

# Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex

Mario Tagliazucchi<sup>a,1</sup>, Orit Peleg<sup>b,1,2</sup>, Martin Kröger<sup>c</sup>, Yitzhak Rabin<sup>d,3</sup>, and Igal Szleifer<sup>a,3</sup>

<sup>a</sup>Department of Biomedical Engineering, Department of Chemistry, and Chemistry of Life Processes Institute, Northwestern University, Evanston, IL 60208; <sup>b</sup>Institute for Neuroinformatics, University of Zurich and Eidgenössische Technische Hochschule Zurich, 8057 Zurich, Switzerland; <sup>c</sup>Polymer Physics, Department of Materials, Eidgenössische Technische Hochschule Zurich, 8093 Zurich, Switzerland; and <sup>d</sup>Department of Physics and Institute for Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel

Edited by Benjamin Widom, Cornell University, Ithaca, NY, and approved January 8, 2013 (received for review July 29, 2012)

The molecular structure of the yeast nuclear pore complex (NPC) and the translocation of model particles have been studied with a molecular theory that accounts for the geometry of the pore and the sequence and anchoring position of the unfolded domains of the nucleoporin proteins (the FG-Nups), which control selective transport through the pore. The theory explicitly models the electrostatic, hydrophobic, steric, conformational, and acid-base properties of the FG-Nups. The electrostatic potential within the pore, which arises from the specific charge distribution of the FG-Nups, is predicted to be negative close to pore walls and positive along the pore axis. The positive electrostatic potential facilitates the translocation of negatively charged particles, and the free energy barrier for translocation decreases for increasing particle hydrophobicity. These results agree with the experimental observation that transport receptors that form complexes with hydrophilic/neutral or positively charged proteins to transport them through the NPC are both hydrophobic and strongly negatively charged. The molecular theory shows that the effects of electrostatic and hydrophobic interactions on the translocating potential are cooperative and nonequivalent due to the interaction-dependent reorganization of the FG-Nups in the presence of the translocating particle. The combination of electrostatic and hydrophobic interactions can give rise to complex translocation potentials displaying a combination of wells and barriers, in contrast to the simple barrier potential observed for a hydrophilic/neutral translocating particle. This work demonstrates the importance of explicitly considering the amino acid sequence and hydrophobic, electrostatic, and steric interactions in understanding the translocation through the NPC.

non-additivity | nuclear transport | disordered protein | coarse grain model | mean force potential

Nucleocytoplasmic transport occurs exclusively through protein pores that perforate the nuclear envelope, the nuclear pore complexes (NPCs) (1). Whereas the NPC is permeable to small molecules (e.g., water, ions) that can diffuse freely through it, bigger cargoes, such as proteins and mRNA, require the assistance of transport receptors (known as karyopherins or “kaps”) to be effectively transported between the cytoplasm and the nucleus. It is challenging to understand how a cargo that is not able to pass through the pore by itself can successfully traverse the pore on forming a substantially larger kap–cargo complex. Because of its importance to the functioning of eukaryotic cells, this apparent paradox has been the focus of attention of numerous studies throughout the past decade (reviewed in refs. 1–9).

There is no universally agreed picture of the detailed mechanism of selective transport through the NPC, although there is broad agreement that a family of proteins called nucleoporins (Nups) is essential for selective transport through the pore (10–14). The folded domains of the Nups form the outer envelope of the NPC (in contact with the nuclear scaffold), and their intrinsically disordered domains protrude into the inner space of the pore.

These disordered domains, known as FG-Nups due to their high content of phenylalanine-glycine residues (FG-repeats), interact with the translocating particles to set up the permeability barrier that controls selective translocation through the NPC. A definitive transport mechanism remains elusive because directly visualizing FG-Nups and the kap–cargo complex within individual NPCs is at the limits of current single-molecule tracking technology (15–17); therefore, theory (18, 19) and computer simulations (20–22) have been used in an attempt to elucidate the essential features of the translocation process. In a recent coarse-grained molecular dynamics (MD) simulation, the kap–FG interaction was found to be highly dynamic and the FG-Nups formed a layer on the pore walls (20). The kap–cargo complex particle interacts with the FG residues in this layer as it diffuses through the channel. Another simulation study suggested that the translocating particle remains bound to the same Nup for its entire trajectory through the NPC (21). The differences between these works arise due to the choice of the molecular model, which, in neither case, considered the specific sequence and length of the FG-Nups and the particular properties of each amino acid in the sequence (e.g., hydrophobicity, charge).

Until recently, it was believed that hydrophobic interactions were solely responsible for the selectivity of the translocation process (11, 14, 20, 23, 24). According to this view, water-soluble proteins generally present a hydrophilic surface and are repelled by the hydrophobic domains of the FG-Nups, but hydrophobic patches on the surface of kaps interact attractively with the FG-Nups. It was assumed that the main role of charged amino acids in FG-repeats (about 15%) is to stabilize the hydrophobic sequences against self-aggregation and collapse. Although this argument suggests that the sign and magnitude of the charges do not play important roles, a recent analysis has shown that kaps and kap–cargo complexes are hydrophobic and highly negatively charged, whereas the unfolded Nup domains have a small net positive charge (25), suggesting that electrostatics may be essential for the selective filtering mechanism. The effect of sequence-dependent electrostatic interactions has not been considered in previous simulations and theories; therefore, its contribution to the overall transport process remains unclear.

The goal of the present work is to address the structure of the FG-Nups within the NPC and the molecular factors that

Author contributions: M.T., O.P., M.K., Y.R., and I.S. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

<sup>1</sup>M.T. and O.P. contributed equally to this work.

<sup>2</sup>Present address: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138.

<sup>3</sup>To whom correspondence may be addressed. E-mail: rabinserious@gmail.com or igalsz@northwestern.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1212909110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1212909110/-DCSupplemental).

determine their interactions with the particle. We study the translocation of different model particles to elucidate the role of the different interactions in the system and understand their interplay. Our predictions are based on a molecular theory developed to study the structure, thermodynamics, and transport behavior of responsive polymers end-grafted to surfaces of arbitrary geometry (26, 27) that is extended here to study the translocation of large particles through a nanopore, the NPC. The theory (*Methods*) is based on a free energy formulation that explicitly treats the size, shape, conformations, and charge state of all the molecular species and accounts for the nontrivial coupling between molecular organization, physical interactions, and chemical equilibrium. Our model for the yeast NPC incorporates the currently available information about the size and shape of the pore and the sequence and tethering position of each individual FG-Nup [according to the model of Alber et al. (28); for details see [Tables S1](#) and [S2](#)]. The geometry of the NPC and the native sequence of the FG-Nups are presented in Fig. 1 *A* and *B*, respectively. Our calculations show that the FG-Nups present a highly inhomogeneous charge distribution: Negative charges are concentrated on pore walls, and positive charges are located at the center of the pore. This result suggests that FG-Nup sequences are optimized to present a positive electrostatic environment along the pore axis to facilitate transport of negatively charged kap–cargo complexes. The systematic calculations presented in this work show that the interactions between hydrophobic/charged translocating particles and the NPC are qualitatively different from those of hydrophobic/neutral or hydrophilic/charged particles.

## Results

### Electrostatic Environment Within the NPC Is Highly Inhomogeneous.

In Fig. 2 (*Left*), we show the calculated density profiles and the electrostatic potential within the NPC obtained in the absence of translocating particles in the pore using the information about the amino acid sequences of the FG-Nups (a summary of the properties of each FG-Nup in the system is provided in [Table S3](#)). The plots show color maps of the total amino acid volume fraction (Fig. 2*A*), the volume fraction of hydrophobic amino acids (Fig. 2*B*), and the electrostatic potential (Fig. 2*C*) along a vertical cut of the pore (i.e., a plane that contains the pore axis; the scheme of the pore is shown in Fig. 1*A*). The total concentration of amino acids and the concentration of hydrophobic amino acids within the NPC are relatively constant, with the exception of the pore's

center, which shows a slightly lower density and a few spots on the pore's surface, where there is an enhanced density. There is a very large concentration of FG-Nup segments outside of the NPC, on both the cytoplasmic and nuclear sides, due to the large volume available for the FG-Nups in the outer regions of the pore; this organization significantly reduces the excluded volume repulsions between the FG-Nups. Interestingly, the electrostatic potential within the NPC (Fig. 2*C*) is highly inhomogeneous and presents pockets of negative electrostatic potential close to the NPC walls, whereas the center of the pore has a positive electrostatic potential.

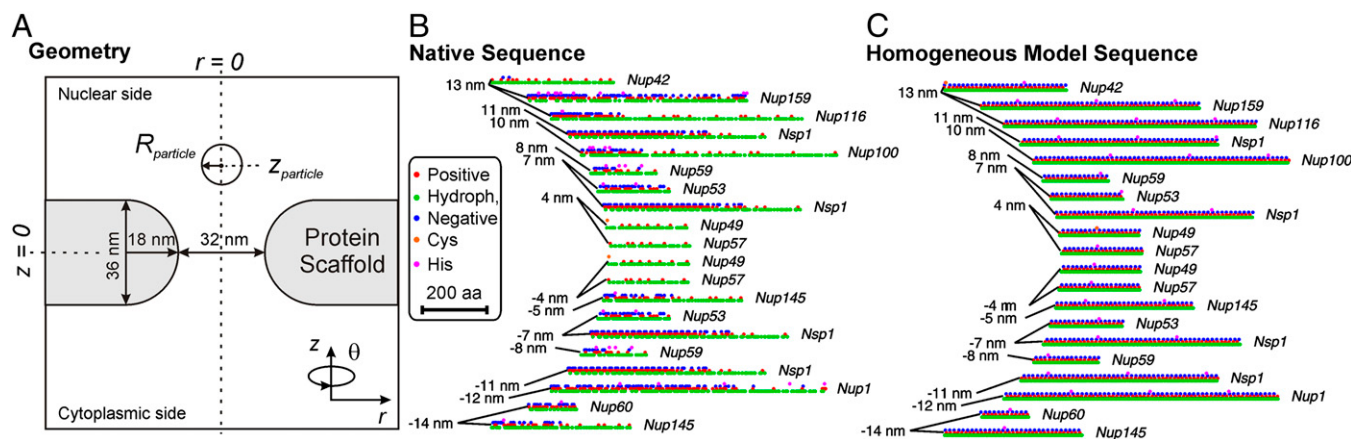
### Positive Electrostatic Environment at the Center of the Pore Is a Direct Consequence of the Native Sequence of the FG-Nups.

To examine the effect of the amino acid sequence of the FG-Nups on charge distribution, we have modified the sequences of each of the FG-Nups from the native yeast sequence to a homogeneous one, which has the same total number of amino acids of each type as the native sequences but distributed homogeneously along each FG-Nup chain (the homogeneous model sequences in Fig. 1*C* and [Table S3](#) illustrate the composition of each FG-Nup). In Fig. 2 *D–F*, we show the results for the homogeneous case. The volume fraction distributions of all amino acids and their hydrophobic subsets are very similar to those of the native yeast sequences, but the electrostatic potential is much more uniform, and with a much lower absolute value, than in the native case. The highly inhomogeneous electrostatic potential in Fig. 2*C* is thus a result of the charge distribution along the FG-Nups due to their native amino acid sequence.

### Electrostatic and Hydrophobic Interactions Between the Translocating Particle and Pore Are Nonadditive.

Our ultimate goal is to understand how the interplay of different interactions allows the translocation of kap–cargo complexes through the pore and blocks the passage of undesired particles. For this purpose, we decided to model translocating particles with well-defined charge and hydrophobicity. It would be possible to generate a particle to model the charge, volume, and hydrophobic segment distribution of a specific protein or kap–cargo complex. However, such calculations would complicate the final goal of this work of elucidating the role of the different interactions in the translocation process. We thus decided to calculate the energetics of translocation of model spherical particles with different surface properties and a radius of 5 nm.

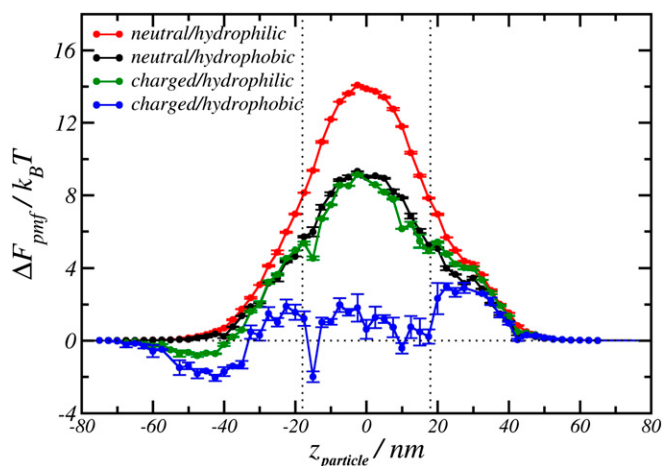
We studied four different particle surfaces: hydrophilic/neutral, hydrophobic/neutral, hydrophilic/charged (with  $-150$  charges per



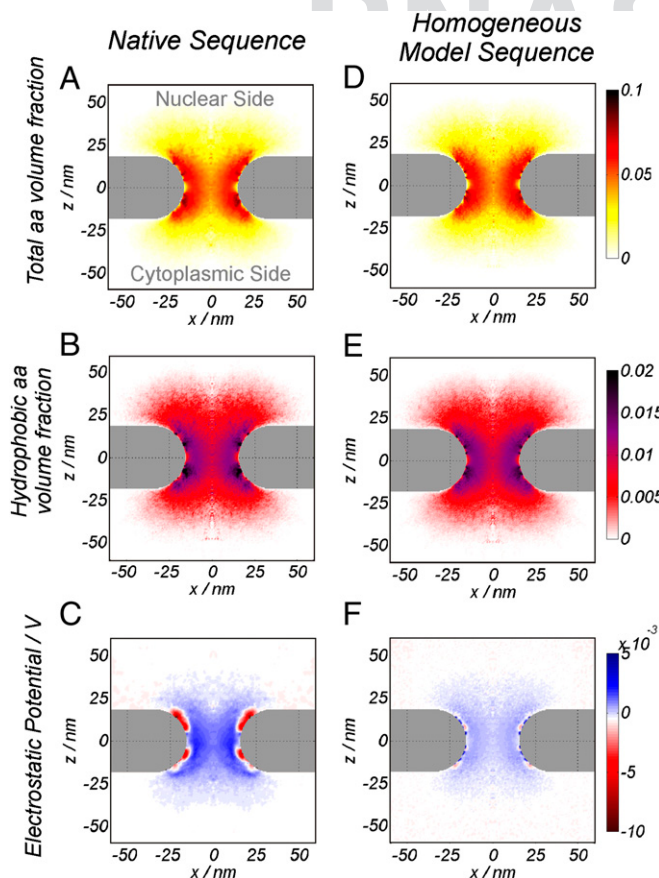
**Fig. 1.** (*A*) Geometry of the model NPC. The pore axis is defined as  $z$ , and the origin is set at the geometrical center of the pore, such that the cytoplasmic and nuclear bulk solutions are located at  $z \rightarrow -\infty$  and  $z \rightarrow \infty$ , respectively. Schematic representations of the amino acid sequences of the FG-Nups for the native (28) (*B*) and homogeneous model (*C*) sequences. The FG-Nups in the homogeneous model sequence contain the same number and type of amino acids as those in the native sequence but in a regular order. The plot shows the different types of amino acids considered in the model: neutral hydrophobic (Hydroph; Ala, Ile, Leu, Phe, Trp, Tyr), Positive (Lys, Arg), Negative (Asp, Glu), Cys, and His (see [Tables S1](#) and [S2](#) and [Fig. S1](#) for model and parameters). For simplicity in the graphical representation, neutral hydrophilic amino acids (Asn, Gln, Gly, Met, Pro, Ser, Thr, Val) are not shown. The figure shows the  $z$ -positions where the chains are anchored to the rigid protein scaffold.

particle), and hydrophobic/charged (also with  $-150$  charges per particle). Our predictions for model cargoes are experimentally testable, for example, by studying the translocation of noble metal nanoparticles and semiconductor quantum dots (QDs) with a well-defined surface chemistry achieved via coating with homogeneous or mixed ligand layers. They are also relevant for the biological problem because they establish the general properties that characterize a translocation-enabled macromolecular complex. The choice of particle charge is based on assuming a charge density of  $-0.5$  charges per square nanometer [for a nanoparticle, this corresponds to  $\sim 1$  charged ligand every 10 ligands on its surface, a reasonable number as measured and predicted for gold nanoparticles (29)].

In Fig. 3, we display the potential of mean force (pmf) as a function of the distance from the center of the particle to the center of the NPC for the four types of translocating particles. The pmf is the effective potential acting on the particle at a given position, averaged over all the degrees of freedom of all the other molecules in the NPC. In other words, the pmf at a given point is the minimal work required to move the translocating particle from the bulk (i.e., very far from the pore) to that point. The pmf is measured with respect to that in the cytoplasmic and nuclear compartments (the pmf in these compartments is the same because both are  $0.15$  M, pH 7.2, 1:1 electrolyte solutions). A particle experiencing a potential barrier several times the thermal energy will be unable to pass through the pore, whereas a particle experiencing a flat potential will translocate through the



**Fig. 3.** Potentials of mean force for model spherical particles ( $R_{\text{particle}} = 5$  nm) at different positions ( $z_{\text{particle}}$ ) along the symmetry axis of the pore (Fig. 1A). Charged particles bear a total charge of  $-150$  |e|. Hydrophobic particles have interaction energies with hydrophilic and hydrophobic segments of  $\epsilon_{\text{particle-hphil}} = 0.16 k_B T$  and  $\epsilon_{\text{particle-hphob}} = 0.27 k_B T$ , respectively ( $\epsilon_{\text{particle-hphil}} = \epsilon_{\text{particle-hphob}} = 0$  for the hydrophilic particle). The vertical dashed lines indicate the position of the entrances to the NPC in the nuclear and cytoplasmic sides. The horizontal dashed line represents zero interaction. Error bars correspond to 1 SD for pmf curves calculated using different sets of chain conformations in the molecular theory (as described in [SI Text](#)).



**Fig. 2.** Molecular organization of the yeast NPC. Total amino acid (aa) volume fraction (A and D), volume fraction of hydrophobic segments (B and E), and electrostatic potential (C and F) for the native and homogeneous model sequences (Fig. 1). The plots show that homogenizing the amino acid sequence affects the electrostatic potential but not the density of amino acids or the density of hydrophobic amino acids.

pore at a rate given by the kinetics of chain rearrangement and/or particle diffusion (modeling the kinetics is beyond the scope of this work).

In Fig. 3, we show that the hydrophilic/neutral translocating particle feels a repulsive (positive pmf) interaction that starts at around 15 nm away from the NPC (the positions of the entrances to the NPC are shown by dashed lines) on the cytoplasmic side and decays on the nuclear side at around 20 nm. The interactions away from the NPC reflect the contribution of FG-Nup conformations that extend away from the pore, as observed in Fig. 24. The repulsive interactions (the only ones relevant for the hydrophilic/neutral particle) arise from two contributions: the osmotic pressure within the pore and the reduction in the number of allowed conformations of the FG-Nups due to the presence of the particle. Henceforth, we will refer to their combination as steric repulsion.

The hydrophobic/neutral nanoparticle (black) curve in Fig. 3 has a shape that is qualitatively very similar to the hydrophilic/neutral curve, with the main difference being the magnitude of the interactions. The weaker repulsion between the NPC and the hydrophobic nanoparticle results from the attractions between the hydrophobic domains of the FG-Nups and the translocating particle. Note, however, that for the strength of hydrophobic interactions used in this calculation, the hydrophobic forces cannot overcome the steric repulsions from the FG-Nups (the effect of the strength of hydrophobic interactions on the pmf is analyzed in [Fig. S2](#)).

The green curve in Fig. 3 shows the pmf acting on the hydrophilic/charged model particle. The curve looks very similar to the hydrophobic/neutral pmf, with one important qualitative difference, namely, that on the cytoplasmic side, we observe the presence of a local minimum. This feature arises from the electrostatic interactions, because there is no observed local minimum for the hydrophobic/neutral and hydrophilic/neutral particles. Interestingly, the attraction first appears when the particle is about 45 nm away from the center of the NPC. Inspection of the electrostatic potential distribution in the NPC in the absence of the nanoparticle (Fig. 2C) shows that the electrostatic potential is almost zero at these distances. Therefore, the attractions arise from the conformational reorganization of the FG-Nups induced by the presence of the negatively charged particle, which attracts the positively

charged FG-Nups at distances from the pore entrance that far exceed the electrostatic screening (about 1 nm for the salt concentration used in this work). In fact, in Fig. S3, we show that inserting a particle on the cytoplasmic side, at  $z = -45$  nm, affects the FG-Nup distribution in a large region between  $-10$  nm  $< z < -60$  nm. Once the hydrophilic/charged translocating particle reaches the region where a relatively high density of the FG-Nups is present (Fig. 2), the pmf becomes repulsive due to the fact that the electrostatic attractions are weaker than the steric repulsions. The quantitative similarity between the black and green curves in Fig. 3 is coincidental, due to the choice of parameters.

A qualitatively different behavior from the other three cases is predicted for the hydrophobic/charged translocating particle (blue curve in Fig. 3). In this case, we see a markedly attractive potential, over 20 nm on the cytoplasmic side, followed by a relatively constant pmf within the NPC, with the exception of the narrow well at around  $-20$  nm and, finally, a repulsive barrier at the exit of the NPC on the nuclear side. An analysis of the different contributions to the pmf (Fig. S4) shows that the narrow well has an electrostatic origin, whereas the repulsive barrier arises from steric and hydrophobic interactions. The effective interaction between the FG-Nups in the NPC and the hydrophobic/charged particle cannot be determined simply from the pmfs of the hydrophilic/charged and hydrophobic/neutral particles (the pmf is nonadditive). For instance, the height of the barrier (maximum of the pmf curve) of the hydrophilic/neutral case is lowered by  $5.0 k_B T$  (where  $k_B T$  is the thermal energy,  $1 k_B T = 2.5$  kJ/mol for  $T = 300$  K) by going to either the hydrophobic/neutral case or the hydrophilic/charged case. However, making the cargo both hydrophobic and charged lowers the barrier by  $12 k_B T$ , which is higher than the sum of the effects of the individual interactions ( $10 k_B T$ ). More importantly, the shape of the pmf acting on the hydrophobic/charged particle is markedly different from that for the hydrophobic/neutral and hydrophilic/charged ones. There is therefore a synergetic effect that arises from the reorganization of the FG-Nups in the pore due to the presence of the translocating particle that depends on the surface properties of the particle. In SI Text, we show systematic calculations of the pmf as a function of hydrophobicity and charge of the translocating particle (Fig. S2). As expected, hydrophobicity and charge have different effects on the pmf. Therefore, particles presenting different surfaces may experience qualitatively different energy landscapes during the translocation process.

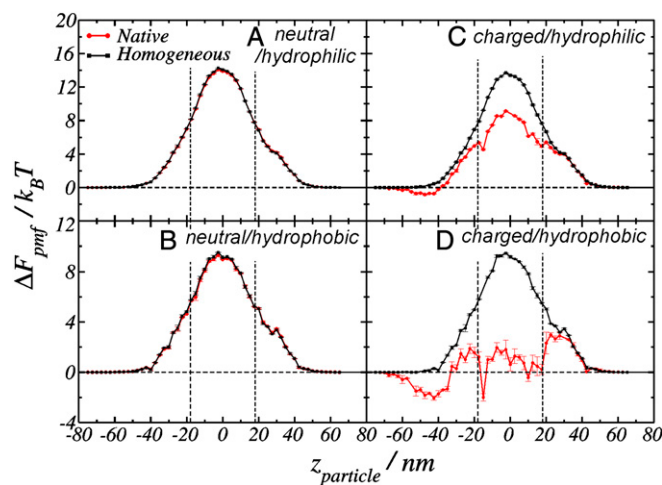
An important conclusion from the pmfs in Fig. 3 for different model particles is that the effective interactions in all these cases cannot be deduced from the knowledge of the spatial organization of the FG-Nups in the NPC in the absence of the translocating objects. In Fig. S3, we show that both the distribution of amino acids and the electrostatic potential within the NPC change on introducing the translocating particle. Another important observation is that the predicted interactions are consistent with experimental observations. Thus, transport receptors (kaps) are negatively charged and hydrophobic (due to hydrophobic pockets on the surface of the kaps) (22), a combination that maximizes the attractive interaction with the pore. Furthermore, a recent study of single-QD tracking in the NPC (17) has found that the kap-capped QDs are attracted to the pore entrance, face a potential barrier on the nuclear side of the channel, and show a Gaussian-like probability distribution (characteristic of a free energy minimum) inside the pore. Inspection of Fig. 3 shows that our results for the pmf acting on the hydrophobic/charged nanoparticle are consistent with these experimental observations. We want to emphasize that a quantitative comparison with the experimental observations has not been attempted in this work, because we did not intend to model the precise charge/hydrophobicity properties of the experimental nanoparticles (e.g., the QDs in the experimental study were modified by the importin- $\beta$  transport receptor).

**Native Sequences of the FG-Nups Are Optimized to Facilitate Transport of Negatively Charged Particles Through the Pore.** To show the further importance of the native amino acid sequence, we have calculated

the pmfs for FG-Nups with a homogeneous model amino acid sequence (corresponding to Figs. 1C and 2D–F). In Fig. 4A and B, we display the pmfs for the hydrophilic/neutral and hydrophobic/neutral model particles for both the native and homogeneous model sequences. The pmfs are essentially identical in both cases. This is not surprising, because these potentials are governed by the steric repulsions and hydrophobic interactions that arise from the distribution of all the amino acids and the hydrophobic ones, respectively; these two distributions are very similar for both types of sequences (Fig. 2A and B vs. Fig. 2D and E). The hydrophilic/charged nanoparticles (Fig. 4C) show important differences in their interactions with the yeast NPC compared with the homogeneous case. For the homogeneous amino acid sequences, the hydrophilic/charged nanoparticles feel a purely repulsive interaction. However, for the native sequence, there is a weak attractive well and the overall potential inside the NPC is lower than that of the homogeneous case. It is clear that these differences arise from the distribution of charged amino acids within the NPC (Fig. 2C vs. Fig. 2F). This difference results in a very dramatic change of the effective interaction potential for the case of hydrophobic/charged translocating particles (Fig. 4D); whereas the homogeneous case presents a purely repulsive pmf, the native NPC shows a complex energy landscape that includes three different regions of the interaction potential as discussed in detail above.

A major conclusion of our work is that transport may be facilitated for negatively charged kap–cargo complexes due not to the overall excess of positive amino acids in the FG-Nups as previously believed (25) (in fact, some FG-Nups have a net negative charge; Table S3) but to the specific charge distribution of the FG-Nups that localizes positive amino acids along the pore axis. It is therefore not surprising that the pmf of the hydrophobic/charged model particles, which bear the closest resemblance to transport receptors and transport receptor–cargo complexes, exhibits the largest contrast between the native and homogeneous model sequences.

**Comparison with Translocation Models.** The most discussed qualitative models for NPC gating are the virtual gate (VG) model (10, 12), the brush model (30), the selective phase (SP) model (11, 23), and the reduction of dimensionality (ROD) model (6, 13). These models differ in the qualitative picture of the distribution of the FG-Nups inside the pore and the mechanism of



**Fig. 4.** Effect of the amino acid sequence on the pmf for different translocating model particles for the NPC bearing the native amino acid sequences (red curves, sequences in Fig. 1B) and the homogenized amino acid sequences (black curves, sequences in Fig. 1C). The vertical dashed lines indicate the position of the entrances to the NPC in the nuclear and cytoplasmic sides. The horizontal dashed line represents zero interaction.

translocation (5, 31). Note that these models provide only a qualitative description of the translocation process, whereas our theory provides quantitative predictions. In the VG model, the FG-Nups block the pore (possibly leaving a narrow channel along the center that allows unhindered translocation of small proteins) and impose an entropic penalty to the passage of large molecules. Interactions between the FG residues and the surface of transport receptors decrease the height of the barrier and allow translocation. The brush model, a variation of the VG model, proposes that the entropic filter is formed by stretched brush-like FG-Nup chains. The FG-Nups in the SP model block the passage through the pore by forming a gel-like structure cross-linked by FG-FG interactions. The main characteristic of the SP model is a gel-like structure, where the mobility of the chains is highly reduced due to transient FG-FG interactions. The transport receptor-cargo complex penetrates the gel by dissociating the cross-links. The ROD model proposes that the FG-Nups form a collapsed layer on pore walls and a channel is formed in the center of the pore. The kap-cargo complexes can slide through the pore by interacting with the FG domains in the collapsed layer (the transport rate is increased by the 1D character of the process).

Our calculations predict that the FG-Nups are close to homogeneously distributed within the pore, in qualitative agreement with both the VG and SP models but not with the ROD model, which postulates a protein-free channel along the pore axis. Note, however, that the ROD model assumes the presence of Kaps within the pore [consistent with experimental findings for the NPC in neuroblastoma cells (32)] that collapse the FG-Nups, as observed in synthetic planar brushes of cys-nup62 (33). In our model, the NPC is free of Kaps and a direct comparison with the ROD model is not possible (to the best of our knowledge, the number of NPC-associated Kaps is not known for yeast). Moreover, in our predictions, the nearly homogeneous distribution of amino acids is not affected (i.e., no collapse to the walls is observed), even when the strength of the hydrophobic interactions in our system is dramatically increased (Fig. S5).

As predicted by our theory, the inner structure of the pore also differs quantitatively from a fully extended brush (proposed in the brush model) because the average end-to-end distance of the grafted chains in our calculation is 6.7-fold smaller than their maximum contour length (Table S4). A further differentiation between the SP and VG models is outside the scope of the theory because the theory makes no predictions about the mobility of the chains. Therefore, we cannot distinguish between the jammed gel-like or fluid brush-like structures proposed by the PS and VG models, respectively. The close to homogeneous amino acid distribution found in our work also disagrees with the extended bundles of FG-Nups observed in MD simulations by Miao and Schulten (22, 34). We attribute this discrepancy to the following:

- i) Difference in the modeled systems: The MD simulations are for a planar surface grafted with 100-aa-long FG-Nups, whereas we model the sequences of the FG-Nups determined for yeast grafted inside a pore that mimics the experimental size and shape of the yeast NPC. We have shown in previous work (27) that increasing chain length and decreasing the radius of curvature of the pore may disfavor the formation of chain aggregates in comparison to the homogeneous system.
- ii) MD simulations in the study by Miao and Schulten (22) are initialized in a fully stretched configuration of the 100-aa model FG-Nups and run for a few microseconds. As those authors point out, the observed morphology may correspond to a kinetically frozen structure instead of an equilibrium structure.
- iii) Our theory neglects inhomogeneities in the angular coordinate, and therefore cannot predict the formation of bundles of chains along this coordinate.

The pmf of a model cargo translocating through the pore predicted by our theory is a free energy criterion that determines what types of objects can be transported. The results of our model agree with previous experimental evidence that suggests (as considered by the VP and SP models) that hydrophobic interactions decrease the translocation energy barrier for kap-cargo complexes. We propose here that the charge distribution along the FG-Nups helps to lower this barrier for negatively charged cargoes. Rout et al. (12) have pointed out that virtual gating can be improved by a nonsymmetrical distribution of FG-domains within the pore. We predict here that the nonsymmetrical distribution of hydrophobic and charged domains and the nonadditivity of interactions give rise to complex translocation potentials, which may help to explain the complex translocation behaviors observed in single-molecule experiments (17).

## Conclusions

In conclusion, we have presented a theoretical study of the structure and translocation of model particles in the yeast NPC system. In this set of calculations, the pore is treated with all the details of the number, sequence, anchoring position, and length of the intrinsic disordered FG-Nups available from experimental observations on yeast. Even so, our model of the FG-Nups is coarse-grained in the sense that it makes no distinction between the hydrophobicities of the different amino acids, does not explicitly incorporate hydrogen bonding, and does not include specific interchain binding [e.g., as observed in the formation of gels with Asn-rich FG sequences (35)]. We have also omitted specific binding interactions between the model particles and the FG domains in the FG-Nups; these interactions have been proposed to play a role in the kap-mediated translocation mechanism (11, 21, 23, 36, 37). The association between the kaps and the FG domains is weak (36, 38) and dynamical (22); therefore, the nature of this interaction is probably between high-affinity ligand-receptor binding and the hydrophobic interaction modeled here. Our work suggests that generic hydrophobicity and negative charge are necessary for translocation of both homogeneous and patched nanoparticles (Fig. S6, in which we show the pmfs for several charge distributions on the surface of the particle). Therefore, it raises the question as to whether highly specific ligand-receptor interactions are needed for successful crossing of the NPC.

Our most important findings are as follows. First, the charge distribution along the FG-Nups (originating from their native sequence) and the tethering position of the FG-Nups create a positively charged environment along the pore axis. This environment could facilitate the passage of the negatively charged kap-cargo complexes. Second, the different interactions in this system are highly nonadditive, and their combination can give rise to complex translocation potentials depending on the properties of the translocating object. This nonadditive behavior results from the reorganization of the flexible FG-Nups in the pore due to the presence of the particle. Moreover, nonadditivity implies that understanding how hydrophobic/neutral and hydrophilic/charged particles interact with the NPC is not enough to conclude how a hydrophobic/charged particle interacts with it. The optimization of the structure is the result of the minimization of the free energy for the overall system and is not reproduced by considering individual contributions only. This is a common theme in soft-matter systems that can change their molecular organization on interaction with the environment, which is also found in many other subcellular systems (e.g., cell membranes).

The mechanism of transport of large molecules through the NPC is complex because it depends on the change in molecular organization of the FG-Nups in the presence of the translocating particle. Our methodology provides a way to study the details of the interactions and the energy landscape within the pore. Future work will focus on the transport of specific protein complexes through the yeast NPC and on transport of kaps through synthetic pores coated with specific molecular modifiers.

## Methods

We model the NPC as a rigid hourglass pore with the dimensions experimentally determined for yeast (28, 39). The disordered proteins, FG-Nups, are tethered to the inner surface of the pore based on the structural model in the study by Alber et al. (28) (additional information is provided in Fig. 1 and *SI Text*). The proteins are modeled with a coarse-grained model that represents each FG-Nup as a freely jointed chain in which each bead models 1 amino acid. Our model accounts for the sequence of each FG-Nup, the anchoring positions of the FG-Nups, and the conformational statistics and appropriate excluded volume of the disordered proteins. The 20 amino acids are divided into six groups: hydrophobic, hydrophilic, positive, negative, cysteine, and histidine. The last two are considered separately due to their  $pK_a$ s and the possibility of charge regulation.

To model the NPC, we use a molecular theory that explicitly considers the shape, size, conformation, charge and charge distribution, and intermolecular and intramolecular interactions of all the molecular species in the system. The theory is formulated by writing down system free energy. In general terms, we write (26, 40, 41):

$$F = -T S_w^{\text{mix}} - T S_A^{\text{mix}} - T S_C^{\text{mix}} - T S_{H^+}^{\text{mix}} - T S_{OH^-}^{\text{mix}} - T S_{\text{FGNup}}^{\text{conf}} + E_{\text{aa-aa}}^{\text{vdW}} + E_{\text{electro}} + F_{\text{acid-base}}^{\text{mix}} + E_{\text{aa-particle}}^{\text{vdW}}$$

where the term  $S_i^{\text{mix}}$  represents the mixing entropy of water molecules, anions, cations, protons and hydroxyl ions ( $i = w, A, C, H^+, \text{ and } OH^-$ , respectively);  $S_{\text{FGNup}}^{\text{conf}}$  is the conformational entropy of the FG-Nup;  $E_{\text{aa-aa}}^{\text{vdW}}$  is the energy of the van der Waals (vdW) attractions between amino acids (the repulsive interactions are modeled in our theory as a packing constraint, as discussed in *SI Text*);  $E_{\text{electro}}$  is the electrostatic energy;  $F_{\text{acid-base}}^{\text{mix}}$  is the free energy associated with the acid-base chemical reactions of the amino acids; and  $E_{\text{aa-particle}}^{\text{vdW}}$  is the energy of the vdW interactions between the amino acids and the translocating particle. Each of these terms is a function of the

distributions of the different molecular species, the probabilities of the different FG-Nup conformations, and the fraction of charged and uncharged amino acids. We find the extremum of  $F$  with respect to these functions to determine the equilibrium structure of the system. In *SI Text*, we present a detailed description of the free energy expression, the molecular model, the minimization procedure, and the numerical calculation. For the pmf calculation, we determine the free energy of the system in the presence of the translocating object and subtract from this value the free energy of the system, where the particle is in the bulk solution (far away from the pore).

**Note Added in Proof.** While this paper was under review, Serdiuk et al. (42) showed that negatively charged sub-3-nm nanoparticles (which lack specific transport receptors on their surface) can translocate through the NPC and localize in the cell nucleus, whereas positively charged nanoparticles of the same size stay in the cytoplasm; these observations are consistent with our predictions.

**ACKNOWLEDGMENTS.** This material is based on work supported as part of the Non-Equilibrium Research Center, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award DE-SC0000989. This research was supported, in part, through the computational resources and staff contributions provided by the Quest high performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. I.S. acknowledges financial support from the National Science Foundation under Emerging Frontiers in Research and Innovation Grant CBET-0937987. Y.R. thanks the US-Israel Binational Science Foundation for financial support. M.K. acknowledges support from the Swiss National Science Foundation through Grant IZ73Z0-128169.

- Fernandez-Martinez J, Rout MP (2012) A jumbo problem: Mapping the structure and functions of the nuclear pore complex. *Curr Opin Cell Biol* 24(1):92–99.
- Elbaum M (2006) Materials science. Polymers in the pore. *Science* 314(5800):766–767.
- Harel A, Forbes DJ (2001) Welcome to the nucleus: CAN I take your coat? *Nat Cell Biol* 3(12):E267–E269.
- Hoelz A, Deblor EW, Blobel G (2011) The structure of the nuclear pore complex. *Annual Review of Biochemistry*, eds Kornberg RD, Raetz CRH, Rothman JE, Thorne JW (Annual Reviews, Palo Alto, CA), Vol 80, pp 613–643.
- Peleg O, Lim RYH (2010) Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. *Biol Chem* 391(7):719–730.
- Peters R (2009) Translocation through the nuclear pore: Kaps pave the way. *Bioessays* 31(4):466–477.
- Terry LJ, Wentz SR (2009) Flexible gates: Dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryot Cell* 8(12):1814–1827.
- Bednenko J, Cingolani G, Gerace L (2003) Nucleocytoplasmic transport: Navigating the channel. *Traffic* 4(3):127–135.
- Görllich D, Mattaj JW (1996) Nucleocytoplasmic transport. *Science* 271(5255):1513–1518.
- Rout MP, et al. (2000) The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J Cell Biol* 148(4):635–651.
- Ribbeck K, Görllich D (2001) Kinetic analysis of translocation through nuclear pore complexes. *EMBO J* 20(6):1320–1330.
- Rout MP, Aitchison JD, Magnasco MO, Chait BT (2003) Virtual gating and nuclear transport: The hole picture. *Trends Cell Biol* 13(12):622–628.
- Peters R (2005) Translocation through the nuclear pore complex: Selectivity and speed by reduction-of-dimensionality. *Traffic* 6(5):421–427.
- Patel SS, Belmont BJ, Sante JM, Rexach MF (2007) Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* 129(1):83–96.
- Yang WD, Gelles J, Musser SM (2004) Imaging of single-molecule translocation through nuclear pore complexes. *Proc Natl Acad Sci USA* 101(35):12887–12892.
- Ma J, Goryaynov A, Sarma A, Yang W (2012) Self-regulated viscous channel in the nuclear pore complex. *Proc Natl Acad Sci USA* 109(19):7326–7331.
- Lowe AR, et al. (2010) Selectivity mechanism of the nuclear pore complex characterized by single cargo tracking. *Nature* 467(7315):600–603.
- Kustanovich T, Rabin Y (2004) Metastable network model of protein transport through nuclear pores. *Biophys J* 86(4):2008–2016.
- Zilman A, Di Talia S, Chait BT, Rout MP, Magnasco MO (2007) Efficiency, selectivity, and robustness of nucleocytoplasmic transport. *PLoS Comput Biol* 3(7):e125.
- Moussavi-Baygi R, Jamali Y, Karimi R, Mofrad MRK (2011) Brownian dynamics simulation of nucleocytoplasmic transport: A coarse-grained model for the functional state of the nuclear pore complex. *PLoS Comput Biol* 7(6):e1002049.
- Mincer JS, Simon SM (2011) Simulations of nuclear pore transport yield mechanistic insights and quantitative predictions. *Proc Natl Acad Sci USA* 108(31):E351–E358.
- Miao L, Schulten K (2010) Probing a structural model of the nuclear pore complex channel through molecular dynamics. *Biophys J* 98(8):1658–1667.
- Ribbeck K, Görllich D (2002) The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *EMBO J* 21(11):2664–2671.
- Bayliss R, et al. (2002) Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats. *EMBO J* 21(12):2843–2853.
- Colwell LJ, Brenner MP, Ribbeck K (2010) Charge as a selection criterion for translocation through the nuclear pore complex. *PLoS Comput Biol* 6(4):e1000747.
- Nap R, Gong P, Szeifer I (2006) Weak polyelectrolytes tethered to surfaces: Effect of geometry, acid-base equilibrium and electrical permittivity. *J Polym Sci B* 44:2638–2662.
- Peleg O, Tagliazucchi M, Kröger M, Rabin Y, Szeifer I (2011) Morphology control of hairy nanopores. *ACS Nano* 5(6):4737–4747.
- Alber F, et al. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450(7170):683–694.
- Wang D, et al. (2011) How and why nanoparticle's curvature regulates the apparent  $pK_a$  of the coating ligands. *J Am Chem Soc* 133(7):2192–2197.
- Lim RYH, et al. (2007) Nanomechanical basis of selective gating by the nuclear pore complex. *Science* 318(5850):640–643.
- Wälde S, Kehlenbach RH (2010) The Part and the Whole: functions of nucleoporins in nucleocytoplasmic transport. *Trends Cell Biol* 20(8):461–469.
- Paradise A, Levin MK, Korza G, Carson JH (2007) Significant proportions of nuclear transport proteins with reduced intracellular mobilities resolved by fluorescence correlation spectroscopy. *J Mol Biol* 365(1):50–65.
- Schoch RL, Kapinos LE, Lim RYH (2012) Nuclear transport receptor binding avidity triggers a self-healing collapse transition in FG-nucleoporin molecular brushes. *Proc Natl Acad Sci USA* 109(42):16911–16916.
- Miao L, Schulten K (2009) Transport-related structures and processes of the nuclear pore complex studied through molecular dynamics. *Structure* 17(3):449–459.
- Ader C, et al. (2010) Amyloid-like interactions within nucleoporin FG hydrogels. *Proc Natl Acad Sci USA* 107(14):6281–6285.
- Bayliss R, et al. (1999) Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP. *J Mol Biol* 293(3):579–593.
- Bayliss R, Littlewood T, Stewart M (2000) Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. *Cell* 102(1):99–108.
- Chailan-Huntington C, Braslavsky CV, Kuhlmann J, Stewart M (2000) Dissecting the interactions between NTF2, RanGDP, and the nucleoporin XFXFG repeats. *J Biol Chem* 275(8):5874–5879.
- Alber F, et al. (2007) The molecular architecture of the nuclear pore complex. *Nature* 450(7170):695–701.
- Tagliazucchi M, de la Cruz MO, Szeifer I (2010) Self-organization of grafted polyelectrolyte layers via the coupling of chemical equilibrium and physical interactions. *Proc Natl Acad Sci USA* 107(12):5300–5305.
- Tagliazucchi M, Rabin Y, Szeifer I (2011) Ion transport and molecular organization are coupled in polyelectrolyte-modified nanopores. *J Am Chem Soc* 133(44):17753–17763.
- Serdiuk T, Alekseev SA, Lysenko V, Skryshevsky VA, Geloën A (2013) Charge-driven selective localization of fluorescent nanoparticles in live cells. *Nanotechnology* 23(31):315101.

# Supporting Information

Tagliazucchi et al. 10.1073/pnas.1212909110

## Theoretical Approach

**Formulation of the Theory.** We divided the nuclear pore complex (NPC) into a rigid scaffold (i.e., the “pore walls”) and the disordered proteins [the FG-nucleoporins (Nups)], which are end-tethered to the inner walls of the scaffold (1, 2). The shape and dimensions of the scaffold are depicted in Fig. 1A. The sequences of the FG-Nups and their grafting positions are shown in Fig. 1B. The system is immersed in a bath containing water molecules (w), protons (H<sup>+</sup>), hydroxyl ions (OH<sup>-</sup>), cations (C), and anions (A). In addition to the FG-Nups and the mobile ions, the system contains a spherical particle of radius  $R_{\text{particle}}$  located at a position  $\mathbf{r}_{\text{particle}}$  (Fig. 1A).

To derive our theoretical framework, we start by writing down the system free energy functional for this system, given by (3):

$$\begin{aligned} \beta F = & \sum_{i=A,C,w} \int \rho_i(\mathbf{r}) [\ln(\rho_i(\mathbf{r})v_w) - 1] d\mathbf{r} \\ & + \sum_{i=H^+,OH^-} \int \rho_i(\mathbf{r}) [\ln(\rho_i(\mathbf{r})v_w) - 1 + \beta\mu_i^0] d\mathbf{r} \\ & + \int \sum_{k=1,N_k} \sigma_k(\mathbf{r}(s)) \sum_{\alpha} P_k(\mathbf{r}(s), \alpha) \ln(P_k(\mathbf{r}(s), \alpha)) ds \\ & + \sum_{l=1,N_l} \sum_{l'=1,N_{l'}} \frac{\beta \varepsilon_{ll'}}{2} \iint g(|\mathbf{r} - \mathbf{r}'|) \langle n_l(\mathbf{r}) \rangle \langle n_{l'}(\mathbf{r}') \rangle d\mathbf{r} d\mathbf{r}' \\ & + \int \left[ \langle \rho_Q(\mathbf{r}) \rangle \beta \psi(\mathbf{r}) - \frac{1}{2} \beta \varepsilon (\nabla \psi(\mathbf{r}))^2 \right] d\mathbf{r} \\ & + \sum_l \int \langle n_l(\mathbf{r}) \rangle \left[ f_l(\mathbf{r}) (\ln(f_l(\mathbf{r})) + \beta\mu_{l,\text{charged}}^0) \right. \\ & \quad \left. + (1 - f_l(\mathbf{r})) (\ln(1 - f_l(\mathbf{r})) + \beta\mu_{l,\text{uncharged}}^0) \right] d\mathbf{r} \\ & + \sum_{l=1,N_l} \beta \varepsilon_{l,\text{particle}} \int g_{\text{particle}}(\mathbf{r}) \langle n_l(\mathbf{r}) \rangle d\mathbf{r}, \end{aligned} \quad [\text{S1}]$$

where  $\beta = 1/k_B T$  and  $\mathbf{r}$  is the position vector.

The first two terms in Eq. S1 represents translational (mixing) entropies of solvent molecules, cations, anions, protons, and hydroxyl ions, where  $\rho_i(\mathbf{r})$  is the number density of species  $i$  at  $\mathbf{r}$  ( $i = A, C, w, H^+$ , and  $OH^-$  for anions, cations, solvent molecules, protons, and hydroxyl ions, respectively) and  $\mu_i^0$  is the standard chemical potential of species  $i$ .

The third term is the conformational entropy of the disordered protein chains, where  $s$  is a parameterization of the membrane area with tethered chains,  $\sigma_k(\mathbf{r})$  is the grafting density of chains of type  $k$  at  $\mathbf{r}$ ,  $P_k(\mathbf{r}, \alpha)$  is the probability of having a chain of type  $k$  anchored at  $\mathbf{r}$  in conformation  $\alpha$ , and  $N_k$  is the number of different FG-Nups in the system. In this term,  $ds$  is the corresponding area element.

The fourth term is the van der Waals (vdW) effective attractive energy between amino acid beads (4), which represents the difference between the segment-segment and segment-solvent vdW attraction energies. In this term,  $\langle n_l(\mathbf{r}) \rangle$  is the number density of the amino acid of type  $l$  ( $N_l$  is the number of different amino acids) at  $\mathbf{r}$ :

$$\langle n_l(\mathbf{r}) \rangle = \int \sum_k \sigma_k(\mathbf{r}(s')) \sum_{\alpha} P_k(\mathbf{r}(s'), \alpha) n_{kl}(\mathbf{r}(s'), \alpha, \mathbf{r}) ds'. \quad [\text{S2}]$$

In Eq. S2,  $n_{kl}(\mathbf{r}(s'), \alpha, \mathbf{r}) d\mathbf{r}$  is the number of amino acids segments of type  $l$  that a protein of type  $k$  tethered at  $\mathbf{r}(s')$  has in

the volume element between  $\mathbf{r}$  and  $\mathbf{r} + d\mathbf{r}$  when it is in conformation  $\alpha$ ,  $ds'$  is the area element, and the integral runs over the area where the FG-Nups are anchored.

The expression  $g(|\mathbf{r} - \mathbf{r}'|)$  is a distance-dependent vdW attractive function of the form:

$$g(|\mathbf{r} - \mathbf{r}'|) = - \left( \frac{a}{|\mathbf{r} - \mathbf{r}'|} \right)^6 \text{ for } a < |\mathbf{r} - \mathbf{r}'| < 1.5\delta \quad [\text{S3}]$$

$$g(|\mathbf{r} - \mathbf{r}'|) = 0 \text{ otherwise,}$$

where  $a$  is the segment length and  $1.5\delta$  is a cutoff parameter. The parameter  $\varepsilon_{ll'}$  (in  $k_B T$  units) determines the strength of the interaction between amino acids of type  $l$  and  $l'$ .

The fifth term in Eq. S1 is the electrostatic contribution to the free energy, where  $\psi(\mathbf{r})$  is the local electrostatic potential,  $\varepsilon$  is the dielectric constant (do not confound with the interaction strength parameter), and  $\langle \rho_Q(\mathbf{r}) \rangle$  is the average charge density at  $\mathbf{r}$ , given by

$$\langle \rho_Q(\mathbf{r}) \rangle = \sum_{i=A,C,H^+,OH^-} \rho_i(\mathbf{r}) q_i + \sum_l \langle n_l(\mathbf{r}) \rangle f_l(\mathbf{r}) q_l + \rho_{Q,\text{particle}}(\mathbf{r}), \quad [\text{S4}]$$

where the sum in the first term on the right hand side (RHS) runs over all charged mobile species,  $q_l$  is the maximum charge of the side chain of the amino acids of type  $l$  (i.e.,  $-1$  for acidic amino acids and  $+1$  for basic amino acids),  $f_l(\mathbf{r})$  is the fraction of amino acids of type  $l$  that are charged at  $\mathbf{r}$  (the summation runs only over chargeable amino acids), and  $\rho_{Q,\text{particle}}(\mathbf{r})$  is the number density of charges that the model particle has at  $\mathbf{r}$  [we assume that these charges do not undergo charge regulation (i.e., they are pH-independent)].

The following contribution to the free energy functional is the mixing entropy between the charged and uncharged states of the side chains bearing acidic and basic functionalities. In this expression,  $\mu_{l,\text{charged}}^0$  and  $\mu_{l,\text{uncharged}}^0$  are the standard chemical potentials of a charged and uncharged amino acid of type  $l$ , respectively, and the summation includes only acidic and basic amino acids. These standard chemical potentials are related to the acid-base equilibrium constants of isolated groups in the bulk by

$$K_{l,a} = C \exp \left[ -\beta\mu_{H^+}^0 - \beta\mu_{l,\text{charged}}^0 + \beta\mu_{l,\text{uncharged}}^0 \right] \quad [\text{S5}]$$

when  $l$  corresponds to an acidic amino acid or

$$K_{l,b} = C \exp \left[ -\beta\mu_{OH^-}^0 - \beta\mu_{l,\text{charged}}^0 + \beta\mu_{l,\text{uncharged}}^0 \right] \quad [\text{S6}]$$

when  $l$  corresponds to a basic amino acid (3). In Eqs. S5 and S6,  $C$  is a constant.

The final term in Eq. S1 represents the attractive vdW interactions between the particle and the FG-Nups. We model that interaction as a well potential; in other words:

$$g_{\text{particle}}(\mathbf{r}) = \begin{cases} -1 & \text{if } R_{\text{particle}} < |\mathbf{r} - \mathbf{r}_{\text{particle}}| < (R_{\text{particle}} + \delta) \\ 0 & \text{otherwise,} \end{cases} \quad [\text{S7}]$$

where  $\mathbf{r}_{\text{particle}}$  is the position of the center of the spherical particle,  $R_{\text{particle}}$  is its radius, and  $\delta$  is the width of the well. The parameter  $\varepsilon_{l,\text{particle}}$  (in  $k_B T$  units) is the depth of the well (i.e., the

strength of the interaction between an amino acid of type  $l$  and the model particle). More specifically,  $\varepsilon_{l,\text{particle}}$  is the free energy difference of replacing the water molecules in a volume  $v_l$  at a distance  $\delta$  from the particle surface by an amino acid of type  $l$  (where  $v_l$  is the molecular volume of an amino acid of type  $l$ ).

The excluded volume repulsive interactions are not explicitly included in Eq. S1. The segment–scaffold, segment–particle, and intramolecular segment–segment interactions are considered in the chain generation process (only conformations that are self-avoiding and do not overlap with the particle or the scaffold are included). We model the intermolecular repulsions through packing constraints at each  $\mathbf{r}$  (3):

$$\sum_{i=A,C,w,H^+,OH^-} \rho_i(\mathbf{r})v_i + \sum_{l=1,N_l} v_l \langle n_l(\mathbf{r}) \rangle = 1, \quad [\text{S8}]$$

where  $v_i$  is the volume of the species  $i$  (anions, cations, water molecules, protons, hydroxyl ions, or amino acids of type  $l$ ).

**Minimization of the Free Energy Functional.** To solve the molecular theory, we minimize the system free energy subjected to the appropriate constraints. The packing constraint, Eq. S8, is enforced by introducing the Lagrange multiplier  $\beta\pi(\mathbf{r})$  [which is a position-dependent osmotic pressure (3, 4)].

The other requirement is the global electroneutrality, given by:

$$\int \langle \rho_Q(\mathbf{r}) \rangle d\mathbf{r} = 0. \quad [\text{S9}]$$

This constraint is enforced by the Lagrange multiplier  $\lambda$ . To fulfill the thermodynamic requirement of constant chemical potential for all the mobile free species everywhere in the system (i.e., at all  $\mathbf{r}$ ), we consider a Legendre transform of the free energy,  $F$ . The presence of two constraints reduces the number of independent thermodynamic variables, and we do not need to consider the chemical potentials of protons, hydroxyls ions, and solvent molecules (3). We thus write a semigrand canonical potential that includes the constraints and chemical potentials of anions and cations:

$$\begin{aligned} \beta W = \beta F + \int \beta \pi(\mathbf{r}) \left[ \sum_{i=A,C,w,H^+,OH^-} \rho_i(\mathbf{r})v_i + \sum_{l=1,N_l} v_l \langle n_l(\mathbf{r}) \rangle - 1 \right] d\mathbf{r} \\ + \lambda \int \langle \rho_Q(\mathbf{r}) \rangle d\mathbf{r} - \beta \mu_C \int \rho_C(\mathbf{r}) d\mathbf{r} - \beta \mu_A \int \rho_A(\mathbf{r}) d\mathbf{r}, \end{aligned} \quad [\text{S10}]$$

and find its extremum with respect to  $\rho_i(\mathbf{r})$  ( $i = w, A, C, H^+, OH^-$ ),  $f_l(\mathbf{r})$ ,  $\psi(\mathbf{r})$ , and  $P_k(\mathbf{r}, \alpha)$ . It can be proven that  $\lambda$ , enforcing the electroneutrality condition, is a constant term added to the electrostatic potential; therefore, it can be included in the choice of boundary conditions (3).

The variation with respect to  $P_k(\mathbf{r}, \alpha)$  leads to the probability distribution function for the protein chains:

$$\begin{aligned} P_k(\mathbf{r}', \alpha) = \frac{1}{\xi_k(\mathbf{r}')} \exp \left[ - \int \left[ \sum_{l=1,N_l} n_{kl}(\mathbf{r}', \alpha, \mathbf{r}) [v_l \beta \pi(\mathbf{r}) \right. \right. \\ \left. \left. + \int \sum_{l'=1,N_{l'}} \beta \varepsilon_{ll'} g(|\mathbf{r} - \mathbf{r}'|) \langle n_{l'}(\mathbf{r}'') \rangle d\mathbf{r}'' + \beta \varepsilon_{l,\text{cargo}} g_{\text{cargo}}(\mathbf{r}) \right] \right. \\ \left. + \sum_l n_{kl}(\mathbf{r}', \alpha, \mathbf{r}) \left[ \ln(f_l(\mathbf{r})) + q_l \beta \psi(\mathbf{r}) + \beta \mu_{l,\text{charged}}^0 \right] d\mathbf{r} \right], \end{aligned} \quad [\text{S11}]$$

where the summation on the third line of the RHS runs only over acidic and basic amino acids. In Eq. S11,  $\xi_k(\mathbf{r}')$  is the partition

function of a chain of type  $k$  tethered at  $\mathbf{r}'$  [normalization constant that ensures  $\sum_\alpha P_k(\mathbf{r}', \alpha) = 1$ ]. Minimization of Eq. S10 with respect to  $\rho_i(\mathbf{r})$  yields the local densities of the anion, cation, proton, hydroxyl, and solvent:

$$\rho_i(\mathbf{r})v_w = \exp[-v_i \beta \pi(\mathbf{r}) + \beta \mu_i - q_i \beta \psi(\mathbf{r})] \quad [\text{S12}]$$

(for  $i = A$  and  $C$ ),

$$\rho_i(\mathbf{r})v_w = \exp[-v_i \beta \pi(\mathbf{r}) - \beta \mu_i^0 - q_i \beta \psi(\mathbf{r})] \quad [\text{S13}]$$

(for  $i = H^+$  and  $OH^-$ ), and

$$\rho_w(\mathbf{r})v_w = \exp[-v_w \beta \pi(\mathbf{r})] \quad [\text{S14}]$$

(for water).

We can rewrite Eqs. S12–S14 by relating the chemical potentials of these species with the bulk densities:

$$\rho_i(\mathbf{r}) = \rho_i^{\text{bulk}} \exp(-v_i \beta [\pi(\mathbf{r}) - \pi^{\text{bulk}}] - q_i \beta [\psi(\mathbf{r}) - \psi^{\text{bulk}}]) \quad [\text{S15}]$$

(for  $i = A, C, H^+$ , and  $OH^-$ ) and

$$\rho_w(\mathbf{r}) = \rho_w^{\text{bulk}} \exp(-v_w \beta [\pi(\mathbf{r}) - \pi^{\text{bulk}}]) \quad [\text{S16}]$$

(for water).

The variation of the potential  $W$ , with respect to  $f_l(\mathbf{r})$  yields:

$$\ln \left( \frac{f_l(\mathbf{r})}{1 - f_l(\mathbf{r})} \right) = -\beta \mu_{H^+}^0 - \beta \mu_{l,\text{charged}}^0 + \beta \mu_{l,\text{uncharged}}^0 - \ln(\rho_{H^+}(\mathbf{r})v_w) \quad [\text{S17}]$$

if  $l$  corresponds to an acidic amino acid or

$$\ln \left( \frac{f_l(\mathbf{r})}{1 - f_l(\mathbf{r})} \right) = -\beta \mu_{OH^-}^0 - \beta \mu_{l,\text{charged}}^0 + \beta \mu_{l,\text{uncharged}}^0 - \ln(\rho_{OH^-}(\mathbf{r})v_w) \quad [\text{S18}]$$

if  $l$  corresponds to a basic amino acid.

These equations can be rewritten to resemble the more familiar acid-base equilibrium equations:

$$K_{l,a}^0 = \rho_{H^+}(\mathbf{r}) \frac{f_l(\mathbf{r})}{1 - f_l(\mathbf{r})} \exp(v_{H^+} \beta \pi(\mathbf{r})) \quad [\text{S19}]$$

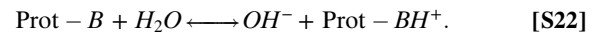
for the dissociation of an acidic amino acid:



or

$$K_{l,b}^0 = \rho_{OH^-}(\mathbf{r}) \frac{f_l(\mathbf{r})}{1 - f_l(\mathbf{r})} \exp(v_{OH^-} \beta \pi(\mathbf{r})) \quad [\text{S21}]$$

for the protonation of a basic amino acid:



Here,  $K_{l,a}^0$  and  $K_{l,b}^0$  are the thermodynamic equilibrium constants for the charging reaction for acidic and basic amino acids, respectively. The latter constants can be multiplied by the factor  $\rho_w^{\text{bulk}}/N_A$  (with  $N_A$  equal to Avogadro's number) to obtain the commonly used equilibrium constants based on molar bulk concentrations (3).

The extremum of  $W$  with respect to  $\psi(\mathbf{r})$  gives rise to a generalized Poisson–Boltzmann equation for the electrostatic potential:

$$\varepsilon \nabla^2 \psi(\mathbf{r}) = -\langle \rho_Q(\mathbf{r}) \rangle, \quad [\text{S23}]$$

where  $\varepsilon$  is the dielectric constant of the system (which is assumed to be  $\mathbf{r}$ -independent)

The boundary conditions for Eq. S23 are:

$$\psi^{bulk} = 0 \quad [\text{S24}]$$

and, at the walls of the scaffold,

$$\nabla_n \psi(\mathbf{r}(\mathbf{s})) = 0, \quad [\text{S25}]$$

where the operator  $\nabla_n$  is the derivative in the direction normal to the surface at point  $\mathbf{r}$ . The condition (Eq. S25) implies that the scaffold walls are uncharged.

**Symmetry Considerations.** Each FG-Nup inside the NPC is present as eight different copies tethered at the same  $z$ -position with an octagonal symmetry (1, 2). We approximate this symmetry by using cylindrical coordinates and assuming that the properties of the system are homogeneous in the angular coordinate (i.e., we assume a  $C_\infty$  rotation axis instead of the real  $C_8$  axis). The properties of the system therefore depend only on  $r$  and  $z$ . This approximation reduces the dimensionality of the numerical problem from three to two dimensions, and thus makes the problem computationally tractable. Based on this approximation, we can integrate  $n_{kl}(\mathbf{r}', \alpha, \mathbf{r})$  over all  $\theta$ :

$$\int_0^{2\pi} n_{kl}(\mathbf{r}', \alpha, \mathbf{r}) r dz dr d\theta = 2\pi r n_{kl}(\mathbf{r}', \alpha, \mathbf{r}) dr dz = n_{kl}^\theta(\mathbf{r}', \alpha, r, z) dr dz, \quad [\text{S26}]$$

where  $n_{kl}^\theta(\mathbf{r}', \alpha, r, z) dr dz$  is defined as the number of segments of type  $l$  that a chain of type  $k$  anchored at  $\mathbf{r}'$  and in conformation  $\alpha$  has in the cylindrical shell segment between  $r + dr$  and  $z + dz$ . Substitution of Eq. S26 into Eq. S2 gives:

$$\langle n_l(r, z) \rangle = \frac{1}{2\pi r} \int \sum_k \sigma_k(\mathbf{r}'(\mathbf{s}')) \sum_\alpha P_k(\mathbf{r}'(\mathbf{s}'), \alpha) n_{kl}^\theta(\mathbf{r}', \alpha, r, z) ds'. \quad [\text{S27}]$$

The integration domain is the inner pore surface; thus, we can write:

$$\langle n_l(r, z) \rangle = \frac{1}{2\pi r} \int_0^{2\pi} \int_0^\infty \int_{-\infty}^\infty \sum_k \sigma_k(\mathbf{r}'(\mathbf{s}')) \times \sum_\alpha P_k(\mathbf{r}'(\mathbf{s}'), \alpha) n_{kl}^\theta(\mathbf{r}', \alpha, r, z) \delta(r' - r_z) r' dz' dr' d\theta', \quad [\text{S28}]$$

where  $r_z$  is the radius of the pore at  $z'$  and  $\delta(x)$  is the Dirac delta function. Each different FG-Nup is expressed as eight copies tethered at a position  $z_k$ . The surface density is thus  $\sigma_k(r', z', \theta') = \delta(z' - z_k) 8 / (2\pi r_k)$ , where  $r_k$  is the radius of the pore at  $z = z_k$ . We can thus write,

$$\langle n_l(r, z) \rangle = \frac{1}{2\pi r} \int_0^{2\pi} \int_0^\infty \int_{-\infty}^\infty \sum_k \frac{8}{2\pi r_k} \times \sum_\alpha P_k(\mathbf{r}'(\mathbf{s}'), \alpha) n_{kl}^\theta(\mathbf{r}', \alpha, r, z) \delta(r' - r_k) \delta(z' - z_k) r' dz' dr' d\theta'. \quad [\text{S29}]$$

Integration over  $r'$ ,  $z'$ , and  $\theta'$  finally yields:

$$\langle n_l(r, z) \rangle = \frac{8}{2\pi r} \sum_k \sum_\alpha P_k(\alpha) n_{kl}^\theta(\alpha, r, z), \quad [\text{S30}]$$

where  $P_k(\alpha)$  is the probability of finding a chain of type  $k$  (tethered at  $r = r_k$  and  $z = z_k$ ) in conformation  $\alpha$ .

The Poisson equation, Eq. S23, in cylindrical coordinates is:

$$\frac{\partial^2 \psi(r, z)}{\partial r^2} + \frac{\partial^2 \psi(r, z)}{\partial z^2} + \frac{1}{r} \frac{\partial \psi(r, z)}{\partial r} = -\frac{\langle \rho_Q(r, z) \rangle}{\varepsilon}. \quad [\text{S31}]$$

**Discretization and Numerical Solution.** To solve the molecular theory, Eqs. S11–S14, S17, and S18 are substituted into the packing constraint and Poisson equations (Eqs. S8 and S23). The resulting equations are discretized in the  $r$  and  $z$  coordinates using cells of dimension  $\delta \times \delta$ , with  $\delta = 0.5$  nm (this value of  $\delta$  was also used in the segment–segment and segment–particle interactions; Eqs. S3 and S7). The number of cells in  $r$  and  $z$  is  $M_r$  and  $M_z$ , respectively. Although the system is infinite in both  $z$  and  $r$  in principle due to the infinite size of the reservoirs, it is necessary only to use  $M_r$  and  $M_z$  large enough to (i) contain every possible chain conformation and (ii) guarantee the asymptotic convergence of the electrostatic potential and the concentrations of the free species to their bulk values. We used the values  $M_r = 500$  and  $M_z = 1,400$  in our calculations (note that only a part of this system is plotted in the color maps in the main text). The equations to be solved are then discretized by replacing integrals by sums, as described in our previous publications. For example, discretization of Eq. S11 yields:

$$P_k(\alpha) = \frac{1}{\xi_k} \exp \left[ -\sum_{j_z=1}^{M_z} \sum_{j_r=1}^{M_r} \left[ \sum_{l=1, N_l} n_{kl}^\theta(\alpha, j_r, j_z) [v_l \beta \pi(j_r, j_z) + \sum_{l'=1, N_{l'}} \left[ \sum_{k_z=1}^{M_z} \sum_{k_r=1}^{M_r} \beta \varepsilon_{ll'} g(|j_z - k_z|, j_r, k_r) \langle n_{l'}(k_r, k_z) \rangle \right] + \beta \varepsilon_{l, \text{cargo}} g_{\text{cargo}}(j_r, j_z) \right] + \sum_l n_{kl}^\theta(\alpha, j_r, j_z) \left[ \ln(f_l(j_r, j_z)) + q_l \beta \psi(j_r, j_z) + \beta \mu_{l, \text{charged}}^0 \right] \right] \quad [\text{S32}]$$

$[n_{kl}^\theta(\alpha, j_r, j_z) = 0$  for cells inside the particle or the scaffold, as discussed in the section on the generation of chains].

In this equation, the discrete indices  $j_r$  and  $j_z$  denote the cylindrical shell section with inner radius  $r = (j_r - 1)\delta$ , outer radius  $r = j_r\delta$ , and delimited by the planes  $z = (j_z - 1)\delta$  and  $z = j_z\delta$ . The coefficients  $g(|j_z - k_z|, j_r, k_r)$  are the discretized functions for  $g(|\mathbf{r} - \mathbf{r}'|)$ . They are obtained by integrating the truncated vdW attractions in the cylindrical shell section (5):

$$g(|j_z - k_z|, j_r, k_r) = \int_{(k_r - 1)\delta}^{k_r\delta} r dr \int_0^{2\pi} d\theta \int_{(j_z - k_z - 1/2)\delta}^{(j_z - k_z + 1/2)\delta} dz h(j_r, z, r, \theta) \quad [\text{S33}]$$

with

$$h(j_r, z, r, \theta) = - \left[ \frac{a}{\left[ (r \cos \theta - (j_r - 1/2)\delta)^2 + (r \sin \theta)^2 + z^2 \right]^{1/2}} \right]^6$$

for

$$a < \left[ (r \cos \theta - (j_r - \frac{1}{2}) \delta)^2 + (r \sin \theta)^2 + z^2 \right]^{1/2} < 1.5 \delta$$

or

$$0 \quad \text{otherwise.} \quad [\text{S34}]$$

The discretization procedure transforms the set of integro-differential equations given by Eqs. S8 and S23 into a system of nonlinear equations solved using standard numerical methods. A typical calculation (i.e., to determine the free energy of the system for a given position of the particle) requires solving  $\sim 2.1 \cdot 10^6$  coupled nonlinear equations with  $\sim 6 \cdot 10^7$  terms per equation. This calculation requires  $\sim 24$  h running on a 20-processor parallel implementation.

### Molecular Model

**Amino Acid Model.** To make the problem computationally tractable, we grouped the amino acids into six categories according to their properties. More specifically, we divided them into neutral hydrophilic amino acids ( $l = 1$ ), neutral hydrophobic amino acids ( $l = 2$ ), amino acids with an acid side chain with  $\text{pK}_a < 7.2$  (Asp and Glu;  $l = 3$ ), amino acids with a basic side chain with  $\text{pK}_a > 7.2$  (Arg and Lys;  $l = 4$ ), amino acid with an acid side chain with  $\text{pK}_a > 7.2$  (Cys;  $l = 5$ ), and amino acid with a basic side chain with  $\text{pK}_a < 7.2$  (His;  $l = 6$ ). Note that amino acids in the third and fourth groups are charged at physiological pH, whereas those in the fifth and sixth groups are not. However, we treat all them on the same basis and consider their acid-base equilibrium properties because the local environment can cause very large shifts in the local pH (3, 6); therefore, an amino acid that would be charged in the bulk solution can be neutral within the pore (or the other way around). We summarize the types of amino acids and their  $\text{pK}_a$  values in Table S1. We consider the volume of all amino acids to be equal to the average volume of  $0.095 \text{ nm}^3$ .

We grouped the vdW interaction coefficients between amino acids ( $\varepsilon_{ll'}$ ) as follows:  $\varepsilon_{ll'} = \varepsilon_{\text{hphil}}$  for interactions between hydrophilic amino acids,  $\varepsilon_{ll'} = \varepsilon_{\text{hphob}}$  for interactions between hydrophobic amino acids, and  $\varepsilon_{ll'} = (\varepsilon_{\text{hphil}} \varepsilon_{\text{hphob}})^{1/2}$  (i.e., Lorentz-Berthelot mixing rule) for the interaction between hydrophilic and hydrophobic amino acids. Table S2 summarizes the interaction coefficients between all amino acid pairs. In our calculations, we chose  $\varepsilon_{\text{hphil}} = 1.0 k_B T$  and  $\varepsilon_{\text{hphob}} = 2.7 k_B T$ . The value for  $\varepsilon_{\text{hphil}}$  is based on the value of  $\varepsilon_{\text{critical}} \sim 0.9 k_B T$ , where  $\varepsilon_{\text{critical}}$  is the  $\varepsilon$  value that is required for microphase separation of a neutral planar brush at vanishing surface coverage, and it is close to the  $\theta$  temperature of the polymer. In other words, we assume that the hydrophilic amino acids are close to the  $\theta$  temperature for the polymer backbone and the hydrophobic amino acids are in a poor solvent. We do not expect that the choice of  $\varepsilon_{\text{hphil}}$  and  $\varepsilon_{\text{hphob}}$  will affect the main conclusions of our work, which, as we explain in the main text, are related to the distribution of charged amino acids within the sequences. For instance, in Fig. S6, we show the effect of  $\varepsilon_{\text{hphil}}$  on the structure and electrostatic potential of the system. The figure shows that the heterogeneous charge distribution is unaffected in the range  $0 < \varepsilon_{\text{hphil}} < 1.5 k_B T$  (in fact, the charge distribution becomes even less homogeneous for  $\varepsilon_{\text{hphil}} > 1.2 k_B T$ ). Future experimental insights on the structure of the NPC will allow us to refine our parameterization of  $\varepsilon_{\text{hphil}}$  and  $\varepsilon_{\text{hphob}}$ . Regarding such parameterization, we note that, in principle, the coefficients  $\varepsilon_{ll'}$  model vdW interactions; however, in complex systems, such as unfolded proteins, they describe the quality of the solvent for the polymer backbone. In other words, any parameterization of  $\varepsilon_{ll'}$  will contain contributions to the quality of the solvent other than vdW forces (e.g., hydrogen bonding or  $\pi$ - $\pi$  stacking), with the

exception of excluded volume and electrostatic interactions, which are explicitly included in our theory.

**Chain Model.** We model the FG-Nups as freely jointed linear chains made of  $N$  hard sphere monomers at fixed bond length. We chose the sequences and grafting positions of the FG-Nups (Fig. 1 and Table S3) according to the yeast NPC model of Rout and colleagues (1, 7), which is the most detailed model in the literature. This model was derived from biophysical and proteomic data, and it is currently the only model that contains information about the localization (anchoring point) of each copy of the FG-chains.

**Chain Generation.** We approximate the sums over all possible chain conformations (i.e., in Eq. S2) by a sum over a representative set of  $3 \cdot 10^6$  conformations for each different FG-Nup. We estimate the error bar in the potential of mean force (pmf) calculations by performing calculations with two to three different sets of chain conformations (we generate different sets of conformations by changing the seed of the random number generation routine in our program). The error bars reported in the pmf curves correspond to 1 SD from the average.

The time required to generate such number of chains for  $151 < n < 857$  is computationally prohibitive using a simple generation routine due to the amino acid-scaffold, amino acid-particle, and intrachain amino acid-amino acid excluded volume requirements; therefore, we have implemented an enriched sampling method.

The chains tethered to the NPC walls and subjected to confinement due to the NPC and particle had been generated monomer by monomer using an enriched samples strategy (8) characterized by two parameters: the length of a partial strand,  $s$ , and the multiplicity, or branching functionality,  $p$ . Within the enrichment idea, we start from simple sampling and counter-balance attrition by adding copies of allowed configurations at a fixed rate. Assume that we already know the attrition would make the sample size shrink as  $\sim \exp(-\lambda n)$ , where  $n$  denotes the number of monomers created following the last branching event. If  $\lambda \approx s^{-1} \ln p$ , we would get a sample of fixed ( $n$ -independent) size if we make  $p$  copies of all surviving chains after each  $s$  newly generated monomer. This can be done breadth-first or depth-first. In the former case, all chains of the entire sample are simulated up to a common length  $n$ ; the entire sample is then copied  $p$  times and the next  $s$  iterations are performed. In a depth-first implementation, only a single configuration is dealt with at each time point and the handling of the copies is done by recursion. This is much faster because it requires fewer memory accesses and much less memory, but one has to know the attrition constant  $\lambda$  in advance. In a breadth-first implementation, on the other hand, the attrition can be learned on the fly, by re-adjusting the sample to a prefixed size at each enrichment step. If attrition is not known a priori, one can still use some learning strategy in a depth-first algorithm. We have combined both methods, use a recursive function call as proposed in by Grassberger (8), and obtain  $\lambda$  once for a given geometry and desired sample size (typically  $S = 10^7 - 10^8$ ); we then use this value in all subsequent runs. The adopted enriched samples method allows the creation of long excluded volume chains in a short time, and it is particularly superior to alternate methods in the presence of strong confinement. Together with the efficient storage of configurations, the method matches our requirements almost perfectly because it is characterized by the most efficient generation and lookup of long chain conformations, whereas it does not spend any time with consuming memory access requests. With minor modification, it allows one to study even longer, and more, chains at either moderate memory access or computational cost.

We consider intrachain self-avoiding interactions during chain generation by considering two amino acids located at positions  $i$  and  $i'$  in the chain to overlap if

$$\begin{aligned} |\mathbf{r}_i - \mathbf{r}_{i'}| < 0.38 \text{ nm} & \text{ if } 2 \leq |i' - i| \leq 5 \\ |\mathbf{r}_i - \mathbf{r}_{i'}| < 0.55 \text{ nm} & \text{ if } |i' - i| > 5 \end{aligned} \quad [\text{S35}]$$

(Fig. S1). With this criterion, the amino acids located less than five bonds apart interact with the average amino acid bond length of 0.38 nm, whereas other amino acids pairs interact with an apparent amino acid diameter of 0.55 nm, which is calculated from the average amino acid volume of 0.095 nm<sup>3</sup> assuming a spherical shape. This strategy of separating between close and far amino acids follows the spirit of the virtual atom molecular mechanics force field model, as described by Korkut and Henrickson (9).

The total number of configurations that can be included is ultimately limited by system's memory. Therefore, we optimized chain storage. We do not need the exact 3D trajectories but their discretized and projected paths on a given 2D regular lattice, because we use the assumption of rotational invariance for the NPC geometry. We have encoded the necessary information about chain configurations, that is, their paths on the 2D lattice ( $r, z$ -space), whereas the underlying conformations are generated off-lattice and remain unaffected by the choice of the lattice. Although the storage of the full trajectories of  $S$  chains with  $N$  monomers requires  $\sim 24 \times N \times S$  bytes (i.e., 96 GB for  $N = 400$  and  $S = 10^7$ ), the storage of their paths on a 2D lattice requires only  $\sim 0.5 \times N \times S$  bytes (i.e., 2 GB for  $N = 400$  and  $S = 10^7$ ).

**Particle Model.** We consider a spherical particle at a position  $z_{\text{particle}}$  and  $r_{\text{particle}} = 0$  ( $r_{\text{particle}}$  cannot take other values because we impose a cylindrical geometry on the system). Ions, solvent, and chains cannot penetrate the particle. The charges of the particle (for nonneutral particles) are homogeneously distributed on its surface. The vdW attractions between the particle and an amino acid segment,  $\varepsilon_{l, \text{particle}}$ , depend on the type of amino acids:  $\varepsilon_{\text{particle}} = h \cdot (\varepsilon_{\text{hphil}} \cdot \varepsilon_{\text{hphob}})^{1/2}$  for hydrophilic amino acids and  $\varepsilon_{\text{particle}} = h \cdot \varepsilon_{\text{hphob}}$  for hydrophobic amino acids, where  $\varepsilon_{\text{hphil}}$  and  $\varepsilon_{\text{hphob}}$  are the strengths of the amino acid-amino acid vdW attractions (section on modeling of amino acids) and  $h$  is a parameter that controls the hydrophobicity of the particle, which is zero for hydrophilic particles and 0.1 for the hydrophobic particle (except for the calculations shown in Fig. S2, where we vary this variable systematically).

### Effects of Hydrophobicity and Charge on the pmf Are Nonequivalent, and Their Combination Produces Complex Translocation Potentials

In this section, we address the effects of the strength of the hydrophobic interactions and of the total charge of a hydrophobic/charged particle on its effective interaction with the NPC. Two cases are considered: one with a fixed strength of hydrophobic interactions and varying charge (Fig. S2A) and another with a fixed number of charges and varying strength of the hydrophobic interactions (Fig. S2B). In the case of varying the nanoparticle charge, there is no change in the height of the barrier on the nuclear side of the NPC (around  $z = 30$  nm). This result and the analysis of the different contributions to the pmf (Fig. S4) suggest that this barrier is governed mainly by steric and vdW contributions. On the other hand, the pmf at the center of the pore and on the cytoplasmic side is affected by both charge and hydrophobic interactions. There is also a sharp minimum around  $z_{\text{cargo}} = -20$  nm, which increases in magnitude for increasing charge of the translocating particle. This minimum is absent in the pmf for the homogeneous model sequence (Fig. 4D); thus, we attribute it to electrostatic interactions between the positively charged terminal segments of the FG-Nups and the negatively charged model particle (Fig. S4).

An important conclusion from Fig. S2A and B is that the effects of increasing hydrophobic or electrostatic interactions on the shape of the pmf are nonequivalent. The combination of these

interactions gives rise to the complex potential energy landscapes observed in these figures. This unique concept indicates that the pmf experienced by kap-cargo complexes with different charge and hydrophobicity properties can be different. For example, according to our calculations, a highly hydrophobic translocating object will get trapped in the center of the pore (blue curve in Fig. S2B), whereas a highly negative one will be attracted to the entrance on the cytoplasmic side, will feel an approximately constant potential inside the pore (with the exception of the sharp minimum at approximately  $-20$  nm), and will be blocked at the nuclear side exit (blue curve in Fig. S24). Note that in vivo, kap-cargo complexes may be released from the traps inside the NPC by binding Ran-GTP, which dissociates the complex and releases the cargo, which can then diffuse into the nucleus (10, 11); this mechanism is not incorporated in our model.

### Molecular Organization Within the NPC Is Modified by the Presence of the Model Particle

The molecular organization within the NPC is modified by the presence of the model particle, as shown in Fig. S3.

### Contributions to the pmf

We have separated the 10 different contributions to the free energy (Eq. S1) into three components and plotted them in Fig. S4 for the translocation of a hydrophobic/charged particle. These three components can be defined as steric, hydrophobic, and electrostatic (with the latter containing the bare electrostatic, ion entropy, and acid-base equilibrium):

$$\begin{aligned} \beta F_{\text{steric}} = & \int \rho_i(\mathbf{r}) [\ln(\rho_w(\mathbf{r})v_w) - 1] d\mathbf{r} \\ & + \int \sum_{k=1, N_k} \sigma_k(\mathbf{r}(s)) \sum_{\alpha} P_k(\mathbf{r}(s), \alpha) \ln(P_k(\mathbf{r}(s), \alpha)) ds \end{aligned} \quad [\text{S36}]$$

$$\begin{aligned} \beta F_{\text{electrostatic}} = & \sum_{i=A, C} \int \rho_i(\mathbf{r}) [\ln(\rho_i(\mathbf{r})v_w) - 1] d\mathbf{r} \\ & + \sum_{i=H^+, OH^-} \int \rho_i(\mathbf{r}) [\ln(\rho_i(\mathbf{r})v_w) - 1 + \beta \mu_i^0] d\mathbf{r} \\ & + \int \left[ \langle \rho_Q(\mathbf{r}) \rangle \beta \psi(\mathbf{r}) - \frac{1}{2} \beta \varepsilon (\nabla \psi(\mathbf{r}))^2 \right] d\mathbf{r} \\ & + \sum_l \int \langle n_l(\mathbf{r}) \rangle \left[ f_l(\mathbf{r}) (\ln(f_l(\mathbf{r})) + \beta \mu_{l, \text{charged}}^0) \right. \\ & \left. + (1 - f_l(\mathbf{r})) (\ln(1 - f_l(\mathbf{r})) + \beta \mu_{l, \text{uncharged}}^0) \right] d\mathbf{r} \end{aligned} \quad [\text{S37}]$$

$$\begin{aligned} \beta F_{\text{Hydrophobic}} = & \sum_{l=1, N_l} \sum_{l'=1, N_{l'}} \frac{\beta \varepsilon_{ll'}}{2} \iint g(|\mathbf{r} - \mathbf{r}'|) \langle n_l(\mathbf{r}) \rangle \langle n_{l'}(\mathbf{r}') \rangle d\mathbf{r} d\mathbf{r}' \\ & + \sum_{l=1, N_l} \beta \varepsilon_{l, \text{cargo}} \int g_{\text{cargo}}(\mathbf{r}) \langle n_l(\mathbf{r}) \rangle d\mathbf{r} \end{aligned} \quad [\text{S38}]$$

### Effect of $\varepsilon_{\text{hphil}}$ on the Total Volume Fraction of Amino Acids, the Volume Fraction of Hydrophobic Amino Acids, and the Charge Distribution in the System

The effect of  $\varepsilon_{\text{hphil}}$  on the total volume fraction of amino acids, the volume fraction of hydrophobic amino acids, and the charge distribution in the system are shown in Fig. S4.

### End-to-End Distance of the FG-Nups Is Much Smaller Than Their Contour Length

We determined the contour lengths in Table S4 as  $N \cdot l$ , where  $N$  is the number of amino acids in the FG-Nup and  $l$  is the segment

length ( $l = 0.38$  nm). The average end-to-end distance for the FG-Nup,  $k$ , was determined as  $\sum_{\alpha} P_k(\alpha) r_{\text{end-to-end}}(k, \alpha)$ , where  $r_{\text{end-to-end}}(k, \alpha)$  is the end-to-end distance of the conformation  $\alpha$  of chain  $k$ . In those cases in which a given FG-Nup is tethered to more than one position within the system (Table S3), we averaged the average end-to-end distances of all copies of that FG-Nup and reported that value in Table S4.

### Effect of Charge Distribution on the pmf

We studied the effect of the distribution of charges on the particle surface on the pmf for particles with a net charge of  $-150$   $|e|$ . We examined three different charge distributions: (i)  $-150$  charges

homogeneously distributed on the entire particle surface, (ii)  $-150$  charges distributed homogeneously on the surface of one hemisphere and zero charges in the other hemisphere, and (iii)  $-200$  charges distributed homogeneously on the surface of one hemisphere and  $+50$  charges distributed homogeneously on the surface of the other hemisphere. Fig. S6 shows the pmf for particles with hydrophobic and hydrophilic surfaces and the native or homogeneous model sequences (in all cases, the hydrophobicity is uniformly distributed over the surface of the particle). The main conclusion of these calculations and the results shown in Fig. 3 is that the effect of charge distribution on the pmf is of minor importance compared with the effect of the total charge.

1. Alber F, et al. (2007) The molecular architecture of the nuclear pore complex. *Nature* 450(7170):695–701.
2. Peleg O, Lim RYH (2010) Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. *Biol Chem* 391(7):719–730.
3. Nap R, Gong P, Szleifer I (2006) Weak polyelectrolytes tethered to surfaces: Effect of geometry, acid-base equilibrium and electrical permittivity. *J Polym Sci B* 44:2638–2662.
4. Szleifer I, Carignano MA (1996) Tethered polymer layers. *Adv Chem Phys* 96:165–260.
5. Peleg O, Tagliazucchi M, Kröger M, Rabin Y, Szleifer I (2011) Morphology control of hairy nanopores. *ACS Nano* 5(6):4737–4747.
6. Tagliazucchi M, Azzaroni O, Szleifer I (2010) Responsive polymers end-tethered in solid-state nanochannels: When nanoconfinement really matters. *J Am Chem Soc* 132(35):12404–12411.
7. Alber F, et al. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450(7170):683–694.
8. Grassberger P (1997) Pruned-enriched Rosenbluth method: Simulations of theta polymers of chain length up to 1,000,000. *Phys Rev E* 56(3):3682–3693.
9. Korkut A, Hendrickson WA (2009) A force field for virtual atom molecular mechanics of proteins. *Proc Natl Acad Sci USA* 106(37):15667–15672.
10. Görlich D, Panté N, Kutay U, Aebersold U, Bischoff FR (1996) Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J* 15(20):5584–5594.
11. Lowe AR, et al. (2010) Selectivity mechanism of the nuclear pore complex characterized by single cargo tracking. *Nature* 467(7315):600–603.

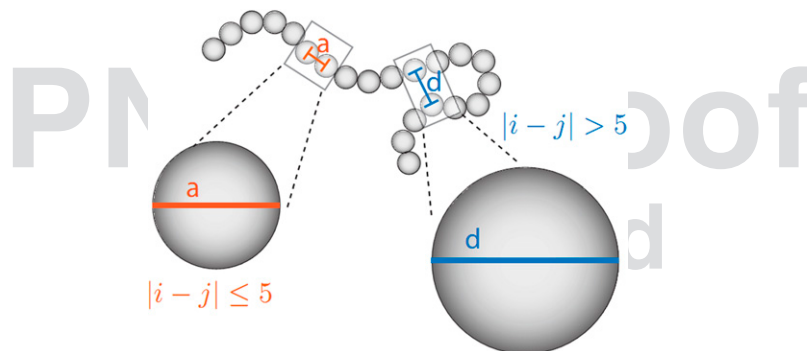
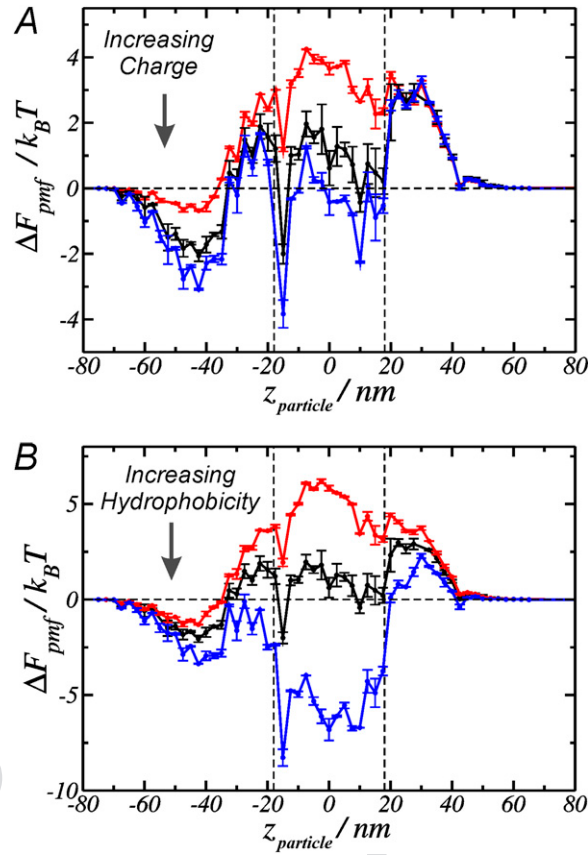
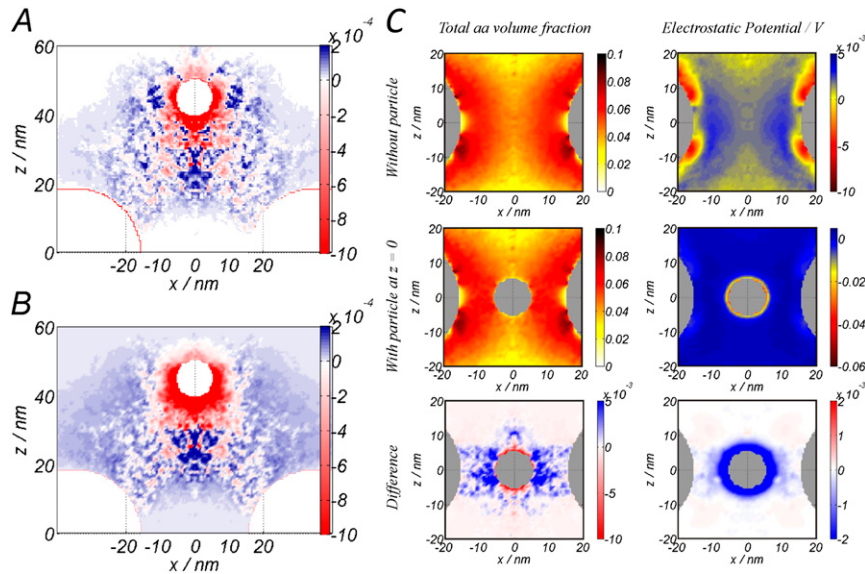


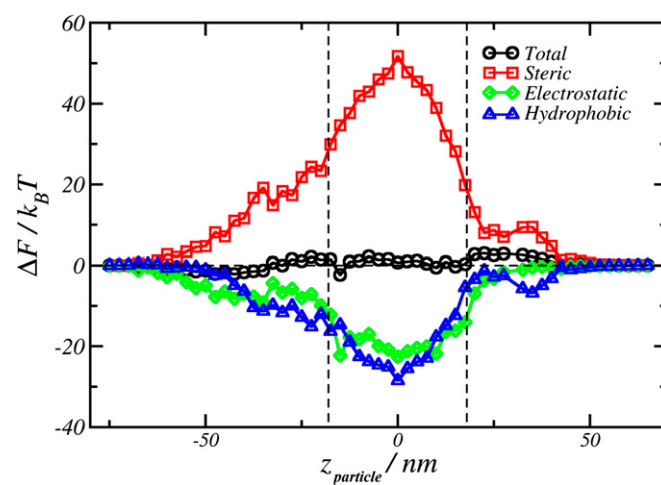
Fig. S1. Overlapping criteria using the chain generation procedure (as defined in Eq. S35),  $a = 0.38$  nm,  $d = 0.55$  nm.



**Fig. S2.** Effect of particle charge (A) and hydrophobicity (B) on the pmf. The particles in A have interaction energies of  $\epsilon_{\text{particle-hphil}} = 0.16 k_B T$  (hydrophilic amino acid/particle interaction) and  $\epsilon_{\text{particle-hphob}} = 0.27 k_B T$  (hydrophobic amino acid/particle interaction) and charges of  $-100 |e|$  (red curve),  $-150 |e|$  (black curve), and  $-175 |e|$  (blue curve). The particles in B have a charge of  $-150 |e|$  and interaction energies that are 50% (red curve), 100% (black curve), and 150% (blue curve) of those used in A. The vertical dashed lines indicate the position of the outer and inner NPC walls. The horizontal dashed line represents zero interaction. Error bars correspond to 1 SD for pmf curves calculated using different sets of chain conformations in the molecular theory.

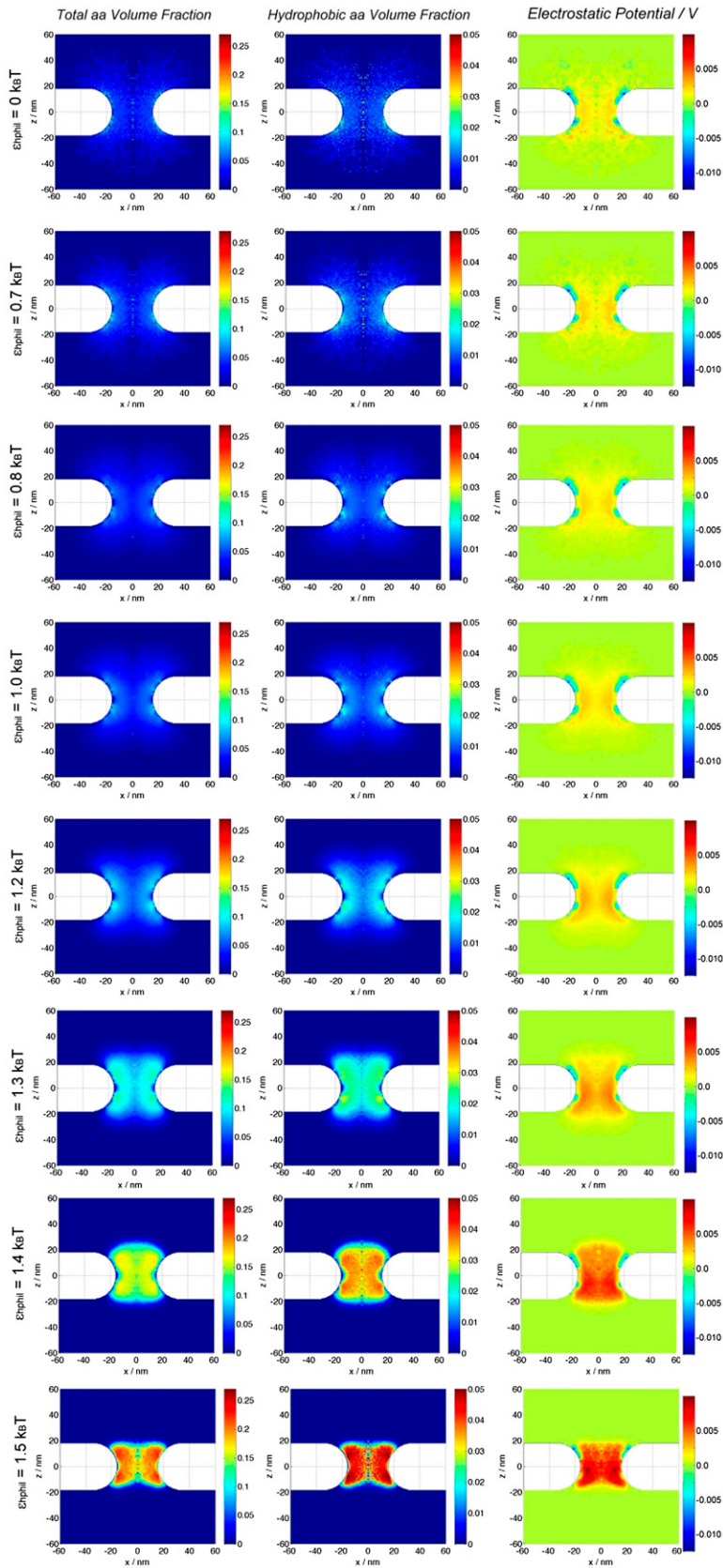


**Fig. S3.** Presence of a hydrophobic/charged particle reorganizes the structure of the NPC. Difference between the volume fraction of the amino acids in the presence and absence of a hydrophobic/charged particle (A) and a hydrophilic/charged particle (B); the spherical particle is centered at  $z = 45$  nm. Positive (blue) and negative (red) values indicate an increase and decrease in density, respectively, on introducing the particle. (C) First two rows show the volume fraction profile and electrostatic potential in the absence and presence of a charged/hydrophobic particle. The third row shows the difference between the calculations in the presence of the particle and those without the particle.

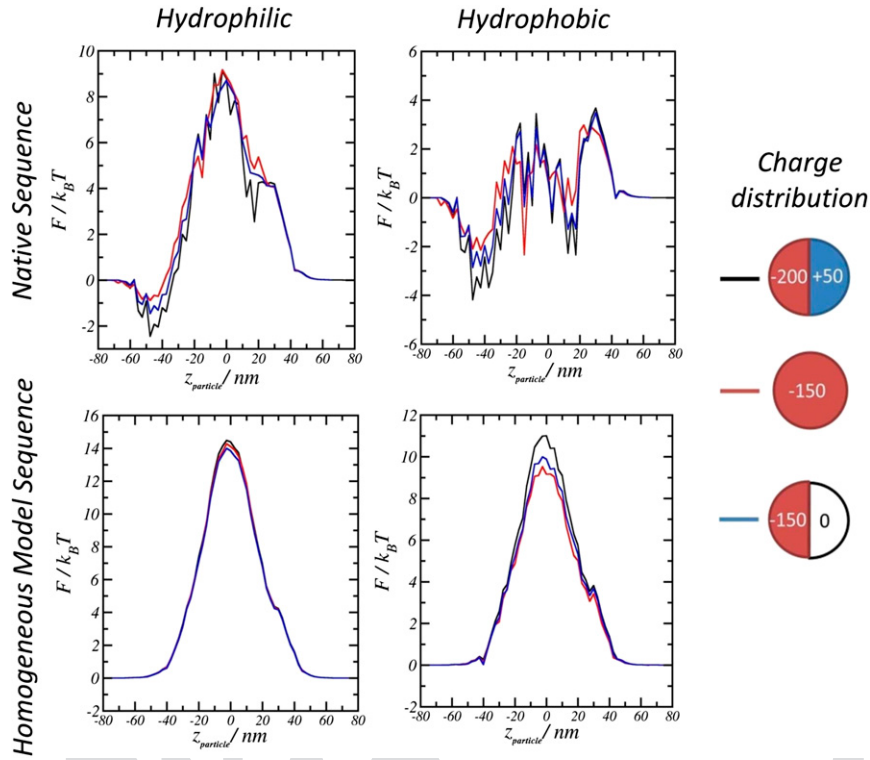


**Fig. S4.** Contributions to the pmf (grouped according to Eqs. S36–S38) for the translocation of a hydrophobic/charged particle.

PNAS proof  
Embargoed



**Fig. S5.** Total amino acid (aa) volume fraction, volume fraction of hydrophobic segments, and electrostatic potential for different values of the vdW attraction strength between hydrophilic segments,  $\epsilon_{\text{hphil}}$ .



**Fig. S6.** Potentials of mean force for the translocation of particles with different charge distributions and hydrophilic (Left) or hydrophobic (Right) surfaces through NPCs with the native (Upper) or homogeneous model (Lower) FG-Nup sequences. The interaction energies between hydrophilic/hydrophobic segments and the particle were the same as in Fig. 3.

**Table S1.** Classification of amino acids into different groups and their acid-base and solubility properties

|         | Amino acids                            | Side chain acid-base properties | Solubility property |
|---------|----------------------------------------|---------------------------------|---------------------|
| $l = 1$ | Asn, Gln, Gly, Met, Pro, Ser, Thr, Val | Neutral                         | Hydrophilic         |
| $l = 2$ | Ala, Ile, Leu, Phe, Trp, Tyr           | Neutral                         | Hydrophobic         |
| $l = 3$ | Asp, Glu                               | Acid, $pK_a = 4$                | Hydrophilic         |
| $l = 4$ | Lys, Arg                               | Base, $pK_a = 11$               | Hydrophilic         |
| $l = 5$ | Cys                                    | Acid, $pK_a = 8.3$              | Hydrophilic         |
| $l = 6$ | His                                    | Base, $pK_a = 6.08$             | Hydrophilic         |

**Table S2.** Values of the interaction coefficients,  $\varepsilon_{ll'}$

|          | $l = 1$                                                 | $l = 2$                                                 | $l = 3$                                                 | $l = 4$                                                 | $l = 5$                                                 | $l = 6$                                                 |
|----------|---------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|
| $l' = 1$ | $\varepsilon_{hphil}$                                   | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   |
| $l' = 2$ | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $\varepsilon_{hphob}$                                   | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ |
| $l' = 3$ | $\varepsilon_{hphil}$                                   | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   |
| $l' = 4$ | $\varepsilon_{hphil}$                                   | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   |
| $l' = 5$ | $\varepsilon_{hphil}$                                   | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   |
| $l' = 6$ | $\varepsilon_{hphil}$                                   | $(\varepsilon_{hphil} \cdot \varepsilon_{hphob})^{1/2}$ | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   | $\varepsilon_{hphil}$                                   |

**Table S3. Properties of the FG-Nups considered in the calculation**

| FG-Nup | Z, nm*         | FG-domain <sup>†</sup> | Length | Positive | Negative | Positive-negative | Hydrophobic | Neutral     |     |     |
|--------|----------------|------------------------|--------|----------|----------|-------------------|-------------|-------------|-----|-----|
|        |                |                        |        |          |          |                   |             | hydrophilic | Cys | His |
| Nup42  | 13             | 1–382                  | 382    | 13       | 2        | 11                | 90          | 277         | 0   | 0   |
| Nup159 | 13             | 387–1071               | 685    | 62       | 100      | –38               | 133         | 380         | 0   | 10  |
| Nup116 | 13             | 172–960                | 789    | 47       | 36       | 11                | 174         | 530         | 0   | 2   |
| Nup100 | 10             | 1–800                  | 800    | 38       | 26       | 12                | 161         | 571         | 0   | 4   |
| Nup49  | –4, 4          | 1–251                  | 251    | 7        | 0        | 7                 | 59          | 184         | 1   | 0   |
| Nup57  | –4, 4          | 1–255                  | 255    | 7        | 0        | 7                 | 50          | 198         | 0   | 0   |
| Nup145 | –5, –14        | 1–433                  | 433    | 31       | 23       | 8                 | 103         | 275         | 0   | 1   |
| Nup53  | –7, 7          | 41–267                 | 227    | 21       | 25       | –4                | 59          | 120         | 0   | 2   |
| Nup59  | –8, 8          | 1–206                  | 206    | 16       | 12       | 4                 | 38          | 136         | 0   | 4   |
| Nsp1   | –11, –7, 7, 11 | 1–617                  | 617    | 73       | 65       | 8                 | 140         | 339         | 0   | 0   |
| Nup1   | –12            | 220–1,076              | 857    | 91       | 72       | 19                | 181         | 508         | 0   | 5   |
| Nup60  | –14            | 389–539                | 151    | 18       | 21       | –3                | 32          | 79          | 0   | 1   |

Length is measured as the number of amino acids. The number of hydrophobic amino acids is the number of Ala, Ile, Phe, Trp, and Val.

\*Anchoring point coordinate Z was defined according to the data presented in figure S26 of ref. 1.

<sup>†</sup>The FG-domains are defined as in the study by Yamada et al. (2), other than Nup53 and Nup59, which were defined as the largest contiguous sequence of amino acids containing the FG motifs FxFG, GLFG, FG, and FxF, as well as their reverse sequences (using this order of priority in case of overlaps), separated by ~100 amino acids and including 10 additional amino acids at the termini of the domain.

1. Alber F, et al. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450(7170):683–694.
2. Yamada J, et al. (2010) A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol Cell Proteomics* 9(10):2205–2224.

**Table S4. Contour length and end-to-end distance for the FG-Nups in the system**

| FG-Nup | Contour length, nm | Average end-to-end distance, nm |
|--------|--------------------|---------------------------------|
| Nup42  | 145.1              | 21.6                            |
| Nup159 | 260.3              | 31.5                            |
| Nup116 | 299.8              | 35.5                            |
| Nup100 | 304                | 37.6                            |
| Nup49  | 95.4               | 15.5                            |
| Nup57  | 96.9               | 15.9                            |
| Nup145 | 164.5              | 24.3                            |
| Nup53  | 86.3               | 15.0                            |
| Nup59  | 78.3               | 14.2                            |
| Nsp1   | 234.5              | 30.8                            |
| Nup1   | 325.7              | 35.8                            |
| Nup60  | 57.4               | 11.1                            |