Supporting Information

Off-Pathway Assembly: A Broad-Spectrum Mechanism of Action for Drugs that Undermine Controlled HIV-1 Viral Capsid Formation

Alexander J. Pak, John M. A. Grime, Alvin Yu, and Gregory A. Voth*
Department of Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, IL, USA

*Corresponding author: gavoth@uchicago.edu

Additional Details on the Coarse-Grained (CG) Model for Capsid (CA)

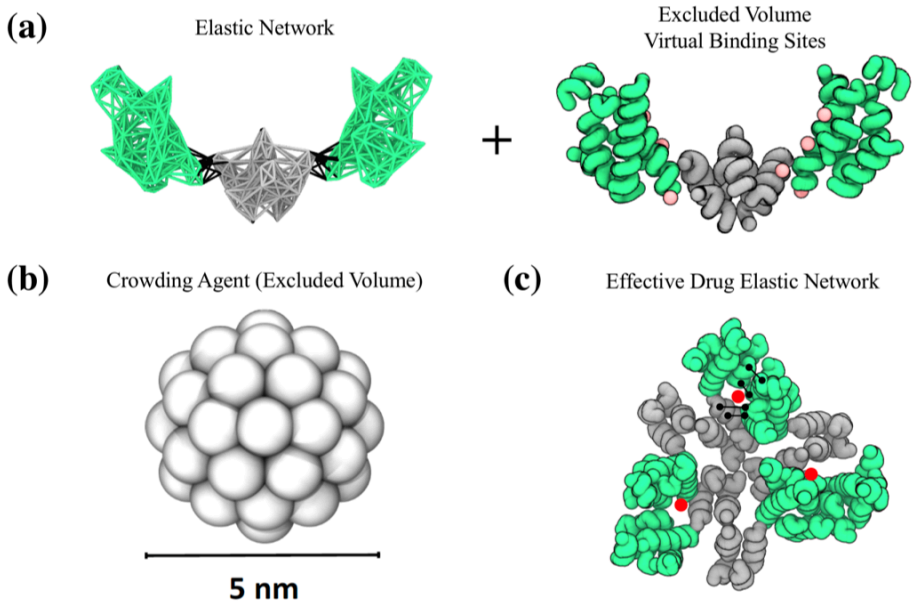

Figure S1. Schematic of the CG models used in this work. (a) The CG model for the CA dimer is composed of two elements: bonded potentials (an elastic network model or ENM) that maintain the secondary and tertiary structure of the CA dimer and nonbonded potentials (both excluded volume and attractive binding potentials). The pink beads represent "virtual sites" that may bind to specific CG sites on other CA dimers; in the "active" ("inactive") state, these virtual sites are described by an attractive interaction (no interaction). These virtual sites recapitulate the anisotropic nature of the binding interactions at the protein-protein interface. The green and grey regions indicate the NTD and CTD domains, respectively. (b) Depiction of the macromolecular crowding agent that is described by excluded volume interactions. (c) Representation of a trimer of dimers (TOD) structure with an additional ENM (black lines) that connect the NTD and CTD domains of one CA dimer to that of an adjacent CA dimer. The red dots indicate the location of the PF74 and GS-CA1 binding pocket. Additional details for (a) and (b) can be found in Ref. 1.

Structural Details of the Coarse-Grained Model in Comparison to All-Atom Simulations

All-atom simulations of a TOD with and without PF74 were prepared using PDB 5L93[2] (PDB 4QNB[3] was used to identify the PF74 binding pocket and orientation). The protein was solvated by water and 0.15 M NaCl in a 17x17x17 nm cubic box with periodic boundary conditions. The protein/PF74 and water force-fields were described using CHARMM36m and TIP3P, respectively, using CGenFF.[4-5] Three of the PF74 dihedrals resulted in large penalty scores (>50) due to mistaken classification during the automated CGenFF procedure; these were manually replaced with analogues from carbohydrates and ribonucleic acids.[6-7] Note that additional parameter optimization was not performed. In total, 365k and 431k atoms were simulated for the apo and holo systems, respectively.

All simulations were performed using GROMACS 5.1.2.[8] After energy minimization, systems were equilibrated for 100 ns in the constant NPT ensemble using a Parrinello-Rahman barostat[9] at 1 atm with a 10 ps damping time, a Nosé-Hoover chain thermostat[10] at 300 K with a 1 ps damping time, and a 2 fs integration timestep. Production runs were performed in the constant NVT ensemble using the same thermostat for 500 ns. Hydrogen bonds were constrained using the LINCS algorithm.[11] Statistics were gathered every 10 ps. For analysis, the entire protein complex was rigidly re-aligned within each frame to remove global translational and rotational fluctuations. The radius of curvature (r_{curv}) of the TOD was calculated using the following expression:

$$r_{curv} = r_{N-C}\frac{r_{N-N}}{r_{N-N}-r_{C-C}} \quad (1)$$

where r_{N-C} is the intra-monomer distance between the center of mass (CoM) of the NTD and CTD domain, r_{N-N} is the inter-monomer distance between the CoMs of the NTDs of adjacent monomers, and r_{C-C} is the inter-monomer distance between the CoMs of the CTDs of adjacent monomers.

In Figure S2, we compare the root-mean-squared deviation (RMSD) of the Cα backbone (using PDB 5L93 as the reference) and radius of curvature of the holo and apo cases at both all-atom and CG resolutions. We find that the mean of the RMSD is not affected by the presence of PF74, which is consistent with previously performed all-atom MD simulations on CA pentamer-of-hexamers[12] and the absence of discernible differences in the solved CA structure of PF74-bound hexamers.[3] A subtle change in the RMSD distribution is observed; the apo case adopts a wider distribution, which indicates that some structural variability from the atomic model is present. More interestingly, we note that while the mean curvature is preserved in the all-atom apo and holo case, the distribution is narrowed in the latter. This difference suggests that PF74 induces a rigidification that resists variable curvature. Furthermore, comparison of the distributions from the all-atom and CG simulations indicates that the CG model qualitatively captures these structural effects.

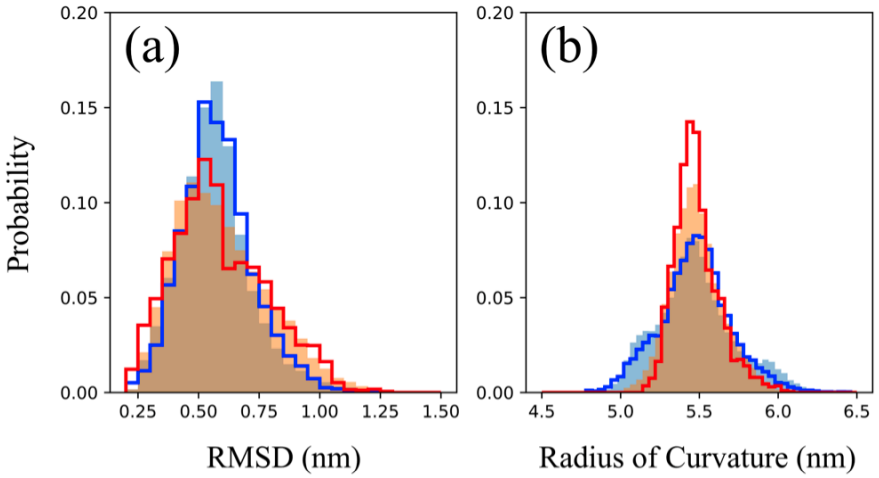

Figure S2. Sampled distributions of the (a) root-mean-squared deviation (RMSD) of the Cα backbone and (b) radius of curvature of the TOD (see above for details). In both panels, the shaded blue (orange) region indicates all-atom data for the drug-free (drug-bound) case, while the solid blue line (red line) indicates the CG data for the drug-free (drug-bound) case.

Additional Details on Analysis Metrics

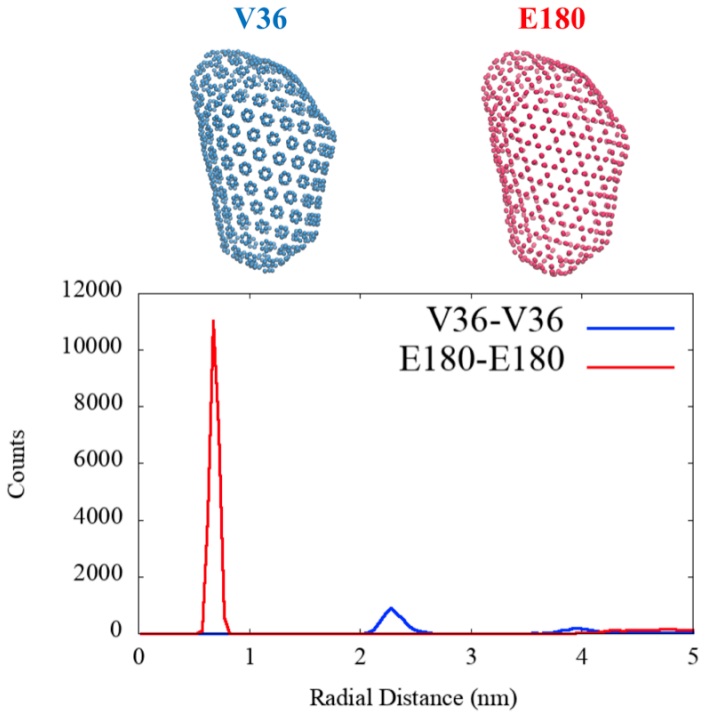

Figure S3. (Top) Schematic of a pre-assembled CG representation of a mature core (mapped to PDB 3J3Y from Ref. 13) that displays residue V36 (blue beads) and residue E180 (red beads). The proximity between these beads represents an intra-hexamer/pentamer contact and a cross-

dimer/hexamer contact, respectively. (Bottom) Non-normalized radial distribution functions between V36 (blue line) and E180 (red line) beads show a distinct peak with large separation from other peaks. The distance criteria used to determine graph connectivity was derived using this data: for simplicity, a pairwise distance below 2.75 nm was considered a successful contact.

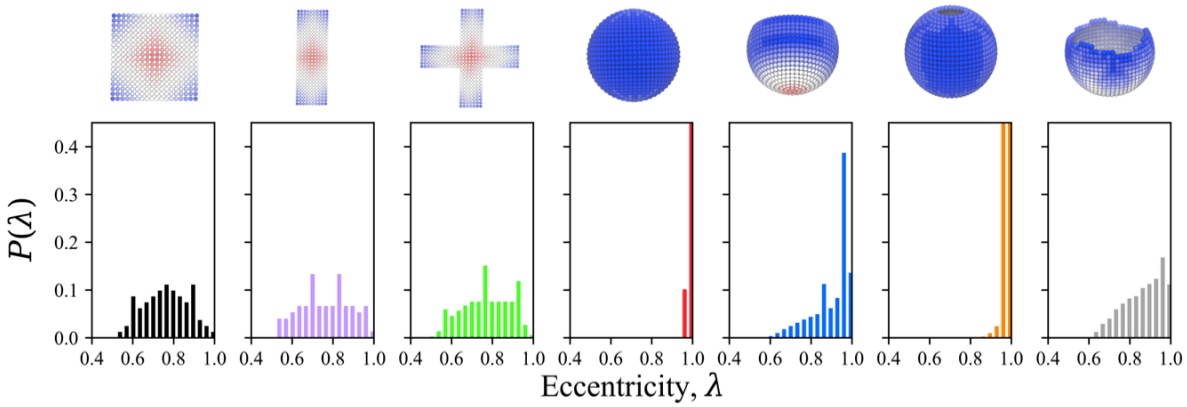

Figure S4. Eccentricity distributions for select lattice morphologies with uniformly spaced nodes. The top panels depict each lattice with the color (from red = 0.50 to blue = 1.00) indicating the eccentricity of the node. Panels 4-7 are spherical lattices that are fully enclosed, perfectly cut along a longitudinal line close to the equator, perfectly cut along a longitudinal line close to the pole, and unevenly cut around a longitudinal line close to the equator.

Table S1. Statistical metrics for distributions depicted in Figure S4 (from left to right)

Panel	1	2	3	4	5	6	7
Std. Dev.	0.11	0.12	0.12	0.01	0.09	0.02	0.10
Skew	0.00	0.00	-0.14	-2.65	-1.17	-1.06	-0.52
Kurtosis	-0.61	-1.00	-0.90	4.98	0.32	2.85	-0.73

We present additional analysis on graph eccentricity and its interpretation through Figure S4 and Table S1. Information embedded in an eccentricity distribution can be used to describe various aspects of a graph. For instance, the minimum (maximum) eccentricity is known as the graph radius (diameter) while the node with the minimum eccentricity is known as the graph center.[14] The interpretations for moments of the distribution – such as the standard deviation, skew, and kurtosis – are less clear but can be rationalized through an empirical exploration of various graph structures. Seven such lattices are presented in Figure S4 (with associated moments in Table S1), which include both flat and curved lattices with variable classes of edges.

Although the lattice states described by the moments of the eccentricity distribution may contain some degeneracy, some general rules of thumb can be suggested. The standard deviation, as an indicator for the breadth of the distribution, may partially describe the uniformity amongst node eccentricities. In a fully-enclosed curved lattice, for example, the maximal paths from any given node are equivalent throughout the lattice. As a result, the distribution is nearly a Dirac delta

function, which has a small standard deviation and large positive kurtosis. On the other hand, a lattice with edges, such as the 2D sheets and cleaved spheres, has greater variability in maximal paths such that the standard deviation is large by comparison. Hence, one may use the standard deviation to delineate lattices with edges from those without.

The skew indicates the asymmetry of the distribution around the mean; a positive (negative) skew indicates that the peak of the distribution is to the left (right) of the mean. In the context of graph eccentricities, the skew is suggestive of the population of nodes that are central (i.e., close to the minimum eccentricity or graph radius) compared to that of nodes that are on the edge (i.e., close to the maximum eccentricity or graph diameter). Hence, a negative skew is suggestive of lattices with a large area of edges compared to that of the center of the lattice.

The kurtosis indicates the "weight" of the tail of the distribution. A positive (negative) kurtosis indicates that the tail of the distribution is short (long). If nodes close to the center or edge of the lattice are uniform in eccentricities, one expects the tail distribution to be uniform such that the kurtosis is positive. On the other hand, if nodes close to the center or edge of the lattice contain a broad distribution of eccentricities, the kurtosis is negative. Hence, lattices with edge nodes that have variable distance to the center of the lattice, such as in the case of the cleaved sphere with uneven edges, can be classified by their negative kurtosis.

Returning to the analysis presented in the main text, we ultimately utilize negative skew and negative kurtosis as qualitative indicators for lattices with anisotropic edges, which contain a high fraction of edges with non-uniform distances to the center (e.g., due to dendritic growth).

Supplemental References

1. Grime, J. M. A.; Dama, J. F.; Ganser-Pornillos, B. K.; Woodward, C. L.; Jensen, G. J.; Yeager, M.; Voth, G. A., Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat Commun* **2016,** *7*.
2. Schur, F. K. M.; Obr, M.; Hagen, W. J. H.; Wan, W.; Jakobi, A. J.; Kirkpatrick, J. M.; Sachse, C.; Kräusslich, H.-g.; Briggs, J. A. G., An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **2016,** *353*, 506-508.
3. Bhattacharya, A.; Alam, S. L.; Fricke, T.; Zadrozny, K.; Sedzicki, J.; Taylor, A. B.; Demeler, B.; Pornillos, O.; Ganser-Pornillos, B. K.; Diaz-Griffero, F.; Ivanov, D. N.; Yeager, M., Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc Natl Acad Sci USA* **2014,** *111*, 18625-18630.
4. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell Jr, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010,** *31*, 671-690.
5. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2016,** *14*, 71.
6. Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D., CHARMM Additive All-Atom Force

Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. *J Chem Theory Comput* **2011,** *7*, 3162-3180.
7. Xu, Y.; Vanommeslaeghe, K.; Aleksandrov, A.; MacKerell Jr, A. D.; Nilsson, L., Additive CHARMM force field for naturally occurring modified ribonucleotides. *J. Comput. Chem.* **2016,** *37*, 896-912.
8. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.
9. Parrinello, M.; Rahman, A., Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980,** *45*, 1196-1199.
10. Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992,** *97*, 2635-2643.
11. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18*, 1463-1472.
12. Perilla, J. R.; Hadden, J. A.; Goh, B. C.; Mayne, C. G.; Schulten, K., All-Atom Molecular Dynamics of Virus Capsids as Drug Targets. *J. Chem. Phys. Lett.* **2016,** *7*, 1836-1844.
13. Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P., Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013,** *497*, 643-646.
14. Takes, W. F.; Kosters, A. W., Computing the Eccentricity Distribution of Large Graphs. *Algorithms* **2013,** *6*.