work was partially supported by National Science Foundation grant No.EF-1038697 (to A.G.), National Science Foundation grant No. DBI-0960480 (to A.G. and M.C.), a James S. McDonnell Foundation Award (to A.G.), National Science Foundation grant No. DMR-1310687 (to R.Z.), and National Institutes of Health grant No. GM077520 (to M.R.).

REFERENCES

1. Lim, R. Y., U. Aebi, and B. Fahrenkrog. 2008. Towards reconciling structure and function in the nuclear pore complex. *Histochem. Cell Biol.* 129:105–116.
2. Alber, F., S. Dokudovskaya, …, M. P. Rout. 2007. The molecular architecture of the nuclear pore complex. *Nature.* 450:695–701.
3. Rout, M. P., J. D. Aitchison, …, B. T. Chait. 2000. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* 148:635–651.

4. Yang, Q., M. P. Rout, and C. W. Akey. 1998. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell.* 1:223–234.
5. Macara, I. G. 2001. Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.* 65:570–594.
6. Denning, D. P., S. S. Patel, …, M. Rexach. 2003. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. USA.* 100:2450–2455.
7. Bayliss, R., T. Littlewood, …, M. Stewart. 2002. GLFG and FxFG nucleoporins bind to overlapping sites on importin-β. *J. Biol. Chem.* 277:50597–50606.
8. Panté, N., and M. Kann. 2002. Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. *Mol. Biol. Cell.* 13:425–434.
9. Grünwald, D., and R. H. Singer. 2010. In vivo imaging of labeled endogenous β-actin mRNA during nucleocytoplasmic transport. *Nature.* 467:604–607.
10. Allen, N. P. C., L. Huang, …, M. Rexach. 2001. Proteomic analysis of nucleoporin interacting proteins. *J. Biol. Chem.* 276:29268–29274.
11. Allen, N. P. C., S. S. Patel, …, M. Rexach. 2002. Deciphering networks of protein interactions at the nuclear pore complex. *Mol. Cell. Proteomics.* 1:930–946.
12. Denning, D. P., and M. F. Rexach. 2007. Rapid evolution exposes the boundaries of domain structure and function in natively unfolded FG nucleoporins. *Mol. Cell. Proteomics.* 6:272–282.
13. Krishnan, V. V., E. Y. Lau, …, M. F. Rexach. 2008. Intramolecular cohesion of coils mediated by phenylalanine-glycine motifs in the natively unfolded domain of a nucleoporin. *PLOS Comput. Biol.* 4:e1000145.
14. Frey, S., R. P. Richter, and D. Görlich. 2006. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science.* 314:815–817.
15. Rout, M. P., J. D. Aitchison, …, B. T. Chait. 2003. Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol.* 13:622–628.
16. Zilman, A., S. DiTalia, …, M. O. Magnasco. 2007. Efficiency, selectivity and robustness of the nuclear pore complex transport. *PLOS Comput. Biol.* 3:e125.
17. Zilman, A., S. DiTalia, …, M. O. Magnasco. 2010. Enhancement of transport selectivity through nano-channels by non-specific competition. *PLOS Comput. Biol.* 6:e1000804.
18. Bickel, T., and R. Bruinsma. 2002. The nuclear pore complex mystery and anomalous diffusion in reversible gels. *Biophys. J.* 83:3079–3087.
19. Osmanovic, D., J. Bailey, …, I. J. Ford. 2012. Bistable collective behavior of polymers tethered in a nanopore. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 85:061917.
20. Mincer, J. S., and S. M. Simon. 2011. Simulations of nuclear pore transport yield mechanistic insights and quantitative predictions. *Proc. Natl. Acad. Sci. USA.* 108:E351–E358.
21. Moussavi-Baygi, R., Y. Jamali, …, M. R. Mofrad. 2011. Brownian dynamics simulation of nucleocytoplasmic transport: a coarse-grained model for the functional state of the nuclear pore complex. *PLOS Comput. Biol.* 7:e1002049.
22. Moussavi-Baygi, R., Y. Jamali, …, M. R. Mofrad. 2011. Biophysical coarse-grained modeling provides insights into transport through the nuclear pore complex. *Biophys. J.* 100:1410–1419.
23. Tagliazucchi, M., O. Peleg, …, I. Szleifer. 2013. Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex. *Proc. Natl. Acad. Sci. USA.* 110:3363–3368.
24. Hills, Jr., R. D., L. Lu, and G. A. Voth. 2010. Multiscale coarse-graining of the protein energy landscape. *PLOS Comput. Biol.* 6:e1000827.
25. Plimpton, S. 1995. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* 117:1–19.
26. Yamada, J., J. L. Phillips, …, M. F. Rexach. 2010. A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics.* 9:2205–2224.
27. Xue, B., R. L. Dunbrack, …, V. N. Uversky. 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim.Biophys. Acta Proteins Proteomics.* 1804:996–1010.
28. Ayton, G. S., W. G. Noid, and G. A. Voth. 2007. Systematic coarse graining of biomolecular and soft-matter systems. *MRS Bull.* 32:929–934.
29. Noid, W. G., J.-W. Chu, …, G. A. Voth. 2007. Multiscale coarse-graining and structural correlations: connections to liquid-state theory. *J. Phys. Chem. B.* 111:4116–4127.
30. Karanicolas, J., and C. L. Brooks, 3rd. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.
31. Milles, S., and E. A. Lemke. 2011. Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153. *Biophys. J.* 101:1710–1719.
32. Tcherkasskaya, O., E. A. Davidson, and V. N. Uversky. 2003. Biophysical constraints for protein structure prediction. *J. Proteome Res.* 2:37–42.
33. Lim, R. Y., N.-P. Huang, …, U. Aebi. 2006. Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl. Acad. Sci. USA.* 103:9512–9517.
34. Ando, D., M. Colvin, …, A. Gopinathan. 2013. Physical motif clustering within intrinsically disordered nucleoporin sequences reveals universal functional features. *PLoS ONE.* 8:e73831.
35. Rubinstein, M., and R. H. Colby. 2003. Polymer Physics. Oxford University Press, Oxford, UK.
36. Peters, R. 2005. Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. *Traffic.* 6:421–427.
37. Grünwald, D., and R. H. Singer. 2012. Multiscale dynamics in nucleocytoplasmic transport. *Curr. Opin. Cell Biol.* 24:100–106.
38. Fischer, H., I. Polikarpov, and A. F. Craievich. 2004. Average protein density is a molecular-weight-dependent function. *Protein Sci.* 13:2825–2828.
39. Akey, C. W. 2010. The NPC-transporter, a ghost in the machine. *Structure.* 18:1230–1232.
40. Beck, M., F. Förster, …, O. Medalia. 2004. Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science.* 306:1387–1390.
41. Colwell, L. J., M. P. Brenner, and K. Ribbeck. 2010. Charge as a selection criterion for translocation through the nuclear pore complex. *PLOS Comput. Biol.* 6:e1000747.
42. Flory, P. J. 1971. Statistical Mechanics of Chain Molecules. Interscience, New York.
43. Sevick, E. 1996. Shear swelling of polymer brushes grafted onto convex and concave surfaces. *Macromolecules.* 29:6952–6958.
44. Liu, S. M., and M. Stewart. 2005. Structural basis for the high-affinity binding of nucleoporin Nup1p to the *Saccharomyces cerevisiae* importin-β homologue, Kap95p. *J. Mol. Biol.* 349:515–525.
45. Isgro, T. A., and K. Schulten. 2005. Binding dynamics of isolated nucleoporin repeat regions to importin-β. *Structure.* 13:1869–1879.
46. Strawn, L. A., T. Shen, …, S. R. Wente. 2004. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat. Cell Biol.* 6:197–206.
47. Kim, J., A. Izadyar, …, S. Amemiya. 2013. Nanoscale mechanism of molecular transport through the nuclear pore complex as studied by scanning electrochemical microscopy. *J. Am. Chem. Soc.* 135:2321–2329.
48. Ma, J., and W. Yang. 2010. Three-dimensional distribution of transient interactions in the nuclear pore complex obtained from single-molecule snapshots. *Proc. Natl. Acad. Sci. USA.* 107:7305–7310.

49. Ma, J., A. Goryaynov, …, W. Yang. 2012. Self-regulated viscous channel in the nuclear pore complex. *Proc. Natl. Acad. Sci. USA.* 109:7326–7331.

50. Yang, W. 2013. Distinct, but not completely separate spatial transport routes in the nuclear pore complex. *Nucleus.* 4:166–175.

51. Yang, W. 2011. 'Natively unfolded' nucleoporins in nucleocytoplasmic transport: clustered or evenly distributed? *Nucleus.* 2:10–16.

52. Williamson, J. R. 2008. Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* 4:458–465.

53. Shulga, N., and D. S. Goldfarb. 2003. Binding dynamics of structural nucleoporins govern nuclear pore complex permeability and may mediate channel gating. *Mol. Cell. Biol.* 23:534–542.

54. Ribbeck, K., and D. Görlich. 2002. The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *EMBO J.* 21:2664–2671.

55. Jäkel, S., J.-M. Mingot, …, D. Görlich. 2002. Importins fulfill a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. *EMBO J.* 21:377–386.

56. Martin, D. K. 2007. Nanobiotechnology of Biomimetic Membranes, *Vol. 1*. Springer, New York.

57. Bartsch, R. A., and J. D. Way. 1996. Chemical separations with liquid membranes: an overview. *In* ACS Symposium Series, *Vol. 642*. ACS Publications, Washington, DC, pp. 1–10.

58. Peer, D., J. M. Karp, …, R. Langer. 2007. Nanocarriers as an emerging platform for cancer therapy. *Nat. Nanotechnol.* 2:751–760.

59. Prokop, A. 2001. Bioartificial organs in the twenty-first century: nanobiological devices. *Ann. N. Y. Acad. Sci.* 944:472–490.

60. Karlsson, M., M. Davidson, …, O. Orwar. 2004. Biomimetic nanoscale reactors and networks. *Annu. Rev. Phys. Chem.* 55:613–649.

61. Jovanovic-Talisman, T., J. Tetenbaum-Novatt, …, B. T. Chait. 2009. Artificial nanopores that mimic the transport selectivity of the nuclear pore complex. *Nature.* 457:1023–1027.

62. Photos, P. J., H. Bermudez, …, D. E. Discher. 2007. Nuclear pores and membrane holes: generic models for confined chains and entropic barriers in pore stabilization. *Soft Matter.* 3:364–371.

Supporting Material for 'Nuclear pore complex protein sequences determine overall copolymer brush structure and function'

David Ando,* Roya Zandi,† Yong Woon Kim,‡
Michael Colvin,▷ Michael Rexach,◇ and Ajay Gopinathan*

*Department of Physics, University of California, Merced, CA, USA; †Department of Physics, University of California, Riverside, CA, USA; ‡Graduate School of Nanoscience and Technology, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea; ▷Department of Chemistry and Chemical Biology, University of California, Merced, CA USA; ◇Department of Molecular, Cell, and Developmental Biology, University of California, Santa Cruz, CA, USA

1 Cylindrical Brush Free Energy

Characterization of the NPC cylindrical polymer brush is achieved by deriving its total free energy per chain:

$$F_T = F_s + F_{ex} + F_{coh} \tag{1}$$

The first term in the RHS of this equation, F_s, is the stretching free energy which we model as originating from stalk part of the NPC brush only. The second term F_{ex} is the excluded volume free energy of this cylindrical brush, and the last term is the cohesive energy F_{coh} of the sticky tip FG domains.

2 Free energy of stalk chain stretching

The per chain stretching free energy is (1):

$$F_s = k_b T(\#blobs_{stalk}) = k_b T \int_{stalk} \mathrm{d}n_b \tag{2}$$

With n_b the number of blobs per nup and where the stretching free energy per nup is equal to Boltzmann's constant times the temperature times the number of blobs in the stalk region of a nup.

The number of blobs in the stalk brush if we have a *flat brush* is:

$$n_b = L/\xi \tag{3}$$

With L the end to end distance of stalk chains and ξ the radius of the stalk brush blobs.

The number of monomers per extended chain is:

$$N = n_b(\xi/a)^{5/3} \tag{4}$$

With a equal to the length of monomers in the chains.

$$\Rightarrow N = (L/\xi)(\xi/a)^{5/3} \tag{5}$$

$$\Rightarrow \xi = (N/L)^{3/2}a^{5/2} \tag{6}$$

Now we can consider the cylindrical brush case. First we consider blob size changes along the stalk brush as a function of the contour length s, which is a function of monomer number m. i.e. m is the mth monomer starting from the direction of the center of the cylinder with $m = 0$ the free end of the stalk chain: In direct analogue to the flat brush case, Eq. 6:

$$\Rightarrow \xi(s) = (\frac{\mathrm{d}m}{\mathrm{d}s})^{3/2}a^{5/2} \tag{7}$$

We can now calculate the change in the number of blobs as a function of s, analogous to Eq. 3, as:

$$\Rightarrow \mathrm{d}n_b = \frac{\mathrm{d}s}{\xi(s)} \tag{8}$$

Which after substitution from Eq. 18 becomes (2):

$$\Rightarrow \mathrm{d}n_b = \mathrm{d}s(\frac{\mathrm{d}s}{\mathrm{d}m})^{3/2}a^{-5/2} \tag{9}$$

We can now solve for the stretching free energy in Eq. 2

$$F_s = k_bT\int_{stalk}\mathrm{d}n_b = N_ck_bT\int \mathrm{d}s(\frac{\mathrm{d}s}{\mathrm{d}m})^{3/2}a^{-5/2} \tag{10}$$

$$= k_bT\int \mathrm{d}m(\frac{\mathrm{d}s}{\mathrm{d}m})(\frac{\mathrm{d}s}{\mathrm{d}m})^{3/2}a^{-5/2} \tag{11}$$

$$= k_bT\int \mathrm{d}m(\frac{\mathrm{d}s}{\mathrm{d}m})^{5/2}a^{-5/2} \tag{12}$$

Which after the mean field approximation of:

$$\frac{\mathrm{d}s}{\mathrm{d}m} = \frac{H}{N} \tag{13}$$

With H the height of the cylindrical stalk brush. Eq. 12 reduces to:

$$F_s = k_bTN(\frac{H}{N})^{5/2}a^{-5/2} \tag{14}$$

Where we can now conclude that the stretching free energy per chain is:

$$F_s = k_bT(\frac{(H/a)^{5/2}}{N^{3/2}}) \tag{15}$$

3 Free energy of excluded volume interactions

For a flat brush the free energy per chain of excluded volume interactions is (1):

$$F_e = \frac{1}{2} k_b T (\#blobs_{stalk}) \rho_{blob} \tag{16}$$

With ρ_{blob} the volume fraction of blobs in the brush.

This can be generalized to a per chain free energy valid for cylindrical brushes:

$$F_s = \frac{1}{2} k_b T \int_{stalk} \mathrm{d}n_b \rho_{blob} \tag{17}$$

With n_b the number of blobs per nup. The local volume faction can be defined as:

$$\rho_{blob} = \frac{\mathrm{d}V_{blobs}}{\mathrm{d}V_{brush}} = \frac{\xi^3 N_c \mathrm{d}n_b}{2\pi s \mathrm{d}s L} \tag{18}$$

Which implies that:

$$F_s = \frac{1}{2} k_b T \int_{stalk} \mathrm{d}n_b \frac{\xi^3 N_c \mathrm{d}n_b}{2\pi s \mathrm{d}s L} \tag{19}$$

Which simplifies to:

$$F_s = \frac{1}{2} k_b T \int_{stalk} \mathrm{d}s \frac{\mathrm{d}n_b^2}{\mathrm{d}s^2} \frac{\xi^3 N_c}{2\pi s L} \tag{20}$$

Which after substitution for $\frac{\mathrm{d}n_b}{\mathrm{d}s}$ by Eq. 8 equals:

$$F_s = \frac{1}{2} k_b T \int_{stalk} \mathrm{d}s \frac{\xi N_c}{2\pi s L} \tag{21}$$

Which after substitution for ξ from Eq. 18:

$$F_s = \frac{1}{2} k_b T \int_{stalk} \mathrm{d}s \frac{(\frac{\mathrm{d}m}{\mathrm{d}s})^{3/2} a^{5/2} N_c}{2\pi s L} \tag{22}$$

Which equals:

$$F_s = \frac{1}{2} k_b T \int_{stalk} \mathrm{d}m \frac{\mathrm{d}s}{\mathrm{d}m} \frac{(\frac{\mathrm{d}s}{\mathrm{d}m})^{-3/2} a^{5/2} N_c}{2\pi s L} = k_b T \int_{stalk} \mathrm{d}m \frac{(\frac{\mathrm{d}s}{\mathrm{d}m})^{-1/2} a^{5/2} N_c}{\pi s L} \tag{23}$$

Which after the mean field approximation of (2):

$$\frac{\mathrm{d}s}{\mathrm{d}m} = \frac{H}{N} \tag{24}$$

and

$$s = R - \frac{H}{2} \tag{25}$$

becomes:

$$F_s = k_b T \int_{stalk} \mathrm{d}m \frac{(\frac{H}{N})^{-1/2} a^{5/2} N_c}{(2R-H)L} = k_b T N \frac{(\frac{H}{N})^{-1/2} a^{5/2} N_c}{(2R-H)L} = k_b T N^{3/2} \frac{H^{-1/2} a^{5/2} N_c}{(2R-H)L} \tag{26}$$

Which after substitution for the number of chains in the brush N_c:

$$N_c = RL/d^2 \tag{27}$$

leads to a per chain excluded volume free energy of:

$$F_{ex} = k_b T (\frac{N^{3/2} a^{5/2}}{d^2}) \frac{R H^{-1/2}}{2R-H} \tag{28}$$

4 Free energy of cohesive interactions

Similar to the flat brush case we can define an energy density of blob interactions given that blob-blob interactions have an energy of $\epsilon k_b T$:

$$f_{coh} = 2\epsilon k_b T c^2 V \tag{29}$$

We have $c = N_c N_b / V$ the concentration of blobs (for N_b the total number of blobs per chain). For each chain or FG nup there exists only one sticky tip, therefore N_b =1 in this case. The volume the sticky tips can take on is approximated as an extended cylindrical region atop the stalk brush region whose volume is $V = 2\pi(R-H)L\delta$, with L equal to the height of the brush region axially along the pore.

N_c is the number of chains determined by the grafting distance d, with

$$N_c = 2\pi RL/d^2 \tag{30}$$

The concentration c is therefore:

$$c = \frac{R}{d^2 \delta (R-H)} \tag{31}$$

Which results in a free energy density of blob interactions of:

$$f_{coh} = 2\epsilon k_b T c^2 V \sim \epsilon k_b T \frac{R^2}{d^4 \delta^2 (R-H)^2} (R-H) L \delta \tag{32}$$

Which can be simplified to:

$$f_{coh} = \epsilon k_b T \frac{R}{d^2 \delta (R-H)} N_c \tag{33}$$

The total cohesive energy *per chain* is then equal to the volume of a sticky tip blob times the cohesive free energy density of all blob interactions divided by the number of chains.

$$F_{coh} = \epsilon k_b T \frac{R}{d^2 \delta (R-H)} N_c \delta^3 / N_c = \epsilon k_b T \frac{R \delta^2}{d^2 (R-H)} \tag{34}$$

5 Free energy for the total brush

We can now solve for the total free energy of the cylindrical brush per chain:

$$F_T = F_s + F_{ex} + F_{coh} \tag{35}$$

$$F_T = k_b T(\frac{(H/a)^{5/2}}{N^{3/2}}) + k_b T(\frac{a^{5/2}N^{3/2}}{d^2})\frac{RH^{-1/2}}{2R-H} + \epsilon k_b T \frac{R\delta^2}{d^2(R-H)} \tag{36}$$

$$= k_b T \left(\frac{(H/a)^{5/2}}{N^{3/2}} + \frac{a^{5/2}N^{3/2}}{d^2}\frac{RH^{-1/2}}{2R-H} + \epsilon \frac{R\delta^2}{d^2(R-H)} \right) \tag{37}$$

6 Brush modeling parameters

We measured the β-CG Model R_g of Nsp1's sticky tip to be 1.8 nm and the R_g of its stalk to be 6.4 nm. For Nup100 we measured the β-CG Model R_g for the sticky tip to be 2.5 nm and its stalk to have an R_g of 2.5 nm. Free energy parameters N (blob number) and a (Kuhn length) were derived from the contour length L (number of AAs * 0.38 nm) and R_g values by solving the simultaneous polymer equations for an excluded volume chain (3):

$$Na = L \tag{38}$$

$$0.377N^{3/5}a = R_g \tag{39}$$

Estimation of the self interaction free energy for the sticky tips of the FG nups was derived from the ratio of the measured R_g in the β-CG Model of the sticky tips to the R_g of an ideal excluded volume chain of the same monomer length.

$$\Delta F_{self} = k_b T \left(\frac{R_g^{ex}}{R_g^{\beta}} \right)^2 \tag{40}$$

We estimated the self interaction energy of the Nsp1 sticky tip to be 4.7 kT. Similarly we estimated the self interaction energy of Nup100's sticky tip to be 8.0 kT. These self interaction energy levels set the energy scale for blob-blob interactions, which we refer to as ϵkT in the paper.

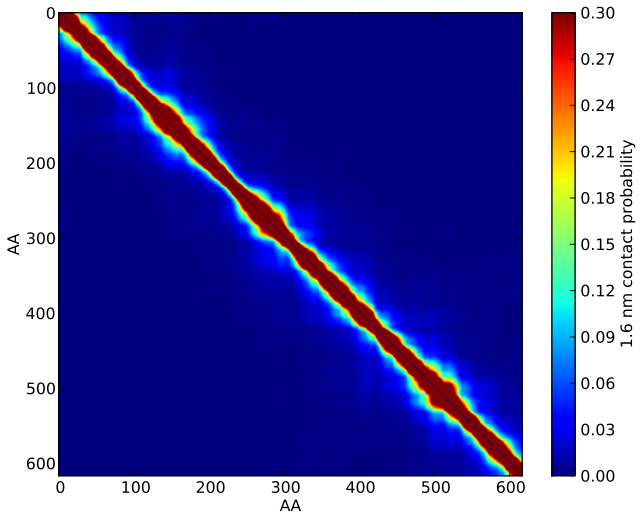

Figure 1: **Contact probability map for a randomly shuffled Nsp1 sequence protein** Probability map of given AAs along the protein chain being closer than 1.6 nm during a 4 μs simulation for a randomly shuffled Nsp1 sequence in the β model shows no block diagonal structure or subsequent biphasicness.

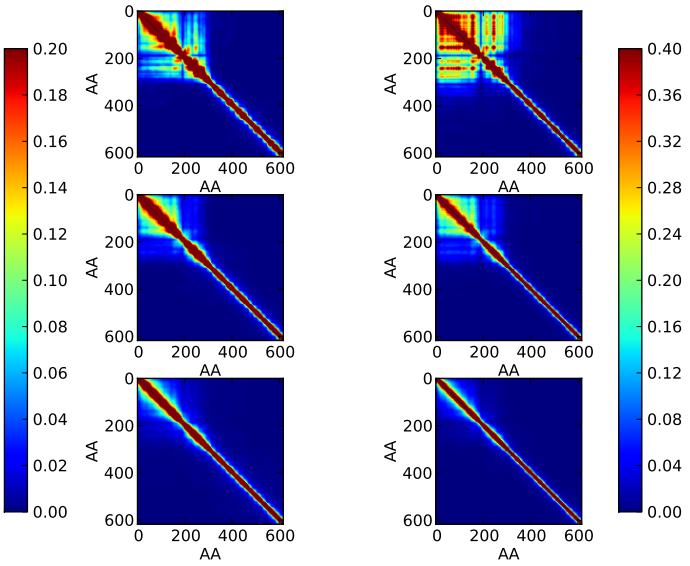

Figure 2: **Contact probability map for individual Nsp1 sequences within ring structures (left) compared to free Nsp1 (right), with the top, center and bottom panels refering to the alpha, beta, and gamma models respectively** The average probability map of given AAs along the protein chain being closer than 1.6 nm during a 4 μs simulation for a Nsp1 sequence grafted to the inner walls of a cylinder as *in vivo* is similar to the contact map of Nsp1 free in solution. The contact map probability of individual free Nsp1s in solution has higher absolute contact probability due to a lack of interaction from other chains which can compete for contact. The contact map from the aggregate simulation demonstrates the same high level of dynamic movement of free Nsp1 even after confinement with partner FG nups in a cylindrical geometry.

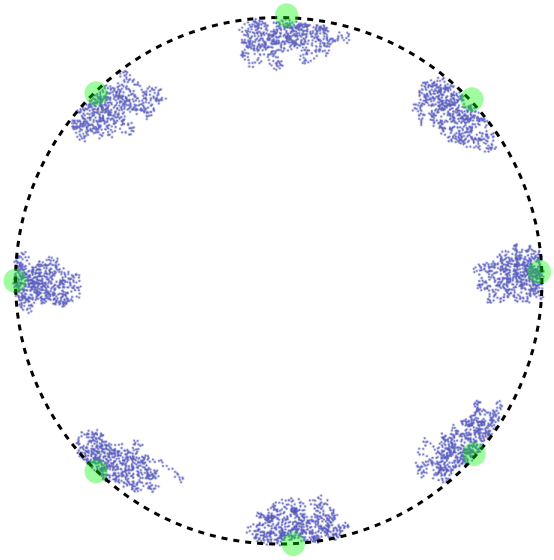

Figure 3: **Ring of Nsp1 initial starting structure** Green circles represent the locations where 8 Nsp1 FG nups were grafted to the inner surface of a cylinder with the initial starting conditions for the ring simulations as shown. Initial positions of the Nsp1 FG nups was located near the wall of the cylinder, yet during simulations the "sticky tips" of the FG nups aggregate towards the center of the cylinder in the α and β models.

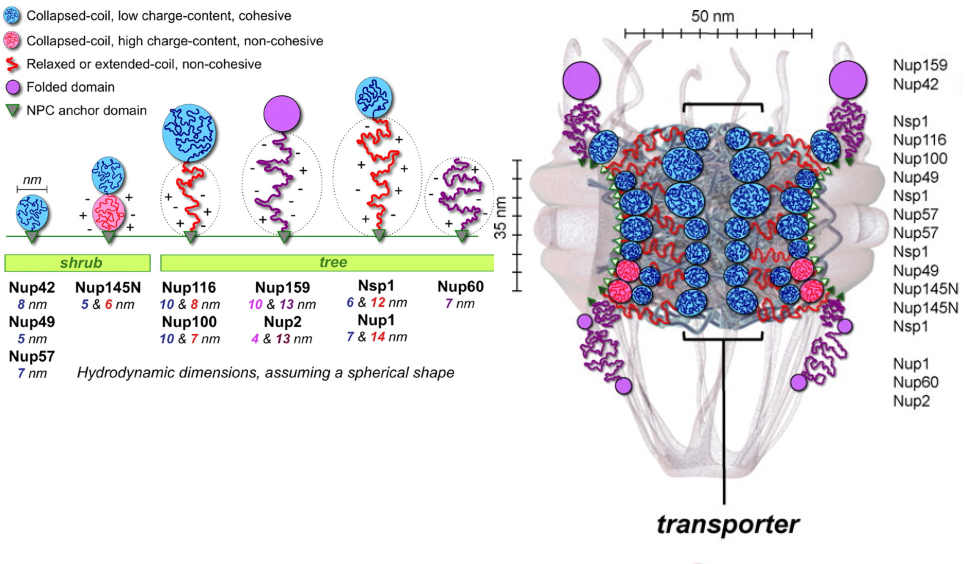

Figure 4: **Left: Diagrams of natively unfolded regions of yeast FG nucleoporins and their hydrodynamic radii.** Right: A diagram of the NPC architecture including the predicted topology and dimensions of yeast FG nucleoporins and their unstructured domains is shown. FG nucleoporins are listed according to the relative location of their C-termini along the z-axis, as determined by immuno-localization (4). These figures and research were originally published in Molecular and Cellular Proteomics. Justin Yamada, Joshua L. Phillips, Samir Patel, Gabriel Goldfien, Alison Calestagne-Morelli, Hans Huang, Ryan Reza, Justin Acheson, Viswanathan V. Krishnan, Shawn Newsam, Ajay Gopinathan, Edmond Y. Lau, Michael E. Colvin, Vladimir N. Uversky, and Michael F. Rexach. A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins. *Molecular and Cellular Proteomics*. 2010; 9:2205-2224.

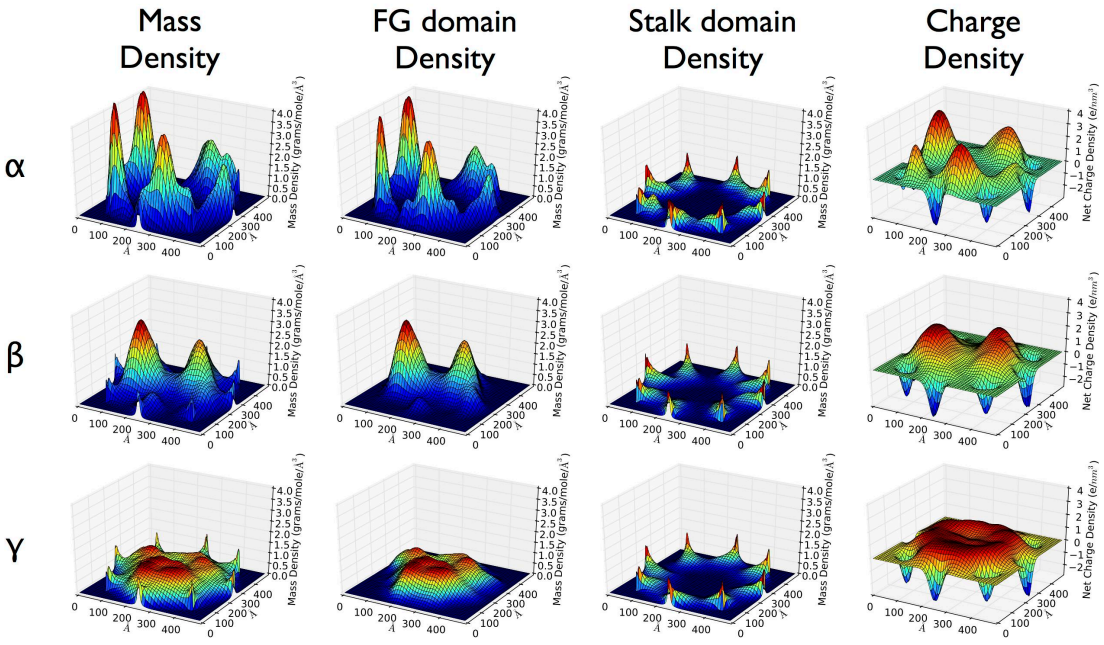

Figure 5: **Density maps for a ring of 8 Nup100s in different modeled scenarios.** Consecutively plotted for all three models are the total mass density within a simulated pore (mg/ml), the mass density of "FG Domains" (mg/ml), the mass density of "Stalk Domains" (mg/ml), and the net charge density (e/nm^3). The total mass density reveals that a dense central plug is unable to form in the α and β models, leaving the pore essentially open. On the other hand the γ model produces a homogenous extended brush resulting in a filled in and closed pore. Positive charge density of the ring of Nup100 nups mirrors closely the mass density of the FG domain, as FG domains are in general positively charged (4).

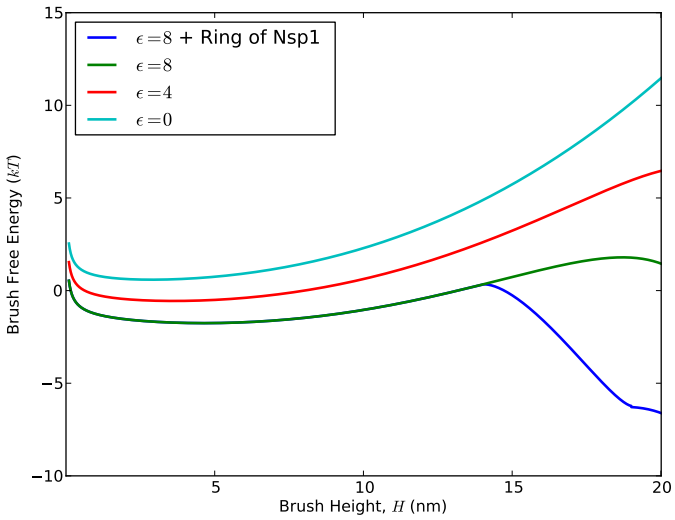

Figure 6: **Free energy of the Nup100 brush as a function of brush height for various values of the blob cohesive energy ($\epsilon k_B T$).** Brush heigh can extend to a maximum of around 20 nm, which is the radius R of the modeled pore minus the size of the sticky tips. When particular transport factors are present which are able to outcompete the inter-FG domain "sticky tips" interactions, the brush is in an open state (light blue). When interactions between sticky tips is able to recover into the several kT range and a closed ring of Nsp1 is adjacent to the Nup100 ring, the pore is able to close with a free energy minimum at $H \sim R - \delta$. We estimated the self interaction energy level of the Nup100 sticky tip to be 8.0 kT, which also sets the energy scale for blob-blob interactions of ϵkT.

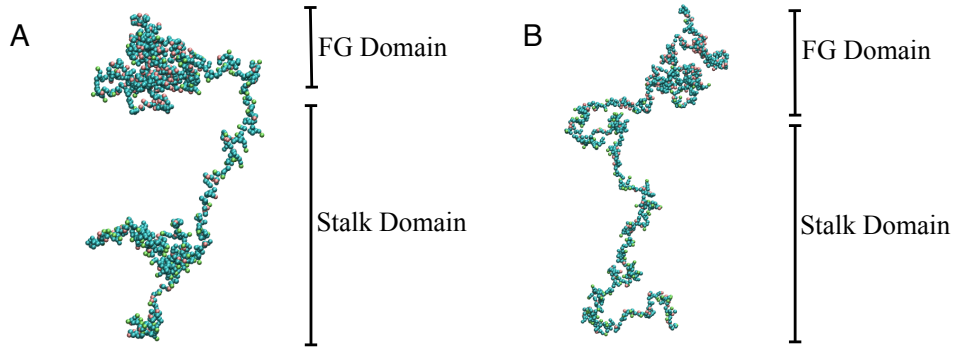

Figure 7: **Simulation snapshots** A) A simulation snapshot of full length Nsp1 showing the biphasic "FG Domain" and "Stalk Domain" structures in the β model. B) A simulation snapshot of full length Nsp1 snapshot as in part A) for the γ model, while the α model snapshot is shown in the main text.

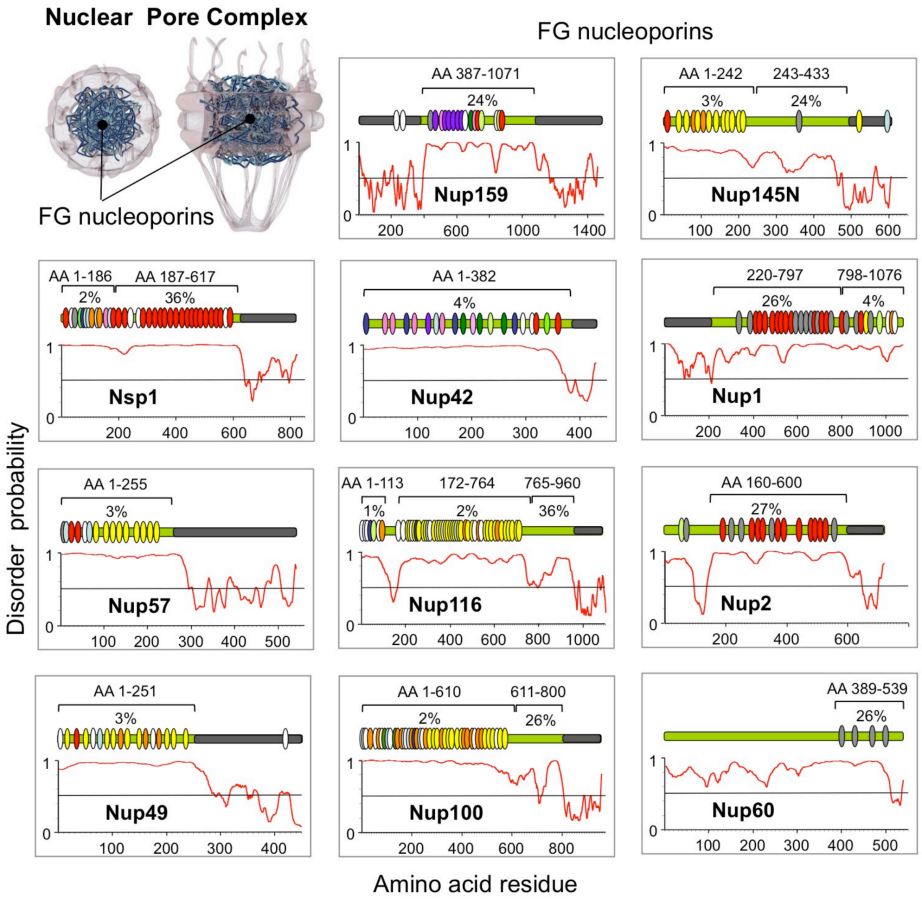

Figure 8: **FG repeat locations among FG nups** The locations of FG motifs and their type are shown for FG nups in S. cerevisiae. GLFG motifs are colored yellow, FxFG red, SPFG dark green, FxFx light gray, SAFG dark blue, PSFG bright green, NxFG light blue, SLFG orange, xxFG white, FxxFG lime green, double FG motifs (SAFGxPSFG) are pink, and the triple FG motifs are purple. This figure was originally published in Molecular and Cellular Proteomics. Justin Yamada, Joshua L. Phillips, Samir Patel, Gabriel Goldfien, Alison Calestagne-Morelli, Hans Huang, Ryan Reza, Justin Acheson, Viswanathan V. Krishnan, Shawn Newsam, Ajay Gopinathan, Edmond Y. Lau, Michael E. Colvin, Vladimir N. Uversky, and Michael F. Rexach. A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins. *Molecular and Cellular Proteomics.* 2010; 9:2205-2224.

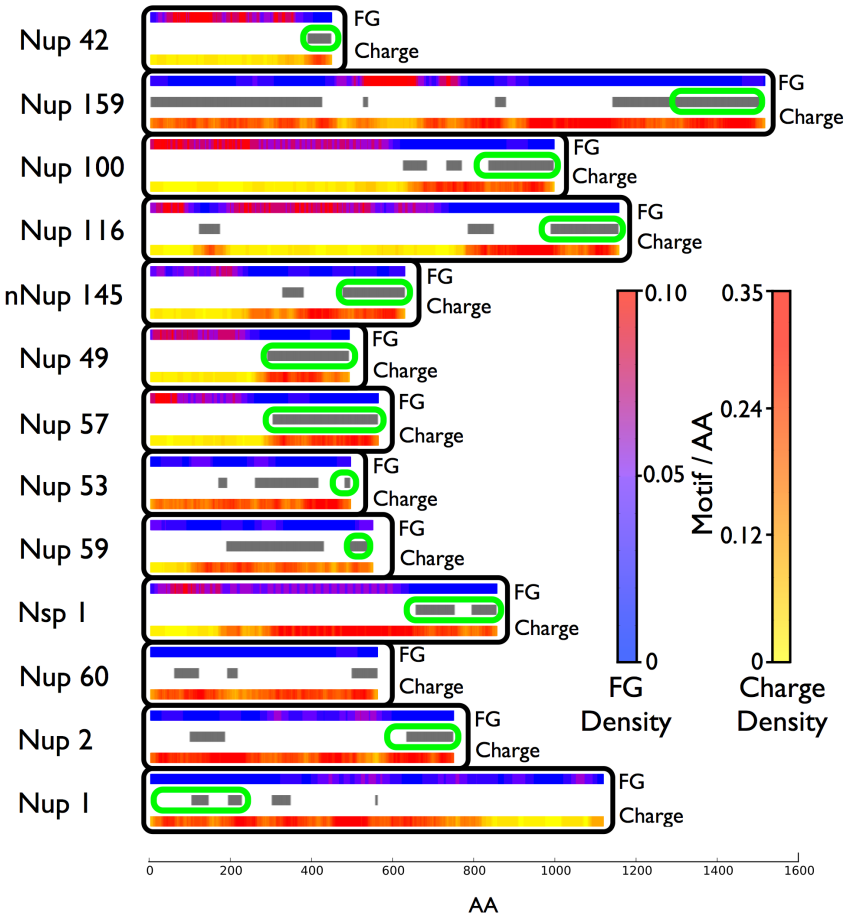

Figure 9: **FG repeat density among FG nups** The spatial distribution of FG motifs and charged AAs for all known FG nups of S. cerevisiae plotted as motif/AA, averaged over 20 nearest AAs. Regions of high FG motif density are shown in pink while regions of low charge density, shown in yellow, roughly correspond spatially throughout the sequences of these nups. Regions of protein which are predicted to form folded structures by the PONDR algorithm are highlighted with grey bars, and known/predicted (4,9) anchor domains circled with green ovals. This figure was originally published in *PLoS one*. Ando, D., M. Colvin, M. Rexach, and A. Gopinathan, 2013. Physical Motif Clustering within Intrinsically Disordered Nucleoporin Sequences Reveals Universal Functional Features, licensed under CC BY 2.5.

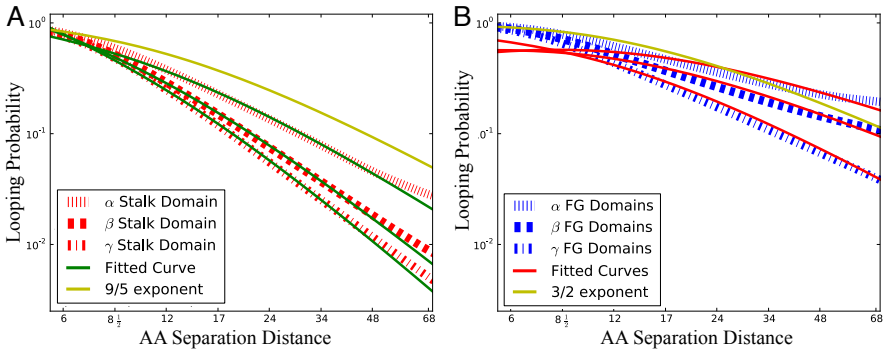

Figure 10: **Looping probabilities** (A) Log-log plot of amino acid contact probability (looping probability) as a function of the amino acid separation distance for stalk domains, averaged over simulated FG nups. The theoretical looping probability for extended coils (<1.6 nm cutoff) has an exponent of $p = 9/5$, and is shown in yellow. For the stalk domains the average fitted scaling exponents were 2.08, 2.58, and 2.82 for the different models α, β, and γ respectively. Stalk domains in these models had extended coil structures. (B) Similar average measured looping probabilities of the FG domains. For the α, β, and γ CG models the average fitted scaling exponents for the FG domains were 1.16, 1.38, 1.81 respectively. A theoretical looping probability with an exponent of 3/2 is shown in yellow, representing the dynamics of relaxed coil polymers. FG domains have looping exponents significantly lower than those of stalk domains, indicating the consistently large difference in ensemble structure between these two domain types. FG domains in the α model had a collapsed coil structure, a relaxed coil structure in the β model, and an extended coil structure in the γ model.

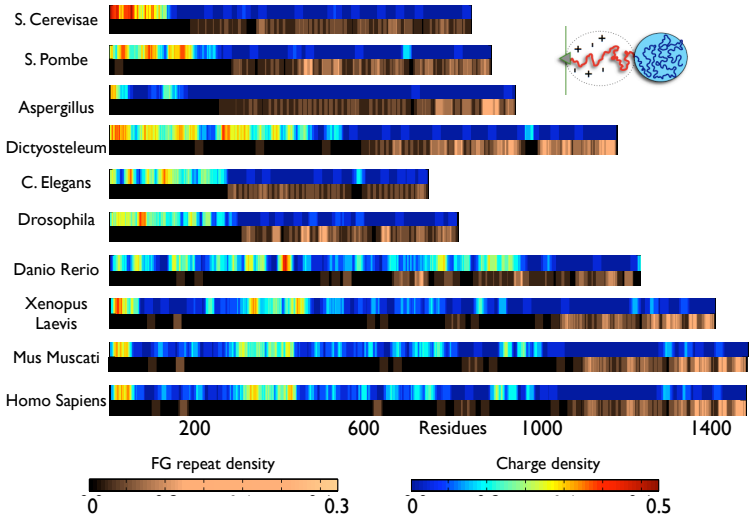

Figure 11: Heat maps showing FG repeat density (AA^{-1}) and charge density (AA^{-1}) across nucleoporins from ten different species. For each organism listed, the densities were measured along the disordered regions of the amino acid sequence of the FG nup which had the most FG motif repeats in that species. There appears to be a property common among all these heat maps, that regions high in FG motifs (pink) "FG domains" are in general disjoint from regions of the sequence high in charged amino acids (red, yellow, and light blue) "Stalk domains" (4). Additionally each FG nup appears to conserve functional features of these domains, such as their orientation, and diblock polymer structure. For comparison, the FG density and motif locations for FG nups from S. cerevisiae can be seen in the Supplementary material, Figs. S8-9 The displayed FG nups are the ones with the most FG repeat per species, while each of these species also contain several other FG Nups whose structure does not fit this paradigm. Uniprot gene identifiers for each FG nup analyzed (from top to bottom) are Q02630, Q9UTK4, B0Y6T9, Q54EQ8, D1MN47, Q9VCH5, B8JIZ8, Q9PVZ2, Q80U93, and P35658 respectively.

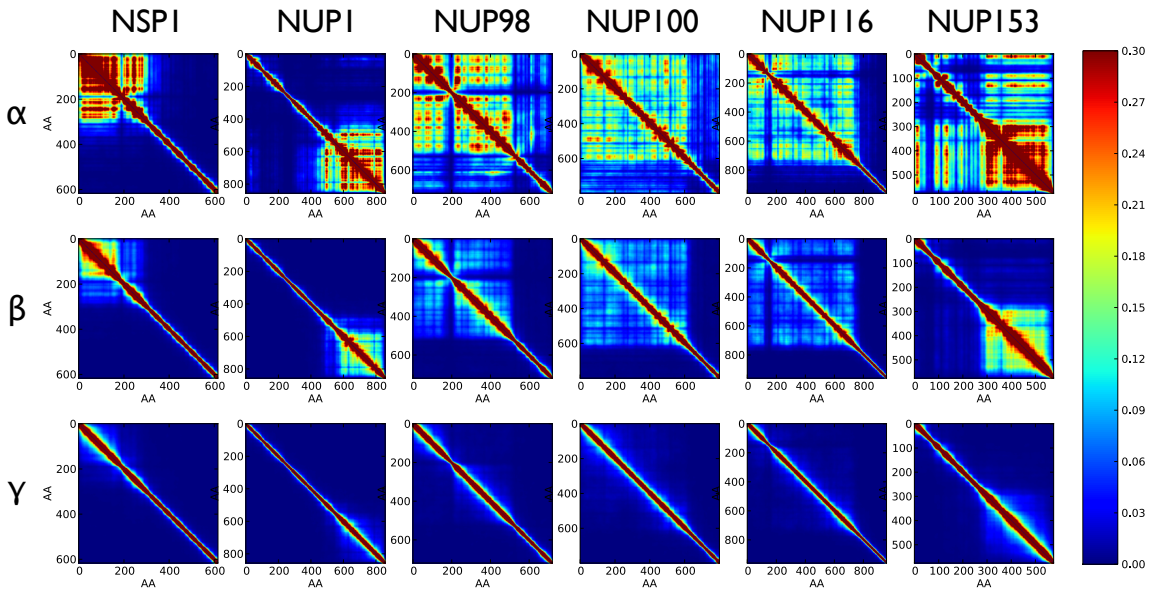

Figure 12: Contact maps for the different CG models α, β, and γ. Contact probability maps show the time averaged contacts between all pairs of amino acids. A block diagonal structure is noticeable in numerous contact maps, with one block for FG domains and diagonal contacts for stalk domains. A diblock structure can be most strongly seen in FG nups simulated under scenario β. Model α shows some block structure but often produces significant contacts between FG and stalk domains, as the two domains often interact. In striking contrast to models α and β, model γ produced nups with little difference in contact probability between FG and stalk domains with the entire FG nups representing unstructured extended homogenous polymers. Amino acid residues shown are with respect to the disordered domains of the simulated FG nups, while full protein AA indexes can be determined by domain definitions in Table S2.

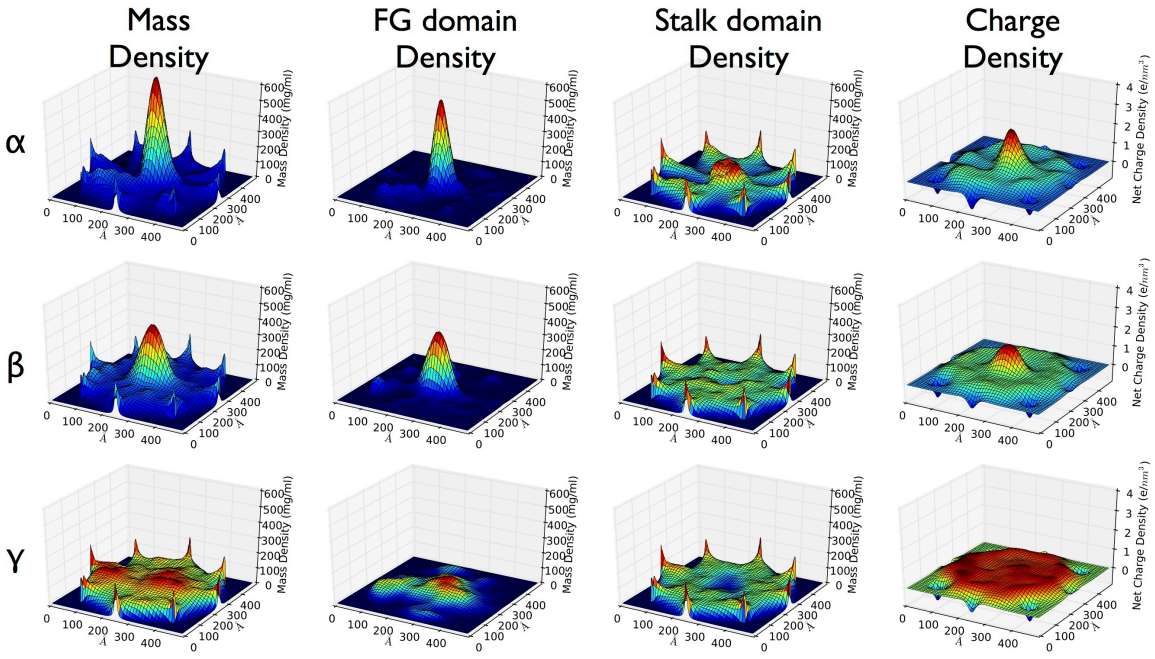

Figure 13: Density maps of a simulated pore containing 8 Nsp1 FG nups. Consecutively plotted for all three models is the total mass density (mg/ml), the mass density of "FG Domains" (mg/ml), the mass density of "Stalk Domains" (mg/ml), and the net charge density (e/nm^3). The total mass density clearly reveals a dense central plug connected by peripheral cables to the pore walls in the α and β models, while the γ model produces a homogenous extended brush. In the α and β models, one block phase (the FG domains) separates along the center of the channel while the other block (the stalk regions) aggregates along the periphery as can be seen in the spatial mass density for these domains. Positive charge density of the ring of Nsp1 nups mirrors closely the mass density of the FG domain, as FG domains are in general positively charged (5). This produces a striking effect in the α and β models, with a strongly positively charged plug region forming which has a low density of charged amino acids (which are predominately located within stalk domains (5)).

Tables

Table 1: **Comparison of Model and Experimental R_g for human Nup153**

Model	Simulation	Experiment
α-CG Model	$R_g = 22.7 \pm 0.1\mathring{A}$	Collapsed Coil Observed (7), implied $R_g = 21.8 \pm 1.8\mathring{A}$ (6)
β-CG Model	$R_g = 45.8 \pm 1.3\mathring{A}$	–
γ-CG Model	$R_g = 57.9 \pm 3.4\mathring{A}$	Relaxed Coil Observed (8), implied $R_g = 54.0 \pm 4.6\mathring{A}$ (6)

Amino acid domain of human Nup153 simulated and analyzed for these comparisons was AAs 899-1475. Experimentally measured properties of disordered FG nups have yet to converge, therefore a number of CG models with different scaling factors (α, β, γ) were developed, with each scaling factor corresponding to a different class of experimental results. The CG model α was scaled such that it produces a collapsed coil (molten globule) (6) for the disordered region of nup153, as predicted by FRET measurements by Milles *et al* (7). CG model γ was scaled such that it produces a relaxed coil (6) for the disordered region of nup153, as predicted by measurements on planar FG nup brushes (8). Bead-halo experiments by Yamada *et al* (4) indicate nups are biphasic, with stalk domains non-interacting while FG domains are "sticky" to other FG domains. A scaling factor β precisely in the middle between α and γ was found to produce such biphasicness, with this scaled CG model labeled β.

Table 2: **Domain Definitions and Disordered Regions used**

FG Nup	Simulated Disordered Region	FG Domain	Stalk Domain
yNsp1	AAs 1-617	AAs 1-186	AAs 187-617
yNup1	AAs 220-1076	AAs 798-1076	AAs 220-797
hNup98	AAs 1-720	AAs 1-485	AAs 486-720
yNup100	AAs 1-800	AAs 1-610	AAs 611-800
yNup116	AAs 172-960	AAs 172-764	AAs 765-960
hNup153	AAs 899-1475	AAs 1195-1475	AAs 899-1194

Human FG nups are prefaced by an 'h', while FG nups from S. cerevisiae are prefaced with an 'y'.

Supporting References

(1) Rubinstein, M., and R. H. Colby, 2003. Polymer physics. *OUP Oxford*.

(2) Sevick, E., 1996. Shear swelling of polymer brushes grafted onto convex and concave surfaces. *Macromolecules* 29:69526958.

(3) Flory, P. J., 1971. Statistical mechanics of chain molecules. *Macmillan*.

(4) Yamada, J., J. L. Phillips, S. Patel, G. Goldfien, A. Calestagne-Morelli, H. Huang, R. Reza, J. Acheson, V. V. Krishnan, S. Newsam, et al., 2010. A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Molecular and Cellular Proteomics* 9:22052224.

(5) Ando, D., M. Colvin, M. Rexach, and A. Gopinathan, 2013. Physical Motif Clustering within Intrinsically Disordered Nucleoporin Sequences Reveals Universal Functional Features. *PloS one* 8:e73831.

(6) Tcherkasskaya, O., E. A. Davidson, and V. N. Uversky, 2003. Biophysical constraints for protein structure prediction. *Journal of proteome research* 2:37–42.

(7) Milles, S., and E. A. Lemke, 2011. Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153. *Biophysical journal* 101:1710-1719.

(8) Lim, R. Y., N.-P. Huang, J. Koser, J. Deng, K. A. Lau, K. Schwarz-Herion, B. Fahrenkrog, and U. Aebi, 2006. Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proceedings of the National Academy of Sciences* 103:95129517.

(9) Denning, D. P., and M. F. Rexach, 2006. Rapid Evolution Exposes the Boundaries of Domain Structure and Function in Natively Unfolded FG Nucleoporins. *Molecular and Cellular Proteomics* 6:272282.