Supplementary data for HIV-1 Tat membrane interaction probed using X-ray and neutron scattering, CD, and MD simulations

Kiyotaka Akabori[1], Kun Huang[2], Bradley W. Treece[1], Michael S. Jablin[1], Brian Maranville[3], Arthur Woll[4], John F. Nagle[1], Angel E. Garcia[2], and Stephanie Tristram-Nagle[1]

[1]Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania, [2]Department of Physics, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180, [3]NIST Center for Neutron Research, 100 Bureau Drive, Stop 6102, Gaithersburg, MD 20899, [4]CHESS, Cornell University, 161 Synchrotron Drive, Ithaca, NY 14853

1. Mosaic spread caused by Tat

As hydration proceeded, the mosaic spread, or degree of misorientation, increased. This was apparent by the increasing lengths of the arcs emanating out from the h=1 and h=2 lamellar orders, shown in **Fig. S1**. Mosaic spread was determined by examining lamellar orders in the LAXS CCD position. Matlab was used to plot the intensity of a lamellar order as a function of η, the angle off the meridian. The intensity data above and below the lamellar order were averaged and subtracted from the intensity at the lamellar order to remove diffuse scattering. The subtracted data were fit with a Voigtian function, setting the Gaussian width to the beam width, and extracting the Lorentzian width (ω_L). Then, the following equation was applied to obtain α, $\omega_L^2 = \alpha^2 + 4(\gamma - \theta_B)^2$, where γ=angle of incidence and θ_B is half the scattering angle of the lamellar order. These results are included in Fig. S2, as a function of time after stopping the helium flow and beginning the hydration. The cooling Peltier was at first set to 200 mAmps, which caused the lamellar D-spacing to increase. It was then decreased to 100 mAmps, which caused less water to condense into the sample, and the lamellar D-spacing decreased, as did the mosaic spread. Thus, Tat's ability to disrupt the bilayers appears to be a reversible effect, dependent on the hydration of the sample.

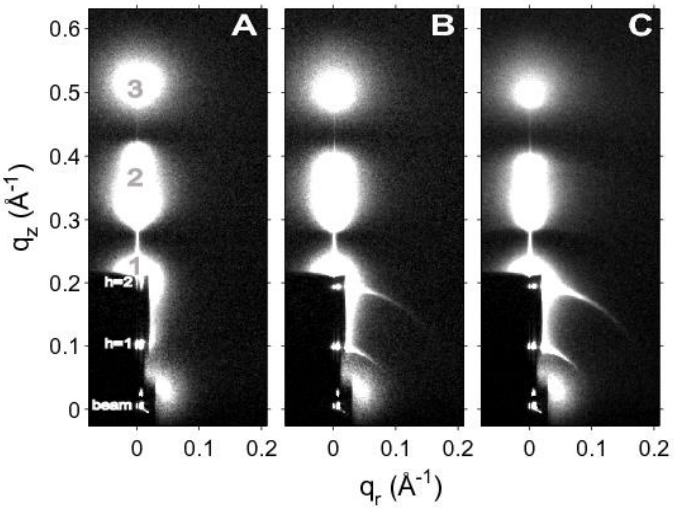

Figure S1. Hydration of DOPC/DOPE (1:1), x=0.034 Tat, 37°C. **A.** Lamellar D-spacing = 60.6 Å, **B.** 65.4 Å, **C.** 65.3 Å. Mosaic spread increases with hydration, shown by visible lengthening of thin arcs near h=1 and h=2 in **B** and **C**. White lobes of diffuse scattering intensity have large grey numbers, while lamellar orders and beam are shown to the left of the Molybdenum beam attenuator (short, dark rectangle).

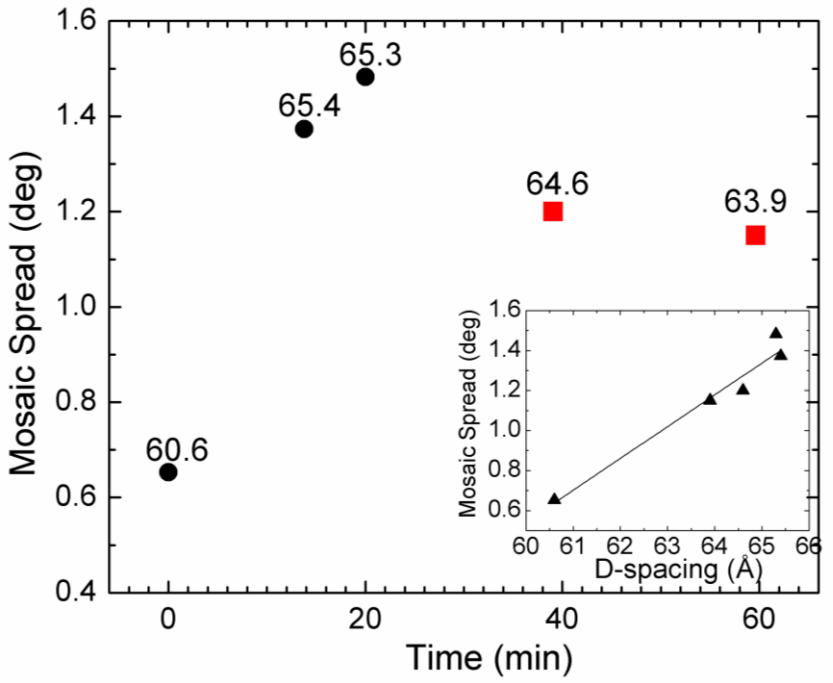

Figure S2. Mosaic spread as a function of time of hydration of DOPC/DOPE (1:1) x=0.034 Tat. Cooling Peltier setting was changed from 200 mAmps (black circles) to 100 mAmps (red squares). Numbers are the lamellar D-spacing (Å). Inset shows that mosaic spread and D-spacing are linearly related.

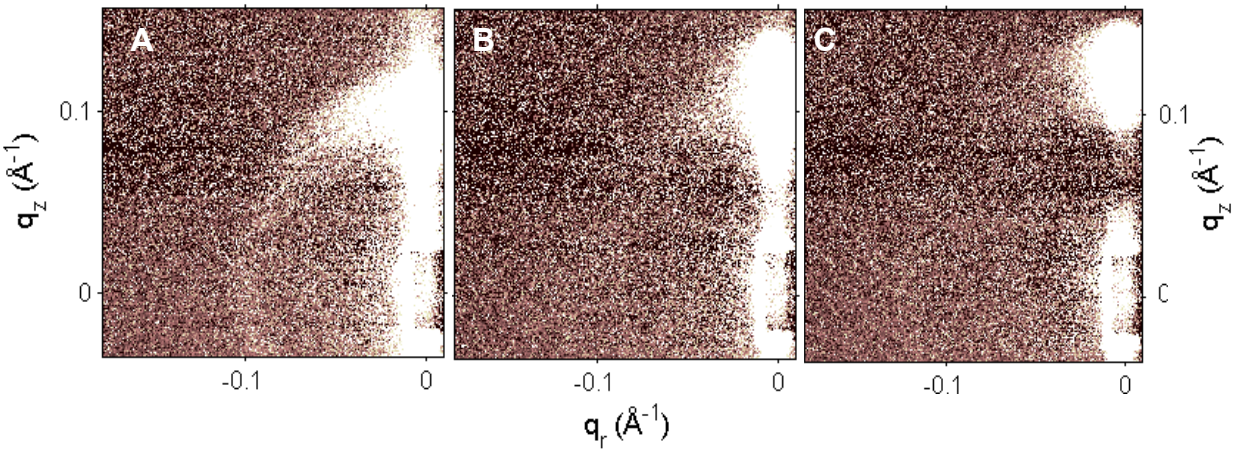

Figure S3. Neutron scattering from stacks of DOPC:DOPE (3:1)/Tat, x=0.059 on a Si substrate hydrated from D_2O vapor. Angle of incidence was 2.4^o and temperature was 37^{o}C. D-spacings were **A.** 63.7 Å, **B.** 57.6, **C.** 49.1 Å. The disappearance of the extensive arc emanating from the first lamellar order as the sample dries, indicates that the misorientation of layers in a stack is a reversible process with hydration, similar to that observed using X-rays.

2. WAXS analysis and results

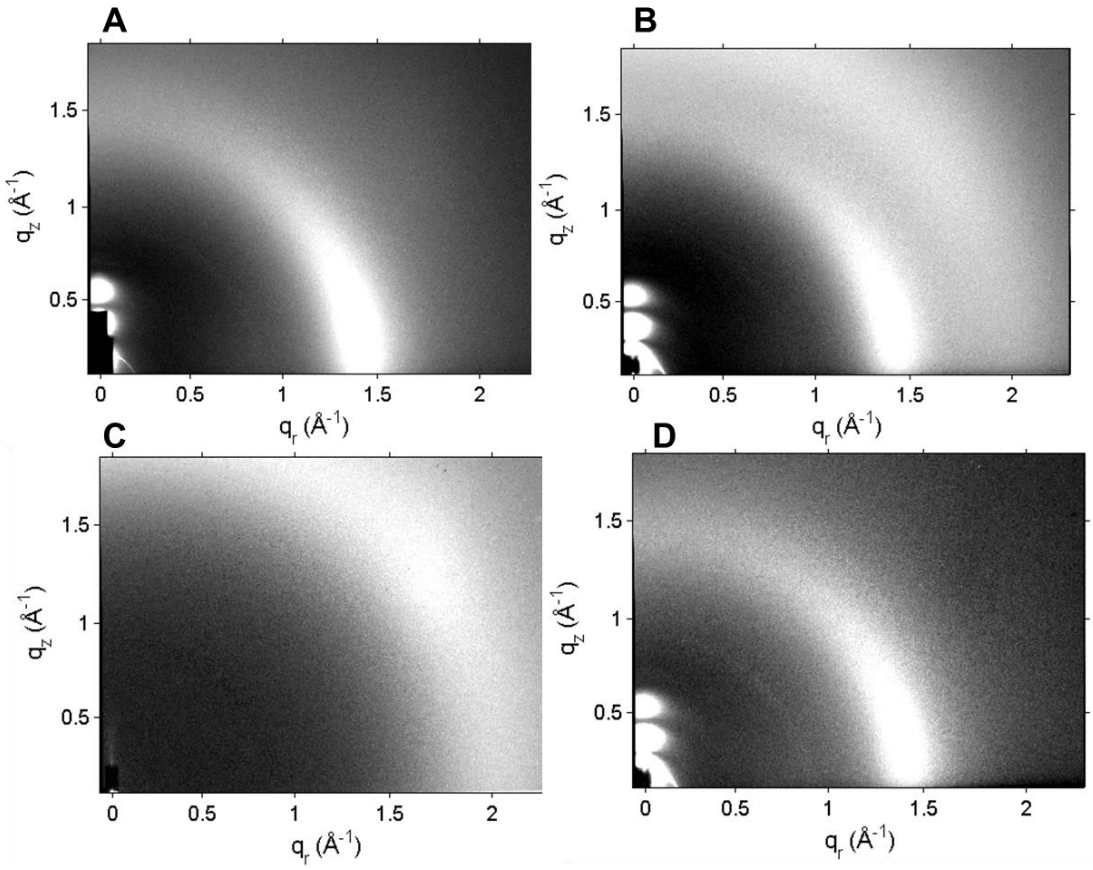

Figure S4. Effect of hydration on WAXS data. **A.** WAXS of a fairly dry sample, lamellar D-spacing = 59.7 Å. **B.** After 15 minutes hydrating with a cooling Peltier setting of 300 mAmps, D = 66.1 Å. **C.** WAXS of water condensed from the vapor onto a silicon wafer. **D.** Sample in **B.** with water removed by subtracting a fraction of condensed water from **C.**

Before the WAXS analysis was carried out, water scattering (shown in **Fig. S4.B,C**) was necessarily removed. Water condenses from the vapor during sample hydration due to the well-insulated hydration chamber and a Peltier cooler located under the sample. As the lamellar D-spacing swells, condensed water appears as a broad, isotropic, diffuse band centered at **q**=2 Å^{-1}. This band is close to the chain correlation WAXS, which is used to obtain the order parameter, S_{xray}, so it is necessary to remove the water scattering prior to analyzing chain WAXS. This is done by quantitatively matching the water scattering in **Fig. S4.B** with a fraction of that in **Fig. S4.C**, and subtracting the water out. We found that hydrated samples are less ordered than drier samples, which suggests that a hydrated sample with the water subtracted out, as in **Fig. S4.D**, is the appropriate one to use for our analysis.

Briefly, WAXS allows calculation of the orientational order parameter, S_{xray}, similar to an NMR order parameter. Mosaic spread α decreases the apparent value of $S_{xray,}$ since misoriented layers will contribute to the diffuse arc emanating up from the equator (see **Fig. S4.B**) (manuscript in preparation). Mosaic spread was determined by examining lamellar orders in the

LAXS CCD position as described in **Section S1**. Matlab was used to plot the intensity of a lamellar order as a function of η, the angle off the meridian. The intensity data above and below the lamellar order were averaged and subtracted from the intensity at the lamellar order to remove diffuse scattering. The subtracted data were fit with a Voigtian function, setting the Gaussian width to the beam width, and extracting the Lorentzian width (ω_L). Then, the following equation was applied to obtain α: $\omega_L^2 = \alpha^2 + 4(\gamma - \theta_B)^2$, where γ=angle of incidence and θ_B=half the scattering angle of the lamellar order.

We used α to correct the decrease of S_{xray} due to mosaic spread, by first plotting S_{xray} values vs. mosaic spread. This was done purely from a theoretical model of how mosaic spread redistributes intensity in the phi direction (manuscript in preparation). We then used these results to interpolate to the value of S_{xray} without mosaic spread given the apparent value of S_{xray} and α for the sample. In order to track the behavior of apparent S_{xray} values over a larger range, mock S_{xray} data were created with different values of S_{xray}, and a Lorentzian weight distribution of smearing the intensity measured from the CCD image was applied to the mock data. S_{xray} was then determined again on the smeared mock data and this new value was assigned to the original S_{xray} value and α. By obtaining smeared S_{xray} values over a wide range of mock data and mosaic spreads, many measured values of S_{xray} and α from the experimental data could be corrected to values of S_{xray} without mosaic spread. The corrections to S_{xray} due to mosaic spread were ~5-8%.

3. Alternative location of Tat from SDP modeling alone

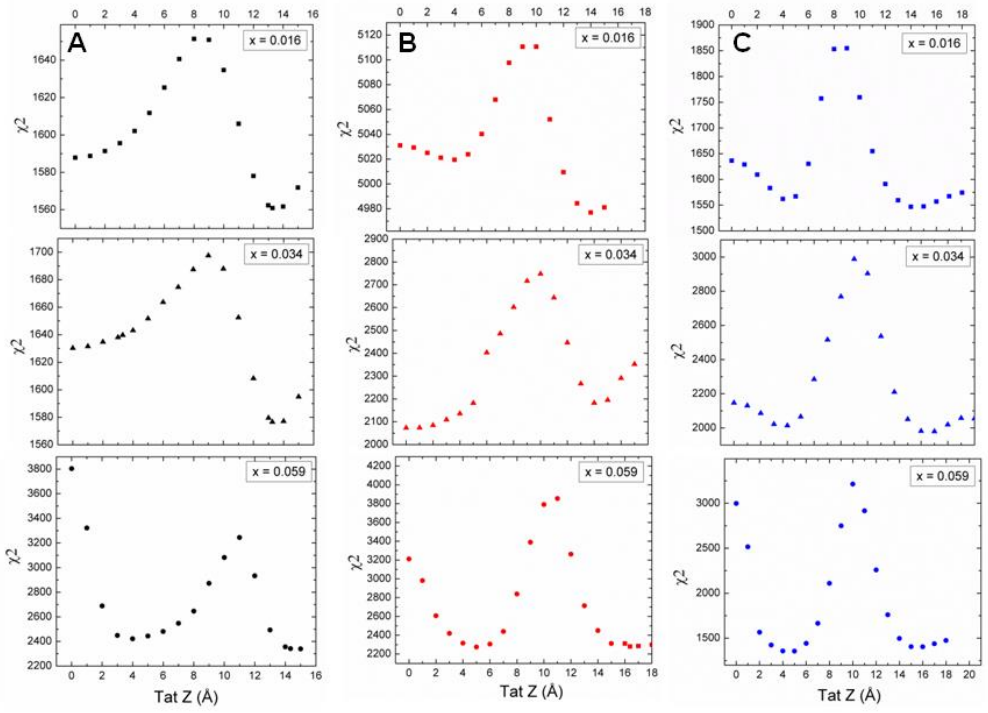

Figure S5. Chi-square as a function of Tat position Z across the bilayer, where 0 = bilayer center. **A**. DOPC, **B**. DOPC:DOPE (3:1), **C**. DOPC:DOPE (1:1). P/(L+P) molar ratios are shown on figures.

In addition to allowing the Tat position to move freely during the SDP fit to the form factor data, we can also fix Tat's position. In order to determine the location of Tat in these membrane mimics, the Tat position Z is fixed in 1 Å steps and the resulting chi-square from the SDP model fit is plotted as shown in **Fig. S5**. As shown, more than one minimum occurred for some of the lipid mixtures and Tat concentrations. In addition to the headgroup location for Tat, there was a second best fit locating Tat in the interior of the hydrocarbon at ~4-5 Å from the bilayer center. This position had an equally good chi-square as the headgroup position in DOPC:DOPE (3:1) and DOPC:DOPE (1:1) at x = 0.059. In general, the Tat interior position was favored at higher Tat concentrations and in mimics with DOPE. For DOPC mixtures, the Tat hydrocarbon position did not fit the data as well as the headgroup position, even at x = 0.059. Thinning of the bilayers was identical with Tat in the hydrocarbon position. Electron density profiles with Tat in the hydrocarbon interior are shown in **Fig. S6**.

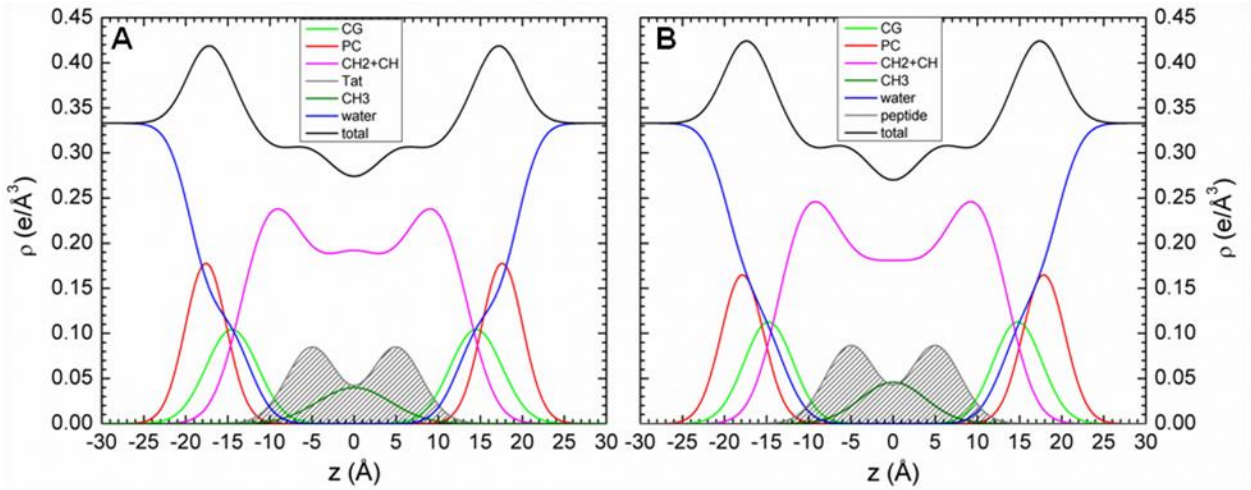

Figure S6. Electron density profiles with Tat in the hydrocarbon interior. **A.** DOPC/DOPE(3:1), x= 0.059 **B.** DOPC/DOPE(1:1), x= 0.059.

Although our SDP model fitting does find these equally low chi-square values when fixing Tat at discrete positions across the bilayer, a free fit does not locate Tat in the hydrocarbon interior. Also, when we compared the form factor from our X-ray data to the form factor from the MD simulations (**Fig. 4**, main paper), we find poor agreement, suggesting that the interior position of Tat is an artifact of the SDP model fitting program. This interior location of Tat, obtained using the SDP model fitting program, was the impetus for placing Tat in the interior in the MD simulation, but we discovered in **Fig.4** (main paper) that the simulation does not support this result. We show it here merely for interest.

4. Circular dichroism (CD) spectroscopy

CD spectroscopy was used to determine the secondary structure of Tat in a fully hydrated lipid environment compared to that solubilized in water. Two cuvette orientations, vertical and inverted, were used to sample different parts of the hydrated film. 8–10 scans were collected with a Jasco 715 on each sample from 260 to 185 nm at 100 nm/minute and averaged. CD data in mean residue ellipiticity were analyzed using the CDSSTR or Continll function in DichroWEB [1] with Basis Set #4 and #SP175 [2]. Uncertainties shown in **Table 1** in the main paper were estimated by averaging multiple fits to the data and taking the average and standard deviation. Precise protein concentrations of the oriented samples were determined using the absorption output of the Jasco, converting to molarity using the WEBsite http://spin.niddk.nih.gov/clore [3]. We also attempted to use the more conventional method of incorporating Tat into small unilamellar vesicles, but this was not successful, since the SUVs were unstable. We therefore used another peptide, a-synuclein (aS) with both the small unilamellar vesicle method and our thin film method and determined that both methods agreed to within 15% for the helical content. In addition, we measured the CD of Tat suspended in 3 ml water (0.05 mg/ml) without lipid.

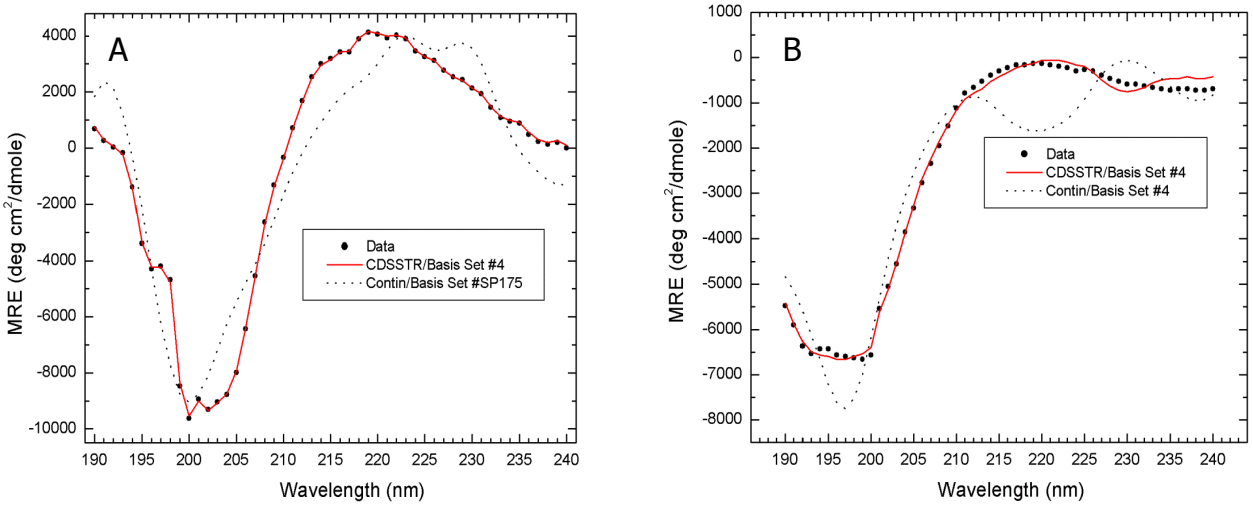

Figure S7. CD spectroscopy of **A**. Tat solubilized in water and **B**. DOPC/DOPE (3:1)/Tat, x = 0.108.

Fig. S7.A shows the mean residue ellipticity (MRE) vs. wavelength of Tat solubilized in water. MRE for DOPC/DOPE (3:1)/Tat, x=0.108 is shown in **Fig. S7.B.** Results of the fits to protein data sets are summarized in the Table 2 in the main paper. As shown in Table 2, there was little difference between the secondary structure of Tat solubilized in water and in an unoriented lipid film. The main 2^{o} structure in each case is β-sheet (includes regular and distorted strands and turns), followed by random coil and <10% α-helix. The random coil may also include a poly(pro)II left-handed helix (P2) [4]. These results suggest that the membrane has little effect on the 2^{o} structure of Tat, and that Tat is not an α-helix in either case.

5. Details for Fig. 9 in main paper (results from MD simulations)

Besides the results shown in the main paper in **Figs. 4, 5, 7, 9** and **Table 3**, we obtained the locations of specific atoms in the bilayer from the simulations. Using the SIMtoEXP program we grouped atoms in different ways, as shown in **Fig. S8**. ARG+ includes only the atoms in the guanidinium group for the 6 arginines in Tat, while ARG indicates the remainder of the atoms in the arginine group, including side chain carbons, alpha-carbon and adjoining peptide carbonyl and nitrogen. The two lysines were treated similarly. As shown, both the charged and uncharged parts of arginines are located closer to the bilayer center than the corresponding lysine parts. The arginines are closer to the unperturbed-by-Tat glycerol-carbonyl group (dotted line), than they are to the unperturbed phosphate groups (solid line) or to the unperturbed choline group (dashed line). However, as is shown in the cartoon in the main paper in **Fig. 9**, the perturbed phosphate groups lie closer to the bilayer center by 3.6 Å. More details of the simulation results are shown in **Table S1** which reports a weighted average based on the χ^2 of the best fit of four simulations to the X-ray data.

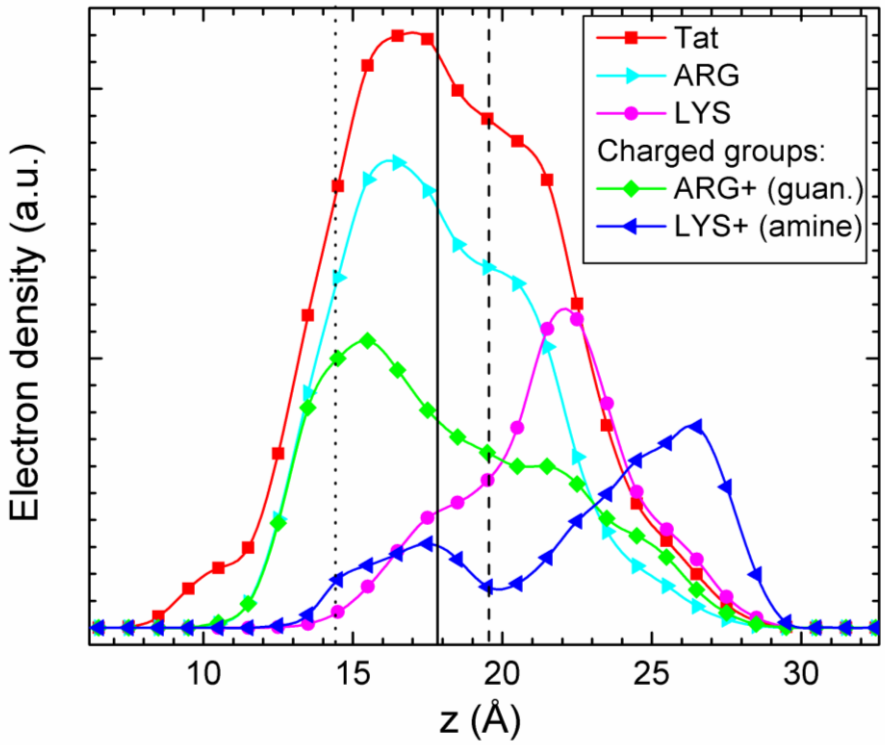

Figure S8. Distributions of components in the MD atomistic simulation of DOPC:Tat 128:2 where the area/lipid (72 Å^2) and Tat position (18 Å) agree best with our X-ray data. Tat indicates the entire Tat molecule, ARG+ indicates the guanidinium groups from 6 arginines, ARG is the rest of the arginine group, LYS+ indicates the amine groups of the 2 lysines, and LYS is the rest of the lysine molecules. The average location of the unperturbed lipid components are indicated for glycerol-carbonyl (dotted line), phosphate (solid line) and choline (dashed line).

Table S1. Results used for Fig. 9 in main paper

# of Tat	A_L	Z_{Tat}	$<D_{PP}>$	D_{PP}	x	Δt	H_{Tat}	R_{Tat}	R_2	Z_{Phos}	$Z_{guanidinium}$
2	72.9	17.1	35.4	32.7	35.6	3.6	8.7	8.3	17.1	14.6	15.1
4	75.2	17.3	34.8	31.9	NA	4.4	7.7	8.8	NA	13.8	15.4

A_L, area/lipid; Z_{Tat}, position of Tat center of mass, $<D_{PP}>$, phosphorus-phosphorus distance averaged over all lipids; D_{PP}, Tat-perturbed phosphate groups; x, = $64*<D_{PP}>-3.5*D_{PP})/60.5$ (thickness away from Tat); Δt, = $<D_{PP}^0>-D_{PP}$ (where $<D_{PP}^0>$ is the value from the pure DOPC simulation); H_{Tat}, Tat height; R_{Tat}, radius of Tat's cylinder; R_2, radius of the calculated in-plane Tat-perturbed region; R_3, effective radius of the simulation box.

6. Additional details concerning the MD simulation methods

All simulations were conducted with a 2 fs time integration step. SETTLE[5] was used to constrain water molecules and LINCS[6] was used to constrain all other bond lengths in the system. VdW interactions were truncated at 1.4 nm with a twin-range cutoff scheme and a dispersion correction was applied to both energy and pressure. Electrostatics interactions were treated with the particle-mesh Ewald (PME) method [7]. The direct term for electrostatics was evaluated within 1.0 nm cutoff and the Fourier term was evaluated with a 0.12 nm grid spacing and a 4^{th} order interpolation. Each system was simulated at 310 K using the V-rescale algorithm[8] with a 0.2 ps time coupling constant. The semi-isotropic parrinello-rahman barostat [9] was used to couple the system at 1 atm in the Z direction with a 5 ps time coupling constant, while the projected area at the X-Y plane was fixed by setting the system compressibility to 0. We inserted the Tats into the system by initially turning off all interactions between Tats and the rest of the system, with Tats constrained at different depths. Then we slowly turned on the interactions to normal strength through thermodynamics integrations. We used umbrella potentials to constrain Tats at desired depths with a force constant of 3000 KJ/mol/nm^2.

References for Supplementary data

[1] A. Lobley, L. Whitmore, B.A. Wallace, DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra, Bioinformatics, 18 (2002) 211-212.

[2] A. Abdul-Gader, A.J. Miles, B.A. Wallace, A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy, Bioinformatics, 27 (2011) 1630-1636. DOI 10.1093/bioinformatics/btr234.

[3] N.J. Anthis, G.M. Clore, Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, Protein Sci, 22 (2013) 851-858. Doi 10.1002/Pro.2253.

[4] N. Sreerama, R.W. Woody, Structural composition of beta(I)- and beta(II)-proteins, Protein Sci, 12 (2003) 384-388. Doi 10.1110/Ps.0235003.

[5] S. Miyamoto, P.A. Kollman, Settle - an analytical version of the shake and rattle algorithm for rigid water models, Journal of Computational Chemistry, 13 (1992) 952-962. DOI 10.1002/jcc.540130805.

[6] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A linear constraint solver for molecular simulations, Journal of Computational Chemistry, 18 (1997) 1463-1472. Doi 10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.0.Co;2-H.

[7] T. Darden, D. York, L. Pedersen, Particle mesh ewald - an N.log(N) method for ewald sums in large systems, J Chem Phys, 98 (1993) 10089-10092. Doi 10.1063/1.464397.
[8] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J Chem Phys, 126 (2007) Doi 10.1063/1.2408420.

[9] M. Parrinello, A. Rahman, Polymorphic transitions in single-crystals - a new molecular-dynamics method, J Appl Phys, 52 (1981) 7182-7190. Doi 10.1063/1.328693.