**Supplemental Information**

**Binding Mechanism of the Matrix Domain of HIV-1 Gag on Lipid Membranes**

V. Monje-Galvan­ and Gregory A. Voth

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

**Table S1.** Matrix bound times and conformation for the 200ns equilibration trajectories

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **System** | **Bound leaflet** | **Conformation** | **1st contact (ns)** | **Myr release (ns)** |
| *Control200* | top | open | 90 1 | 112 2 |
| *Control-1200* | bottom | open | 68 1 | 80 2 |
| *Inner200* | bottom | open | 52 1 | 145 2 |
| *Inner-1200* | top | blocked | 70 1 | *no release* |
| *Raft200* | top | blocked | 25 1 | *no release* |
| *Raft-1200* | top | open | 35 1 | 75 2 |

**Table S2.** Simulation details for the microsecond trajectories of MA proteins.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Monomers*** | | | | |
| **System** | **length (ns)** | **Bound leaflet** | **Conformation** | **Myr location** |
| *Controlpre* | 5000 | top | open | *pre-inserted* |
| *Control-1pre* | 3000 | top | open | *pre-inserted* |
| *Control-2pre* | 5000 | top |  | *pre-inserted* |
| *Inner* | 5000 | bottom | open | exposed |
| *Inner-1* | 1600 | bottom | open | exposed |
| *Innermyr* | 5000 | top | open | partially-inserted |
| *Raft* | 5000 | top | blocked | hydrophobic cavity |
| *Raft-1* | 3000 | top | blocked | hydrophobic cavity |
| *Raftpre* | 5000 | top | open/blocked | *pre-inserted* |
| ***Three units*** | | | | |
| **System** | **length (ns)** | **Bound leaflet** | **Conformation** | **Myr insertion** |
| *inner-Ltrimer* | 1000 | top | trimer | yes (2) |
| *inner-Lmono* | 1000 | top | 3 monomers | none |
| *raft-Ltrimer* | 1000 | top | trimer | none |
| *raft-Lmono* | 1000 | top | 3 monomers | none |

The following subscripts to the membrane models have been selected as notation: ‘*200*’ for the 200 ns trajectories (*raft200, see also Table S1*); “*pre*” for the trajectories starting from a pre-inserted Myr configuration (*raftpre, see also Table S2*); “*myr*” for the system starting with a partially inserted Myr (*innermyr*); “*mono*” and “*trimer*”for the systems simulating separated MA units vs. a formed trimer, respectively; the suffixes “1” & “2” to denote replicates; and the suffix “L” to denote larger membranes with three MA units on the surface.

**Table S3.** Analyses performed on the systems to quantify protein binding, Myr insertion, protein conformational changes, membrane mechanical/structural properties. (The trajectories were re-centered around the protein and the binding leaflet using GROMACS(1) *trajectorv* tool prior to computing the respective quantities.)

|  |  |  |
| --- | --- | --- |
| **Analysis** | **Description** | **Systems analyzed** |
| **Distances** ()  *MDAnalysis(2, 3)* | In most of the cases the z-component of the distance between the center-of-mass (com) of molecules was computed. The bilayer center was determined as the average position between the phosphorus atoms of each leaflet. For the Lid-H1 distance, the absolute value of the distance is reported. | *All systems* |
| **Self-distance** ()  *MDAnalysis* | We defined this quantity as the z-component of the difference between the position of the last and first alpha carbons of a sequence of residues in the protein. For H1, the difference was computed between the com of residues 18 and 8 as shown with the black bashed lines in the sketch to the left.  For the HBR, the middle residue (25) in the sequence was taken as the “last” residue for this calculation since this region is a U-loop; the first residue was defined as the com between the actual first and last residues in the sequence (19 & 29). This distance is shown with the blue dashed lines in the sketch.  A positive value indicates the last residue in the sequence is higher than the first; a value of zero indicates the section is essentially flat. | *Inner, inner1, inner-Ltrimer* |
| **Cosine of the angle**  *MDAnalysis* | The absolute value of the cosine of the angle between a vector defined for the Myr tail and the unit vector in the z-direction is computed and reported. We took the absolute value to present a unified measurement of the orientation of Myr in the bilayer given the protein indistinctly binds the top or bottom leaflets of our symmetric models. The Myr vector is defined from the com of C2 to the com of C14, the first and last atoms of Myr.  A value of 1 indicates Myr is aligned with the z-axis, parallel to the bilayer normal; a value of 0 indicates it is perpendicular to it, i.e. is aligned to the membrane surface. | *Inner, inner1* and *inner-Ltrimer* systems; all *controlpre* and *raftpre* systems. The *innermyr* replica is also reported as reference. |
| **Lipid counts**  *MDAnalysis* | Count of the number of phosphorus atoms from PIP2 and DOSP lipid species within 1 nm of the alpha carbons of the protein. | *All systems* |
| **Lipid densities**  *MDTraj (4)* | 2D histograms of the x-y position of each phosphorus atom in the lipid species of interest computed during the last 50ns of simulation for the short trajectories, and the last 300 ns for the microsecond trajectories. | *All systems* |
| **SCD parameters**  *VMD (5)* | The lipid tail order parameters serve to characterize the structure and order of the lipid tails inside the bilayer. They are computed from simulation data measuring the angle between every C-H bond in the lipid tails and the bilayer normal (z-axis) according to:  Higher values indicate a more ordered bilayer; lower values are expected for the last carbons in the lipid tails and regions with double bonds. Increased sterol concentration tends to increase order in the bilayer (6). The order parameters were computed over the last 200ns of the equilibrated trajectory. | *Inner, inner1, inner-Ltrimer* |
| **Area per lipid** (2/lipid)  *GROMACS* | The average surface area of the x-y plane (parallel to the membrane surface) was computed from the simulation box dimensions over the last 50ns of simulation of the membrane-only equilibration runs of the small membrane patches. The value was divided among the lipids in each leaflet (150) to obtain the average APL for each model and is reported on Table S1. | *Control, inner, raft* (small membrane patches) |
| **RMSF**  *MDAnalysis* | The root-mean-square-fluctuations (RMSF) were computed per residue using the *RMSF* tool from *MDAnalysis* to measure the protein changes during Myr insertion. During the analysis, the protein is aligned to the its structure in the first frame of the trajectory to avoid including translation and rotation of the protein in the calculation. | *All systems* |
| **tICA**  *VMD, MDtraj, MDAnalysis, PyEmma (7)* | Time-lagged independent component analysis is a reduction of dimensionality technique to obtain the slowest independent component in a given process as linear combinations of *n* features. These components are identified by solving the eigenvalue-eigenvector problem for the co-variance matrix of the desired process (8-11). To characterize the Myr insertion in our simulations we examined the distances between every carbon in the Myr tail and the in the main regions of the protein, i.e. H1-H5 and the HBR (all the other coiled regions were not considered). This feature was selected as the best description for Myr insertion after comparison with other features such as the cartesian coordinates of the protein, distances among Myr carbon atoms, and the protein RMSD. The trajectory was pre-processed and the protein aligned to its position in the first frame of the trajectory using VMD to prevent translational and rotational degrees of freedom from interfering in tICA. The lag-time for this analysis was set to 50 after analyzing the decay of eigenvalues with respect to lag time for the two slowest tICs. The coefficients of the corresponding eigenvectors, a.k.a. tIC loads, were also examined to determine the distances that contribute more to the respective tIC, i.e. the slowest motions during Myr insertion. | *Inner, Inner-1, inner-L* w/trimer systems, and the small *control* and *raft* systems with no Myr insertion. |
| **tIC contributions to Myr conformation** (energy/kT)  *MDAnalysis, PyEmma, Seaborn (12)* | An estimate of the relative contributions of a given metric to the state or conformation of Myr in the system, we computed the logarithm of the histogrammed data of (i) the C2 and C14 distances from the bilayer center, see Fig. 4 in the main manuscript; and (ii) the slowest independent components identified from *tICA.* The estimates are given in units of kT (7), a factor of 2.6 for T=310K would render the units in KJ/mol. To facilitate the comparison among computed values from different trajectories we set the minimum to zero. | *All systems* |

**References for the Supplemental Information**

1. Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1-2:19-25.

2. Gowers, R. J., M. Linke, J. Barnoud, T. J. Reddy, M. N. Melo, S. L. Seyler, J. D. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney, and O. Beckstein. 2016. MDAnalysis: A Python Package for the Rapida Analysis of Molecualr Dynamics Simulations. Proceedings of the 15th Python in Science Conference:98-105.

3. Michaud-Agrawal, N., E. J. Denning, T. B. Woolf, and O. Beckstein. 2011. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. Journal of Computational Chemistry 32(10):2319-2327.

4. McGibbon, Robert T., Kyle A. Beauchamp, Matthew P. Harrigan, C. Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, L.-P. Wang, Thomas J. Lane, and Vijay S. Pande. 2015. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophysical Journal 109(8):1528-1532.

5. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14(1):33-38, 27-38.

6. Monje-Galvan, V., and J. B. Klauda. 2015. Modelling Yeast Organelle Membranes and How Lipid Diversity influences Bilayer Properties. Biochemistry 54:6852-6861.

7. Scherer, M. K., B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. Noé. 2015. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. Journal of Chemical Theory and Computation 11(11):5525-5542.

8. Brotzakis, Z. F., and M. Parrinello. 2019. Enhanced Sampling of Protein Conformational Transitions via Dynamically Optimized Collective Variables. Journal of Chemical Theory and Computation 15(2):1393-1398.

9. Pérez-Hernández, G., and F. Noé. 2016. Hierarchical Time-Lagged Independent Component Analysis: Computing Slow Modes and Reaction Coordinates for Large Molecular Systems. Journal of Chemical Theory and Computation 12(12):6118-6129.

10. Pérez-Hernández, G., F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé. 2013. Identification of slow molecular order parameters for Markov model construction. The Journal of Chemical Physics 139(1):015102.

11. M. Sultan, M., and V. S. Pande. 2017. tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. Journal of Chemical Theory and Computation 13(6):2440-2447.

12. Wakson, M., O. Botvinnik, D. O'Kane, P. Hobson, J. Ostblom, S. Lukauskas, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. d. Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, A. Miles, Y. Ram, T. Brunner, T. Yarkoni, M. Lee Williams, C. Evans, C. Fitzgerald, Brian, and A. Qalieh. 2018. mwaskom/seaborn: v0.9.0 (July 2018). Zenodo.