

## SI Appendix

### HIV-1 Capsid Shape, Orientation, and Entropic Elasticity Regulate Translocation into the Nuclear Pore Complex

Arpa Hudait and Gregory A. Voth<sup>#</sup>

Department of Chemistry, Chicago Center for Theoretical Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, IL 60637, USA.

<sup>#</sup>Corresponding author: gavoth@uchicago.edu

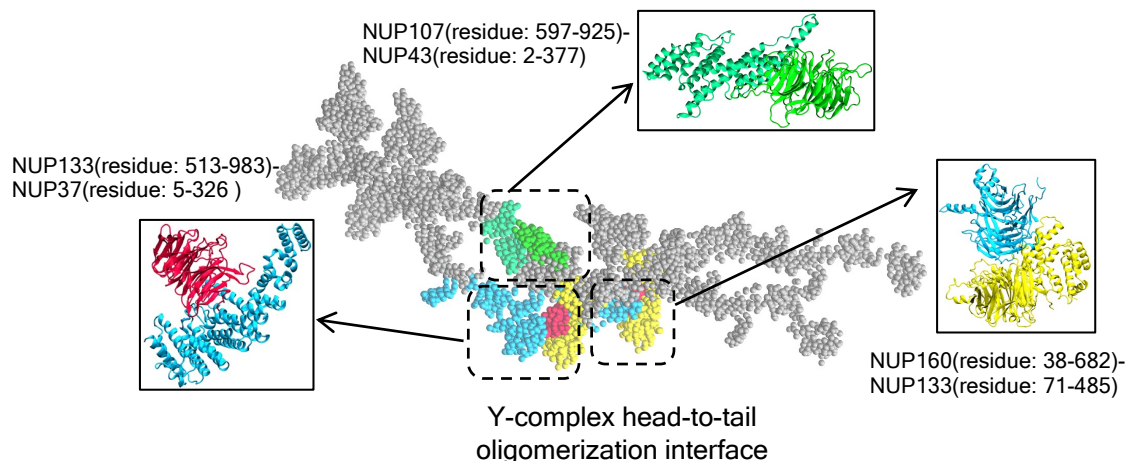
## SI Appendix Movie Legends

**SI Movie 1.** Translocation of the cone-shaped capsid into the Nuclear Pore Complex central channel.

**SI Movie 2.** Translocation of the pill-shaped capsid into the Nuclear Pore Complex central channel.

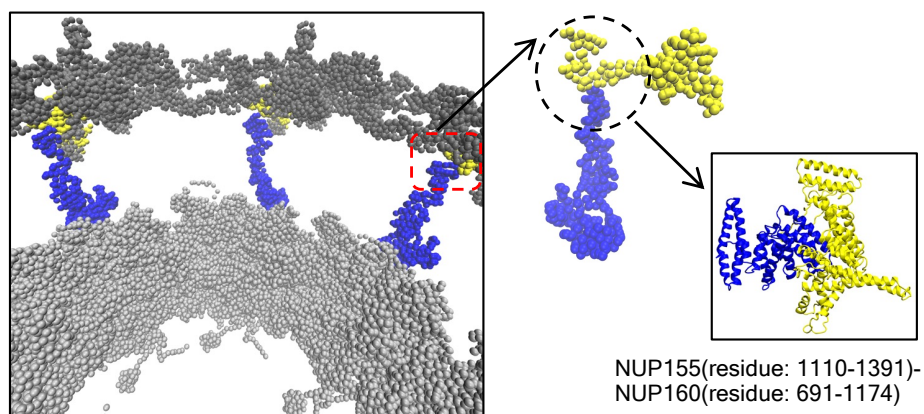
## SI Appendix Figures

### CG Y-complex dimer



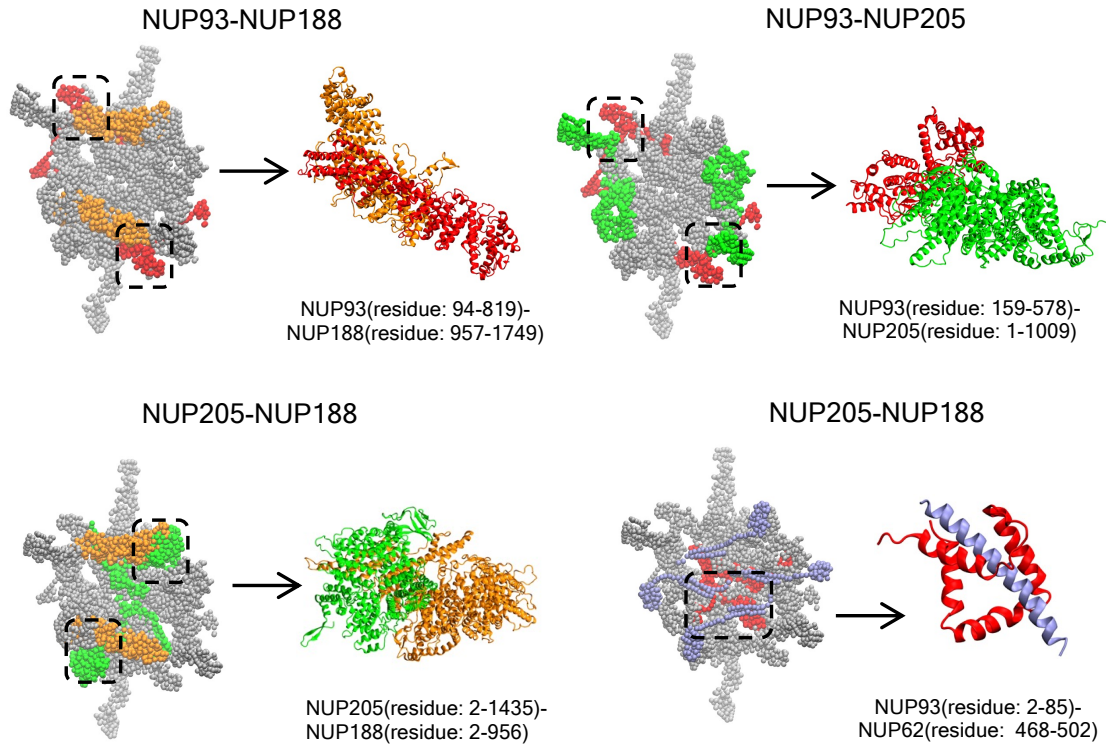
**Fig. S1. Representative depiction of the head-to-tail oligomerization interface between two adjacent Y-complex dimers.** The location of the binding interface between NUP160-NUP133, NUP133-NUP37, and NUP107-NUP43 complexes in the Y-complex dimer are labeled. The corresponding atomistic heterodimer complexes from which the CG molecular interactions are derived are shown in cartoon representation in the inset. Each atomistic heterodimer complex was evolved for 1000 ns to derive the corresponding CG molecular interactions. The NUPs at the binding interface are represented in the same color in Fig. 1 in the main manuscript. The rest of the CG beads of the Y-complex dimer is shown in gray spheres.

### Inner-Outer ring binding interface

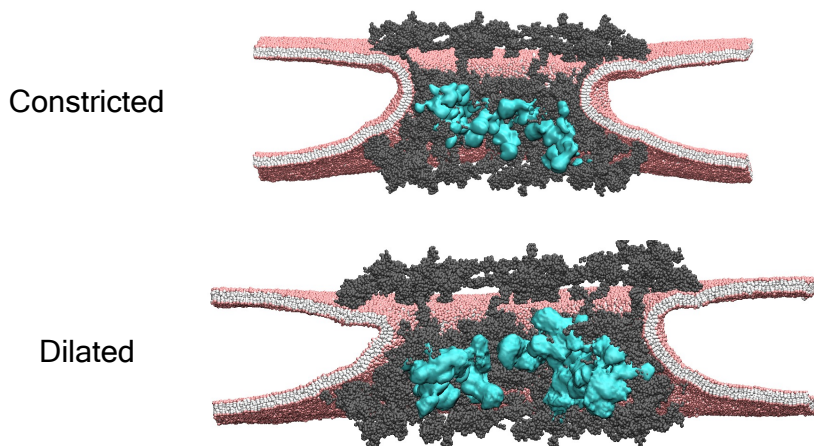


**Fig. S2. Representative depiction of the inner-outer ring binding interface.** 16 copies of NUP155 (blue spheres) are vertically oriented to bind NUP160 (yellow spheres). 8 copies of NUP155 form a binding interface with NUP133 at the cytoplasmic and nuclear rings, respectively. The atomistic NUP155-NUP160 complex from which the CG non-bonded attractive interactions are derived is shown in cartoon representation in the inset. The atomistic NUP155-NUP133 heterodimer complex was simulated for 1000 ns to derive the corresponding CG molecular interactions. The NUPs at the binding interface are represented in the same color in Fig. 1 in the main manuscript. The rest of the CG beads are shown in gray spheres.

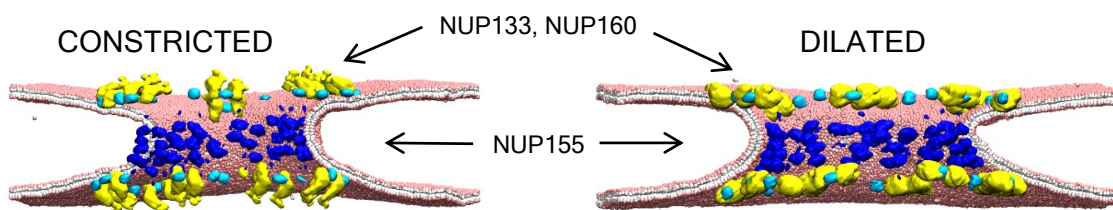
## Inner ring binding interfaces



**Fig. S3. Representative depiction of the inner ring binding interfaces.** The location of the binding interface between NUP93-NUP188, NUP93-NUP205, NUP205-NUP188, and NUP93-NUP62 in an inner ring spoke is shown (left panel). The corresponding atomistic heterodimer complexes from which the CG molecular interactions are derived is shown in the right panel in the cartoon representation. Each atomistic heterodimer complex was simulated for 1000 ns to derive the corresponding CG molecular interactions. The NUPs at the binding interface are represented in the same color in Fig. 1 in the main manuscript. The rest of the CG beads of the inner ring spoke are shown in gray spheres.

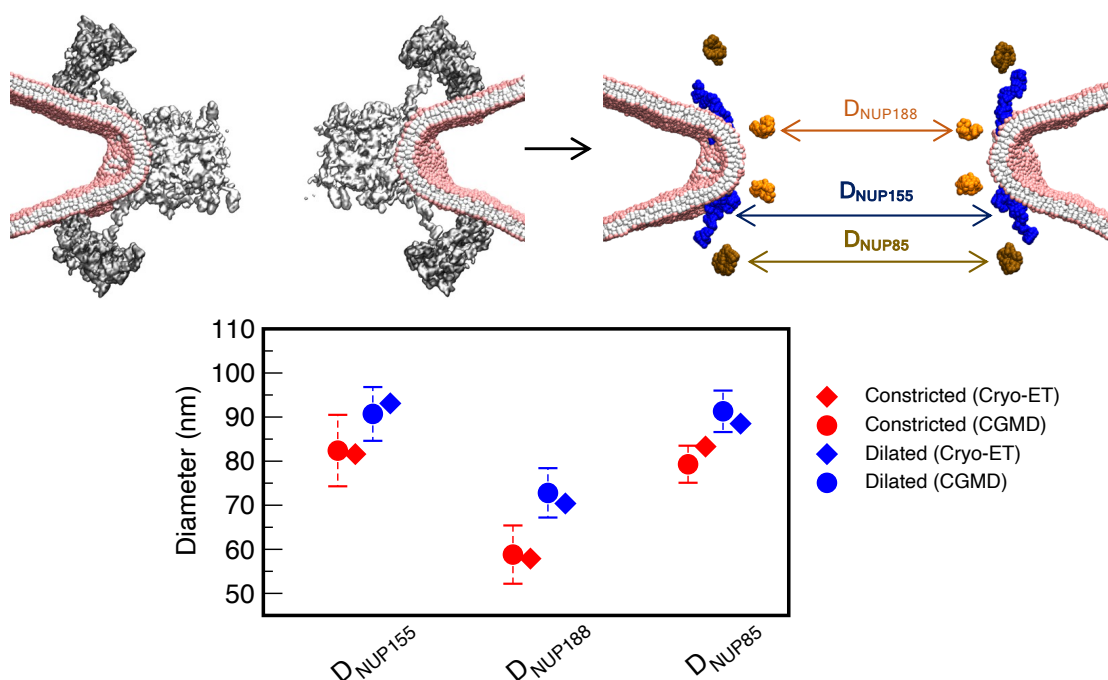


**Fig. S4. Time-averaged densities of the FG-NUPs of the NPC.** The density of the NUP54-58-62 heterotetramer complex were calculated from the final  $150 \times 10^6 \tau_{CG}$  of the CG MD simulations. The averaged density of the NUP54-58-62 heterotetramer complex is shown in cyan isosurface overlaid on the NPC. Rest of the NPC components are shown in gray spheres. The headgroup of the 4-site CG lipid is shown in pink spheres. The interfacial bead and two tail beads of the CG lipid are shown in white spheres.

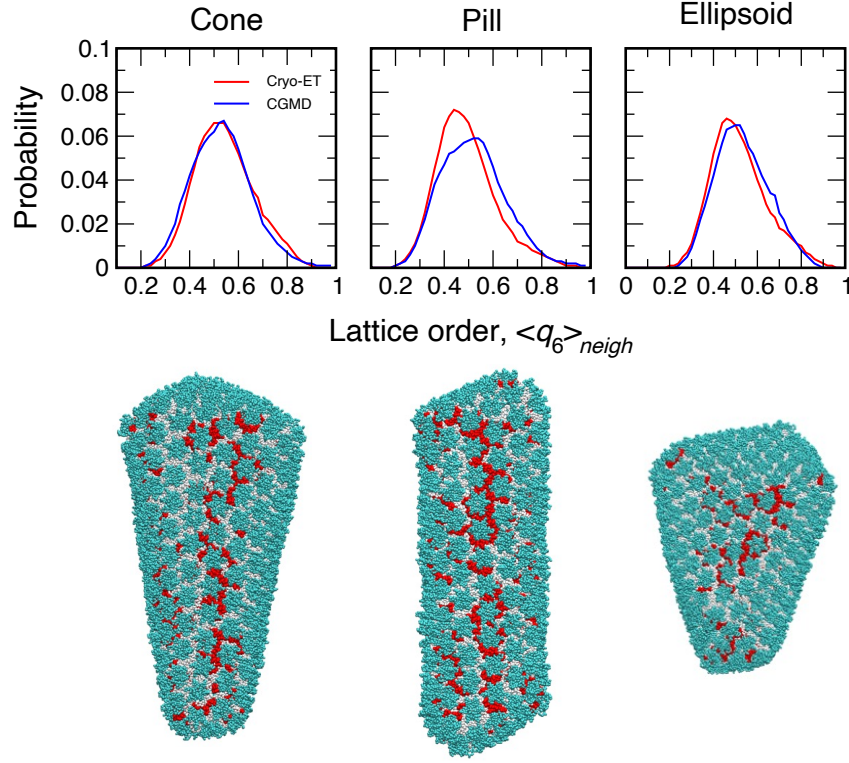


**Fig. S5. Time-averaged densities of the membrane-binding domains of the NPC.** The densities of each membrane-binding  $\beta$ -propeller domain are shown for the constricted and dilated NPC overlaid on the lipid bilayer. The densities were calculated from the final  $150 \times 10^6 \tau_{CG}$  of the CG MD simulations. The lipid bilayer is vertically sliced to show the density of NUP155  $\beta$ -propeller domains (blue isosurface) bound to the membrane around the central channel.  $\beta$ -propeller domains of NUP133 and NUP160 are shown in cyan and yellow isosurface, respectively. The headgroup of the 4-site CG lipid is shown in pink spheres. The interfacial bead and two tail beads of the CG lipid are shown in white spheres.

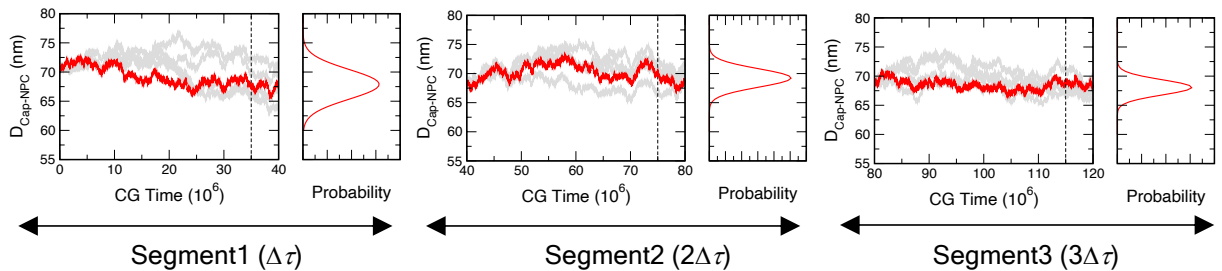




**Fig. S6. Analysis of NPC diameters measured between selected NUPs at the outer (cytoplasm and nuclear) and inner rings.** The outer ring diameters ( $D_{NUP85}$ ) are estimated between the geometric center of NUP85 (inner tip of the Y-complex dimer). At the inner ring, diameters are measured between the geometric center of membrane-bound  $\beta$ -propeller domains of NUP155 ( $D_{NUP155}$ ) and NUP188 ( $D_{NUP188}$ ). The upper left panel shows the cutaway view of the NPC and lipid showing the outer (CR and NR) and inner rings. The upper left panel shows the same cutaway view highlighting only the NUP85, NUP155, and NUP188 proteins, between which the corresponding ring diameters are measured. The bottom panel shows the measured diameters for the constricted and dilated CG NPC models calculated from the final  $150 \times 10^6 \tau_{CG}$  of the CG MD simulations. The same diameters measured for the reference cryo-ET structures are shown in the same plot.

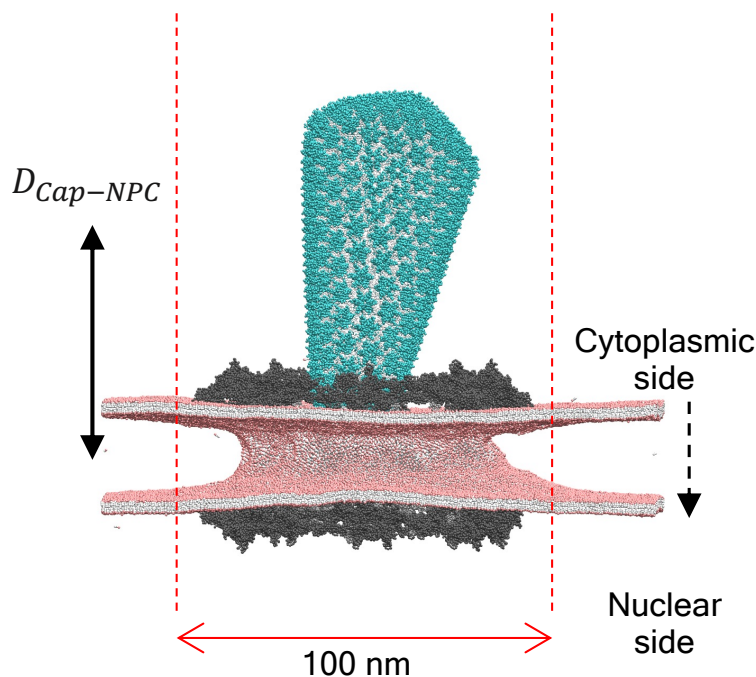


**Fig. S7. Lattice order of the freely diffusing capsid and comparison with cryo-ET structure.** Steinhardt's neighbor-averaged local bond order parameter ( $\langle q_6 \rangle_{neigh}$ ) of HIV-1 capsid for the initial cryo-ET structure and from CG MD simulations (upper panel). The probability distribution of  $\langle q_6 \rangle_{neigh}$  for the CG simulations is calculated from the final  $100 \times 10^6$  CG MD timesteps,  $\tau_{CG}$ . The cone, pill, and ellipsoid capsids seen in the lower contain 1314, 1122, and 1266 CA monomers, respectively. The lower panel also depicts lattice disorder of the capsid lattice of the cone, pill, and ellipsoid morphology at the end of  $400 \times 10^6 \tau_{CG}$  of CG simulations (labeled as CG MD). The CTD domain of the CA monomers with  $\langle q_6 \rangle_{neigh} < 0.4$  is colored in red. The CTD domain of the rest of the CA monomers is colored in white. The NTD domain of all the CA monomers is represented as cyan spheres.

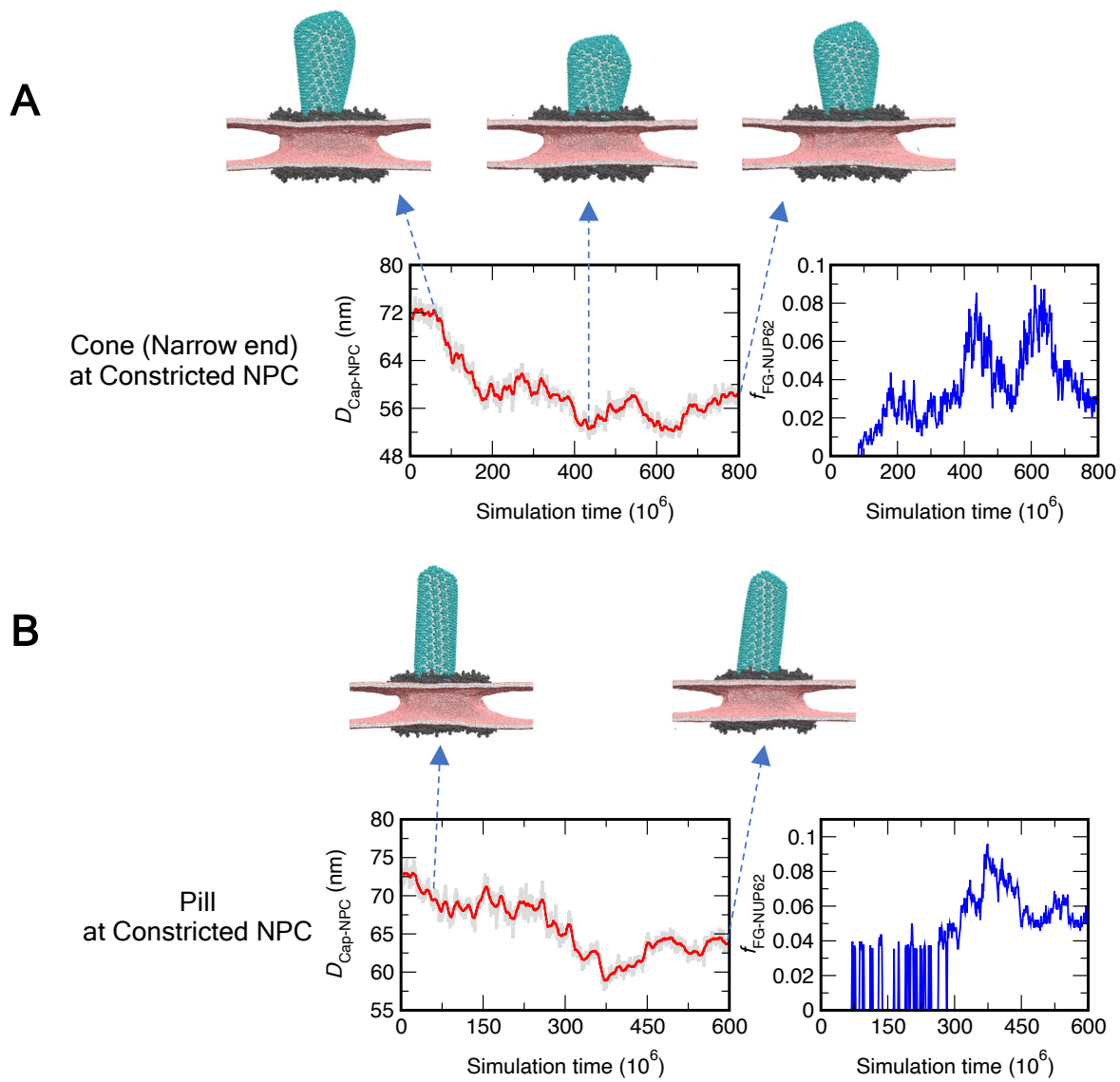


**Fig. S8. Schematic of the path sampling methodology to simulate translocation dynamics of HIV-1 capsid.** For each segment, 5 parallel simulations (each  $40 \times 10^6 \tau_{CG}$ ) were evolved. From the probability distribution of  $D_{Cap-NPC}$  (calculated from final  $5 \times 10^6 \tau_{CG}$ ), the simulation

trajectory was selected for which the endpoint (solid red line) is closest to the mean of the distribution. The endpoint of this trajectory was then used as the initial configuration for simulations in the next segment. The selected trajectories were appended to create the final translocation trajectory. The depicted segments correspond to the first 3 segments of the translocation trajectory of the HIV-1 capsid cone (narrow end pointing to the central channel) at the dilated NPC.



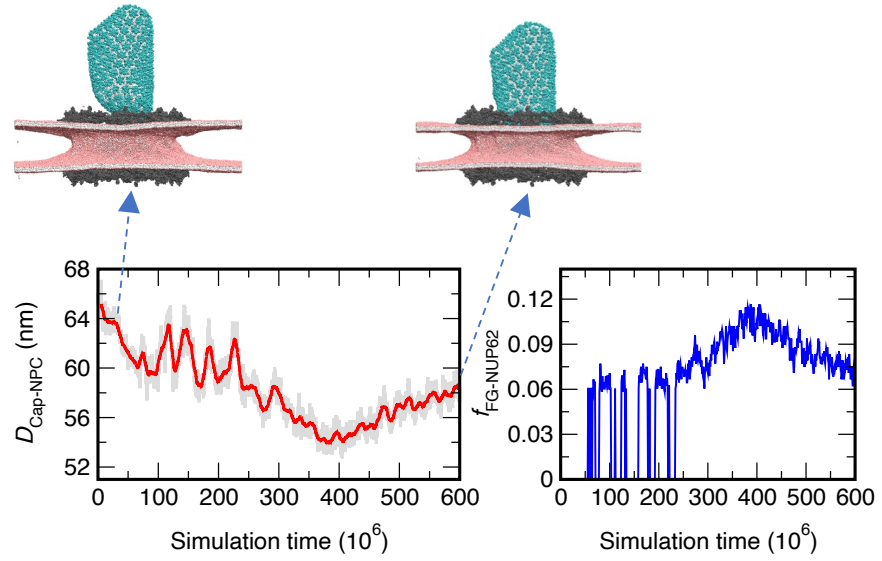
**Fig. S9. Capsid translocation simulation setup.** A representative snapshot of an initial CG MD simulation configuration (side view) of the conical capsid placed at the cytoplasmic end of the NPC with the narrow end pointing to the central channel. The distance ( $D_{Cap-NPC}$ ) between the geometric center of the capsid and NPC inner ring is labeled. The dotted red lines indicate the location of cylindrical confinement of 100 nm diameter. The cytoplasmic and nuclear sides of the NPC are also labeled.



**Fig. S10. Translocation dynamics of HIV-1 capsid into the central channel of the constricted NPC.** Translocation dynamics time series plot (A and B) that depicts the distance ( $D_{Cap-NPC}$ ) between the geometric center of the capsid and NPC inner ring (left panel in red), and the fraction of FG sites of NUP62 ( $f_{FG-NUP62}$ ) bound to CA (right panel in blue). The snapshots depict the cone-shaped (approaching from the narrow end) and pill-shaped capsid at the central channel of the constricted NPC at different points of the translocation dynamics trajectory. The capsid, NPC, and lipids are shown in the color scheme in **Fig. 2** of the main text.

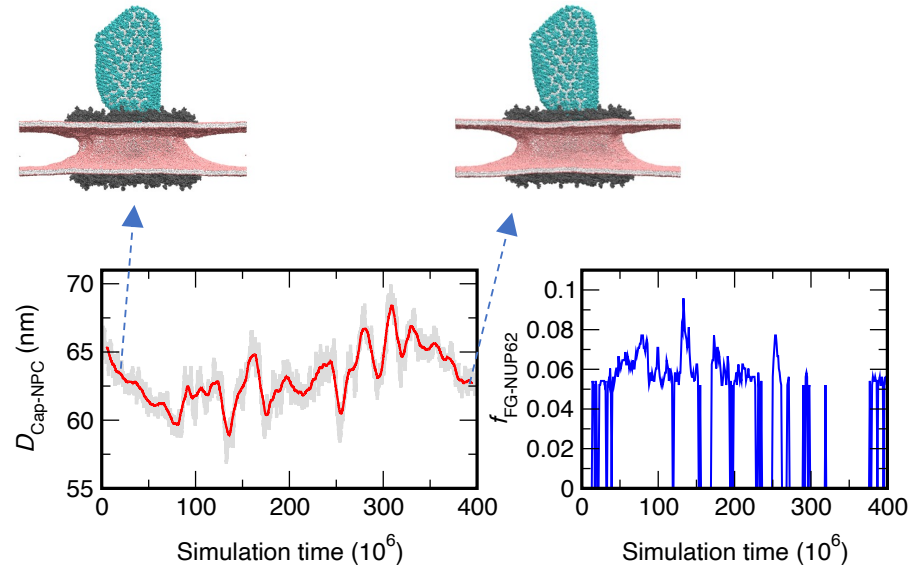
**A**

Ellipsoid  
at Dilated NPC



**B**

Ellipsoid  
at Constricted NPC

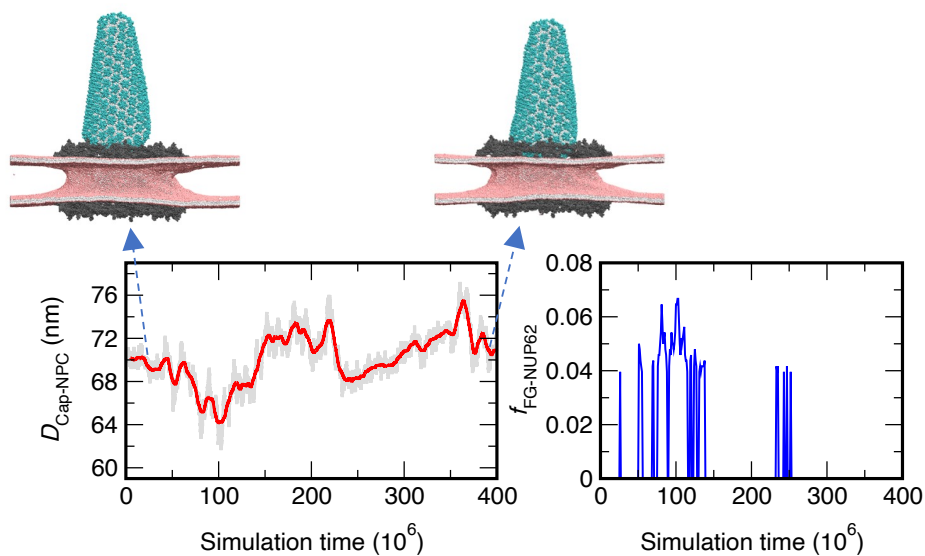


**Fig. S11. Translocation dynamics of ellipsoid-shaped capsid into the NPC central channel.** Translocation dynamics time series plot of the ellipsoid capsid into the dilated (A) and constricted (B) NPC. The capsid, NPC, and lipids are shown in the color scheme in **Fig. 2** of the main text.



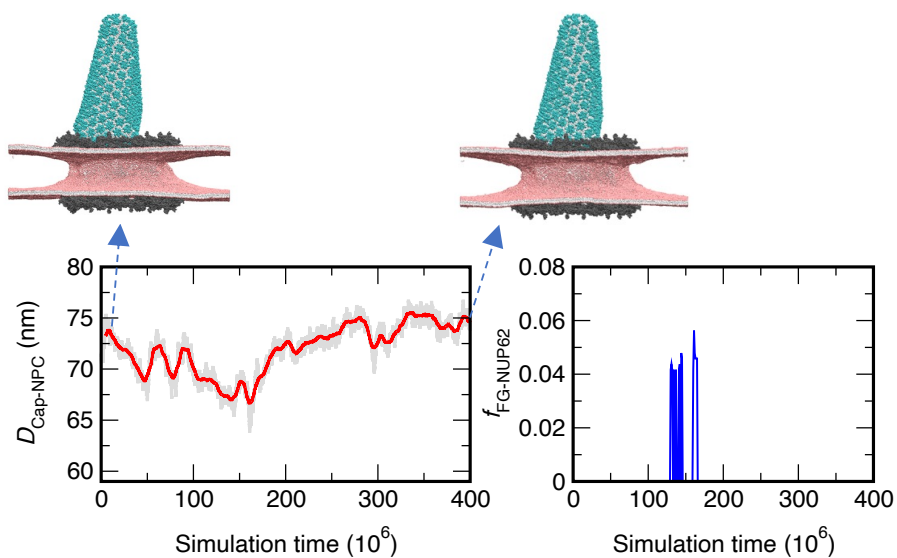
**A**

Cone (Wide end)  
at Dilated NPC

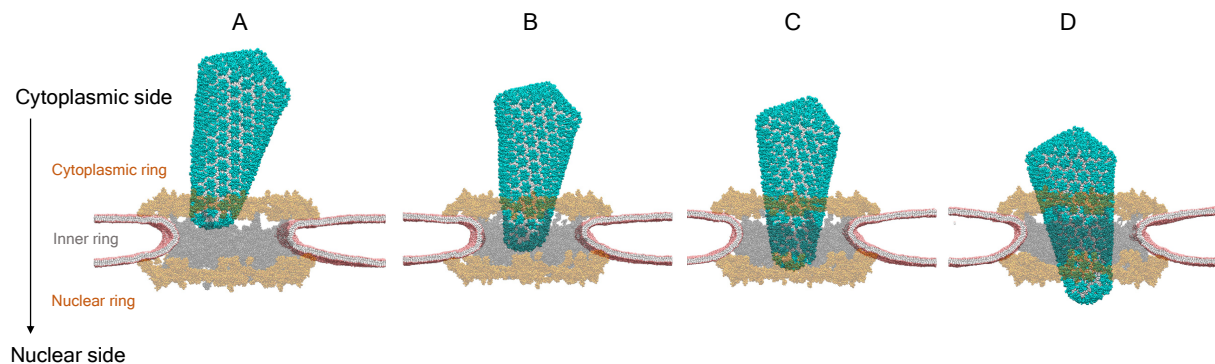


**B**

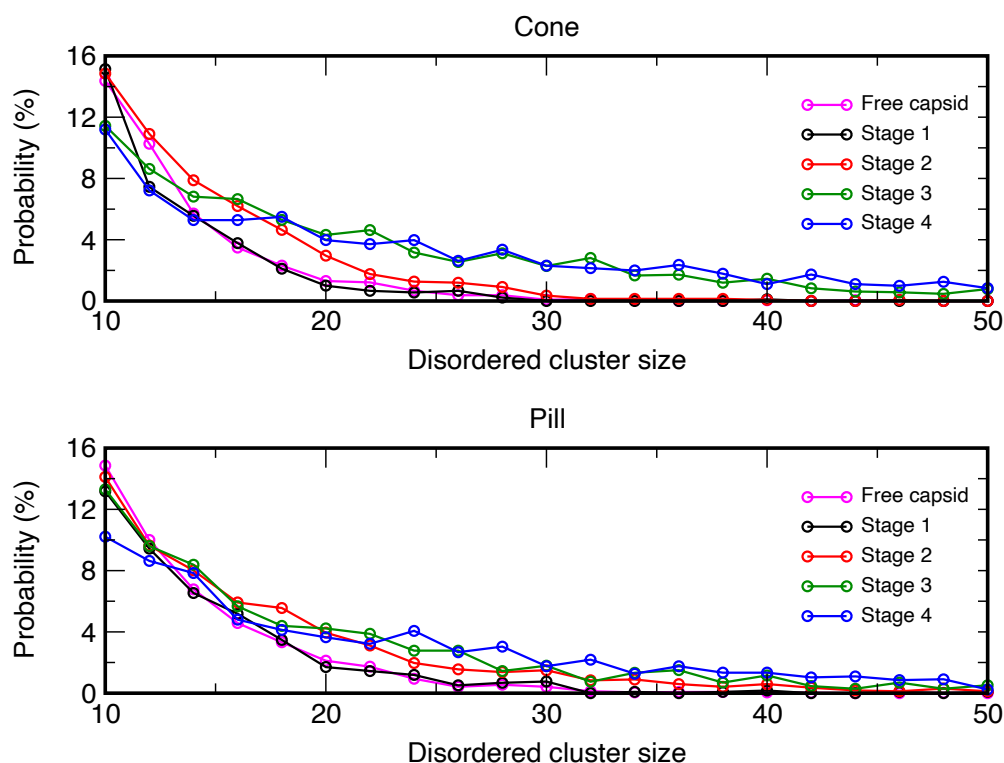
Cone (Wide end)  
at Constricted NPC



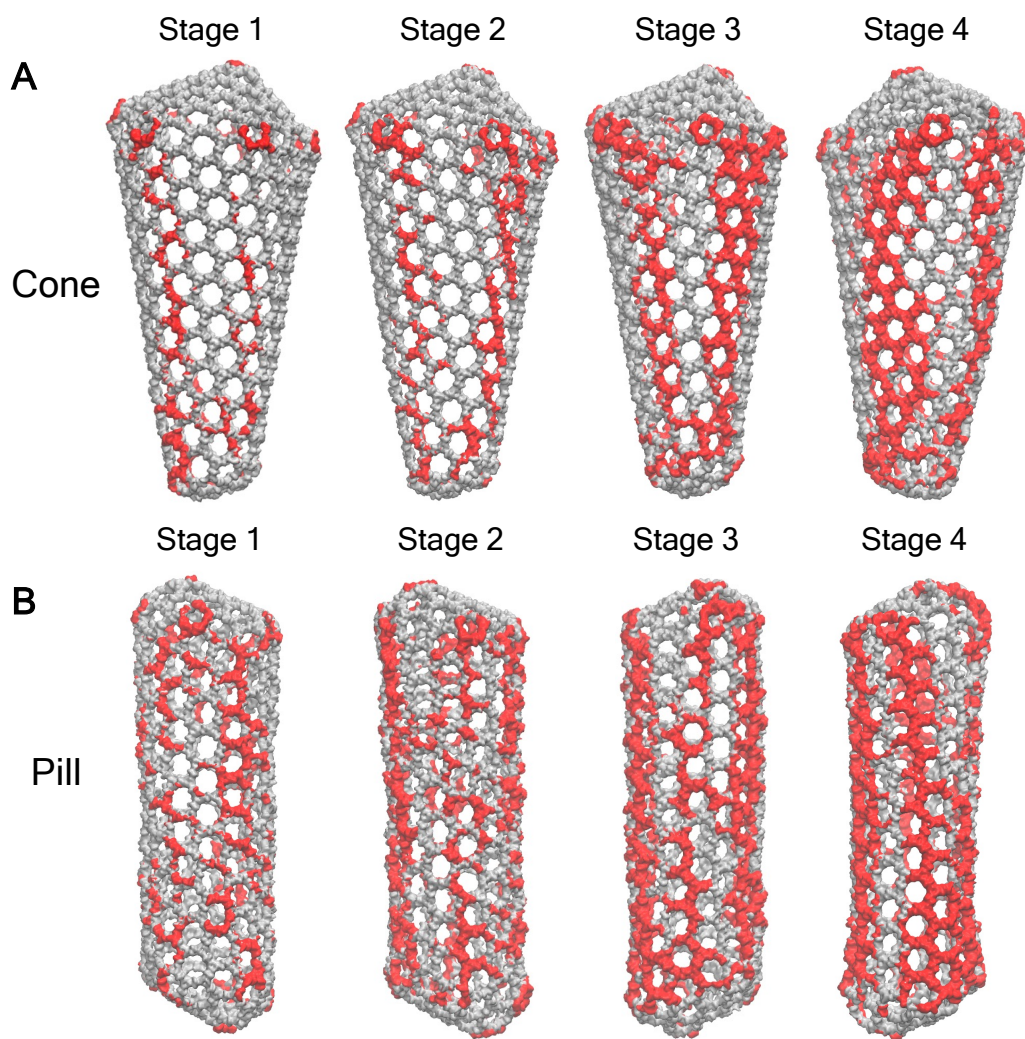
**Fig. S12. Translocation dynamics of cone-shaped capsid (approaching from the wide end) into the NPC central channel.** Translocation dynamics time series plot of the cone-shaped capsid into the dilated (A) and constricted (B) NPC. The capsid, NPC, and lipids are shown in the color scheme in **Fig. 2** of the main text.



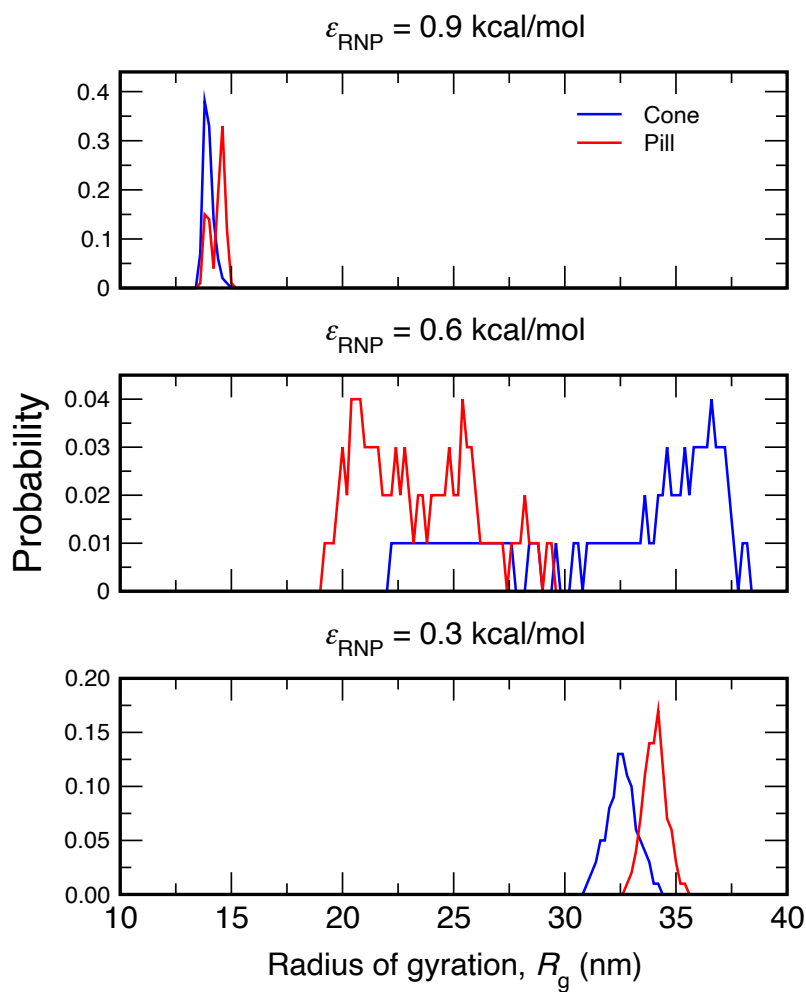
**Fig. S13. Translocation dynamics of the cone-shaped capsid (approaching from the narrow end) into the NPC central channel.** The NPC is shown in semi-transparent representation. The cytoplasmic and nuclear rings are shown in orange. The inner ring is shown in silver. The capsid, and lipids are shown in the color scheme in **Fig. 2** of the main text. In each snapshot (a cutaway sideview of the lipid bilayer is shown). In the panel A, the tip of the capsid makes first encounter with the FG-NUPs of the IR. The majority of the capsid (except the tip) in this stage is at the cytoplasmic side and do not make contact with the NPC. In panel B and C, the tip progressively advances to the nuclear ring, and makes extensive contacts with the FG-NUPs of the central channel. In the panel D, the narrow end of the capsid has translocated beyond the nuclear ring. In this stage, the majority of the capsid is in direct contact with the NPC. Further translocation of the capsid into the nuclear basket will require the presence of NUP153 chains. However, a high-resolution structure of nuclear basket scaffold is currently not available and hence not included in our current CG model.



**Fig. S14. Distribution of the size of disordered CA cluster for the cone-shaped (A) and pill-shaped (B) at different stages of translocation into the dilated NPC central channel.** A particular CA monomer is classified as disordered if  $\langle q_6 \rangle_{neigh} < 0.4$ . Connected clusters of at least 10 CA monomers are considered.

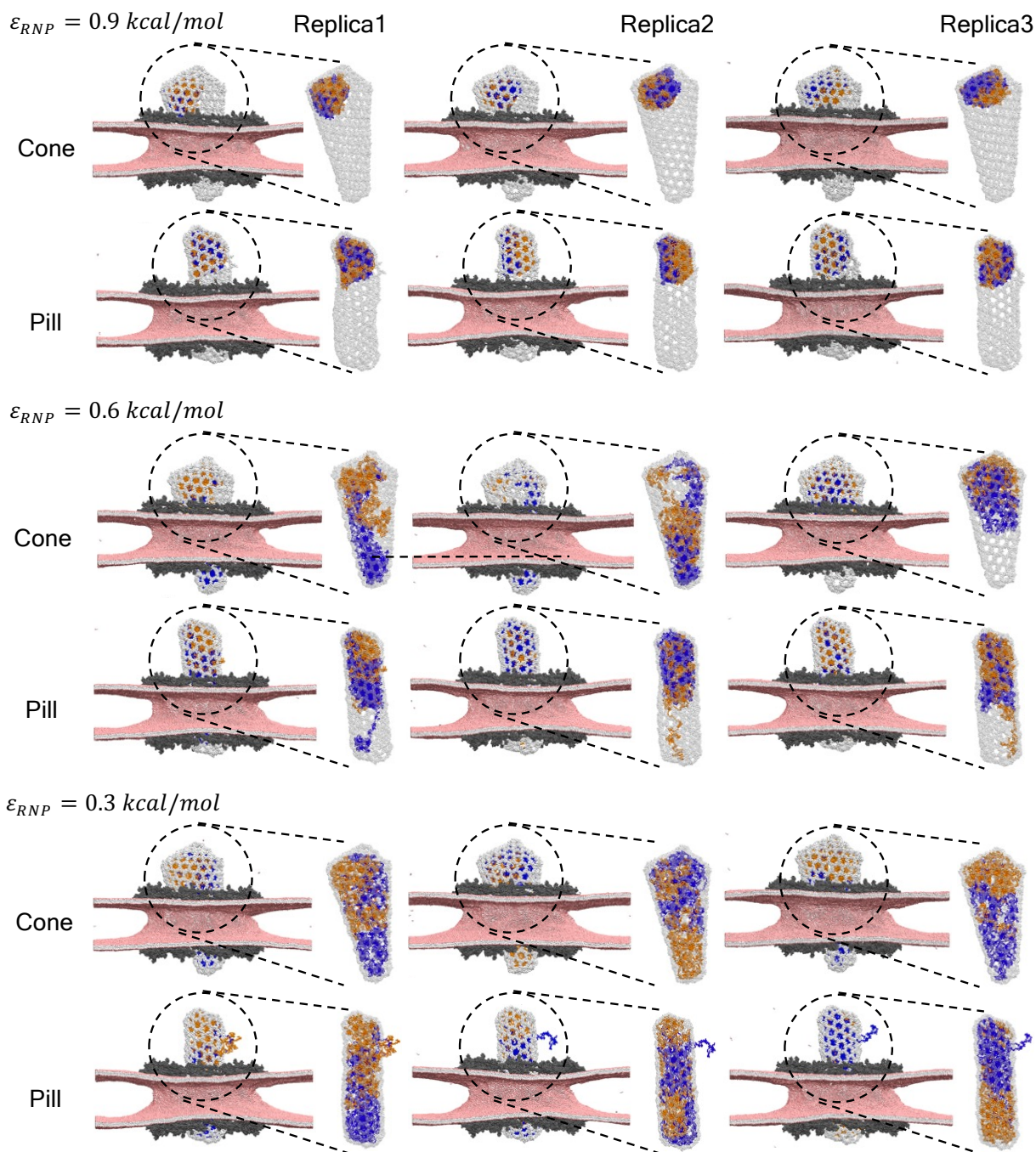


**Fig. S15. Time-averaged capsid lattice disorder at different stages of entry into the NPC central channel for the cone (A) and pill (B).** The CTD domain of capsid is shown in silver. The CTD domain of each CA monomer with  $\langle q_6 \rangle_{neigh} < 0.4$  is shown in red. Only the CTD domain of the capsid is shown to highlight the striated disordered patterns (shown in red) at the lattice. We note that, the highlighted disordered domains (in red) are present in over 50% of the sampled trajectory frames.



**Fig. S16. Probability distribution of the radius of gyration ( $R_g$ ) of the RNP complex in the interior of the capsid.** From top to bottom,  $\epsilon_{RNP}$  is varied from 0.9 to 0.3 kcal/mol, modifying the RNP condensation states. For each  $\epsilon_{RNP}$ , the  $R_g$  distribution is shown in blue and red for the cone-shaped and pill-shaped capsid, respectively.

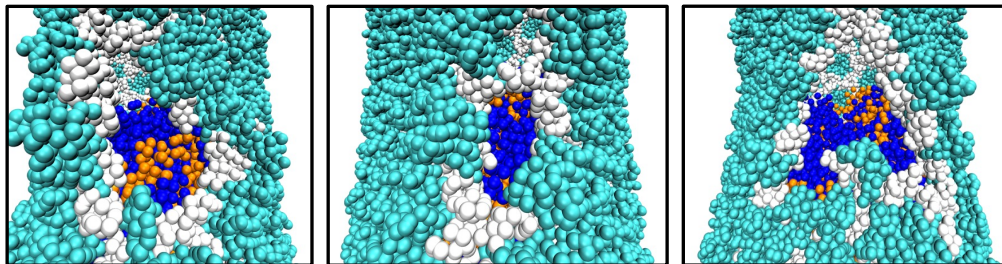




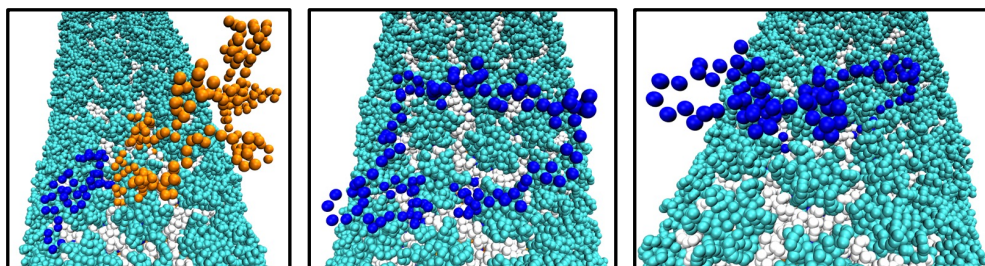
**Fig. S17. Snapshots of the RNP complex in the interior of the capsid.** The two RNP chains are shown in blue and orange spheres. The three panels next to each  $\epsilon_{RNP}$  correspond to 3 replica simulations. The NPC, and lipids are shown in the color scheme as in **Fig. 2**. The CTD domain of the capsid is shown in silver. Only the CTD domain is shown for clear visualization of the condensation state of the internal RNP complex.



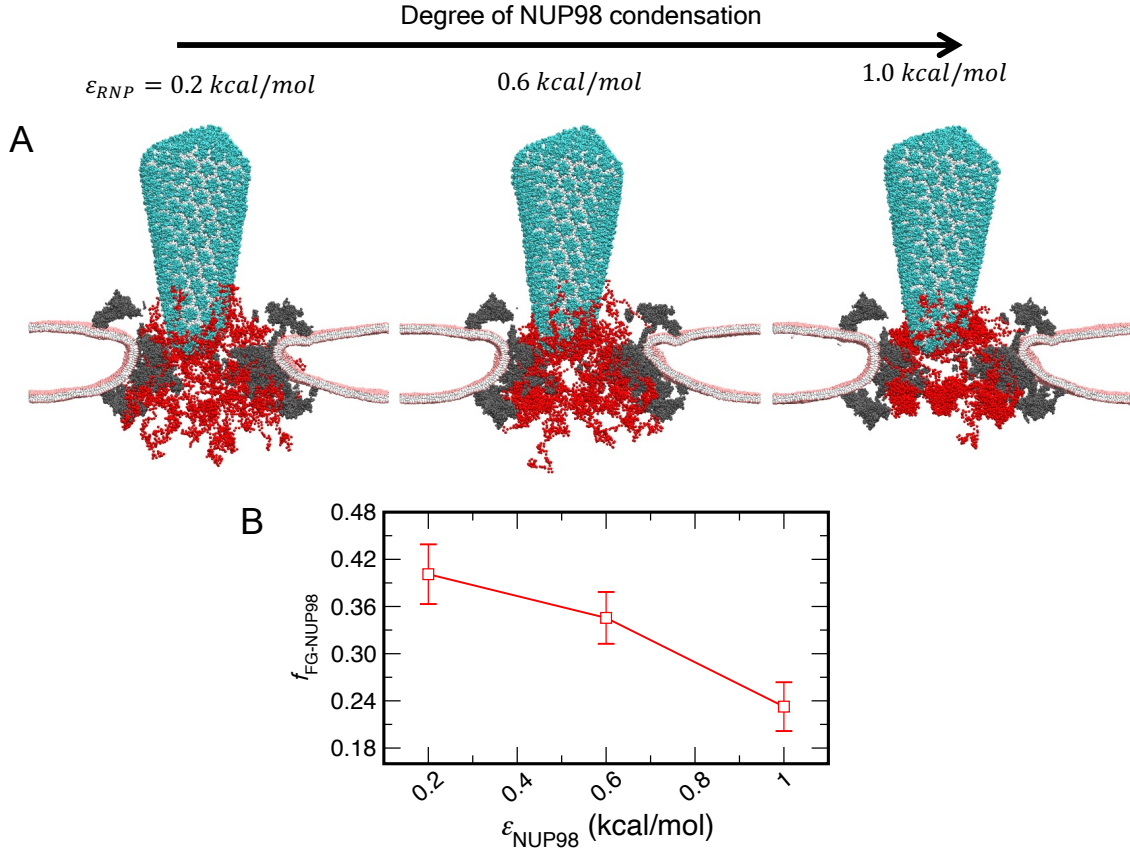
$\varepsilon_{RNP} = 0.9 \text{ kcal/mol}$  (Condensed)



$\varepsilon_{RNP} = 0.3 \text{ kcal/mol}$  (Uncondensed)

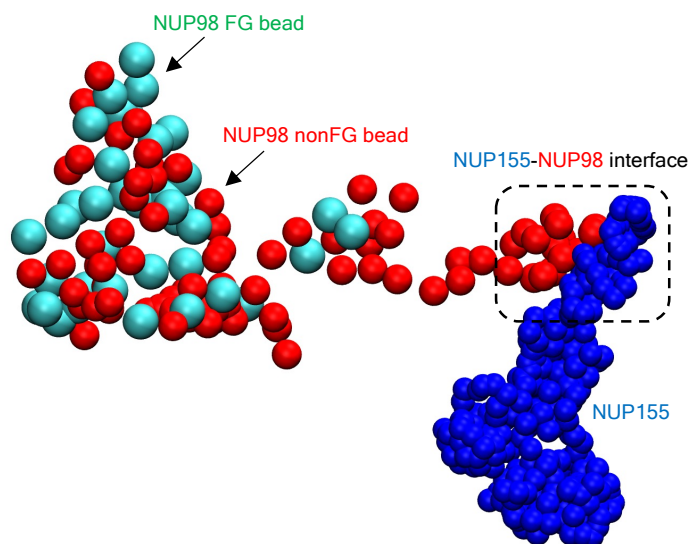


**Fig. S18. Snapshots of the rupture in the capsid lattice and extrusion of the RNP chains for the pill-shaped capsid.** For  $\varepsilon_{RNP} = 0.9 \text{ kcal/mol}$  (upper panel), condensation of the RNP complex creates rupture of the capsid lattice in the vicinity of the globular condensate. For  $\varepsilon_{RNP} = 0.3 \text{ kcal/mol}$  (lower panel), the segments of the uncondensed RNP complex protrudes out of the capsid lattice from the CA lattice rupture.

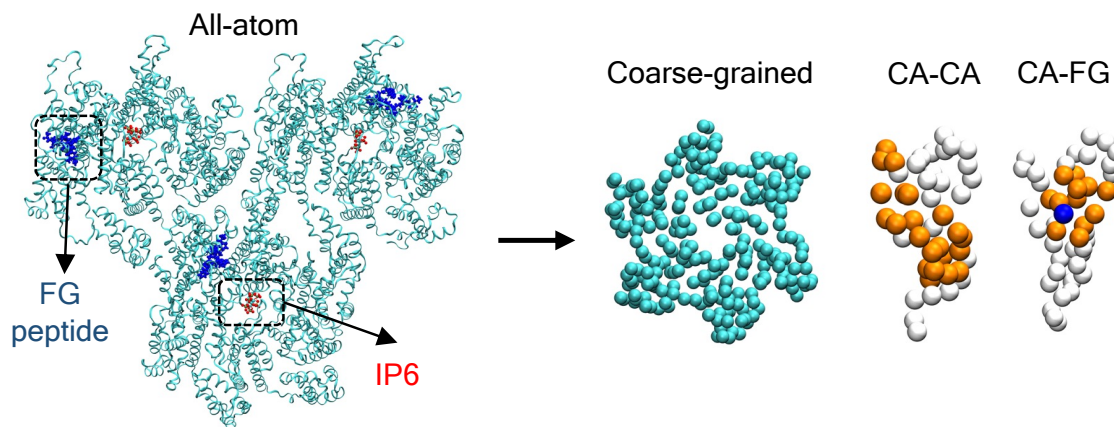


**Fig. S19. Condensation state of NUP98 at the NPC<sub>NUP98</sub> central channel and capsid tip association.** (A) In each snapshot, a cutaway sideview of the NPC and lipid is shown with the same colors as in **Fig. 4**. The NUP98 chains are shown in red spheres. From left to right,  $\epsilon_{NUP98}$  is varied from 0.2 to 1.0 kcal/mol, modulating the NUP98 condensation state. For each snapshot, the association of the capsid tip to NUP98 mesh is shown. (B) The fraction of the FG sites of NUP98 chains bound to CA ( $f_{FG-NUP98}$ ). Note that each NUP98 consists of 12 FG-motif sites, and there are 48 NUP98 chains tethered to NPC<sub>NUP98</sub>. The standard deviation is calculated over 3 replica simulations for each  $\epsilon_{NUP98}$ . For the weaker  $\epsilon_{NUP98}$  interactions, as the tip of the capsid approaches the central channel, FG-motifs of the NUP98 chains contacts the capsid. The binding of the multiple NUP98 FG-motifs immerses the narrow end of the capsid in the unstructured mesh-like environment at the NPC central channel and promotes translocation of the capsid (**Fig. 19A**). In contrast, for the stronger  $\epsilon_{NUP98}$  interactions, the tip of the capsid encounter local condensates of NUP98 chains in which the FG-motifs bind the CA lattice with significantly reduced propensity compared to the weaker  $\epsilon_{NUP98}$  interactions (**Fig. 19B**). Therefore, the lack of FG-motif binding to the constituent CA monomers regulated by the NUP98 condensation state impedes the association of the narrow end of the cone-shaped capsid to the NPC central channel.

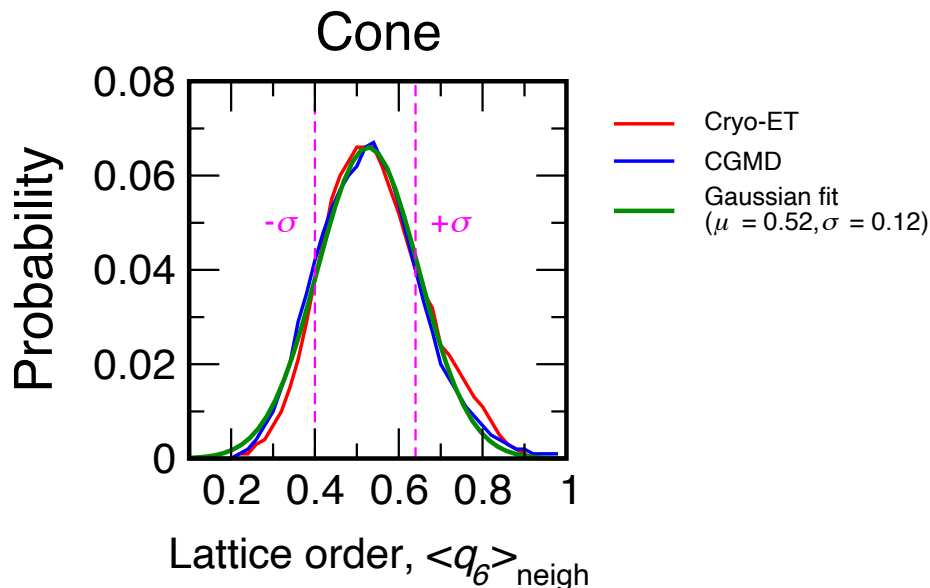
## NUP98 CG model



**Fig. S20. CG model of disordered NUP98.** The FG and non-FG CG beads of a NUP98 chain is shown in cyan and red spheres respectively. We also highlight the NU155-NUP98 binding interface based on the PDB: 7R5J (1). NUP155 is shown in blue spheres. The snapshot corresponds to  $\epsilon_{NUP98} = 0.3$  kcal/mol.



**Fig. S21. Atomistic simulation setup for developing “bottom-up” CG model of CA.** A representative snapshot of the AA 18-mer CA complex (cyan cartoon) complexed with 3 FG peptides (residues SP SGVFTFGAN, shown in blue spheres) is shown in the left panel. The top right panel shows the representative snapshot of a CG CA hexamer (cyan spheres). The top right-most right panel shows (as orange spheres) the CG sites of a CA monomer involved in CA-CA and CA-FG associative interactions. The CG sites not involved in associative interactions in each case are shown as white spheres. The CG FG site associated with the CA monomer is shown as a blue sphere.



**Fig. S22. Gaussian fitting of the probability distribution of  $\langle q_6 \rangle_{\text{neigh}}$  of cone-shaped capsid.** The  $\langle q_6 \rangle_{\text{neigh}}$  probability distribution of the initial cryo-ET structure, CG MD simulations and the gaussian fitting is shown in red, blue and green, respectively. The region between  $\mu - \sigma$  and  $\mu + \sigma$  are shown with dotted magenta lines.

## SI Appendix Methods

### S1. Deriving coarse-grained (CG) model and molecular interactions from atomistic simulations using “bottom-up” methods.

We systematically derived CG models and interactions of NUP monomers and subcomplexes from reference all-atom (AA) MD simulations following a three-step hierarchical methodology. First, CG beads of NUP monomers were mapped from the corresponding AA MD trajectories using the Essential Dynamics Coarse-Graining (EDCG) method (2). Using the derived CG map, we represented intra-protein interactions using heterogeneous elastic network models (hENMs) (3). Then attractive short-ranged CG inter-protein interactions were derived from AA MD trajectories of respective NUP subcomplexes using relative entropy minimization (4). These methods have been successfully utilized before to derive “bottom-up” CG models of various multicomponent protein assemblies (5-7). The details of the coarse-grained modeling procedures are described below.

#### 1. CG mapping from reference atomistic simulations.

To define the CG sites of NUP monomers from AA simulation trajectories, we used Essential Dynamics coarse-graining (EDCG) (2). EDCG is a “bottom-up” coarse-graining procedure in which  $C_\alpha$  sites of a number of consecutive all-atom residues are grouped to a CG site such that the principal modes of motion sampled during atomistic simulations are conserved. In EDCG, the mapping operator ( $\mathbf{M}_R^N: \mathbf{r}^n \rightarrow \mathbf{R}^N$ ), is variationally optimized using simulated annealing to obtain the global minimum of the target residual ( $\chi^2$ ). Here,  $\mathbf{r}^n$  and  $\mathbf{R}^N$  are the configurations of the atomistic and coarse-grained trajectories, respectively. The AA to CG mapping function is adjusted during the optimization to minimize the target residual ( $\chi^2$ ):

$$\chi^2 = \frac{1}{3N} \sum_{I=1}^N \langle \sum_{i,j \in I, j \geq i} |\mathbf{r}_i - \mathbf{r}_j|^2 \rangle_t \quad (1)$$

where,  $N$  is the total number of CG sites. Here  $i, j$  are the unique pairs in the group of all-atom residues that are part of the CG site,  $I$ .  $\mathbf{r}_i = \mathbf{x}_i - \langle \mathbf{x}_i \rangle_t$  is the displacement of atom  $i$  from the atom's mean position,  $\langle \mathbf{x}_i \rangle_t$ . When atoms  $i, j$  move in correlated fashion in the atomistic trajectory, then  $|\mathbf{r}_i - \mathbf{r}_j|^2$  is small. For  $N$  CG beads ( $N$  all-atom residue segments),  $N - 1$  segment boundaries are initially defined along the primary amino acid sequence. At each step of simulated annealing, a new CG map is defined by moving the position of the segment boundary atoms along the primary sequence. The new map is either accepted or rejected according to a Metropolis-Hastings criterion. The new map is accepted if the value of the residual is less than its predecessor ( $\chi_1^2 < \chi_0^2$ ). If the new CG map has a higher residual, then the map is accepted with a probability  $\exp(-\Delta\chi^2/T)$ . Here  $T$  is the coupling to a temperature parameter that is initially high to allow the boundary atoms to move randomly and allow escape from local minima and is gradually lowered during the optimization. All CG maps of NUP monomers were generated from the final 500 ns of AA MD trajectories and had an average resolution of  $\sim 5$   $C_\alpha$  residues per CG site.

## 2. Intra-protein bonded interactions derived from atomistic MD simulations.

Intra-protein bonded interactions were defined using heterogeneous elastic network model (hENM), which captures the flexibility of the protein (3). In the hENM method, initially, harmonic bonds with identical force constant ( $k_{ij}$ ) are assigned between the central CG site ( $i$ ) and all the neighboring CG sites ( $j$ ) within a user-specified distance cutoff ( $r_{cut}$ ). The harmonic force constant ( $k_{ij}$ ) for each bond is iteratively optimized until the fluctuations in the CG model converge to that of reference AA MD trajectory, i.e.,

$$\frac{1}{k_{ij}^{n+1}} = \frac{1}{k_{ij}^n} - \alpha (\langle r_{ij}^2 \rangle_{CG} - \langle r_{ij}^2 \rangle_{AA}) \quad (2)$$

where  $k_{ij}^n$  and  $k_{ij}^{n+1}$  is the harmonic force constant at the iteration step  $n$  and  $n + 1$  step, respectively.  $\langle r_{ij}^2 \rangle = \langle (\mathbf{x}_{ij} - \langle \mathbf{x}_{ij} \rangle)^2 \rangle$  is the mean-squared fluctuation for each CG site pair  $i, j$ .  $\alpha$  is a parameter that controls the magnitude of the adjustment for each iteration. For all NUP monomer CG models, we used a cutoff distance ( $r_{cut}$ ) of 3 nm. The optimization procedure to derive the force constant ( $k_{ij}$ ) was performed from the final 200 ns of AA MD trajectories.

## 3. Interprotein short-ranged attractive interactions from reference atomistic MS simulations.

Inter-protein attractive interactions between CG sites of NUP subcomplexes are modeled with pairwise short-ranged Gaussian potentials ( $E_{gauss}$ ),

$$E_{gauss}(r_{ij}) = \frac{H_{ij}}{\sigma_{ij}\sqrt{2\pi}} \exp\left(-\frac{(r_{ij}-r_{0,ij})^2}{2\sigma_{ij}^2}\right) \quad (3)$$

where  $r_{ij}$  and  $\sigma_{ij}$  are the mean and standard deviation of a CG site pair  $i, j$ . The constant  $A_{ij}$  is optimized using Relative Entropy Minimization (REM) (4). The constant  $A_{ij}$  (referred to as  $\lambda$  in equation 4) is iteratively optimized using Newton-Raphson method as follows,

$$\lambda_{n+1} = \lambda_n - \chi \frac{\left(\frac{\partial S}{\partial \lambda}\right)}{\left(\frac{\partial^2 S}{\partial \lambda^2}\right)} \quad (4)$$

$\chi$  is the “learning rate” of the iterative optimization procedure. Note that the “learning rate” ( $\chi$ ) is varied throughout the optimization procedure.



To perform the REM iteration cycle, we first created a CG NUP subcomplex by mapping the final configuration from the AA MD trajectory. All the NUP subcomplexes used to derive inter-protein interactions consist of two NUP monomers that are in direct contact in the cryo-electron tomography (cryo-ET) structural data (1). For each iteration, a CG simulation of the NUP subcomplex was evolved for  $25 \times 10^6$  CG MD timesteps at 300 K. Total 500 step iteration cycle was performed as changes between successive steps are effectively zero within 500 steps. Learning rate ( $\chi$ ) is varied as 0.5 (1-100 steps), 0.1 (101-200 steps), 0.01 (201-500 steps).

## **S2. Atomistic simulation of NUP monomers and heterodimers.**

We performed 1000 ns long AA MD simulations of NUP monomers and heterodimers in solution to derive the CG molecular model and interactions. We solvated the initial protein structure (monomers and heterodimers) in a cubic box of TIP3P water. The solvated system was built such that there is at least 1.2 nm layer of water between the protein surface and the edge of the simulation cell. To achieve a physiological ion concentration of 150 mM, we added  $\text{Na}^+$  and  $\text{Cl}^-$  ions to the solvated system by replacing randomly selected TIP3P water molecules. The simulation system was then energy minimized using the steepest descent method until the target maximum force was less than 239 kcal/mol/nm. The system was equilibrated by applying harmonic positional restraints (spring constant value of 239 kcal/mol/nm<sup>2</sup>) on protein heavy atoms for 500 ps in the constant NVT ensemble. The temperature of the system was maintained at 310 K using stochastic velocity rescaling thermostat with a time constant of 1 ps. A second constrained equilibration was done for 800 ps in the constant NPT ensemble at 310 K and 1 bar using the isotropic Parrinello-Rahman barostat with a 10 ps time constant. In this equilibration procedure, the harmonic positional constraint on protein heavy atoms was progressively decreased every 200 ps ( $k = 239, 119.5, 23.9, 11.9$  kcal/mol/nm<sup>2</sup>). The production runs were performed in the constant  $N_p T$  ensemble at 310 K and 1 bar. The temperature was maintained using Nose-Hoover chain thermostat with a 2 ps time constant. The pressure of the simulations was maintained with an isotropic Parrinello-Rahman barostat with a 10 ps time constant (8, 9). All simulations were performed with periodic boundary conditions in  $x$ ,  $y$ , and  $z$  directions. The protein was modeled with the CHARMM36m force field (10), and water was modeled with TIP3P parameters (11), using a 2 fs MD timestep. LINCS algorithm was used to constrain the bonds between heavy and hydrogen atoms (12). Electrostatic interactions were computed using the particle mesh Ewald method with a cutoff of 1 nm (13). The van der Waals force was truncated smoothly to zero between 1.0 and 1.2 nm. All atomistic simulations were performed using the Gromacs 2019 MD package (14).

## **S3. CG model of the NPC.**

We first determined the CG mapping of NUP monomers from AA MD simulation trajectories using the Essential Dynamics Coarse-Graining (EDCG) method (2) to identify the group of consecutive residues that maximized the collective motions of the mapped CG protein with the motions of the corresponding AA trajectory. All NUP monomers were mapped such that the average AA to CG mapping resolution is  $\sim 5$  amino acid residues per CG site. We then determined the intramonomer bonding topology from the same AA MD trajectory by creating a heterogeneous elastic network model (hENM) (3). To create the intramonomer bonding topology, we considered all CG sites within 3 nm of a central CG site. To evaluate whether the CG hENM models recapitulate the AA fluctuations, we compared the root mean square (RMS) fluctuations of each CG NUP monomer with the reference AA MD simulations. To model NUP subcomplexes (Y-complex dimer, NUP54-NUP58-NUP62, and NUP155-NUP93) we created a weak harmonic bonding network with a force constant of  $0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  between neighboring monomers with a distance cutoff of 3 nm. The intermonomer force constant value was phenomenologically chosen to preserve the shape of each subcomplex.

Intermolecular non-bonded CG interactions between NUP monomers were modeled using a combination of repulsive excluded volume and attractive interactions. The repulsive excluded volume interaction ( $E_{excl}$ ) was modeled with a soft cosine potential,

$$E_{excl}(r_{ij}) = A \left( 1 + \cos \left( \frac{\pi r_{ij}}{r_c} \right) \right) \quad (1)$$

where  $r_{ij}$  is the pairwise distance between CG site types  $i$  and  $j$ . The value of  $A$  was set to 15 kcal/mol for all  $ij$  pairs. The distance cutoff ( $r_c$ ) for the excluded volume interactions was set at 1.0 nm. Attractive interactions between NUP subcomplexes at the key binding interactions were modeled with pairwise Gaussian potential ( $E_{gauss}$ ),

$$E_{gauss}(r_{ij}) = \frac{H_{ij}}{\sigma_{ij}\sqrt{2\pi}} \exp \left( -\frac{(r_{ij}-r_{0,ij})^2}{2\sigma_{ij}^2} \right) \quad (2)$$

where  $r_{0,ij}$  and  $\sigma_{ij}$  are the mean and standard deviation of the distance between CG site types  $i$  and  $j$ . First, we determined the CG site types  $i$  and  $j$  that are in close contact in the mapped AAMD trajectories of NUP heterodimer complexes with a distance cutoff of 1.75 nm and standard deviation less than or equal to 0.12 nm. For all the subset of CG  $ij$  pairs in close contact,  $r_{0,ij}$  is determined from the mapped AA MD trajectories from the first peak of the pair correlation functions. For all CG  $ij$  pairs  $\sigma_{ij}$  value of 0.12 nm was used. The constant  $H_{ij}$  for each binding interface was optimized using Relative Entropy Minimization from the atomistic trajectories of the respective all-atom subcomplexes (4). The derived attractive interactions ( $E_{attr}$ ) were validated by comparing the pair correlation functions between selected CG ( $ij$ ) pairs of the CG NPC model and AA simulations of the corresponding heterodimer complexes at the binding interfaces.

Membrane association interactions between the CG sites of the NPC and CG lipid head group were modeled using a 12-6 Lennard-Jones potential ( $E_{scLJ}$ ) with a modified soft-core (15),

$$E_{scLJ}(r) = 4\epsilon\lambda^n \left[ 1/(\alpha_{LJ}(1-\lambda)^2 + (r/\sigma)^6)^2 - 1/(\alpha_{LJ}(1-\lambda)^2 + (r/\sigma)^6) \right] \quad (3)$$

where,  $n = 2$ ,  $\alpha_{LJ} = 0.5$ ,  $\alpha = 0.6$ , and  $\sigma = 1.5$  nm. A minimum protein-lipid interaction strength ( $\epsilon = 1.5$  kcal mol<sup>-1</sup>) was identified and used in our simulations. Specifically, protein-lipid interactions were added to CG sites of the  $\beta$ -propeller domain of NUP155, NUP133, and NUP160.

The CG NUP98<sub>1-620</sub> chains consist of 124 beads (**Fig. S20**). The CG model of the folded GLEBS domain (residue 157-213) was derived from AA MD simulation trajectory using EDCG and hENM protocols. The initial atomistic configuration of the GLEBS domain was generated using AlphaFold2 (16). The rest of the CG NUP98 (residue 1-156 and residue 214-620) was represented as a linear polymer chain to mimic the disordered conformation. Each CG bead of the polymer chain is linked with a flexible harmonic with equilibrium bond length of 2 nm and harmonic force constant of 0.5 kcal/mol Å<sup>-2</sup>. The CG beads of the FG-rich region (residue: 1-156 and residue: 214-480) interact within the same chain and with other chains through 12-6 Lennard-Jones potential ( $E_{scLJ}$ ) with a modified soft-core. Here,  $n = 2$ ,  $\alpha_{LJ} = 0.5$ ,  $\alpha = 0.6$ , and  $\sigma = 1.25$  nm are the parameters used. The strength of the inter- and intra-chain interactions ( $\epsilon_{NUP98}$ ) between the FG-rich fragments of NUP98 is varied from 0.3 to 1.5 kcal/mol. The interaction between FG-motif sites of NUP98 and CA monomers is the same as between FG-motif sites of NUP62 and CA and is modeled with attractive pairwise Gaussian interactions ( $E_{gauss}$ ). All other intermolecular non-bonded interactions between NUP98 and other components are modeled with repulsive excluded volume interactions. In the NPC<sub>NUP98</sub> model, each NUP98 chain was anchored to NUP155 in the dilated NPC model based on cryo-ET structure (1).

#### S4. CG model of Lipids.

A 4-site CG lipid model was used consisting of 1 head, 1 interfacial, and 2 hydrophobic tail beads using the same functional form as in ref. (17). We note that an identical lipid bilayer model was used to simulate other complex biological systems, such as immature virion assembly and SARS-CoV-2 virion (5, 6, 18).

### S5. Construction of NPC-lipid composite system.

To create the NPC-lipid composite system two preequilibrated CG lipid bilayers were first aligned coplanar (in the  $x$  and  $y$  direction) to the cytoplasmic and nuclear rings of the CG NPC model. All CG lipids within 0.5 nm of the cytoplasmic and nuclear rings were deleted to avoid unphysical contact. Similarly, the CG lipids located in the NPC channel (coplanar to CR and IR) were also deleted. A cylindrical membrane pore was also created, which wraps around the NPC IR. The composite membrane system was first minimized and then equilibrated for  $10 \times 10^6 \tau_{CG}$  by integrating only the CG lipids (the position of the NPC CG beads was constrained to the initial position) to allow relaxation. The system was then further equilibrated by  $10 \times 10^6 \tau_{CG}$ . In these simulations, the membrane binding  $\beta$ -propeller domain of NUP155, NUP133, and NUP160 were also integrated along with the lipids, while the position of the remainder of the NPC was kept constrained. In  $20 \times 10^6 \tau_{CG}$  of equilibration, all  $\beta$ -propeller domains of NUP155, NUP133, and NUP160 were associated with the lipid. The final dimension of the simulation cell was 200 nm in both the  $x$  and  $y$  direction. In the  $z$  direction, the dimension of the simulation cell was 300 nm.

### S6. CG model of HIV-1 Capsid, CA-CA, and CA-FG interactions.

We derived a “bottom-up” CG model of CA from AA MD trajectories of a composite system of three CA hexamers complexed with IP6 and FG peptides. The reference AA MD trajectory (1  $\mu$ s simulation time) from which the CG interactions were derived is a 3 CA hexamer complex with three 11-residue FG peptides, each bound to a CA hexamer (**Fig. S21**). The CG CA monomer model was derived from the AA MD trajectory using the EDCG method (2). The CA monomer contains 46 CG sites with a resolution of  $\sim 5$  AA residues per CG site. The number of CG sites was chosen to balance the computational efficiency of simulating a full capsid and least-squared error in the principal component subspace. We mapped each FG peptide into a single CG site by calculating the center of mass of the 12-residue peptide bound to the CA monomer. The initial configuration of the FG peptide in the composite AA complex was generated from PDB 5TSX. After CG mapping of each component, the intramonomer bonding topology of the CA monomer was generated using hENM with a cutoff of 3 nm (3).

Intermolecular non-bonded CG interactions between CA monomers were modeled using a combination of repulsive excluded volume ( $E_{excl}$ ) and attractive pairwise Gaussian interactions ( $E_{gauss}$ ). The value of  $A$  in  $E_{excl}$  was 15 kcal/mol for all  $ij$  pairs of CA. The distance cutoff ( $r_c$ ) for the excluded volume interactions was set at 1.0 nm. To derive the attractive interactions ( $E_{gauss}$ ) we assumed that all CG beads of CA monomers that are in close contact participate in associative interactions. We classified close contact between CA CG sites in the CG mapped AA MD trajectory of the three CA hexamer complex as the  $ij$  pairs within a distance cutoff of 1.75 nm and a standard deviation of less than 0.15 nm. The parameters  $r_{0,ij}$  and  $\sigma_{ij}$  were determined by a fit to the first peak of the pair correlation function between CG  $ij$  pairs through least-squares regression. The constant  $H_{ij}$  for CG  $ij$  pairs was optimized using Relative Entropy Minimization from the corresponding AA MD trajectory (4).

The derived CG parameters were evaluated by simulating full capsids of different shapes (conical, pill, and ellipsoid) and monitoring whether the capsid lattice order was maintained during the timescales of the simulations. The comparison of the distribution of the  $\langle q_6 \rangle_{neigh}$  values for each CA monomer calculated from the CG simulations, and initial cryo-ET configuration for all

capsid shapes simulated are shown in **Fig. S7**. The final snapshot for each capsid depicts the lattice disorder classified as CA monomers with  $\langle q_6 \rangle_{neigh} < 0.4$ . The depicted domains in **Fig. S7** are the capsid lattice's intrinsic disorder (19).

Attractive interaction between the CG FG site and CA monomer was derived using the identical protocol described earlier. For all CG  $ij$  pairs for the CA-FG attractive interactions ( $E_{gauss}$ ) the  $\sigma_{ij}$  value of 0.12 nm was used.

### S7. CG trajectory analysis.

(a) *Capsid-FG association*: We identify CA-FG association by using the following distance-based criteria:

$$d_{CA-FG} = \left| \frac{1}{N_{CA}} \sum_{i=1}^N r_{CA(i)} - r_{FG} \right| \quad (4)$$

where  $d_{CA-FG}$  is the distance between the selected FG site and center of mass of the CA monomer calculated over selected indices. Here,  $r$  is the coordinate of the CG sites and  $N_{CA}$  is the total number of CG sites considered for the distance calculation. For the calculation, CA monomer CG site indices 9-14, 21-23, and 26-28 are considered. Note that a CA monomer consists of 46 CG sites. A CA monomer and FG site of NUP62 is associated if  $d_{CA-FG} < 3$  nm.

(b) *Capsid lattice order analysis*: We quantified the capsid lattice order by the neighbor averaged Steinhardt bond order parameter (20, 21). The Steinhardt bond order parameters determine the order of the local environment using an algorithm based on spherical harmonics. To calculate the averaged Steinhardt bond order parameter, we considered the CG site index 21, which is the closest site to the geometric center of the CA monomer. Considering the selected CG site for each CA monomer of the viral capsid, the neighbor averaged Steinhardt bond order parameter

$$\langle q_6 \rangle_{neigh} = \left( \frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{q}_{lm}|^2 \right)^{1/2} \quad (5)$$

where  $\bar{q}_{lm}(i) = \frac{1}{N_B(i)} \sum_{k=0}^{N_B(i)} q_{lm}(i)$  and  $l = 6$ . Here  $N_B$  is the number of neighbors of CA monomer  $i$ , and the reference CA monomer itself. The term  $q_{lm} = \frac{1}{N_B} \sum_{i=1}^{N_B} Y_{lm}(\theta_i(r), \phi_i(r))$ , and  $Y_{lm}(\theta_i(r), \phi_i(r))$  are spherical harmonics of rank  $l$  and  $m$ , where  $\theta_i(r)$  and  $\phi_i(r)$  are the polar angles of each of the  $N_B$  bonds between the central CA monomer and neighboring CA monomers. We note that the closest 3 neighbors of CA monomer  $i$  are considered for the calculation of  $\langle q_6 \rangle_{neigh}$ . For a specific CA monomer, the higher the value of  $\langle q_6 \rangle_{neigh}$ , the lattice contiguous to the CA monomer is more ordered. Conversely, the lower values of  $\langle q_6 \rangle_{neigh}$  indicate perturbation from the ideal lattice packing and therefore indicate more disorder and a likely weakening of the capsid structural integrity.

To determine the cutoff for deviation from ideal lattice packing, we first fit the probability distribution of the  $\langle q_6 \rangle_{neigh}$  values (**Fig. S7**) for the freely diffusing cone-shaped capsid (no RNP inside) to a gaussian functional form (**Fig. S22**). The fitting resulted in the mean ( $\mu = 0.52$ ) and standard deviation ( $\sigma = 0.12$ ). The choice of cutoff corresponds to all the  $\langle q_6 \rangle_{neigh}$  values below  $(\mu - \sigma) < 0.4$ . For the freely diffusing cone-shaped capsid (no RNP) this choice results in ~16% of CA monomers deviating from ideal lattice packing, and hence classified as disordered. Note, the freely diffusing capsid is unstressed, in contrast to the stressed lattice when the capsid is docked at the NPC central channel.

(c) *Largest disordered cluster calculation*: To investigate the formation of disordered domains of capsid lattice during translocation through the NPC central channel, we performed clustering analysis. First, we identified all CA monomers with  $\langle q_6 \rangle_{neigh} < 0.4$ . Then for a particular CA

monomer with  $\langle q_6 \rangle_{neigh} < 0.4$  we identified any closest 3 neighbors with  $\langle q_6 \rangle_{neigh} < 0.4$ . The largest disordered cluster was defined as the cluster containing the most contiguous CA monomers with  $\langle q_6 \rangle_{neigh} < 0.4$  for each configuration. The radius of gyration ( $R_g$ ) of the largest cluster was calculated for the largest connected disordered CA domain.

### **S8. CG MD simulation details of capsid translocation into the NPC central channel and freely diffusing capsid.**

Capsid translocation simulations were performed with a path-sampling approach as noted earlier. The simulations were discretized into consecutive ( $n$ ) segments. For each segment, 5 replica simulations, each  $40 \times 10^6$  CG timesteps long, were performed. For each segment, the probability distribution of  $D_{Cap-NPC}$  calculated from final  $5 \times 10^6$  CG timesteps for all 5 replica simulations were fitted to a normal distribution. The endpoint of a trajectory that yielded the  $D_{Cap-NPC}$  value closest to the mean of the cumulative probability distribution was used to initialize the next simulation segment (**Fig. S8**). Finally, the selected trajectory from each segment was concatenated to create the complete trajectories of capsid translocation.

Simulations of freely diffusing capsids were performed by initially placing a capsid in a simulation cell of dimension 200 nm in  $x$ ,  $y$ , and  $z$  directions. The initial position of the capsid was adjusted such that the geometric center of the capsid was aligned with the center of the simulation cell. These simulations were performed in the constant  $NVT$  ensemble. The temperature of the system was maintained using a Langevin thermostat at 300 K with a coupling constant of 2000  $\tau_{CG}$ .

### **S9. CG model of Ribonucleoprotein Complex (RNP).**

The CG RNP in our simulations was modeled as a linear polymer chain containing 3000 beads. Each RNP polymer in our simulation minimally represents a 9-kb RNA genome complexed with nucleocapsid proteins. The mass of each CG site in the RNP was set to 500 Da. Each CG bead of the RNP was linked with flexible harmonic bonds with a harmonic force constant of 0.5 kcal/mol  $\text{\AA}^{-2}$ . The CG beads of the RNP interact with each other through a 12-6 Lennard-Jones potential ( $E_{scLJ}$ ) with a modified soft-core. Here,  $n = 2$ ,  $\alpha_{LJ} = 0.5$ ,  $\alpha = 0.6$ , and  $\sigma = 1.25$  nm are the parameters used. The strength of the RNP-RNP interaction ( $\epsilon$ ) was varied from 0.3 to 0.9 kcal/mol, which allows for modulating the condensation state of the RNP complex as noted earlier. Intermolecular non-bonded CG interactions between CA and RNP polymer were modeled with a combination of the repulsive excluded volume ( $E_{excl}$ ) and attractive pairwise Gaussian interactions ( $E_{gauss}$ ). The parameters for the repulsive excluded volume interactions are identical to that between CA and CA. The RNP polymer interacts with two C-terminal CG beads of the CA through attractive pairwise Gaussian interactions ( $E_{gauss}$ ) to emulate electrostatic interactions between the RNP and charged residues in the C-terminal end of CA ( $H_{ij} = -4.5$  kcal/mol,  $r_{0,ij} = 1.25$  nm, and  $\sigma_{ij} = 0.12$  nm).

### **SI Appendix References:**

1. S. Mosalaganti *et al.*, AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **376**, eabm9506 (2022).
2. Z. Zhang *et al.*, A Systematic Methodology for Defining Coarse-Grained Sites in Large Biomolecules. *Biophys. J.* **95**, 5073-5083 (2008).
3. E. Lyman, J. Pfaendtner, G. A. Voth, Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **95**, 4183-4192 (2008).



4. M. S. Shell, The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **129**, 144108 (2008).
5. A. J. Pak, A. Yu, Z. Ke, J. A. G. Briggs, G. A. Voth, Cooperative multivalent receptor binding promotes exposure of the SARS-CoV-2 fusion machinery core. *Nature Communications* **13**, 1002 (2022).
6. A. J. Pak, M. Gupta, M. Yeager, G. A. Voth, Inositol Hexakisphosphate (IP6) Accelerates Immature HIV-1 Gag Protein Assembly toward Kinetically Trapped Morphologies. *Journal of the American Chemical Society* **144**, 10417-10428 (2022).
7. A. Hudait, J. H. Hurley, G. A. Voth, Dynamics of upstream ESCRT organization at the HIV-1 budding site. *Biophysical Journal* **122**, 2655-2674 (2023).
8. D. J. Evans, B. L. Holian, The Nose–Hoover thermostat. *The Journal of Chemical Physics* **83**, 4069-4074 (1985).
9. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182-7190 (1981).
10. J. Huang *et al.*, CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71-73 (2017).
11. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).
12. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).
13. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092 (1993).
14. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
15. T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, W. F. van Gunsteren, Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters* **222**, 529-539 (1994).
16. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
17. J. M. A. Grime, J. J. Madsen, Efficient Simulation of Tunable Lipid Assemblies Across Scales and Resolutions. *arXiv e-prints*, arXiv:1910.05362 (2019).
18. A. Yu *et al.*, A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys. J.* **120**, 1097-1104 (2021).
19. S. Mattei, B. Glass, W. J. H. Hagen, H.-G. Kräusslich, J. A. G. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434-1437 (2016).
20. W. Lechner, C. Dellago, Accurate determination of crystal structures based on averaged local bond order parameters. *The Journal of Chemical Physics* **129**, 114707 (2008).
21. P. J. Steinhardt, D. R. Nelson, M. Ronchetti, Bond-orientational order in liquids and glasses. *Physical Review B* **28**, 784-805 (1983).