Defining the HIV Capsid Binding Site of Nucleoporin 153

Shunji Li [1], Jagdish Suresh Patel [1,2], Jordan Yang [3], Angela Marie Crabtree [1], Brenda M. Rubenstein [3,5], Peik Karl Lund-Andersen [1], Frederick Marty Ytreberg [2,4], Paul Andrew Rowley [1]*

[1] Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA
[2] Institute for Modeling Collaboration and Innovation, University of Idaho, Moscow, ID 83844, USA
[3] Department of Chemistry, Brown University, Providence, RI 02912, USA
[4] Department of Physics, University of Idaho, Moscow, ID 83844, USA
[5] Center for Computational Molecular Biology, Brown University, Providence, RI 02912, USA

*Correspondence: prowley@uidaho.edu;

Abstract

The interaction between the HIV-1 capsid (CA) and human nucleoporin 153 (NUP153) is vital for delivering the HIV-1 preintegration complex into the nucleus via the nuclear pore complex. The interaction with CA requires a phenylalanine/glycine-containing motif in the C-terminus of NUP153. This study used molecular modeling and biochemical assays to determine the amino acids of NUP153 that are essential for its interactions with CA. Molecular dynamics, FoldX, and PyRosetta simulations delineated the minimal CA binding motif of NUP153 based on the known structure of NUP153 bound to the HIV-1 CA hexamer. Computational predictions were experimentally validated by testing the interaction of NUP153 with CA using an *in vitro* binding assay and a cell-based TRIM-NUP153C restriction assay. This multidisciplinary approach identified eight amino acids from P1411 to G1418 that stably engage with CA, with significant correlations between molecular models and empirical experiments. Specifically, P1411, V1414, F1415, T1416, F1417, and G1418 were confirmed as critical amino acids required to interact NUP153 with CA.

IMPORTANCE: Human immunodeficiency virus (HIV) can infect non-dividing cells by interacting with host nuclear pores. The host nuclear pore protein NUP153 directly interacts with the HIV capsid to promote viral nuclear entry. This study used a multidisciplinary

approach combining computational and experimental techniques to map the essential amino acids of NUP153 required for HIV capsid interaction. This approach revealed that the HIV capsid interacts specifically with only six amino acids of NUP153, suggesting other FG-containing motifs could also interact with the capsid. Based on molecular modeling, naturally occurring polymorphisms in human and non-human primates would be predicted to prevent NUP153 interaction with capsid, potentially protecting from HIV infection.

Keywords: HIV; NUP153; Capsid; Molecular modeling; FoldX; PyRosetta.

Introduction

In eukaryotic cells, the nuclear envelope compartmentalizes the cytoplasm from the nucleoplasm. It is a physical barrier that must be traversed by viruses requiring access to the nucleus during their lifecycle, particularly when infecting non-dividing cells (1, 2). Access to the nucleus is through membranous pores in the nuclear envelope that are each stabilized by a large assemblage of ~30 different nucleoporin proteins called the nuclear pore complex (NPC) (3). The NPC regulates nucleocytoplasmic transport with a selectively permeable barrier of unstructured filamentous nucleoporins that fill the nuclear pore and project from its surface. These filamentous nucleoporins contain an abundance of phenylalanine/glycine (FG) repeats that create a hydrophobic barrier to prevent the free diffusion of large macromolecules. Cellular proteins interact directly with nucleoporins to enable the nuclear ingress and egress of specific cellular cargos and interaction with FG nucleoporins is important for the efficient trafficking of macromolecules that are larger than ~40 kDa.

Lentiviruses require the NPC to transport viral proteins and nucleic acids during the infection of non-dividing cells. Specifically, the HIV-1 genome is delivered to the NPC encapsidated in a fullerene cone constructed of capsid (CA) hexamers and pentamers (4, 5). The CA multimer has been observed docking with the surface of the NPC, but there is currently much debate on the exact mechanism of HIV-1 nuclear ingress (6). Many HIV-1 proteins have been shown to traffic to the nucleus, but the CA plays a dominant role in enabling the infection of non-dividing cells (7, 8). Genome-wide RNA interference screens have identified several nucleoporins required to complete the HIV-1 lifecycle (9–12). Of all the nucleoporins depleted from human cells in large-scale screens, NUP153 was consistently important for HIV-1 infection. NUP153 depletion resulted in up to a 100-fold drop in HIV-1 infectivity, reduced nuclear import of cDNA, and integration (13–17). The importance of NUP153

during viral nuclear ingress extends to other primate lentiviruses (18). However, it is less critical for lentiviruses that infect other mammals, such as equine infectious anemia virus and feline immunodeficiency virus (16).

NUP153 has an overall disordered structure and is anchored by its N-terminal domain to the nuclear basket. The NUP153 C-terminal domain (NUP153C) is rich in FG motifs that can project into the cytoplasm and nucleoplasm (19). The FG-rich C-terminus of NUP153 is required for CA binding (20), with a motif at amino acid positions 1407-1422 (TNNSPSGVFTFGANSS) playing a dominant role in this interaction (Figure 1A) (16, 21). NUP153C interaction does not occur with CA monomers and is specific to a hydrophobic pocket at the interface between adjacent monomers of the CA hexamer (21). Mutations within CA that disrupt this pocket can prevent HIV-1 infection of non-dividing cells and nuclear ingress of the preintegration complex. The central FTFG sequence of this motif (amino acids 1415-1418) is crucial for the interaction with a hydrophobic pocket formed between CA monomers. Moreover, mutation of the amino acids F1415, T1416, and F1417 in NUP153 interferes with the CA interaction (16, 21). Similarly, other host proteins also interact with CA at the same interface as NUP153, including CPSF6 and SEC24C (22, 23). All these proteins insert phenylalanine sidechains into the same hydrophobic pocket but have differences in the surrounding amino acid sequence. CA-targeting small molecules PF74, BI-2, and Lenacapavir also insert phenyl groups into the same binding pocket (24–26). Although there is an abundance of phenylalanine residues in many nucleoporins, there is a specific interaction of the NUP153 motif 1407-1422 with CA (16, 21). However, the sequence-specific determinants of this interaction have not been rigorously identified.

This study aimed to use molecular modeling combined with laboratory techniques to characterize the known interaction between NUP153 and HIV-1 CA. Molecular modeling has been utilized in many biological systems to answer fundamental questions regarding protein folding and function (27–29) and provides detailed information about how protein residues interact with a binding partner at the atomic scale. There are 29 FG repeats within NUP153, and it is unclear what makes the motif at positions 1407-1422 unique in its specific interaction with CA. Molecular dynamics (MD) simulations and *in silico* mutagenesis were used to determine the residues required for CA interaction with NUP153. These modeling predictions were validated by assaying mutant NUP153 and its interaction with CA in cell-based and *in vitro* CA pulldown assays. We find that modeling predictions correlate well with

empirical studies. The stable interaction of specific residues of NUP153 with CA and their sensitivity to mutation has enabled the determination of the specific sequence motif within NUP153 required for interaction with CA.

Results

Molecular modeling identified residues essential for CA-binding in NUP153.

The FG-containing motif of NUP153 interacts with a hydrophobic pocket formed by two adjacent CA monomers (Figure 1B) (21). The first molecular modeling approach involved conformational sampling of this CA-NUP153C complex via molecular dynamics (MD) simulations (30). Employing MD allowed us to investigate subtle conformational changes during the simulation and provided information on the stability of the interaction of NUP153C with CA. The root mean square deviation (RMSD) calculated using the NUP153 peptide backbone atoms confirmed that amino acids P1411-G1418 were stably associated with CA during 100 ns MD simulations (Figure 1C). Root mean square fluctuations (RMSF) analysis indicated that amino acids P1411-G1418 fluctuated less while in the binding pocket of CA during MD simulation (RMSF < 2.5 Å) (Figure 1D). Larger RMSF values of amino acids 1407-1410 and 1419-1422 indicated that they did not stably interact with CA during these simulations. The protein-protein binding affinity prediction tools FoldX and PyRosetta were then employed to assess the effects of amino acid substitutions in NUP153C on the binding stability using the $\Delta\Delta G_{bind}$ value, where $\Delta\Delta G_{bind} = \Delta G_{bind}$ (mutant) - ΔG_{bind} (wild-type). The FoldX analysis combined the FoldX software with MD simulations to compute $\Delta\Delta G_{bind}$ values for all possible 19 amino acid substitutions at each site in the NUP153 motif (27). Overall, a negative $\Delta\Delta G$ value suggests the binding is stabilized by an amino acid substitution, whereas a positive value indicates destabilization. Five amino acid residues, P1411, V1414, F1415, F1417, and G1418, were considered critical binding sites for the CA interaction, as substituting these residues resulted in more positive $\Delta\Delta G_{bind}$ values (Figure 1E). These residues are part of the region in NUP153C (P1411-G1418) that is stably associated with CA during MD simulations. The $\Delta\Delta G_{bind}$ values were compared to each substitution's volume and hydrophobicity values (Table 1). Increased CA binding (decreasing $\Delta\Delta G_{bind}$ values) was significantly correlated with the mutations that altered the volumes of the amino acid sidechains. For the residues F1415-F1417, increasing sidechain volume increased binding, whereas smaller sidechain volume at residues G1413, V1414, and G1418

increased binding (Table 1). CA binding at position T1416 was correlated with increased sidechain volume and hydrophobicity (Table 1).

To validate the predictions made by MD+FoldX, we selected two substitutions at each position from P1411-G1418 for analysis with PyRosetta. These 16 mutations represented substitutions with high or low $\Delta\Delta G_{bind}$ values. Overall, PyRosetta agreed with the MD+FoldX predictions for mutations at positions G1413, T1416, and G1418 and predicted larger $\Delta\Delta G_{bind}$ values for the destabilizing mutants of P1411, S1412, V1414, F1415, and F1417 (Figure 1F). Comparing the predictions of MD+FoldX and PyRosetta resulted in a strong positive correlation (Pearson's correlation coefficient r = 0.88, p <0.0001) (Figure 1G).

Molecular modeling predicts the effects of mutations in NUP153C on the CA interaction as measured by co-sedimentation. Molecular modeling predictions suggest that specific amino acids were more important for the NUP153C interaction with CA. To validate modeling predictions from both MD+FoldX and PyRosetta, 16 mutations in the central PSGVFTFG motif (residues 1411-1418) were created in NUP153C to represent eight substitutions with high $\Delta\Delta G_{bind}$ and eight with low $\Delta\Delta G_{bind}$. Each mutant NUP153C was expressed in a HEK293T human cell line as a TRIM domain fusion and a C-terminal HA tag. Cell lysates containing NUP153C were used to determine interaction with recombinant purified multimeric CA tubes (Figures 2A and 2B). To assemble these tubes, CA monomers with engineered cysteine mutations were cross-linked to form hexamers (Figure 2A) and assembled into higher-order tubular structures (Figure 2B). The NUP153C-CA interaction was determined based on the fraction of NUP153C that bound and co-sedimented with multimeric CA tubes (Figure 2C). Wild-type NUP153C was efficiently bound to CA tubes, with 46% detected in the pellet fraction (Figure 2C and 2D). The binding of NUP153C with CA depended on the formation of the multimeric CA tubes as the reduction of the cystine bonds disassembled the CA tubes and localized NUP153C to the supernatant fraction (Figure 2C). A mutant NUP153C with a deletion of the entire interaction motif (ΔP1411-G1418) resulted in only 26% of NUP153C binding to CA (Figure 2D). As predicted by molecular modeling, the substitutions with low $\Delta\Delta G_{bind}$ (S1412M, G1413M, F1415M, F1417Y, and G1418A) did not reduce the binding with CA compared to wild-type NUP153C (Figure 2D). The mutations G1413M, F1415M, and F1417Y significantly increased the binding of NUP153C to CA tubes compared to the wild-type. Destabilizing mutations that had the highest $\Delta\Delta G_{bind}$ values (average of 4.62 ± 4.98 kcal/mol MD+FoldX, 7.295 ± 7.349 kcal/mol

PyRosetta) displayed reduced binding to CA tubes similar to the deletion mutant (ΔP1411-G1418) (Figure 2D). The mutants S1412P and G1413W appeared to bind CA similar to wild type NUP153C, reflecting their lower $\Delta\Delta G_{bind}$ values (1.413/0.544 kcal/mol MD+FoldX and 5.919/1.159 kcal/mol PyRosetta, respectively). The conservative mutations P1411M, V1414I, and T1416M did not bind with CA regardless of their lower $\Delta\Delta G_{bind}$ values (Figure 2D). These results suggested that P1411, V1414, F1415, T1416, F1417, and G1418 are necessary for CA interaction as they appeared most sensitive to mutation. S1412 or G1413 were less critical for CA interaction and could tolerate mutations predicted to increase $\Delta\Delta G_{bind}$. A comparison of these binding data with modeling predictions shows a significant negative correlation (Pearson's correlation coefficient -0.54 (p<0.001) and -0.39 (p<0.01), for MD+FoldX and PyRosetta, respectively) (Figure 1E).

Mutations predicted to disrupt NUP153C-CA interaction prevent HIV-1 restriction by TRIM-NUP153C. Modeling predictions made by MD+FoldX and PyRosetta were further scrutinized by testing the interaction between NUP153C and CA using a cell-based assay (Figure 3A) (16). HEK293T cells were transiently transfected with NUP153C with an N-terminal fusion to the TRIM domain from the Rhesus Macaque TRIM5α restriction factor. Cell lines expressing TRIM-NUP153C were challenged with HIV-GFP pseudotyped with VSV-G (Figure 3A). Interaction between TRIM-NUP153C and the CA resulted in the restriction of viral replication and a ~2-fold drop in GFP positive cells, to 53%, compared to no TRIM control (Figure 3B). Of the eight NUP153C mutants that were predicted not to affect CA interaction (low $\Delta\Delta G_{bind}$ values), five displayed wild type-like CA interaction reducing HIV-1 transduction to 69% (SD ±11.27) (P1411M, S1412M, G1413M, V1414I, and F1415M) (Figure 3B). The remaining three mutations were active in HIV-1 restriction but only reduced transduction to 81.21% (SD ± 4.87) (T1416M, F1417Y, and G1418A), indicating a loss of CA interaction. Five of the eight disruptive mutations with high $\Delta\Delta G_{bind}$ values reduced transduction to 87.36% (SD ± 6.74) (V1414W, F1415G, T1417R, F1417G, and G1418Y) (Figure 3B). The less disruptive mutations with high $\Delta\Delta G_{bind}$ values demonstrated almost wild-type-like restriction of HIV (P1411Y, S1412P, and G1413W). Measuring the expression of the TRIM-NUP153C mutants indicated that the majority were expressed to a similar level within HEK293T cells (Figure 3B, bottom). When there was reduced expression of NUP153C, there did not appear to be a reduction in HIV restriction (P1411Y, P1411M, G1413W, G1413M). Of all the mutants tested, only V1414W showed decreased expression and a concomitant decrease in HIV restriction. As demonstrated with

the co-sedimentation assay, we found a significant correlation between modeling predictions and TRIM-NUP153C restriction (Pearson's correlation coefficient 0.66 (p<0.001) and 0.57 (p<0.001) for MD+FoldX and PyRosetta, respectively) (Figure 3C).

Discussion

Using a multidisciplinary approach combining molecular modeling and empirical laboratory assays, we have identified the critical amino acid residues in the C-terminal domain of NUP153 required for CA interaction. We have determined that the minimal CA interaction motif is likely eight amino acids in length based on MD simulations and *in silico* mutagenesis using two different modeling software programs (MD+FoldX and PyRosetta). Specifically, we find that amino acids 1411-1418 (PSGVFTFG) remain stably associated with CA during MD simulations. Predictions made by both modeling approaches were in general agreement that the mutation of six of these eight residues would be disruptive to CA interaction. An *in vitro* co-sedimentation assay and a cell-based TRIM-NUP153C restriction assay were used to determine the biological relevance of the modeling predictions. Overall, MD+FoldX and PyRosetta predictions correlate well with empirical data, confirming the specific binding of CA to NUP153C. These data suggest that the minimal required sequence for the binding of NUP153C to CA is PxxVFTFG, where x is any amino acid.

Validation of MD+FoldX $\Delta\Delta G_{bind}$ predictions using PyRosetta highlights a strong agreement between two different tools using different energy functions. MD+FoldX seemed to discern CA-interacting from non-interacting mutants better than PyRosetta, as confirmed by experimental data (Figures 2E and 3C). This difference between the models is likely because PyRosetta uses a single experimental structure of the NUP153-CA complex to predict $\Delta\Delta G_{bind}$ values. In contrast, the MD+FoldX approach uses multiple snapshots extracted from MD simulations. Improved performance of MD+FoldX in predicting $\Delta\Delta G_{bind}$ values also highlighted the importance of incorporating conformational sampling. Interestingly, both methods did poorly in predicting the effects of mutating T1416. Indeed, both methods suggest small $\Delta\Delta G_{bind}$ values for T1416M and T1416R, indicating binding stabilization. Conversely, experimental data has shown that the mutations T1416M and T1416R (this study) and T1416A (21) disrupt CA interaction. This discrepancy between modeling and empirical data is likely due to the water-mediated interaction of T1416 with CA residues R173 and E63 (21). These modeling inaccuracies are expected because one of the significant

limitations of fast protein-protein binding affinity prediction tools (FoldX and PyRosetta) is that they ignore the explicit presence of bridging water molecules (31).

The hydrophobic pocket in the CA hexamer has been reported to accommodate the host factors CPSF6 and SEC24C (21–23). The same pocket has also been targeted by small molecules such as PF74 and BI-2 (21, 24), and the antiviral drug Lenacapavir and its derivatives (25, 32). These different proteins and small molecules adopt slightly different conformations within the CA pocket, but with a common feature of a phenyl group insertion into the binding pocket with hydrophobic interactions with CA. The phenyl group required to interact NUP153 with CA is the sidechain of F1417, which overlays with CA-bound CPSF6 residue F321 and Sec24C residue F236. In addition, the backbone amides of F1417, F321, and F236 all form a hydrogen bond with the sidechain of CA N57. However, bulky sidechain substitutions appear to be tolerated at F1417, as demonstrated by a significant correlation between increasing sidechain size and CA interaction (Table 1). Mutations at position F1417 were predicted to be deleterious to CA interaction with the lowest $\Delta\Delta G_{bind}$ of 0.94 kcal/mol for the conservative mutation F1417Y. The average $\Delta\Delta G_{bind}$ value of all mutations at F1417 was 3.71 kcal/mol (SD ± 1.44), demonstrating that F1417 is crucial for CA interaction. The importance of F1417 is confirmed by the empirical findings, except that the F1417Y substitution improved CA binding. Similarly, modeling predicted that mutation of F1415 was also deleterious to CA interaction, but the average $\Delta\Delta G_{bind}$ value (1.85 kcal/mol; SD ± 0.78) was lower than that of F1417. Similar to F1417Y, the bulky F1415M mutation increased CA binding in both empirical assays and would be predicted to make hydrophobic contact with P38 of CA. The sulfur-containing sidechain of F1415M into this hydrophobic binding pocket of CA is similarly positioned to the sulfonyl group of Lenacapavir that hydrogen bonds with CA S41 and N57 (33). Both S1412 and G1413 were anchored to CA during MD simulations and were insensitive to mutations based on modeling predictions, and this agreed with empirical data. The ability of S1412 to tolerate mutation reflects the interaction with CA residue Q176 via the main chain. Similarly, G1413 can be substituted for tryptophan (G1413W) without altering CA interaction because the bulky sidechain would be exposed to the solvent. Even though these larger amino acids are accommodated at these positions, there is still a preference for small volume sidechains (Table 1).

Genetic variation in host factors hijacked by viruses can protect from viral infection and result in signatures of positive selection and NUP153 is evolving rapidly in primates (34).

However, the CA interaction site is 100% identical in 35 primate species except in gorilla (*Gorilla gorilla*) and drill (*Mandrillus leucophaeus*). The fixed S1412P substitution in gorilla NUP153 would likely have a minor effect on CA interaction ($\Delta\Delta G_{bind}$ 1.41 kcal/mol). Drill NUP153 has a more deleterious G1418S substitution ($\Delta\Delta G_{bind}$ 2.27 kcal/mol). This substitution has a polar sidechain that is less likely to be accommodated within the hydrophobic CA pocket and limits the flexibility of the polypeptide backbone. Interestingly, a simian-human immunodeficiency virus (SHIV) chimera with the CA from SIV_{mnd1} (infecting *Mandrillus sphinx*) does not require NUP153 when infecting human cells (18). It is tempting to speculate that the incompatibility between SIV_{mnd1} and human NUP153 could be due to the adaptation of the SIV_{mnd1} CA to accommodate S1418 in this primate NUP153 or to circumnavigate NUP153 entirely. In humans, there were no high-frequency single nucleotide polymorphisms (SNPs) that would alter the amino acid sequence of the CA interaction site in NUP153 (gnomAD database (35). Two low-frequency non-synonymous SNPs were found at the same positions as in gorilla and drill resulting in the mutations S1412A and G1418V. These mutations in these individuals could potentially disrupt CA interaction and provide protection from HIV infection ($\Delta\Delta G_{bind}$ 0.09 and 3.29 kcal/mol, respectively). Despite ongoing pressure from primate lentiviruses, there may have been selection against non-synonymous mutations at the CA interacting FG-containing motif of NUP153.

The minimal PxxVFTFG sequence required for CA interaction could mean that other FG repeats within NUP153 and other nucleoporins could bind CA. This is supported by the fact that only a complete deletion of the FG-region of NUP153 will abolish CA binding (20) and non-synonymous substitutions and small deletions do not completely perturb CA interaction (16). The relevance of other FG repeats in NUP153 for CA interaction remains to be thoroughly investigated.

Materials and Methods

Structure preparation for molecular modeling

The X-ray crystal structure of the HIV-1 CA hexamer interacting with human NUP153C was downloaded from Protein Data Bank (PDB ID:4U0D) (21). 3D coordinates file was modified to remove all but six chains of CA monomer and six chains of NUP153C. MODELLER software altered engineered residues in CA protein to wild-type and built the missing residues in all the chains to complete the experimental structure (36).

Molecular dynamics simulations

The complete structure of the HIV CA hexamer bound to NUP153C was used as a starting structure for the MD simulation. The input structure was subjected to MD simulation using the protocol reported in our previous study (37). Briefly, the AMBER99SB*-ILDNP force field and the GROMACS 5.1.2 software package were used for generating topology files and performing simulations (38, 39). The final production simulation was run for 100 ns, and snapshots were saved every 1 ns resulting in 100 snapshots for the protein complex. The MD trajectory was visualized using the VMD software package and analyzed using the *grmsf* module available in the GROMACS package to calculate the root mean square fluctuation (RMSF) of all the atoms in each residue in the NUP153 motif during the simulation (40).

Mutagenesis of NUP153C by FoldX and PyRosetta

MD snapshots of the HIV hexamer CA – NUP153 complex were analyzed using the FoldX software to estimate the relative binding affinities ($\Delta\Delta G_{bind}$) for all possible mutations at each site in the NUP153 motif. As with our previous studies (27, 37), our FoldX analysis protocol involved processing each snapshot six times in succession using the RepairPDB command to energy minimize the snapshot and the BuildModel command to generate all possible 19 single mutations at each site in the NUP153 motif. The binding affinity (ΔG_{bind}) was subsequently estimated using the AnalyseComplex command. $\Delta\Delta G_{bind}$ for each mutation was calculated by taking the difference between mutated and wild-type ΔG_{bind} values. We then averaged $\Delta\Delta G_{bind}$ values across all individual snapshot estimates for each mutation. To estimate $\Delta\Delta G_{bind}$ values for all possible 19 mutations at each amino acid site of NUP153, we performed 30,400 FoldX calculations (16 NUP153 residues × 19 possible mutations at each site × 100 MD snapshots). Finally, we obtained 304 averaged $\Delta\Delta G_{bind}$ values for all possible mutations of the NUP153 motif (see S1 File). PyRosetta-4 was used to compute the difference in binding stability scores between 16 selected mutant and wild-type structures (PDB: 4U0D) (41). The score is designed to capture the change in thermodynamic binding stability caused by the mutation (42). First, we repacked all sidechains sampled from the 2010 Dunbrack rotamer library (43) and applied the quasi-Newton minimization method via the 'dfpmin' algorithm in PyRosetta (44) with a tolerance of 0.001 (45) and the REF2015 scoring function (46) and allowing both the backbone torsion and sidechain angles to move. This procedure was performed ten times, and the lowest-scoring structure was selected for introducing mutations and subsequent binding stability calculations. Next, each missense mutation was introduced into the model of NUP153. All residues within a 10 Å distance of

the mutated residue's center were repacked, followed by a Monte Carlo sampling coupled with a quasi-Newton minimization of the backbone and all sidechains. We performed ten independent simulations of 5,000 Monte Carlo cycles each. To compute binding energy, we first scored the total energy of a bound state structure, separated CA and NUP153C, and then scored the unbound state total energy. The binding energy (ΔG_{bind}) is computed by subtracting the unbound state total energy from the bound state total energy. This procedure was performed ten times, and the predicted $\Delta\Delta G_{bind}$ was obtained by taking the average of the three lowest scoring structures. All molecular modeling data can be found in File S1.

Plasmids construction and mutagenesis

The plasmid pLPCX-TRIM-NUP153C(human)-HA encoding the TRIM domain from TRIM5α of *Rhesus macaque* (residues 1 to 304) fused to the HA-tagged human NUP153 C-terminal domain (896 to 1475) was obtained from the Engelman laboratory. TRIM-NUP153C-HA was amplified by PCR and sub-cloned to the Gateway™ entry vector pCR8 to create plasmid pUI034. Gateway™ cloning introduced the gene into the custom destination vector pCDNA3-GW, following the manufacturer's instructions (Thermo Fisher). Site-directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis Kit, following the manufacturer's instructions (New England Biolab). All primers used for cloning and site-directed mutagenesis can be found in Table S1 and a list of plasmids used in Table S2.

Cells and virus

HEK293T cells (ACS-4500™, ATCC) were maintained at 37°C with 5% CO_2 in Dulbecco's Modified Eagle Medium (Sigma-Aldrich #D6429) supplied with 10% Fetal Bovine Serum (Sigma-Aldrich), 2 mM L-Glutamine (VWR #L0131-0100), and 1% Penicillin/Streptomycin solution (Corning, #30–002). Single-cycle HIV-1 virus with a GFP reporter gene was generated using a 10 mm dish by co-transfecting HEK293T cells with 4 µg psPAX2 (Addgene #12259), 4 µg pLJM1-EGFP (Addgene #19319), and 4 µg pCMV-VSVG (Addgene #8454) using Lipofectamine 3000 following the manufacturer's instruction (Invitrogen). After 48 h, the supernatant was passed through a 0.45 µm filter and stored at -80°C.

Purification and in vitro assembly of CA hexamers and tubes

The expression of HIV-1 CA protein was adapted from Pornillos et al. (47). *E. coli* BL21(DE3)pLysS was transformed with pET11a-HIV-NL4-3 encoding CA with the mutations A14C and E45C. Transformed bacteria were cultured in LB media with ampicillin (15 µg/mL) and chloramphenicol (100 µg/mL) at 37°C until OD_{600} 0.8. Expression of CA was induced with a final concentration of 1 mM IPTG and incubated at 37°C for 4 h. Cells were centrifuged at 4,500 *x g* for 20 min at 4 °C. Cells were suspended in lysis buffer (50 mL for 4L; 50 mM Tris-Cl pH 8.0, 50 mM NaCl, 100 mM β-ME, protease inhibitor cocktail tablets [Sigma-Aldrich #11836153001]). The cell suspension was incubated on ice for 20 min with the addition of 1 g of lysozyme and 50 U of Benzonase® (EMD Millipore #70746-3). The cell suspension was subjected to sonication (MICROSON™ XL Ultrasonic cell disruptor) for 10 s at 80% of maximum output power for a total processing time of 5 min. Between pulses, samples were allowed to cool for 30 s on ice. The cell lysate was clarified by centrifugation (27,000 *x g* for 1 h at 4°C) and incubated with supersaturated ammonium sulfate (final concentration 25% of the total volume) on ice for 20 min. Precipitated CA was collected by centrifugation at 9,000 *x g* for 20 min at 4°C. The pellet was suspended in dialysis buffer (20 mM MOPS pH 6.8, 20 mM β-ME), transferred to a 3.5K MWCO dialysis cassette (Thermo Fisher #PI66110), and dialyzed against 1 L dialysis buffer for 16 h.

CA was purified using ion-exchange chromatography (ÄKTA start protein purification system, Cytiva). Dialysed lysate was centrifuged at 20,000 *x* g for 10 min at 4°C, passed through a 0.45 µm filter, and applied to a HiTrap SP FF (Cytiva #17-5054-01) column connected to a HiTrap Q FF column (Cytiva #17-5156-01). Fractions were collected at ~25% of the sodium chloride gradient (~0.25 M) and assayed for purity by SDS-PAGE (Fig S1A). Eluted CA protein was dialyzed using a 10K MWCO cassette (Thermo Fisher #PI66130). To assemble CA oligomers, the cassette was sequentially incubated at 4ºC in assembly buffer (25 mM Tris-Cl pH 8.0, 1M NaCl) supplemented with 20 mM, 2 mM, and 0.2 mM β-ME for 8 h, 24 h, and 48 h, respectively. The efficiency of CA assembly into hexamers was assessed by SDS-PAGE, and assembly into multimeric tubes was confirmed by transmission electron microscopy (Franceschi Microscopy & Imaging Center, Washington State University).

CA co-sedimentation assay

This assay was adapted from Selyutina et al. (48). Approximately 60,000 HEK293T cells were seeded into each well of a 12-well dish. After incubation for 24 h, cells were transfected with 500 ng of pCDNA3-TRIM-NUP153C using 1.5 µL of TransIT®-293 transfection reagent (Mirus Bio). Cells were incubated for 24 h before harvesting by scraping

into 100 µL of CA binding buffer (10 mM Tris, pH 7.4, 1.5 mM $MgCl_2$, 10 mM KCl, 1X Halt™ protease and phosphatase inhibitor cocktail [Thermo Fisher #PI78440]). Cell lysates were mixed for 15 min at 4ºC before centrifugation at 21,000 *x g* for 15 min at 4°C. Clarified cell lysates were collected, and the protein content was normalized to 1.5 mg/mL by Bradford assay. 20 µL of CA tubes (~112 pmol) and 80 µL of whole-cell lysate were mixed and incubated at room temperature for 1 h. The reaction was centrifuged for 8 min at 21,000 *x g* at 4°C. 15 µL of the supernatant was collected and compared to samples that were not centrifuged using Western dot blotting.

Western blotting

Samples were separated by SDS-PAGE and transferred to nitrocellulose membranes by Trans-Blot® Turbo™ Transfer System (1.0 A, 25V, 15 min). Alternatively, samples were loaded onto the nitrocellulose membrane using a Bio-Dot® microfiltration apparatus (Bio-Rad). Membranes were blocked with 3% non-fat milk in TBS with 0.1% Tween-20 (TBST) for 1 h. To probe for the HA tag, the membrane was incubated with rat anti-HA-HRP (3F10, Sigma-Aldrich #12013819001; 1 in 2,000 dilution) for 1 h. To probe for CA and actin, membranes were incubated with rat anti-p24 antibody (ARP-64571; NIH HIV Reagent Program; 1 in 5,000 dilution) or rat anti-actin antibody (clone C4, VWR #10221-880; 1 in 500 dilution) for 1 h. Membranes were washed with 5 mL of TBST three times for 5 min each with gentle rocking. The anti-p24 blots were transferred to a new tray and probed against goat anti-rat antibody (Thermo Fisher #62-652-0; 1 in 4,000 dilution) for 40 min. Blots were visualized, and signals were quantified using Amersham™ Imager 600. Exposure time was adjusted manually to ~10 seconds for anti-HA blots, ~4 s for anti-p24 blots, and ~10 seconds for anti-actin blots.

TRIM-NUP153C-mediated restriction and cell flow cytometry

This assay was adapted from Matreyek et al. (16). 60,000 HEK293T cells were seeded and transfected with TRIM-NUP153C as described in the CA co-sedimentation assay. 24 h post-transfection, cells were transduced with HIV-GFP. Media was discarded 24 h post-transduction, and fresh media was added to the wells. 48 h after transduction, cells were treated with 0.25% trypsin (VWR #16777-202) and centrifuged at 2,000 *x g* for 3 min at room temperature. Cell pellets were suspended, fixed with 300 µL of Dulbecco's phosphate-buffered saline (DPBS; VWR #45000-434) containing 1% paraformaldehyde (Electron Microscopy Sciences #15710), and incubated at 4°C for 1 h. Cells were centrifuged at 2,000 *x*

g for 3 min, and the cell pellet was suspended in 500 µL DPBS. This step was repeated twice, and the final pellet was suspended in 100 µL flow cytometry buffer (DPBS with 4% FBS) and transferred to a 96-well U bottom assay plate (CELLTREAT #229590). Fluorescent cells were quantified using the CytoFLEX S instrument (Beckman Coulter).

Author Contributions: S.L., J.S.P., J.Y., and P.A.R. wrote the manuscript.

Funding: Research reported in this publication was supported by the National Science Foundation EPSCoR Research Infrastructure Improvement Program: Track-2, award number OIA-1736253, Randall Women in Science research grant (University of Idaho), the Institute for Health in the Human Ecosystem (IHHE) research grant (University of Idaho), and by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number P20GM104420. Computer resources were provided in part by the Research Computing and Data Services of the Institute for Interdisciplinary Data Sciences, sponsored by the National Institutes of Health (NIH P30GM103324). This research also used the computational resources provided by the high-performance computing center at Idaho National Laboratory, which is supported by the Office of Nuclear Energy of the U.S. DOE and the Nuclear Science User Facilities under Contract No. DE-AC07-05ID14517. The funders had no role in study design, data collection and analysis, publishing decisions, or manuscript preparation.

Acknowledgment

We would like to thank Dr. Lee Fortunato for the advice and technical support in creating HIV-GFP. We also thank Yesol Sapozhnikov, Dr. Craig Miller, and Dr. Mohamed Megheib for their help with statistical analyses. The following reagent was obtained through the NIH HIV Reagent Program, Division of AIDS, NIAID, NIH: Monoclonal Anti-Human Immunodeficiency Virus Type 1 (HIV-1) p24 Gag (#24-2, produced *in vitro*), ARP-6457, contributed by Dr. Michael H. Malim.

1. Fay N, Panté N. 2015. Nuclear entry of DNA viruses. Front Microbiol 6:467.

469 2. Cohen S, Au S, Panté N. 2011. How viruses access the nucleus. Biochimica et biophysica
470 acta 1813:1634–1645.

471 3. Hampoelz B, Andres-Pons A, Kastritis P, Beck M. 2019. Structure and Assembly of the
472 Nuclear Pore Complex. Annu Rev Biophys 48:1–22.

473 4. Pornillos O, Ganser-Pornillos BK, Yeager M. 2011. Atomic level modeling of the HIV
474 capsid. Nature 469:424–427.

475 5. Briggs JAG, Grünewald K, Glass B, Förster F, Kräusslich H-G, Fuller SD. 2006. The
476 Mechanism of HIV-1 Core Assembly: Insights from Three-Dimensional Reconstructions of
477 Authentic Virions. Structure 14:15–20.

478 6. Guedán A, Caroe ER, Barr GCR, Bishop KN. 2021. The Role of Capsid in HIV-1 Nuclear
479 Entry. Viruses 13:1425.

480 7. Yamashita M, Emerman M. 2004. Capsid Is a Dominant Determinant of Retrovirus
481 Infectivity in Nondividing Cells. J Virol 78:5670–5678.

482 8. Yamashita M, Perez O, Hope TJ, Emerman M. 2007. Evidence for direct involvement of
483 the capsid protein in HIV infection of nondividing cells. PLoS pathogens 3:1502–1510.

484 9. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J,
485 Elledge SJ. 2008. Identification of Host Proteins Required for HIV Infection Through a
486 Functional Genomic Screen. Science 319:921–926.

487 10. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GMC, Irelan JT, Chiang C, Tu BP,
488 Jesus PDD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhen KL,
489 Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK.
490 2008. Global Analysis of Host-Pathogen Interactions that Regulate Early-Stage HIV-1
491 Replication. Cell 135:49–60.

492 11. Yeung ML, Houzet L, Yedavalli VSRK, Jeang K-T. 2009. A genome-wide short hairpin
493 RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1
494 replication. J Biol Chem 284:19463–19473.

12. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, Espeseth AS. 2008. Genome-Scale RNAi Screen for Host Factors Required for HIV Replication. Cell host & microbe 4:495–504.

13. Nunzio FD, Danckaert A, Fricke T, Perez P, Fernandez J, Perret E, Roux P, Shorte S, Charneau P, Diaz-Griffero F, Arhel NJ. 2012. Human Nucleoporins Promote HIV-1 Docking at the Nuclear Pore, Nuclear Import and Integration. PloS one 7:e46037-15.

14. Nunzio FD, Fricke T, Miccio A, Valle-Casuso JC, Perez P, Souque P, Rizzi E, Severgnini M, Mavilio F, Charneau P, Diaz-Griffero F. 2013. Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. Virology 440:8–18.

15. Kane M, Rebensburg SV, Takata MA, Zang TM, Yamashita M, Kvaratskhelia M, Bieniasz PD. 2018. Nuclear pore heterogeneity influences HIV-1 infection and the antiviral activity of MX2. Elife 7:e35738.

16. Matreyek KA, Yucel SS, Li X, Engelman A. 2013. Nucleoporin NUP153 Phenylalanine-Glycine Motifs Engage a Common Binding Pocket within the HIV-1 Capsid Protein to Mediate Lentiviral Infectivity. PLoS pathogens 9:e1003693.

17. Matreyek KA, Engelman A. 2011. The Requirement for Nucleoporin NUP153 during Human Immunodeficiency Virus Type 1 Infection Is Determined by the Viral Capsid. J Virol 85:7818–7827.

18. Mamede JI, Sitbon M, Battini J-L, Courgnaud V. 2013. Heterogeneous susceptibility of circulating SIV isolate capsids to HIV-interacting factors. Retrovirology 10:77.

19. Fahrenkrog B, Maco B, Fager AM, Köser J, Sauder U, Ullman KS, Aebi U. 2002. Domain-specific antibodies reveal multiple-site topology of Nup153 within the nuclear pore complex. J Struct Biol 140:254–267.

20. Buffone C, Martinez-Lopez A, Fricke T, Opp S, Severgnini M, Cifola I, Petiti L, Frabetti S, Skorupka K, Zadrozny KK, Ganser-Pornillos BK, Pornillos O, Nunzio FD, Diaz-Griffero F. 2018. Nup153 Unlocks the Nuclear Pore Complex for HIV-1 Nuclear Translocation in Non-dividing Cells. J Virol JVI.00648-18-78.

21. Price AJ, Jacques DA, McEwan WA, Fletcher AJ, Essig S, Chin JW, Halambage UD, Aiken C, James LC. 2014. Host Cofactors and Pharmacologic Ligands Share an Essential Interface in HIV-1 Capsid That Is Lost upon Disassembly. PLoS pathogens 10:e1004459-17.

22. Rebensburg SV, Wei G, Larue RC, Lindenberger J, Francis AC, Annamalai AS, Morrison J, Shkriabai N, Huang S-W, KewalRamani V, Poeschla EM, Melikyan GB, Kvaratskhelia M. 2021. Sec24C is an HIV-1 host dependency factor crucial for virus replication. Nat Microbiol 6:435–444.

23. Price AJ, Fletcher AJ, Schaller T, Elliott T, Lee K, KewalRamani VN, Chin JW, Towers GJ, James LC. 2012. CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. PLoS pathogens 8:e1002896.

24. Lamorte L, Titolo S, Lemke CT, Goudreau N, Mercier J-F, Wardrop E, Shah VB, Schwedler UK von, Langelier C, Banik SSR, Aiken C, Sundquist WI, Mason SW. 2013. Discovery of novel small-molecule HIV-1 replication inhibitors that stabilize capsid complexes. Antimicrob Agents Ch 57:4622–31.

25. Link JO, Rhee MS, Tse WC, Zheng J, Somoza JR, Rowe W, Begley R, Chiu A, Mulato A, Hansen D, Singer E, Tsai LK, Bam RA, Chou C-H, Canales E, Brizgys G, Zhang JR, Li J, Graupe M, Morganelli P, Liu Q, Wu Q, Halcomb RL, Saito RD, Schroeder SD, Lazerwith SE, Bondy S, Jin D, Hung M, Novikov N, Liu X, Villaseñor AG, Cannizzaro CE, Hu EY, Anderson RL, Appleby TC, Lu B, Mwangi J, Liclican A, Niedziela-Majka A, Papalia GA, Wong MH, Leavitt SA, Xu Y, Koditek D, Stepan GJ, Yu H, Pagratis N, Clancy S, Ahmadyar S, Cai TZ, Sellers S, Wolckenhauer SA, Ling J, Callebaut C, Margot N, Ram RR, Liu Y-P, Hyland R, Sinclair GI, Ruane PJ, Crofoot GE, McDonald CK, Brainard DM, Lad L, Swaminathan S, Sundquist WI, Sakowicz R, Chester AE, Lee WE, Daar ES, Yant SR, Cihlar T. 2020. Clinical targeting of HIV capsid protein with a long-acting small molecule. Nature 584:614–618.

26. Blair WS, Pickford C, Irving SL, Brown DG, Anderson M, Bazin R, Cao J, Ciaramella G, Isaacson J, Jackson L, Hunt R, Kjerrstrom A, Nieman JA, Patick AK, Perros M, Scott AD, Whitby K, Wu H, Butler SL. 2010. HIV Capsid is a Tractable Target for Small Molecule Therapeutic Intervention. PLoS pathogens 6:e1001220-10.

551 27. Patel JS, Quates CJ, Johnson EL, Ytreberg FM. 2019. Expanding the watch list for
552 potential Ebola virus antibody escape mutations. Plos One 14:e0211093.

553 28. Yang J, Naik N, Patel JS, Wylie CS, Gu W, Huang J, Ytreberg FM, Naik MT, Weinreich
554 DM, Rubenstein BM. 2020. Predicting the viability of beta-lactamase: How folding and
555 binding free energies correlate with beta-lactamase fitness. Plos One 15:e0233509.

556 29. Miller CR, Johnson EL, Burke AZ, Martin KP, Miura TA, Wichman HA, Brown CJ,
557 Ytreberg FM. 2016. Initiating a watch list for Ebola virus antibody escape mutations. PeerJ
558 4:e1674-17.

559 30. 2002. Understanding Molecular Simulation: From Algorithms to ApplicationsSecond
560 Edition. Academic Press.

561 31. Gonzalez TR, Martin KP, Barnes JE, Patel JS, Ytreberg FM. 2020. Assessment of
562 software methods for estimating protein-protein relative binding affinities. Plos One
563 15:e0240573.

564 32. Vidal SJ, Bekerman E, Hansen D, Lu B, Wang K, Mwangi J, Rowe W, Campigotto F,
565 Zheng J, Kato D, Chandrashekar A, Barrett J, Patel S, Wan H, Anioke T, Mercado NB,
566 Nkolola JP, Ferguson MJ, Rinaldi WJ, Callebaut C, Blair W, Cihlar T, Geleziunas R, Yant
567 SR, Barouch DH. 2022. Long-acting capsid inhibitor protects macaques from repeat SHIV
568 challenges. Nature 601:612–616.

569 33. Bester SM, Wei G, Zhao H, Adu-Ampratwum D, Iqbal N, Courouble VV, Francis AC,
570 Annamalai AS, Singh PK, Shkriabai N, Blerkom PV, Morrison J, Poeschla EM, Engelman
571 AN, Melikyan GB, Griffin PR, Fuchs JR, Asturias FJ, Kvaratskhelia M. 2020. Structural and
572 mechanistic bases for a potent HIV-1 capsid inhibitor. Science 370:360–364.

573 34. Meyerson NR, Rowley PA, Swan CH, Le DT, Wilkerson GK, Sawyer SL. 2014. Positive
574 selection of primate genes that promote HIV-1 replication. Virology 454–455:291–298.

575 35. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL,
576 Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA,
577 Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K,
578 Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE,

579 Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware
580 JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly
581 S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R,
582 Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts
583 K, Tolonen C, Wade G, Talkowski ME, Salinas CAA, Ahmad T, Albert CM, Ardissino D,
584 Atzmon G, Barnard J, Beaugerie L, Benjamin EJ, Boehnke M, Bonnycastle LL, Bottinger
585 EP, Bowden DW, Bown MJ, Chambers JC, Chan JC, Chasman D, Cho J, Chung MK, Cohen
586 B, Correa A, Dabelea D, Daly MJ, Darbar D, Duggirala R, Dupuis J, Ellinor PT, Elosua R,
587 Erdmann J, Esko T, Färkkilä M, Florez J, Franke A, Getz G, Glaser B, Glatt SJ, Goldstein D,
588 Gonzalez C, Groop L, Haiman C, Hanis C, Harms M, Hiltunen M, Holi MM, Hultman CM,
589 Kallela M, Kaprio J, Kathiresan S, Kim B-J, Kim YJ, Kirov G, Kooner J, Koskinen S,
590 Krumholz HM, Kugathasan S, Kwak SH, Laakso M, Lehtimäki T, Loos RJF, Lubitz SA, Ma
591 RCW, MacArthur DG, Marrugat J, Mattila KM, McCarroll S, McCarthy MI, McGovern D,
592 McPherson R, Meigs JB, Melander O, Metspalu A, Neale BM, Nilsson PM, O'Donovan MC,
593 Ongur D, Orozco L, Owen MJ, Palmer CNA, Palotie A, Park KS, Pato C, Pulver AE,
594 Rahman N, Remes AM, Rioux JD, Ripatti S, Roden DM, Saleheen D, Salomaa V, Samani
595 NJ, Scharf J, Schunkert H, Shoemaker MB, Sklar P, Soininen H, Sokol H, Spector T,
596 Sullivan PF, Suvisaari J, Tai ES, Teo YY, Tiinamaija T, Tsuang M, Turner D, Tusie-Luna T,
597 Vartiainen E, Vawter MP, Ware JS, Watkins H, Weersma RK, Wessman M, Wilson JG,
598 Xavier RJ, Neale BM, Daly MJ, MacArthur DG. 2020. The mutational constraint spectrum
599 quantified from variation in 141,456 humans. Nature 581:434–443.

600 36. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, Pieper U,
601 Sali A. 2006. Comparative Protein Structure Modeling Using Modeller. Curr Protoc
602 Bioinform 15:5.6.1-5.6.30.

603 37. Bazurto JV, Nayak DD, Ticak T, Davlieva M, Lee JA, Hellenbrand CN, Lambert LB,
604 Benski OJ, Quates CJ, Johnson JL, Patel JS, Ytreberg FM, Shamoo Y, Marx CJ. 2021. EfgA
605 is a conserved formaldehyde sensor that leads to bacterial growth arrest in response to
606 elevated formaldehyde. Plos Biol 19:e3001208.

607 38. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison
608 of multiple Amber force fields and development of improved protein backbone parameters.
609 Proteins Struct Funct Bioinform 65:712–725.

610 39. Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. 2005.
611 GROMACS: Fast, flexible, and free. J Comput Chem 26:1701–1718.

612 40. Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. J Mol
613 Graphics 14:33–38.

614 41. Chaudhury S, Lyskov S, Gray JJ. 2010. PyRosetta: a script-based interface for
615 implementing molecular modeling algorithms using Rosetta. Bioinformatics 26:689–691.

616 42. Das R, Baker D. 2008. Macromolecular Modeling with Rosetta. Annu Rev Biochem
617 77:363–382.

618 43. Shapovalov MV, Dunbrack RL. 2011. A smoothed backbone-dependent rotamer library
619 for proteins derived from adaptive kernel density estimates and regressions. Struct Lond Engl
620 1993 19:844–58.

621 44. Gray JJ, Chaudhury S, Lyskov S. 2009. The PyRosetta Interactive Platform for Protein
622 Structure Prediction and Design. CreateSpace Independent Publishing Platform; 2nd edition.

623 45. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman KW,
624 Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F,
625 Ban Y-EA, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popović Z,
626 Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D,
627 Bradley P. 2011. Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the
628 Simulation and Design of Macromolecules. Methods Enzymol 487:545–574.

629 46. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov
630 MV, Renfrew PD, Mulligan VK, Kappel K, Labonte JW, Pacella MS, Bonneau R, Bradley P,
631 Dunbrack RL, Das R, Baker D, Kuhlman B, Kortemme T, Gray JJ. 2017. The Rosetta All-
632 Atom Energy Function for Macromolecular Modeling and Design. J Chem Theory Comput
633 13:3031–3048.

634 47. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua Y, Whitby FG, Stout CD, Sundquist
635 WI, Hill CP, Yeager M. 2009. X-Ray Structures of the Hexameric Building Block of the HIV
636 Capsid. Cell 137:1282–1292.

637 48. Selyutina A, Bulnes-Ramos A, Diaz-Griffero F. 2018. Binding of host factors to
638 stabilized HIV-1 capsid tubes. Virology 523:1–5.

639

640

641

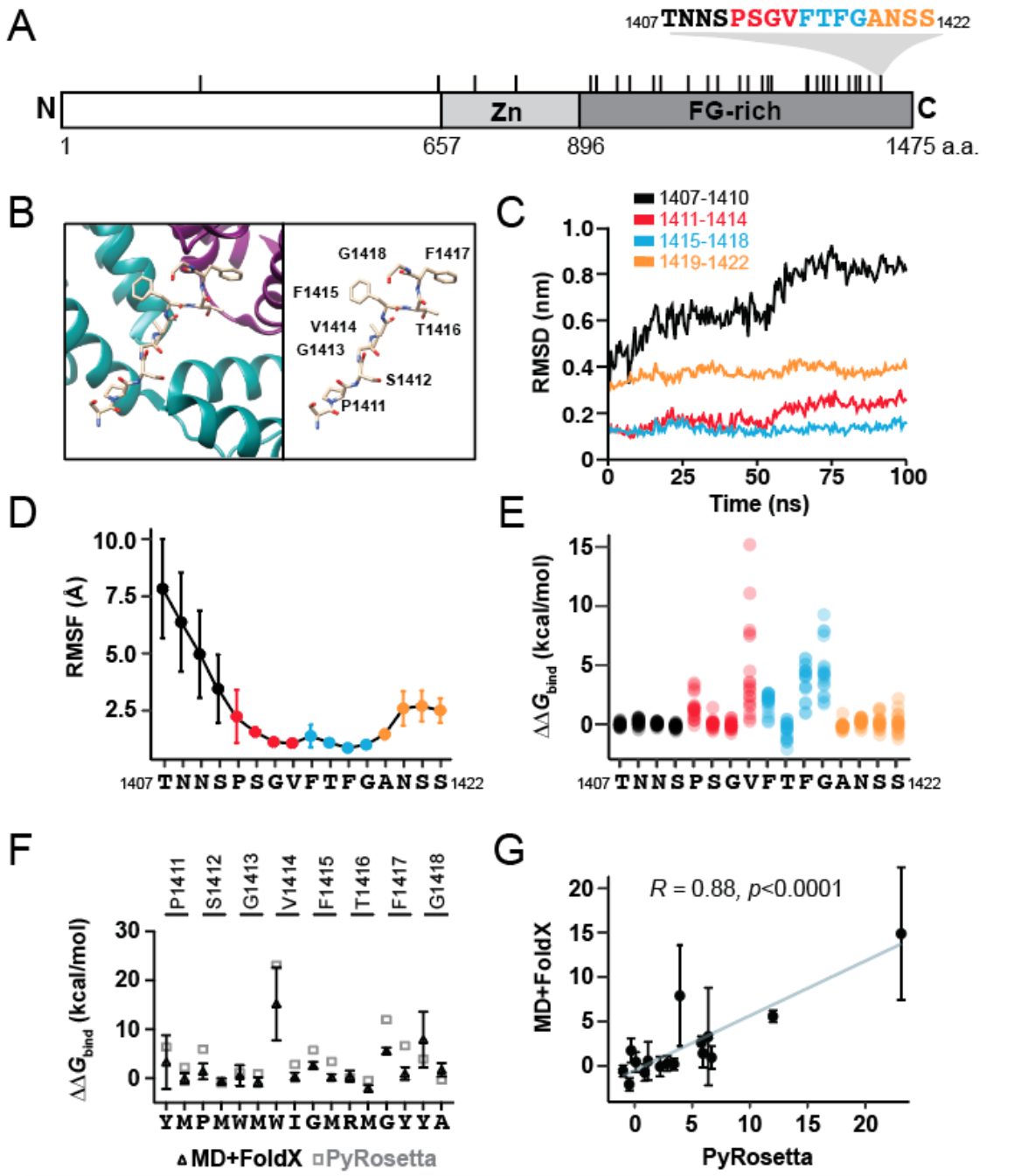

Figure 1. Molecular modeling of NUP153C-CA interaction defines a central eight amino acids that are stably associated with CA. (A) Domain diagram of the NUP153 protein. Tick marks represent FG repeats, with an expanded view of the FG motif that interacts directly with the HIV-1 CA (colored in relation to panels C-E). The numbering of NUP153 represents amino acid residues at the termini and domain boundaries. Zn - zinc finger domain. (B) *Left:* A structural representation of the HIV-1 CA hexamer bound by the NUP153C peptide (PDB: 4U0C). Magenta: monomer A; Green: monomer B. *Right:* A labeled representation of NUP153C without the CA structure. (C) RMSD of different regions of the NUP153C peptide

651 during a 100 ns MD simulation. (D) A plot of the average RMSF of each amino acid residue
652 in six copies of the NUP153C residues 1407-1422 bound to the hexamer CA during a 100 ns
653 MD simulation (n= 6, error bars are standard deviations). (E) $\Delta\Delta G_{bind}$ calculated by the
654 MD+FoldX approach for all possible amino acid substitutions at each position in NUP153C
655 (residues 1407-1422). (F) A comparison of the $\Delta\Delta G_{bind}$ values calculated by MD+FoldX and
656 PyRosetta for 16 NUP153C mutations. Error bars are standard deviations. (G) Correlation
657 plot of $\Delta\Delta G_{bind}$ estimated by MD+FoldX and PyRosetta, where the trendline shows the linear
658 relationship between predicted $\Delta\Delta G_{bind}$ values from the two different methods.
659 Corresponding *R* and *p* values are displayed.
660

661

	Pearson correlation coefficient	
Residue	**Hydrophobicity**	**Volume**
P1411	-0.368	0.4068
S1412	-0.4023	-0.4348
G1413	-0.4193	0.4738*
V1414	-0.2894	0.6345**
F1415	-0.3076	-0.7852**
T1416	-0.6186**	-0.4828*
F1417	-0.4132	-0.7489**
G1418	-0.2615	0.7582**

662

663 **Table 1. The volume of amino acid sidechains in NUP153C is more important for CA**
664 **interaction than hydrophobicity.** Pearson correlation coefficient values comparing the
665 $\Delta\Delta G_{bind}$ MD+FoldX against the hydrophobicity and volume of amino acid sidechains. * p
666 <0.05, ** p <0.01.

667

668

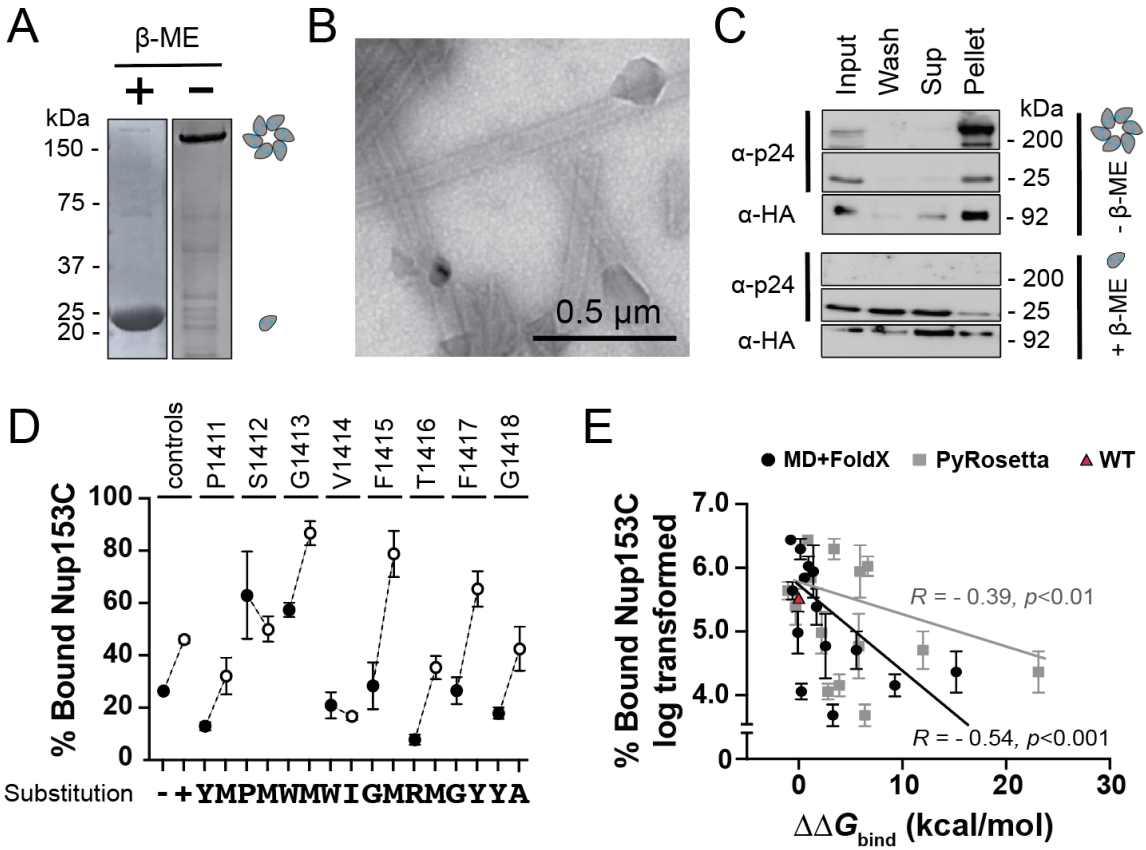

Figure 2. Molecular modeling predicts the effects of NUP153C mutations on CA interaction as measured by co-sedimentation. (A) SDS-PAGE of the monomeric and the cross-linked hexameric CA. (B) Transmission electron micrograph of CA tubes assembled from cross-linked CA hexamers. (C) A Western blot of a co-sedimentation assay using CA monomers (with the reducing agent β-mercaptoethanol (β-ME)) and CA tubes (without β-ME). Input: 10% of the total reaction volume. NUP153C was detected with an anti-HA antibody. (D) NUP153C mutants co-sedimented with CA tubes. Western blot signals were normalized to input. Error bars are standard deviations. Control reactions included the wild-type NUP153C (+) and NUP153C with a deletion of the CA interaction motif Δ1411-1418 (-). White data points represent mutations with low $\Delta\Delta G_{bind}$ values, and black circles represent mutations with high $\Delta\Delta G_{bind}$ values for each site. (E) Evaluation of the $\Delta\Delta G_{bind}$ calculated by either MD+FoldX (black) or PyRosetta (gray) versus the experimental dataset. The Pearson's correlation was calculated for each modeling method against three independent replicates of experimental data.

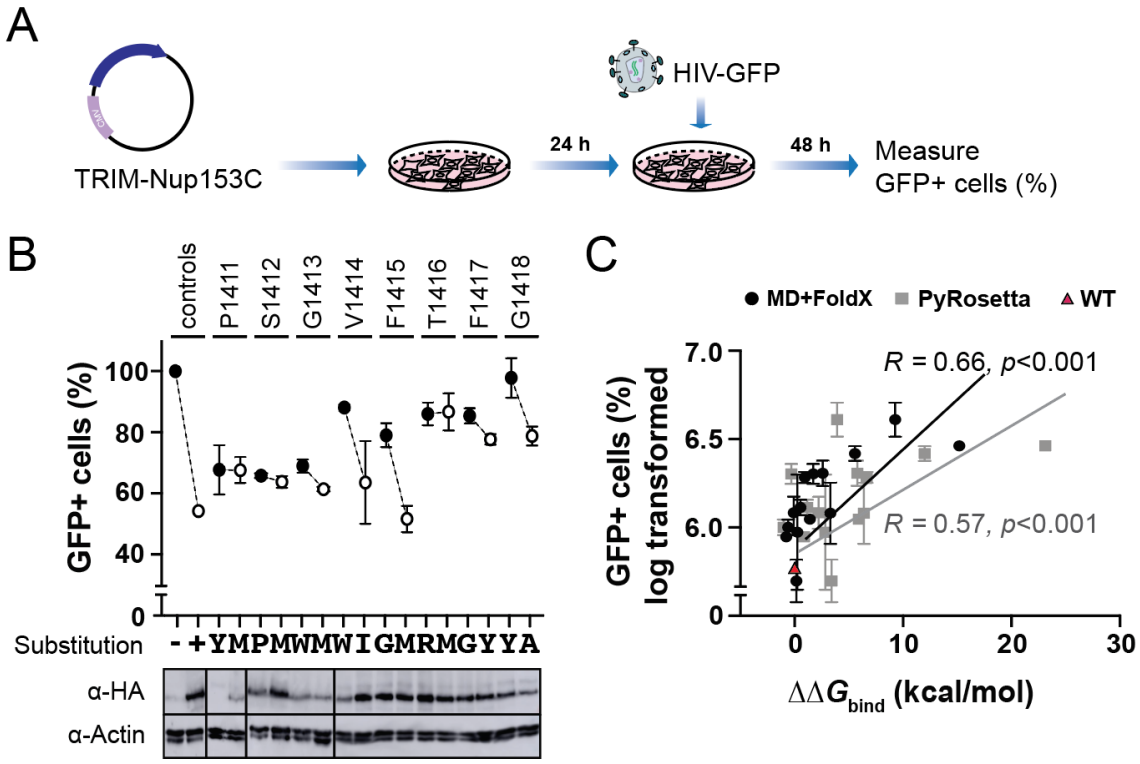

Figure 3. Molecular modeling predicts the effects of mutations in NUP153C on the CA interaction as measured by TRIM-NUP153C restriction. (A) A schematic representation of the workflow of a TRIM-NUP153C restriction assay. (B) *Top.* NUP153C interaction with capsid was measured by the degree of HIV-1 restriction by transient expression of TRIM-NUP153C in HEK293T cells. The relative percentage of GFP-positive cells, indicating HIV-1 transduction, was measured by flow cytometry. White data points represent mutations that had low $\Delta\Delta G_{bind}$ values, and black circles represent mutations that had high $\Delta\Delta G_{bind}$ values for each site. *Bottom.* The expression of each mutant TRIM-NUP153C was examined by Western blot. (C) The correlation of the empirical data presented in (B) with the $\Delta\Delta G_{bind}$ as calculated by either MD+FoldX or PyRosetta.