Supporting Material

Title: Early Stages of the HIV-1 Capsid Protein Lattice Formation

Author Information: John M. A. Grime,[†] and Gregory A. Voth,[†‡]*

Author Affiliation:

[1]Computation Institute
Argonne National Laboratory
Lemont, IL 60439

[2]Department of Chemistry, James Franck Institute, and Computation Institute
The University of Chicago
5735 S. Ellis Ave
Chicago, Illinois 60637

RMSD comparison for C_α atoms in α-helices of experimental CA monomer structures. As the CG CA monomer structure is generated using a "template" all-atom structure from experimental data (see Methods), it is necessary to determine that such a template monomer is representative of the HIV-1 CA protein in both pentameric and hexameric arrangements.

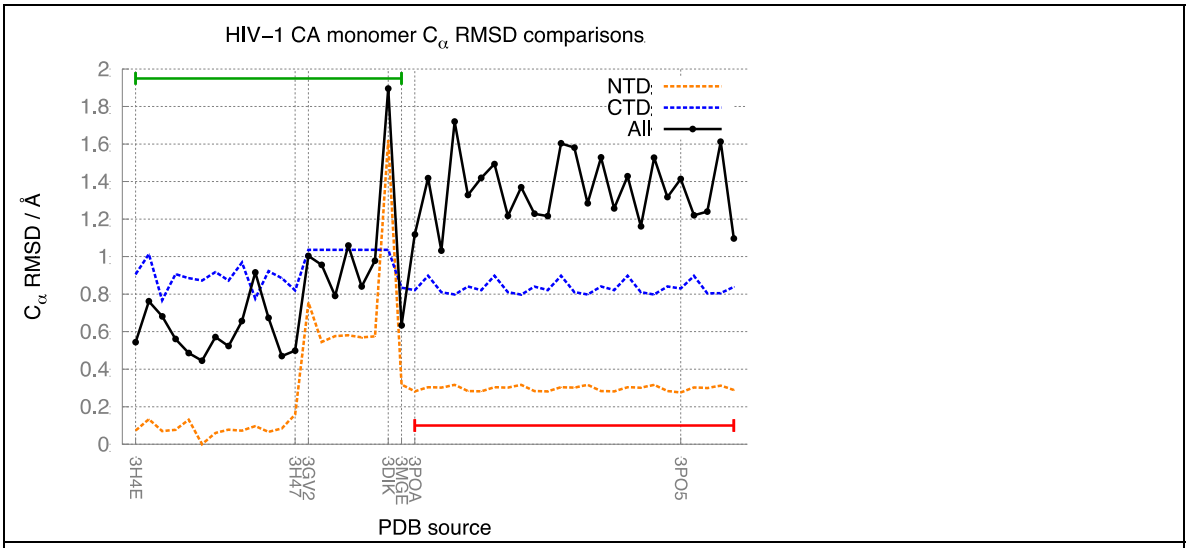

Fig. S1

Fig. S1 presents the RMSD values calculated across C_α atoms of the α-helices in the template CA monomer and a representative set of experimental CA structures. RMSD values of the NTD α-helices alone shown depicted as an orange dashed line, RMSD values of the CTD α-helices alone as a blue dashed line. The RMSDs of the total structures are a solid black line. CA monomers from experimental hexamers are spanned using a horizontal green line, with monomers from experimental pentamers spanned using a horizontal red line. The peak in RMSD values for the 3DIK structure reflects the comparatively low resolution of that data (9Å), but in all cases the basic structure of the CA monomer is conserved across both hexameric and pentameric assemblies.

Maximum number of trimer-of-dimers, pentamers and hexamers detected in quasi-2D planar simulations when NTD-NTD and NTD-CTD interaction strengths, ε_{NN} and ε_{NN}, are independently varied. To investigate the effects of the NTD-NTD and NTD-CTD interaction strengths (ε_{NN} and ε_{NC}) on the formation of mature-style p6 lattice structures using the CG CA dimer model, a series of CG MD simulations were performed with 56 CG CA dimers weakly constrained to a quasi-2D planar surface. Both flexible and rigid dimer interfaces were considered, with the surface coverage ρ varied from 0.38 to 0.6 (see Methods in main manuscript) and ε_{NN} and ε_{NN} varied independently for each ρ.

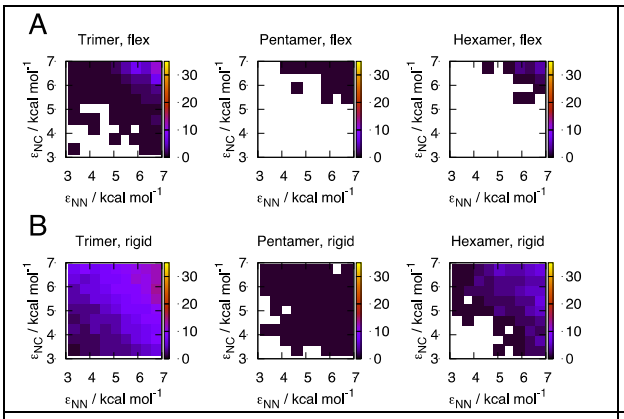

Fig S2

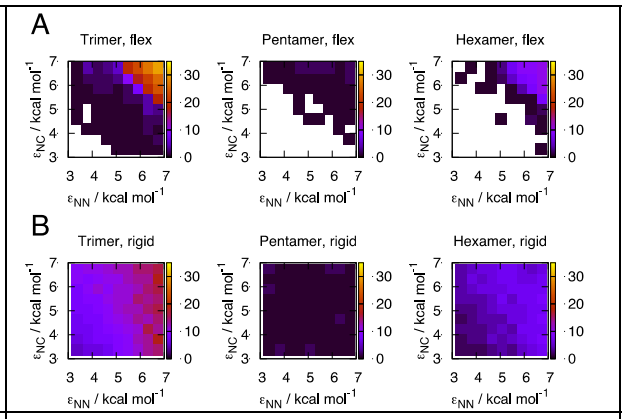

Fig S3

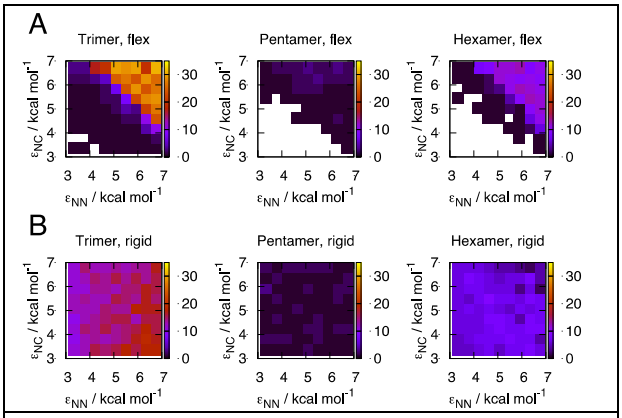

Fig S4

Figs. S2–S4 (above) illustrate the maximum number of trimer-of-dimers, pentamer and hexamer structures detected at any point during the quasi-2D planar CG MD simulations. Three selected surface coverage values are shown: ρ = 0.38 (**Fig. S2**), ρ = 0.5 (**Fig. S3**), and ρ = 0.6 (**Fig. S4**). **Figs. S2 *A*, S3 *A*,** and **S4 *A*** show data for flexible dimer interfaces, and **Figs. S2 *B*, S3 *B*** and **S4 *B*** show data for rigid dimer interfaces. The color bars indicate the maximum number of structures detected. Note that for flexible dimer interfaces at low ρ (e.g. **Fig. S2 *A***) pentamers and hexamers only appear for relatively large NTD-NTD and NTD-CTD interaction strengths, whereas these structures appear for smaller NTD-NTD and NTD-CTD interaction strengths in the presence of a rigid dimer interface (**Fig. S2 *B***). Trimer-of-dimers structures, however, form over a wide range of NTD-NTD and NTD-CTD interaction strengths for ρ = 0.38.

Total number of unique trimer-of-dimers, pentamer and hexamer structures detected in CG MD simulations on a quasi-2D spherical surface for ε/k_BT = 6.71. A "unique" structure (trimer-of-dimers, pentamer or hexamer) is determined in the CG MD simulations by the internal CG MD monomer indices of the CG CA proteins from which the structure is composed. Structures are therefore distinct when their component CA monomer indices differ, and hence the total number of unique trimer-of-dimers, pentamer and hexamer structures formed can be recorded in the simulation. This approach allows the initial formation environment of each structure to be studied while removing noise in the data from oscillatory formation/breakup/re-formation patterns in individual structures. An example of such an

oscillatory pattern might be where a hexamer forms and thermal fluctuations proceed to nudge a single NTD from the central hexamer ring slightly outside the cutoff used for structure identification (see Methods). The hexamer structure is still largely intact at this point, and subsequent thermal motions can easily move the NTD back into the cutoff radius of the structural identification routine. Such an occurrence should not be recorded as an initial formation event, and can be excluded from the data by cross-referencing the CG monomer indices in the hexamer against previously detected structures.

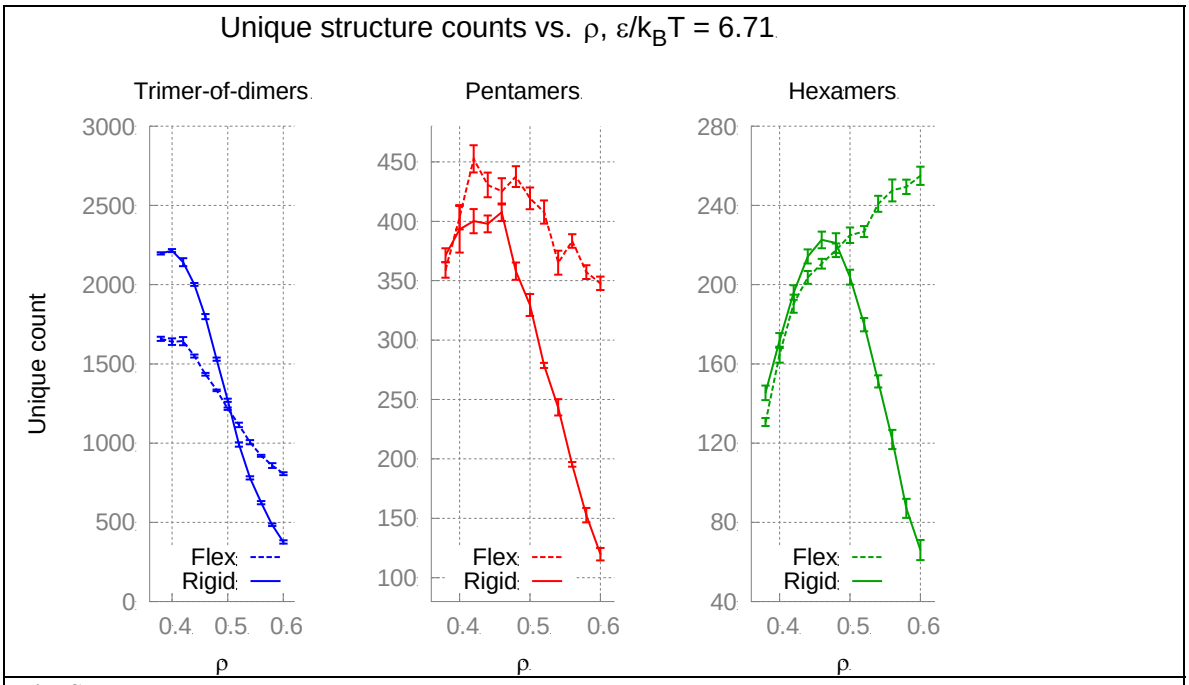

Fig. S5

Fig. S5 displays the mean total number of unique structures detected in 5 repeat CG MD simulations for each surface coverage value ρ, with error bars indicating the standard error of the mean estimated from the repeat simulations. Note that the number of *unique* trimer-of-dimers and pentamer structures detected at lower values of ρ are significantly higher than the average number of those structures present in the simulations as a function of time (see **Fig. 4** in the main text), indicating that relatively large numbers of unstable trimer-of-dimers and pentamer structures form and dissociate during the CG MD simulations.

Average structure counts as a function of simulation time and total number of unique structures detected in CG MD simulations on a quasi-2D planar surface for ε/k_BT = 6.71. To further examine the influence of surface curvature on the phenomena described in the main text, 10 repeat CG MD simulations of were run with NTD-NTD and NTD-CTD interaction strengths and surface coverage ρ identical to the spherical simulations in **Fig. 4** of the main text and **Fig. S5**, but using 56 CG CA dimers constrained to a quasi-2D planar surface instead of a quasi-2D spherical surface. The results are presented in **Fig. S6**.

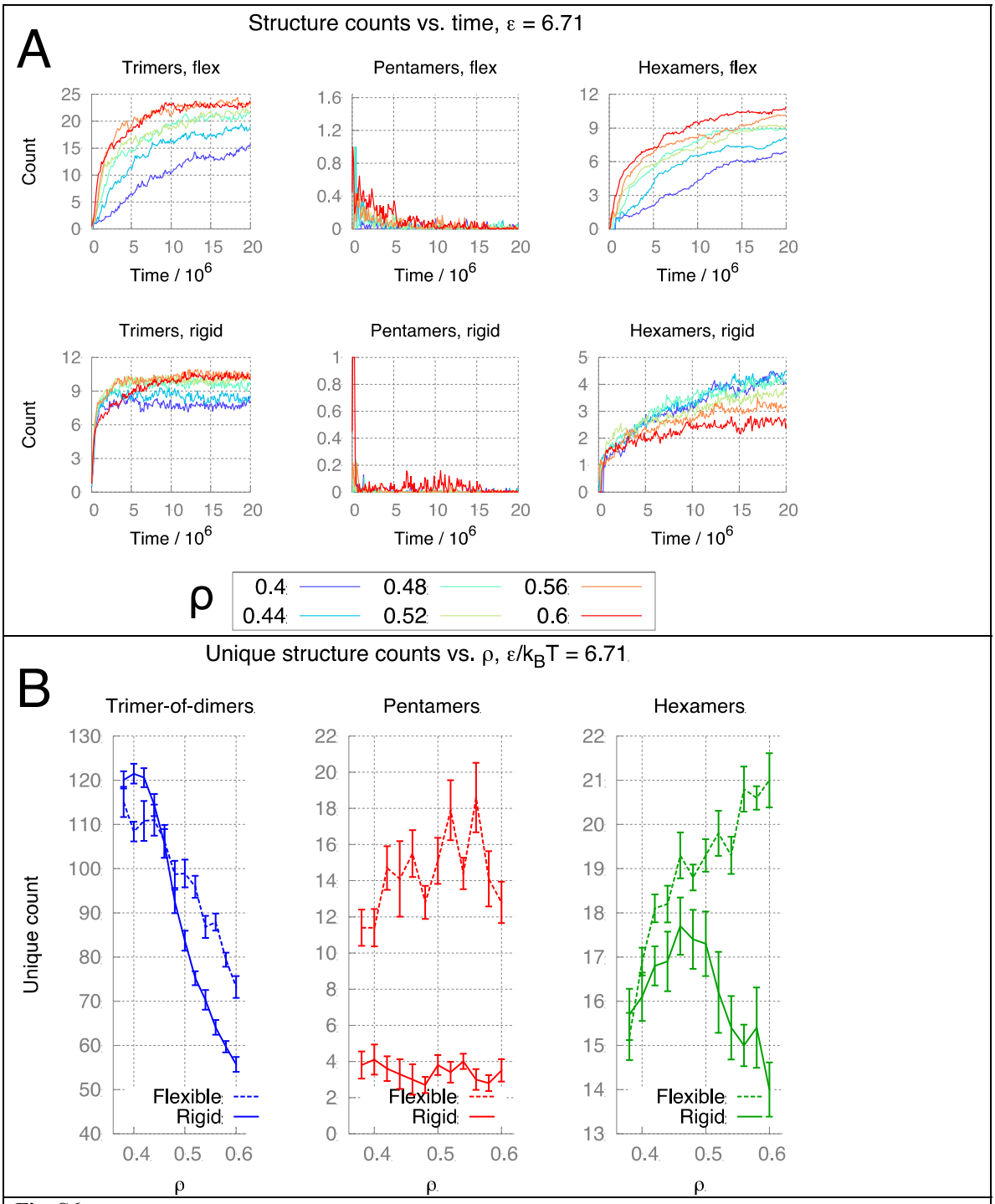

Fig. S6

The qualitative behaviors of the average number of trimer-of-dimers and hexamers present in the CG MD simulations as a function of time (**Fig. S6** ***A***) are similar to those observed for the quasi-2D spherical surfaces (see **Fig. 4** in the main text), indicating that the formation of these structures is relatively insensitive to local curvature. The average number of pentamers in the planar CG MD simulations relative to the number of trimer-of-dimers and hexamer structures (as well as the number of unique pentamers identified, **Fig. S6** ***B***) are significantly reduced,

however, indicating that *pentamer formation is suppressed in the absence of local curvature*. The average pentamer and hexamer counts in **Fig. S6** ***A*** are plotted on the same relative scale as in **Fig. 4** to aid comparison to the relative average pentamer and hexamer counts in **Fig. 4. E**rror bars shown in **Fig. S6** ***B*** indicate standard error of the mean estimated from the 10 CG MD simulations.

Full formation probabilities *P* for trimer-of-dimers, pentamer and hexamer structures in the quasi-2D spherical surface CG MD simulations.

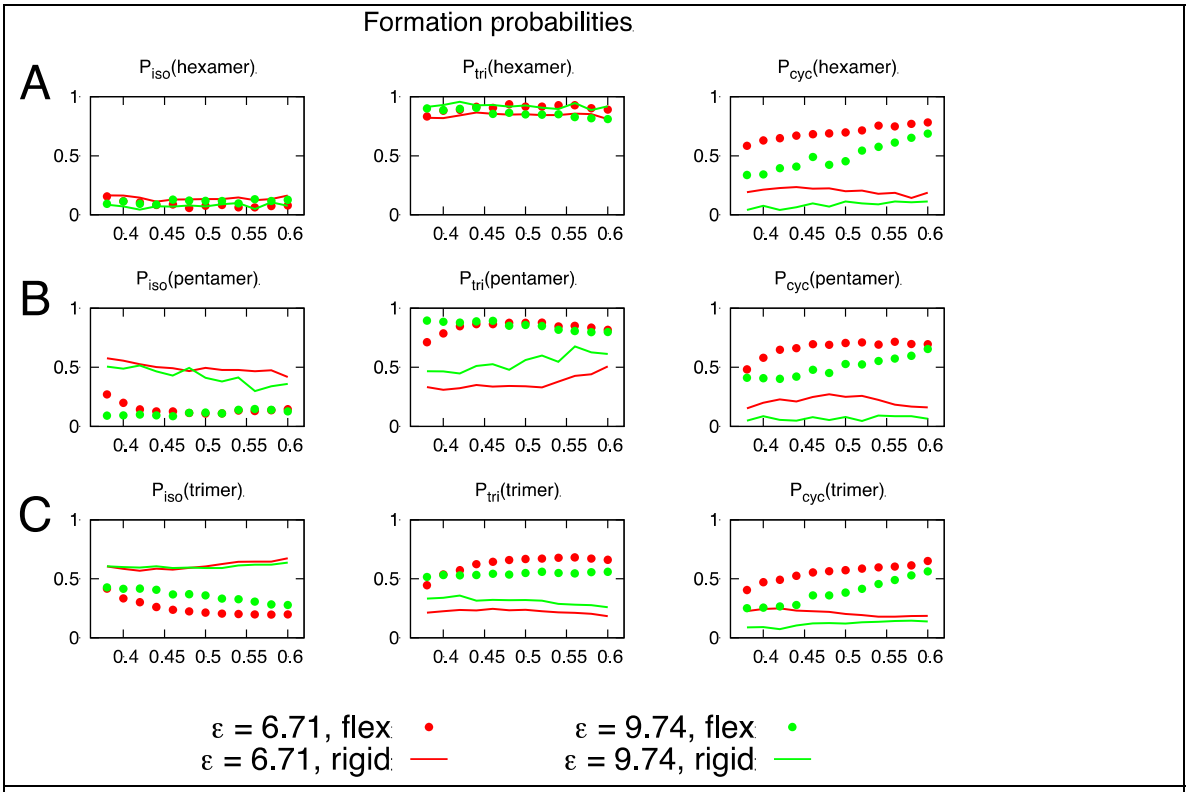

Fig. S7

Fig. S7 presents the full set of formation probabilities P calculated for each of the structures of interest in the main text (see **Fig. 6** of the main text). The formation probabilities are calculated only from the first appearance of each unique structure (see **Fig. S5** for a discussion of determining "unique" structures), and hence rapid oscillatory formation and breakup of particular structures is deliberately excluded from these data.

Proportion of CG CA dimers aggregated into a protein "raft" via native-style NTD-NTD or NTD-CTD contacts as a function of time for the quasi-2D spherical surface CG MD simulations. We define a "raft" as a set of CG CA dimers aggregated via native-style NTD-NTD and NTD-CTD contacts, containing at least one of the structural motifs considered in the manuscript (i.e. a triangle-of-dimers, pentamer or hexamer). Note that these aggregates can be largely disordered, and not conform to contiguous mature CA p6 lattice, but the dimers in the raft are nonetheless linked by native-style NTD-NTD and NTD-CTD contacts.

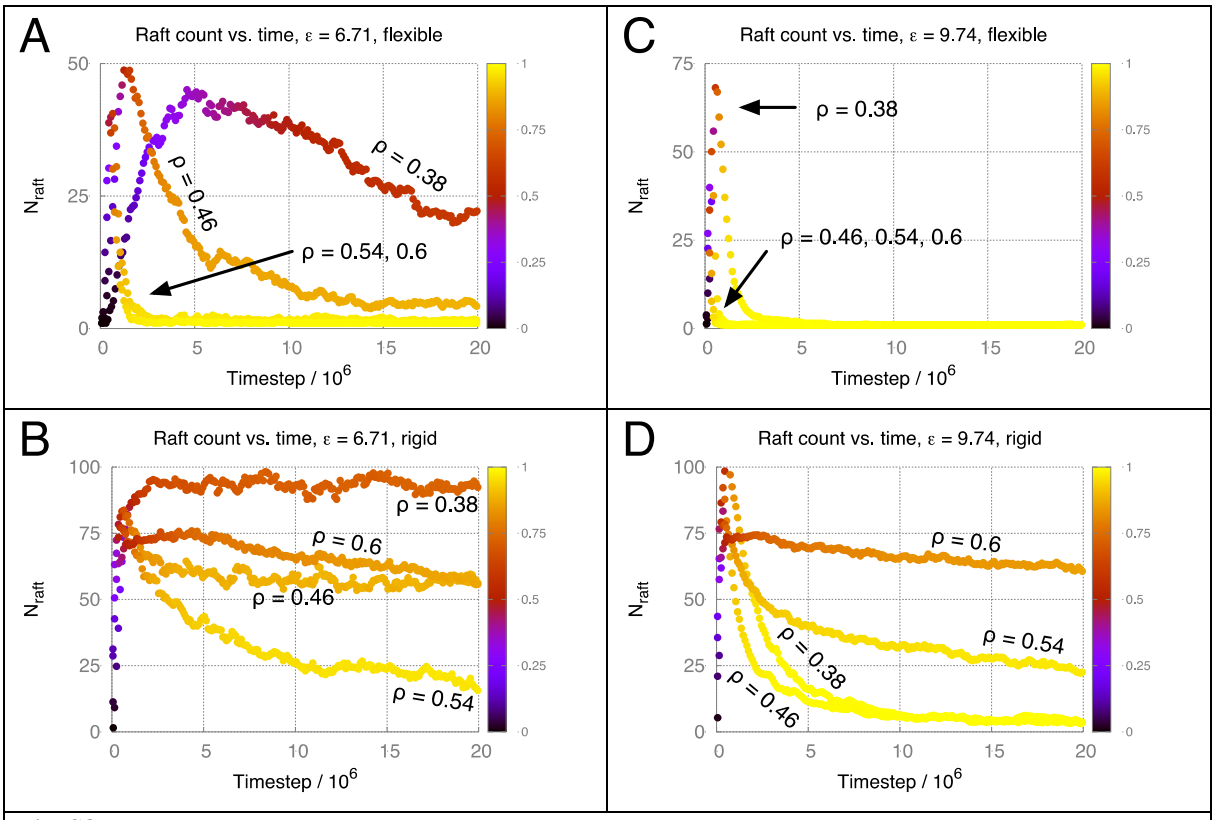

Fig. S8

Fig S8 *A–D* above show the number of rafts, N_{raft}, as a function of simulation time, with the color indicating the proportion of the total system dimers associated with a raft (*black = 0.0, bright yellow = 1.0*). **Fig. S8 *A*** and ***B*** contain the results of selected ρ using ε/k_BT = 6.71 for flexible and rigid dimer interfaces respectively, while **Fig. S8 *C*** and ***D*** contain the results for ε/k_BT =9.74. Note that for the completely rigid dimer interfaces, as the surface protein density increases beyond ρ = ~0.52, the N_{raft} remains high as a function of simulation time and the proportion of protein associated with a raft (i.e. making native-style NTD-NTD and NTD-CTD contacts) decreases. This indicates jamming of the system, as the dimers are unable to reorganize into larger mature p6 lattice regions in the restricted space and surface curvature available, instead persisting as a large number of small, isolated rafts.

Total number of unique trimer-of-dimers, pentamer and hexamer structures detected in CG MD simulations on a quasi-2D spherical surface for ε/k_BT = 9.74.

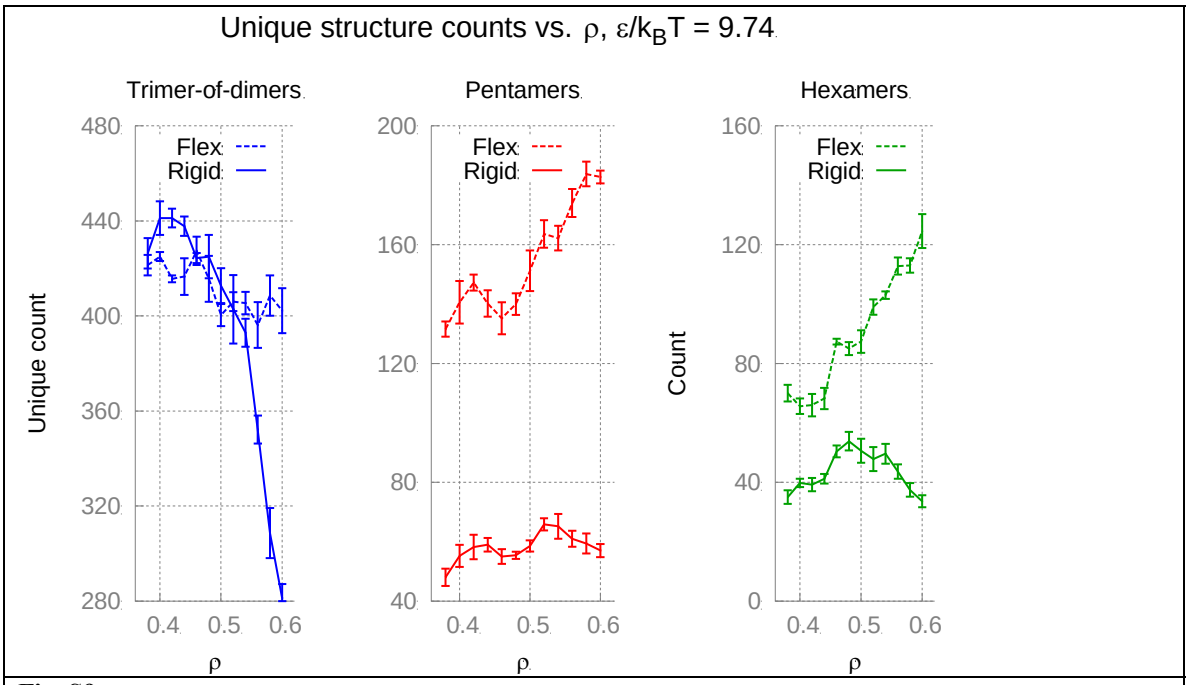

Fig. S9

Same explanation as for **Fig. S5**, with $\varepsilon/k_BT = 9.74$.

Snapshots from final configurations of CG MD simulations using a quasi-2D spherical surface for $\varepsilon/k_BT = 9.74$.

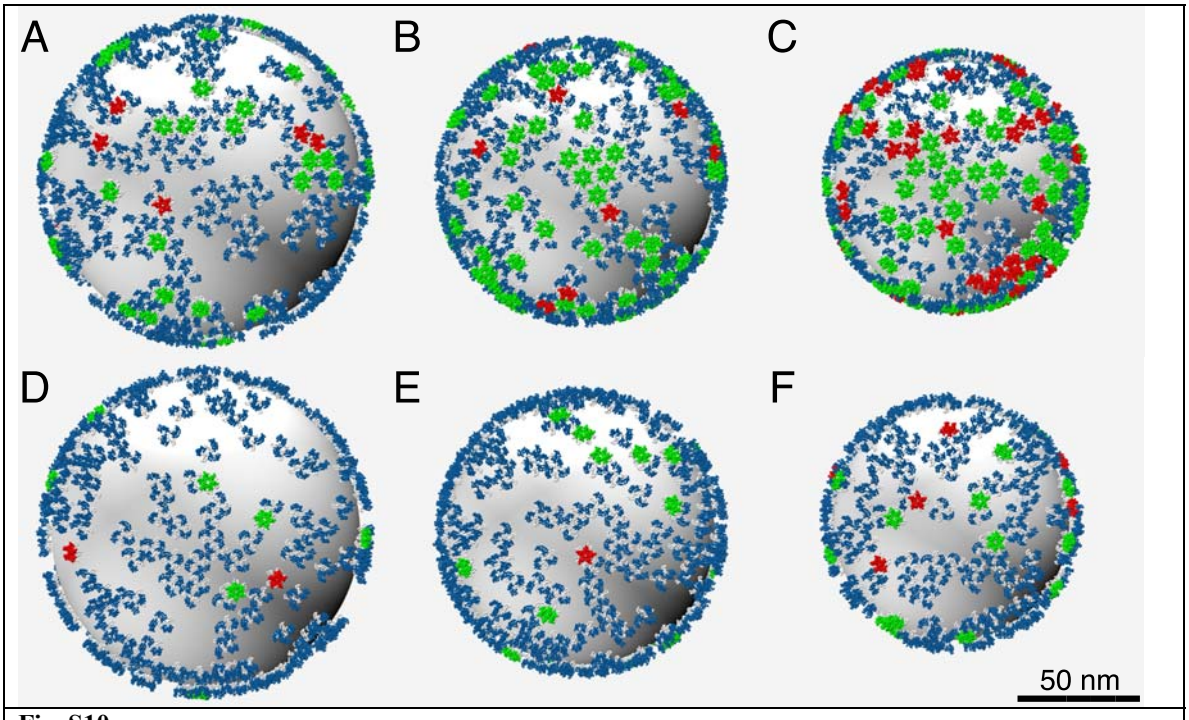

Fig. S10

Same as **Fig. 5** of the main text, with $\varepsilon/k_BT = 9.74$.

Supporting References

(1) Pornillos, O., B. K. Ganser-Pornillos, B. N. Kelly, Y. Hua, F. G. Whitby, C. D. Stout, W. I. Sundquist, C. P. Hill, and M. Yeager. 2009. X-ray structures of the hexameric building block of the HIV capsid. *Cell.* 137:1282-1292.

(2) Ganser-Pornillos, B. K., A. Cheng, and M. Yeager. 2007. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell.* 131:70-79.

(3) Pornillos, O., B. K. Ganser-Pornillos, S. Banumathi, Y. Hua, and M. Yeager. 2010. Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus. *J. Mol. Biol.* 401:985-995.

(4) Pornillos, O., B. K. Ganser-Pornillos, and M. Yeager. 2011. Atomic-level modelling of the HIV capsid. *Nature.* 469:424-427.

Supporting Figure Legends

Fig S2. Maximum structure counts for **ρ = 0.38**. **Top row, *A*** flexible dimer model. **Bottom row, *B*** rigid dimer model. White regions denote no structures recorded.

Fig S3. Maximum structure counts for **ρ = 0.5**. **Top row, *A*** flexible dimer model. **Bottom row, *B*** rigid dimer model. White regions denote no structures recorded.

Fig S4. Maximum structure counts for **ρ = 0.6**. **Top row, *A*** flexible dimer model. **Bottom row, *B*** rigid dimer model. White regions denote no structures recorded.

Fig. S5. Mean total number of unique structures detected over the course of 5 repeat CG MD simulations using a quasi-2D spherical surface for $\varepsilon/k_BT = 6.71$. Results for flexible CG CA dimer interfaces are depicted as dashed lines, rigid dimer interfaces as solid lines. Error bars indicate standard error of the mean, estimated over the 5 repeat simulations.

Fig. S6. Average number of structures present as a function of simulation time (***A***) and total number of unique structures detected (***B***) for 10 repeat CG MD simulations using a quasi-2D planar surface with $\varepsilon/k_BT = 6.71$. Error bars in ***B*** indicate the standard error of the mean, estimated over the 10 repeat simulations, with results for flexible and rigid CG CA dimer interfaces depicted as dashed and solid lines respectively.

Fig. S7. Full formation probabilities for all three structural motifs discussed in the manuscript. **Top row, *A***: hexamer formation probabilities. **Central row, *B***: pentamer formation probabilities. **Bottom row, *C***: trimer-of-dimers formation probabilities.

Fig. S8. Proportion of total system protein connected together via native-style NTD-NTD and NTD-CTD contacts into "rafts" as a function of simulation time. Color denotes the proportion of the total system content that is raft-associated. **Left column, *A* and *B***: $\varepsilon/k_BT = 6.71$, **right column, *C* and *D***: $\varepsilon/k_BT = 9.74$. **Top row, *A* and *C***: flexible dimer interface, **bottom row, *B* and *D***: completely rigid dimer interface using the 2KOD dimer interface structure.

Fig. S9. Mean total number of unique structures detected in 5 CG MD simulations using a quasi-2D spherical surface and $\varepsilon/k_BT = 9.74$. Results for flexible CG CA dimer interfaces are depicted as dashed lines, rigid dimer interfaces as solid lines. Error bars indicate the standard error of the mean estimated over the 5 repeat simulations.

Fig. S10. Selected snapshots of final configurations from spherically constrained simulations for $\varepsilon/k_BT = 9.74$. **Top row:** flexible dimer model, $\rho = 0.38$ (***A***), 0.54 (***B***) and 0.6 (***C***). **Bottom row:** rigid dimer model, $\rho = 0.38$ (***D***), 0.54 (***E***) and 0.6 (***F***). Color scheme as in **Fig. 5** from main manuscript.