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.

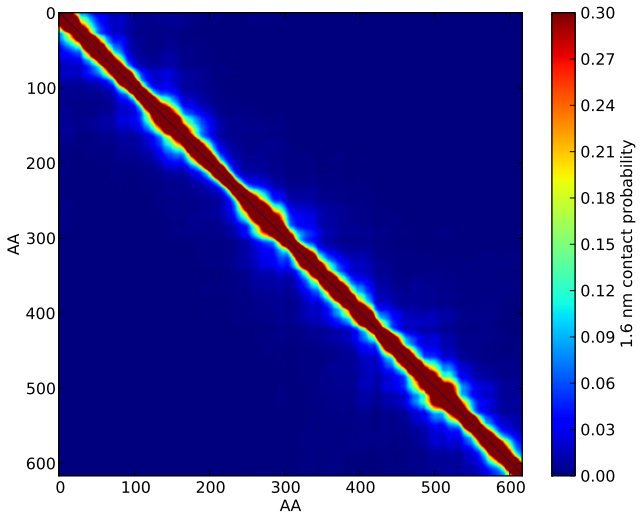

Figure 1: **Contact probability map for a randomly shuffled Nsp1 sequence protein** Probability map of given AAs along the protein chain being closer than 1.6 nm during a 4 μs simulation for a randomly shuffled Nsp1 sequence in the β model shows no block diagonal structure or subsequent biphasicness.

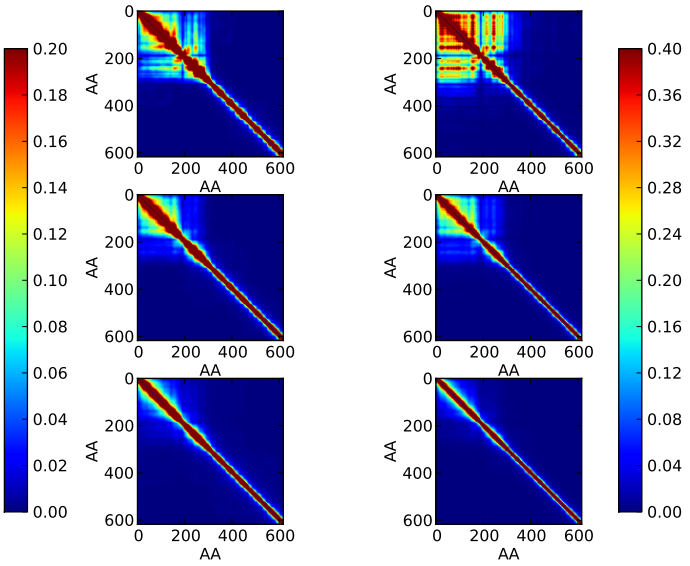

Figure 2: **Contact probability map for individual Nsp1 sequences within ring structures (left) compared to free Nsp1 (right), with the top, center and bottom panels refering to the alpha, beta, and gamma models respectively** The average probability map of given AAs along the protein chain being closer 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.

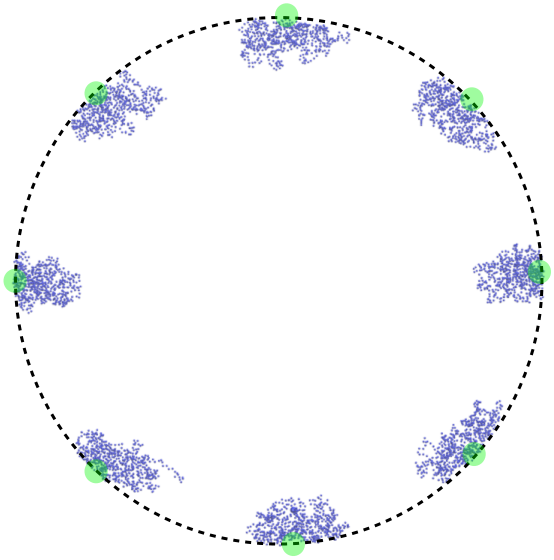

Figure 3: **Ring of Nsp1 initial starting structure** Green circles represent the locations where 8 Nsp1 FG nups were grafted to the inner surface of a cylinder with the initial starting conditions for the ring simulations as shown. Initial positions of the Nsp1 FG nups was located near the wall of the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

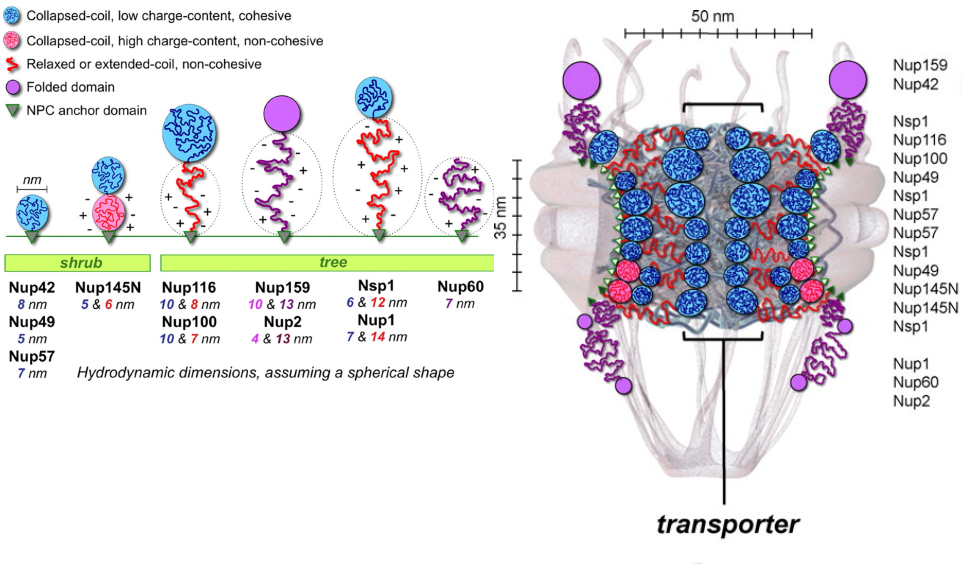

Figure 4: **Left: Diagrams of natively unfolded regions of yeast FG nucleoporins and their hydrodynamic radii.** Right: A diagram of the NPC architecture including the predicted topology and dimensions of yeast FG nucleoporins and their unstructured domains is shown. FG nucleoporins are listed according 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.

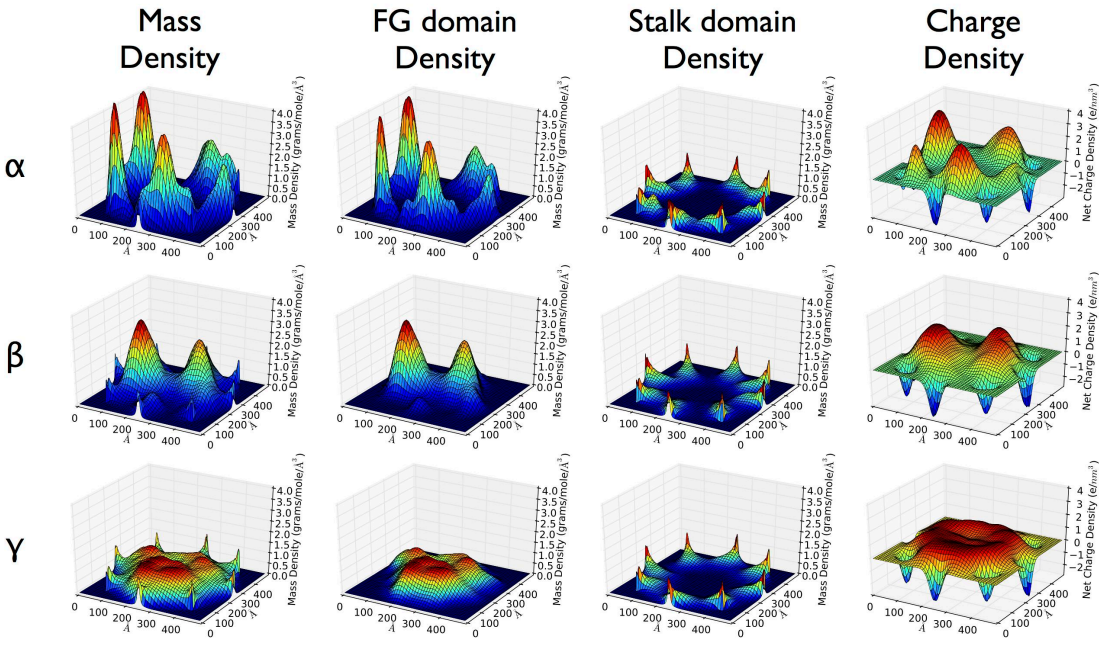

Figure 5: **Density maps for a ring of 8 Nup100s in different modeled scenarios.** Consecutively plotted for all three models are the total mass density within a simulated pore (mg/ml), the mass density of "FG Domains" (mg/ml), the mass density of "Stalk Domains" (mg/ml), and the net charge density (e/nm^3). 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).

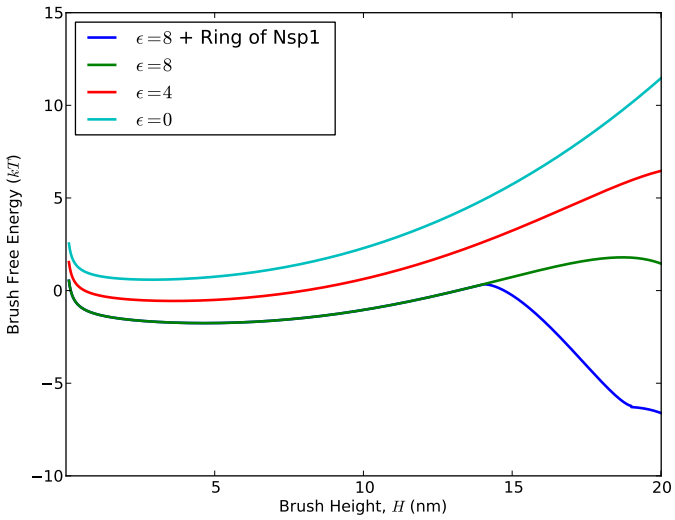

Figure 6: **Free energy of the Nup100 brush as a function of brush height for various values of the blob cohesive energy ($\epsilon k_B T$).** Brush heigh can extend to a maximum of around 20 nm, which is the radius R of the modeled pore minus the size of the sticky tips. When particular transport factors 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.

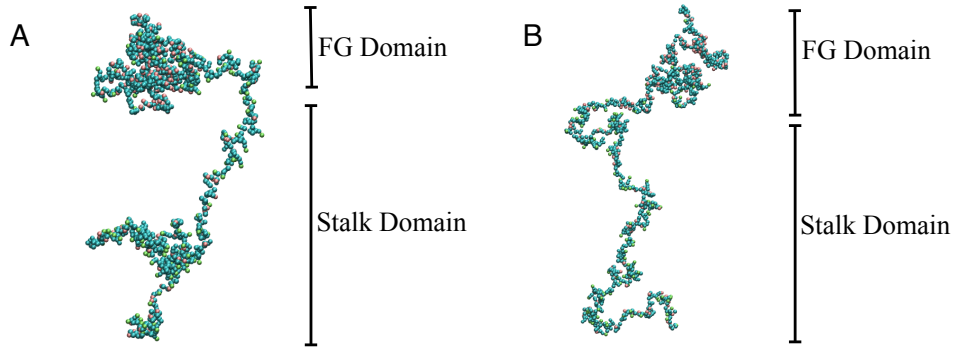

Figure 7: **Simulation snapshots** A) A simulation snapshot of full length Nsp1 showing the biphasic "FG Domain" and "Stalk Domain" structures in the β model. B) A simulation snapshot of full length Nsp1 snapshot as in part A) for the γ model, while the α model snapshot is shown in the main text.

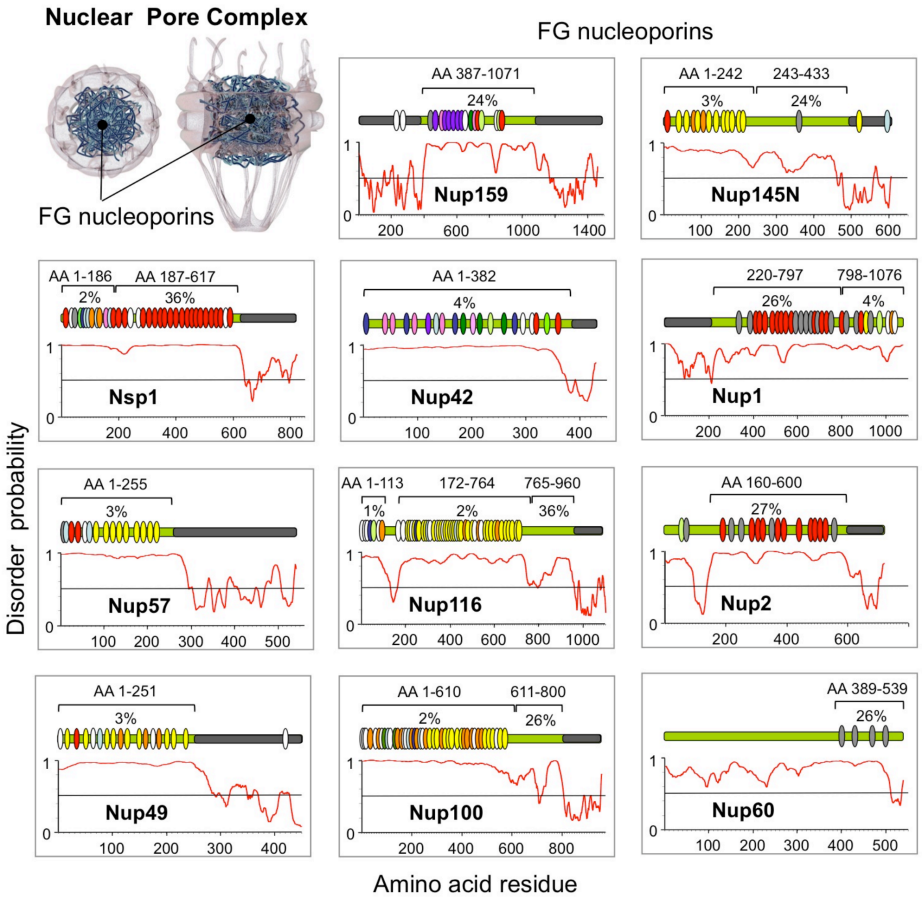

Figure 8: **FG repeat locations among FG nups** The locations of FG motifs and their type are shown for FG nups in S. cerevisiae. GLFG motifs are colored yellow, FxFG red, SPFG dark green, FxFx light gray, SAFG dark blue, PSFG bright green, NxFG light blue, SLFG orange, xxFG white, FxxFG lime green, 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.

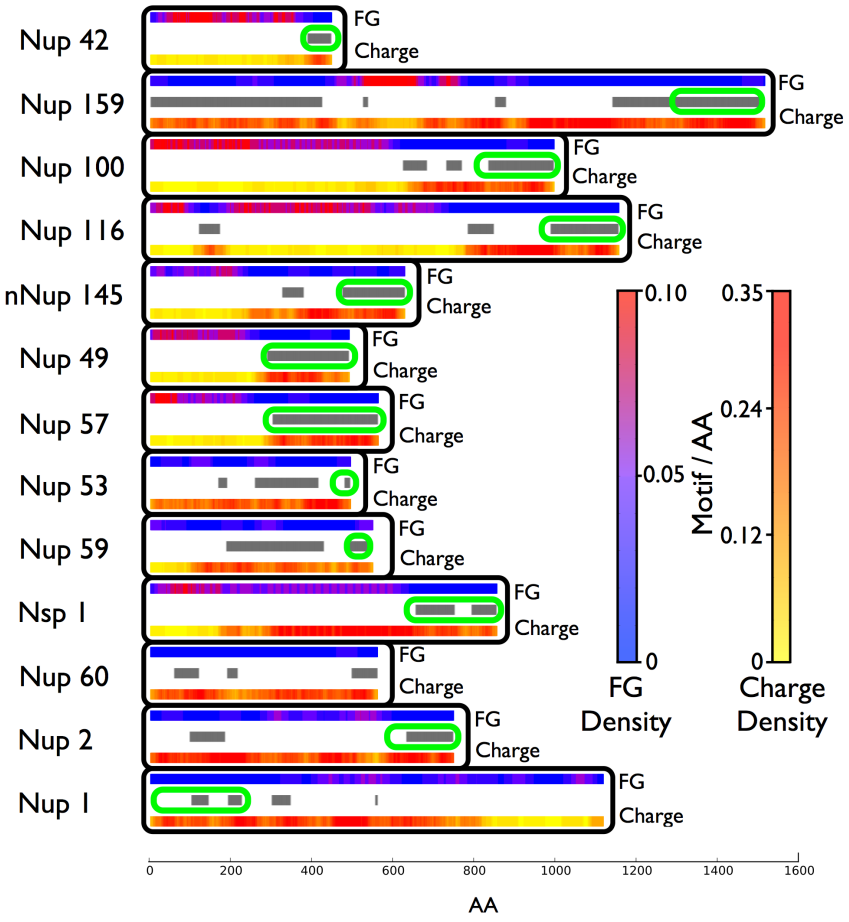

Figure 9: **FG repeat density among FG nups** The spatial distribution of FG motifs and charged AAs for all known FG nups of S. cerevisiae plotted as motif/AA, averaged over 20 nearest AAs. Regions of high FG motif density are shown in pink while regions of low charge density, shown in yellow, roughly 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.

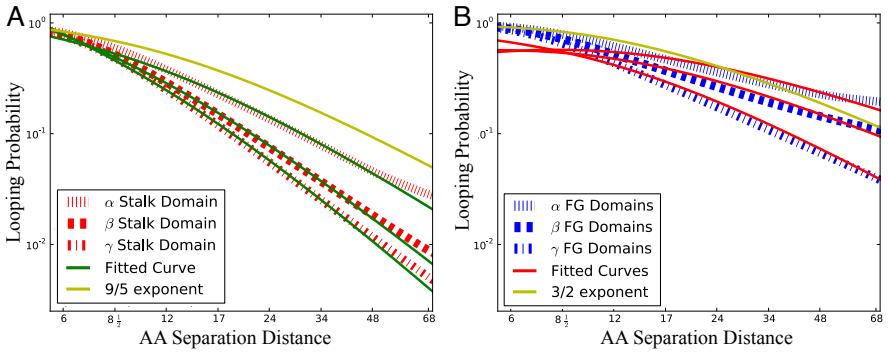

Figure 10: **Looping probabilities** (A) Log-log plot of amino acid contact probability (looping probability) as a function of the amino acid separation distance for stalk domains, averaged over simulated FG nups. The theoretical looping probability for extended coils (<1.6 nm cutoff) has an exponent 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.

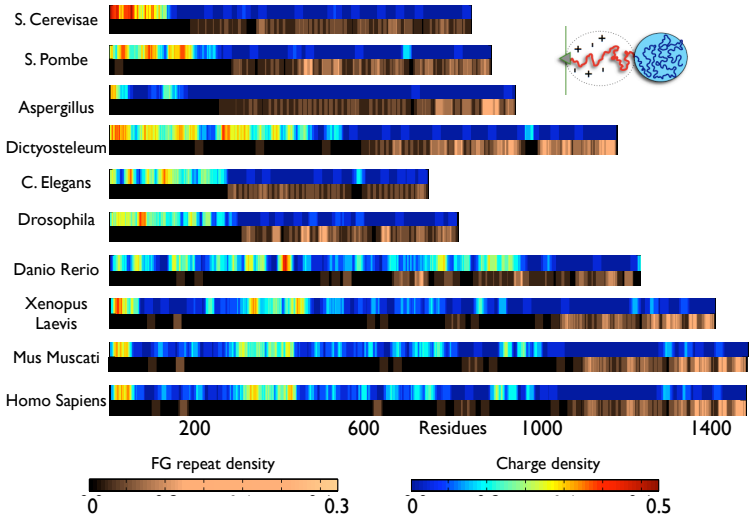

Figure 11: Heat maps showing FG repeat density (AA^{-1}) and charge density (AA^{-1}) across nucleoporins from ten different species. For each organism listed, the densities were measured along the disordered regions of the amino acid sequence of the FG nup which had the most FG motif repeats in that 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.

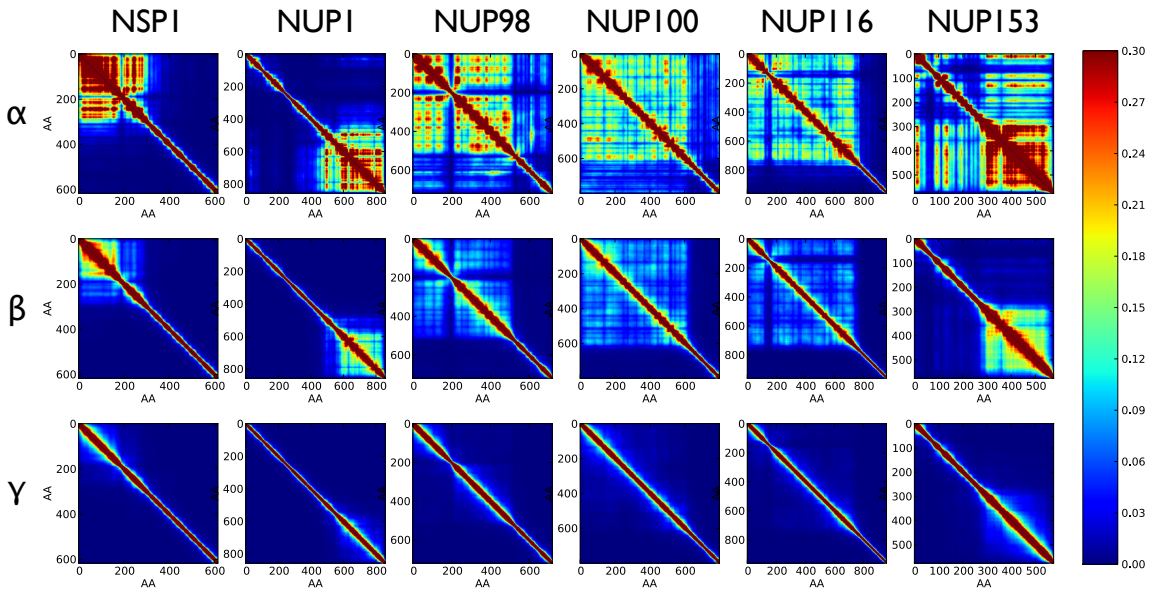

Figure 12: Contact maps for the different CG models α, β, and γ. Contact probability maps show the time averaged contacts between all pairs of amino acids. A block diagonal structure is noticeable in numerous contact maps, with one block for FG domains and diagonal contacts for stalk domains. A diblock 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.

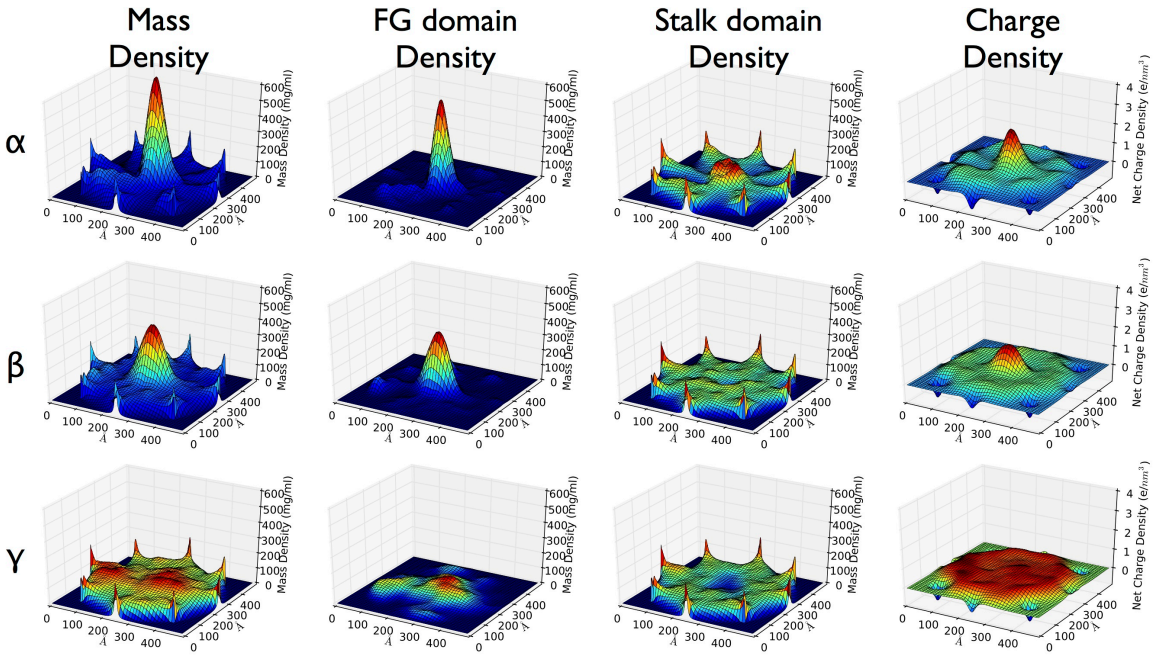

Figure 13: Density maps of a simulated pore containing 8 Nsp1 FG nups. Consecutively plotted for all three models is the total mass density (mg/ml), the mass density of "FG Domains" (mg/ml), the mass density of "Stalk Domains" (mg/ml), and the net charge density (e/nm^3). The total mass density clearly 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.