June 2, 2014

1 Analysis of Diffuse Scattering

1.1 Theory

This section briefly describes NFIT analysis.

1.2 Results

Show X-ray data. Show fitting boxes. Show the Kc values. Also, show the resultant form factors, which qualitatively show the membrane thinning.

2 Modeling the Bilayer Structure

In the case of X-rays, the features with the most contrast are the electron-dense headgroups, providing the head-head spacing D_{HH}, as well as the terminal methyl groups in the bilayer center. Modeling of the bilayer structure was done similarly to the SDP model [1].

Parsing of DOPC into lipid components is shown in Fig 1. The phosphate/choline (PC) and carbonyl/glycerol (CG) groups together make up the lipid headgroup whereas the hydrocarbon chain region is divided into two components, the methylene and methine groups combination (CH_2+CH) and terminal methyl group (CH_3). We combine methylene (CH_2) and methine groups (CH) in order to avoid proliferation of fitting parameters.

(Also, create a table for basic parameters like ρ).

2.1 Functional forms

Our model for electron density profile (EDP) of Tat/lipid bilayer system consists of six structrual subgroups (Fig. **??**). The volume probability distributions of components PC, CG, terminal methyl groups (CH_3), and Tat are described by Gaussian functions,

$$P_i(z) = \frac{c_i}{\sqrt{2\pi}}\left(\exp\left\{-\frac{(z+z_i)^2}{2\sigma_i^2}\right\} + \exp\left\{-\frac{z-z_i)^2}{2\sigma_i^2}\right\}\right), \qquad (1)$$

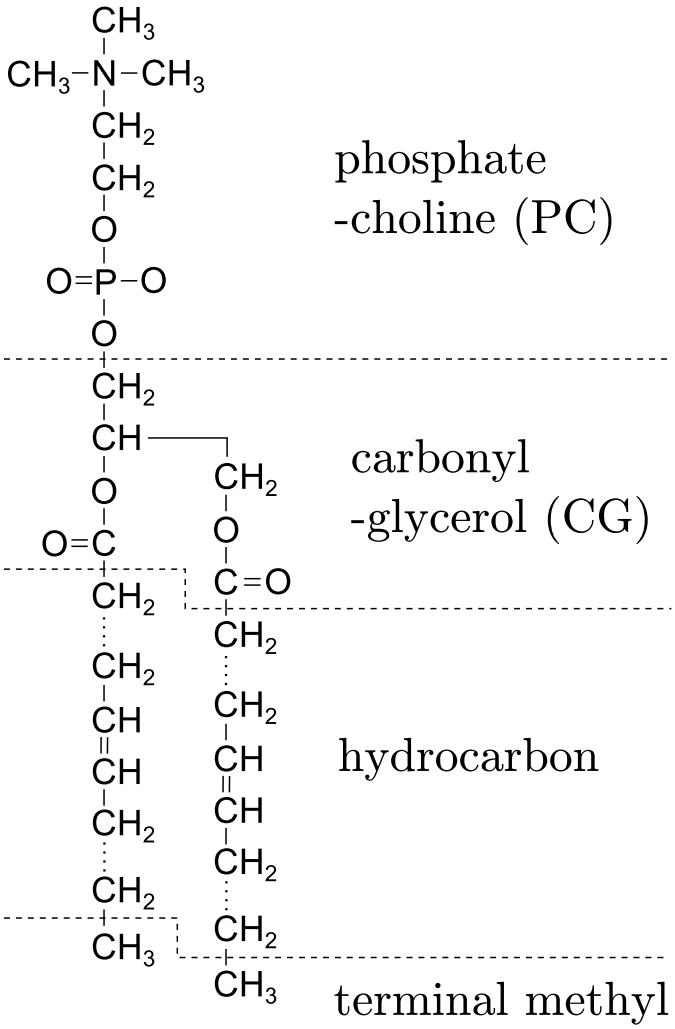

Figure 1: Schematic of DOPC showing each lipid component. The dash lines show where the lipid is divided into different components. The lipid headgroup is divided into two components, phophate-choline and carbonyl-glycerol. The hydrocarbon chain region is also divided into two components, methelence+methine and terminal methyl groups.

where c_i is an integrated area underneath the curve and the two parts of the expression describe the two bilayer leaflets. The calculation of c_i is explained below.

The hydrocarbon chain region (HC) is represented by error functions,

$$P_{\mathrm{HC}}(z) = \frac{1}{2}[\mathrm{erf}(z, -z_{\mathrm{HC}}, \sigma_{\mathrm{HC}}) - \mathrm{erf}(z, z_{\mathrm{HC}}, \sigma_{\mathrm{HC}})], \tag{2}$$

where

$$\mathrm{erf}(z, z_i, \sigma_i) = \frac{2}{\sqrt{\pi}} \int_0^{\frac{z-z_i}{\sqrt{2\sigma}}} dx\, e^{-x^2}. \tag{3}$$

The volume probability distribution for the methylene and methine groups combination can then be expressed as

$$P_{\mathrm{CH_2+CH}}(z) = P_{\mathrm{HC}}(z) - P_{\mathrm{CH_3}}(z). \tag{4}$$

This definition satisfies spatial conservation in the central hydrocarbon region where the total probability P_{HC} equals one. We also consider an alternative definition,

$$P_{\mathrm{CH_2+CH}}(z) = P_{\mathrm{HC}}(z) - P_{\mathrm{CH_3}}(z) - P_{\mathrm{Tat}}(z), \tag{5}$$

which allows Tat to be located inside the hydrocarbon chain region. The spatial conservation requires the water volume probability distribution to be

$$P_{\mathrm{W}}(z) = 1 - P_{\mathrm{PC}}(z) - P_{\mathrm{CG}}(z) - P_{\mathrm{Tat}}(z) - P_{\mathrm{HC}}(z) \quad \text{(model A)} \tag{6}$$

$$P_{\mathrm{W}}(z) = 1 - P_{\mathrm{PC}}(z) - P_{\mathrm{CG}}(z) - P_{\mathrm{HC}}(z) \quad \text{(model B)} \tag{7}$$

Because X-rays measure the contrast between the bilayer and surrounding solvents, water, the experimental form factor is compared to the water subtracted model form factor,

$$F(q_z) = 2 \int_0^{\frac{D}{2}} dz \left(\sum_i (\rho_i - \rho_{\mathrm{W}}) P_i(z) \right) \cos(q_z z), \tag{8}$$

where i = PC, CG, Tat, CH+CH_2, and CH_3.

2.2 Constraints

The height of the hydrocarbon chain error function is fixed to one by imposing spatial conservation, whereas the mean position of the terminal methyls is

fixed to $z_{\mathrm{CH_3}} = 0$ by symmetry arguments. The total lipid volume V_{L} is fixed to the experimentally measured value. The headgroup volume V_{HL} was determined to be 331 Å^3 for gel phase phosphatidylcholine bilayers [**?**], and we assume the same volume for the fluid phase bilayers.

We define the ratios of component volumes that control volume allocation: in the headgroup region,

$$R_{\mathrm{PC}} = \frac{V_{\mathrm{PC}}}{V_{\mathrm{HL}}}, \quad R_{\mathrm{CG}} = \frac{V_{\mathrm{CG}}}{V_{\mathrm{HL}}}, \quad R_{\mathrm{Tat}} = \frac{V_{\mathrm{Tat}}}{V_{\mathrm{HL}}}, \tag{9}$$

and in the hydrocarbon chain region,

$$r = \frac{V_{\mathrm{CH_3}}}{V_{\mathrm{CH_2+CH}}}, \quad r_{12} = \frac{V_{\mathrm{Tat}}}{2V_{\mathrm{CH_2+CH}}}. \tag{10}$$

These volumetric parameters satisfy the following equality,

$$1 = R_{\mathrm{PC}} + R_{\mathrm{CG}} + R_{\mathrm{Tat}} \quad \text{and} \quad V_{\mathrm{L}} - V_{\mathrm{HL}} = 2\left(16V_{\mathrm{CH_2+CH}} + rV_{\mathrm{CH_2+CH}}\right). \tag{11}$$

The component volumes constraint the height of the Gaussians as

$$c_{\mathrm{PC}} = \frac{V_{\mathrm{PC}}}{A_{\mathrm{L}}\sigma_{\mathrm{PC}}} \tag{12}$$

$$c_{\mathrm{CG}} = \frac{V_{\mathrm{CG}}}{A_{\mathrm{L}}\sigma_{\mathrm{CG}}} \tag{13}$$

$$c_{\mathrm{CH_3}} = \frac{2V_{\mathrm{CH_3}}}{A_{\mathrm{L}}\sigma_{\mathrm{CH_3}}} \tag{14}$$

$$c_{\mathrm{Tat}} = \frac{V_{\mathrm{Tat}}}{A_{\mathrm{L}}\sigma_{\mathrm{Tat}}} \tag{15}$$

where A_{L} is area per lipid.

The ratio R_{CG} of the carbonyl/glycerol volume to the headgroup volume V_{HL} was suggested to be 0.41 [2], so we constrain the CG volume to 135.7 Å^3 and the phosphate/choline volume to 195.3 Å^3.

At 30 ℃, the volume of DOPC is 1303 Å^3 and the headgroup volume 331 Å^3, so the volume of hydrocarbon chain region is $1303 - 331 = 972$ Å^3 [2]. The ratio r of the volumes of the chain terminal methyl (CH_3) to the chain methylenes (CH_2) was suggested to be 1.95, and the ratio r_{12} of the volumes of the chain methines (CH) to the chain methylenes 0.91.

Because there are 14 CH_2 groups, 2 CH groups, and 1 CH_3 group in each DOPC hydrocarbon chain, we have $2 \times (14V_{CH_2} + 2V_{CH} + V_{CH_3}) = 972$ Å^3. Using $V_{CH_3}/V_{CH_2} = 1.95$ and $V_{CH}/V_{CH_2} = 0.91$, we get $V_{CH_2} = 27.3$ Å^3, $V_{CH} = 24.9$ Å^3, and $V_{CH_3} = 53.3$ Å^3. These calculated volumes lead to $V_{CH_3}/V_{CH_2+CH} = 1.97$. At 37 ℃, the volume of DOPC was measured to be 1313.5 Å^3, so we have $2 \times (16V_{CH_2+CH} + V_{CH_3}) = 1313.5 - 331$. Assuming that the ratio V_{CH_3}/V_{CH_2+CH} at 37 ℃ is the same as that at 30 ℃ gives $V_{CH_2+CH} = 27.3$ Å^3 and $V_{CH_3} = 53.9$ Å^3. We constrain these volumes to the calculated values in our model.

	DOPC	62:1		28:1		16:1	
		A	B	A	B	A	B
V_L	1314	1344	1344	1380	1380	1432	1432
V_{HL}	331	362	331	397	331	450	331
V_{Tat}	0	30.5	30.5	66.1	66.1	119	119
R_{PC}	0.59	0.54	0.59	0.49	0.59	0.43	0.59
R_{CG}	0.41	0.38	0.41	0.34	0.41	0.30	0.41
R_{Tat}	0	0.08	0	0.17	0	0.27	0
r_{12}	0	0	0.558	0	1.21	0	2.17
r	1.97	1.97	1.97	1.97	1.97	1.97	1.97

Table 1: Volumetric constraints. A and B refer to two different models described in the text.

2.3 Results

Using the model described above, we fitted our measured X-ray form factors. Figure 2 shows that the best fit to the DOPC form factor was very good. Figure 3 shows the model electron density profile derived from the best fit. (Table ? shows the best fit parameters for DOPC bilayers. Explain the constraints on the distance between the headgroup components.)

For fitting the form factors for systems with Tat, we fixed the width of Tat Gaussian because we found that the best fits of this width was always unphysically too small. We fitted with three different values of widths, 2.5, 3.0, and 3.5, to study the range of variation due to the Tat width. The choice was made based on MD simulation results. (Check this again) Figure 4 shows the best fit to the form factor for DOPC:Tat (62:1). (Show all the results

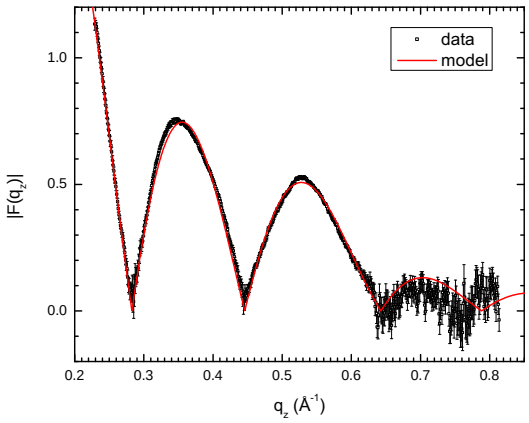

Figure 2: DOPC form factor

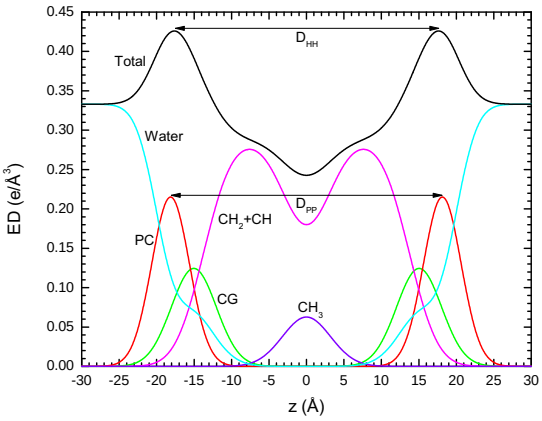

Figure 3: DOPC electron density (ED) profile

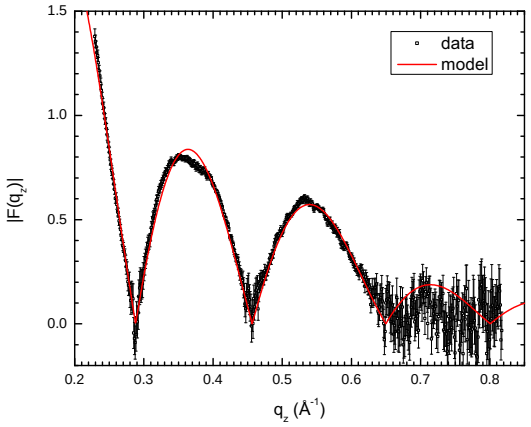

Figure 4: DOPC:Tat (62:1) form factor

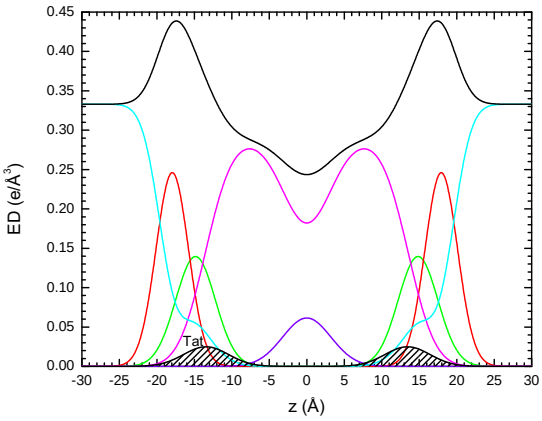

Figure 5: DOPC:Tat (62:1) ED profile. Colors are the same as in Fig. 3

in a composite figure. Show results for DOPC:DOPE bilayers.)

Figure 6 shows χ^2 as a function of fixed Tat position, z_{Tat}. Generally, two minima were found. For DOPC bilayers, z_{Tat} in the headgroup region yielded the best fit. (Show plots for DOPC bilayers) However, as Fig. 7 shows,

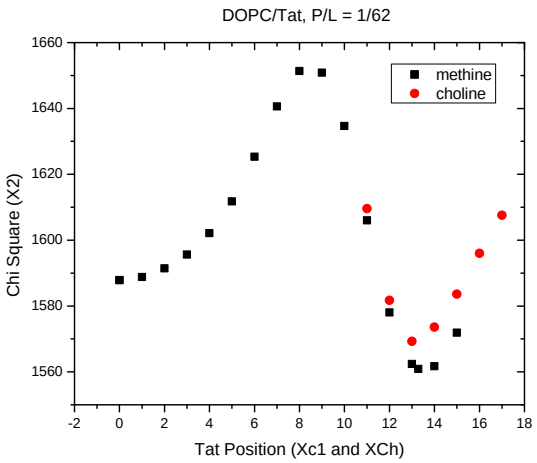

Figure 6: χ^2 as a function of z_{Tat} for DOPC:Tat (62:1).

the fits are are equally good with Tat inside and outside the chain region for DOPC:DOPE bilayers. (Show a bunch more plots for DOPC:DOPE membranes) While modeling suggested that Tat could be in the tail region, this position seems energetically unfavorable as Tat is a hydrophilic molecule. With a help of MD simulations, we were able to discard the interior position as an artifact of our modeling. The simulation results are discussed later.

3 Volume Measurements

(Maybe do this in appendix?)

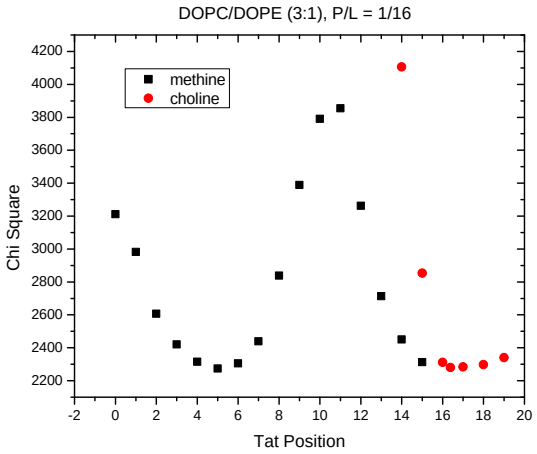

Figure 7: χ^2 as a function of z_{Tat} for DOPC:DOPE (3:1) with $x_{\text{Tat}} = 1/17$, where $x_{\text{Tat}} = Tat/(Tat + Lipid)$.

4 MD simulation

Three systems with different petpide concentrations (DOPC:Tat = 128:0, 128:2, and 128.4) were studied with Gromacs 4.6.1 package [3]. DOPCs were modeled by Slipid force field [4, 5] and HIV-Tats were modeled by Amber99SB [6]. The systems were simulated at 310 K with a constant area in the x-y plane. The z direction was coupled to 1 atm with constant pressure. The center of mass (COM) distance between each peptide and the bilayer was constrained by an umbrella potential with a force constant of 3000 kJ/mol/nm^2. Each system was explored 18 independent simulations as a combination of 3 different constant area and 6 different peptide insertion depths (except the pure DOPC system).

$$z_{\text{cm}} = \frac{\sum_{i=1}^{N} m_i z_i}{\sum_{i=1}^{N} m_i} \tag{16}$$

The center of mass constraint was applied through an external force field, which derives from an added pontential energy of the system. The pontential is like a spring, where it depends on the deviation of the distance between

the center of mass of Tat and DOPC from a prefered value, z_0,

$$U(z_1^{\mathrm{Tat}}, \ldots, z_1^{\mathrm{DOPC}}, \ldots) = -\frac{1}{2}k\big(z_{\mathrm{cm}}^{\mathrm{Tat}} - z_{\mathrm{cm}}^{\mathrm{DOPC}} - z_0\big)^2 \tag{17}$$

Then, $-\partial U/\partial z_i$ is equal to the external force acting on atom, i. Before applying this constraint, Tats were attached to the bilayer from the water region. During the first 20 ns for pre-equilibration, Tats were allowed to change their configuration, which resulted in different configurations for each Tat when attached to the bilayer. It is possible that the Tats configuration inside the bilayer at the end of simulations was affected by this initial configuration of Tats. Instead of preparing many simulations with different initial Tat configurations, we avaraged over all Tats present in the system. We also performed many simulations with different A_L and z_{Tat} to investigate how robust some of the Tat structural features are. Many simulation results are shown in the appendix of this thesis.

4.1 Local Thinning of Membranes

The SIMtoEXP program only gives the average quantities for each leaflet. While our x-ray data are sensitive to the avarage bilayer electron density, local information of Tat-DOPC interactions can be obtained from MD simulations. In this section, we discuss a method to extract a local membrane thickness.

The presence of Tat may result in compression of lipid bilayer along z-direction. If so, the phosphorus-phosphorus distance of the bilayer near Tat (denoted by DPP' in the figure below) may be different from that away from Tat (DPP). For small Tat concentration, DPP would be the same as that of pure DOPC if the distance from all Tats is large enough. Of course, for our concentrations, the thinning effect may extend throughout the bilayer because the lateral effect of Tat might have a larger lateral decay length than the distance between Tats. Whether that is the case or not, one would expect that the thickness near the Tats is smaller than the average thickness.

So, we want to measure DPP'. First, let us define what we mean by lipids close to Tat. As in Fig. 8, we imagine a cylinder around Tat and pick up all the phosphorus atoms within it. Approximating Tat as a cylinder with its height given by the FWHM of its electron density distribution, its radius $R_{\mathrm{Tat}} = 9$ Å comes from the volume of Tat $= 1876$ Å^3 and $h_{\mathrm{Tat}} = 7.6$ Å measured from one of the simulations. Let us define the lateral center of the cylinder in some way - a crude approximation would put it at the arginine in

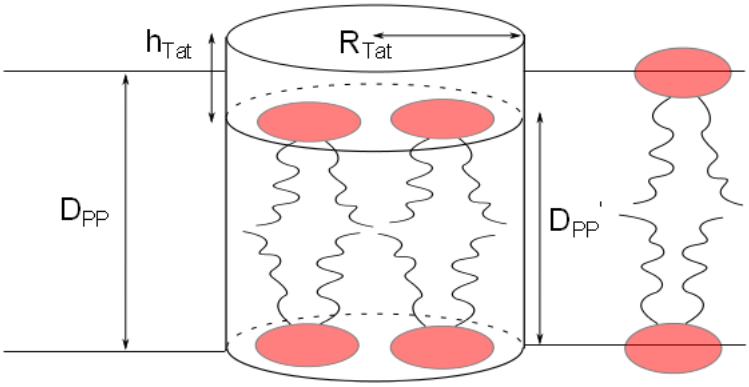

Figure 8: test

the middle of the amino acid sequence. Then let us define DPP' using only those lipids whose phosphorus atoms lie within these 9Å cylinders around the tats. Then $D_{\mathrm{PP}} = z^{+}_{\mathrm{phos}} - z^{-}_{\mathrm{phos}}$ where z^{+}_{phos} and z^{-}_{phos} are the average Z of the n1 (n2) lipids in the upper and lower monolayer, respectively. To be more precise, assume that the arginine in the middle of the amino acid sequence is at the center of the cylinder. For a refined method, we could find the center of mass of each Tat and use them as the lateral center of cylinders (instead of a particular carbon atom in an arginine).

The algorithm for doing this is straightforward. For each time frame (snapshot), the positions (X_i ,Y_i, Z_i) of each Tat, i, are listed. We choose phosphorus atoms whose (X, Y) lateral position lies within 9 Å of any one of the Tat's lateral position. Then, Z position of the chosen phosphorus atoms are placed in a list. Then, z_{phos} are calculated from the list. The number of selected phosphorus atoms in each monolayer was also recorded. This value gives local lateral depletion if the Tat cylinders are assumed not to overlap. After averaging over many snapshots, there will hopefully be decent statistics for local thinning and local difference between Tat and phosphorus Z positions, and maybe local lateral depletion if overlaps are taken into account.

4.2 Linear Model

As described in the previous section, a Tat is modeled as a cylinder with its radius equal to R_1, height h_{Tat}, and volume V_{Tat} such that $R_1 = \sqrt{V_{\mathrm{Tat}}/(\pi h_{\mathrm{Tat}})}$.

Let $h(r)$ represent the phosphorus height profile of a leaflet. The two leaflets are assumed to be decoupled. In the linear model, lipids are seprated into three regions: suppressed, boundary, and unperturbed (see Fig. for the definitions of various quantities). In the suppressed region, lipids are uniformly

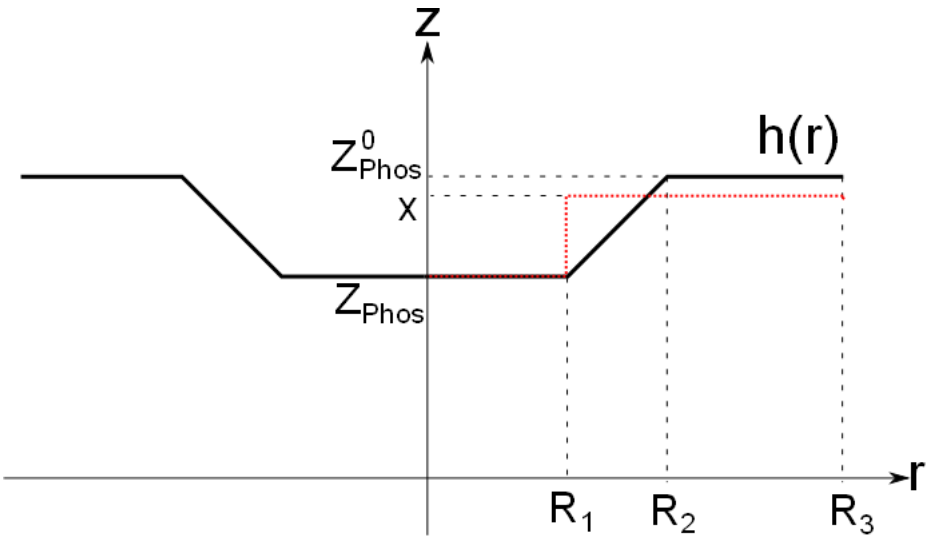

Figure 9: test

compressed by Tat toward the center of the bilayer; $h(r)$ is a constant, z_{phos}. In the boundary region, $h(r)$ is assumed to linearly increase with the distance r from the center of the Tat. In the unperturbed region, lipids do not interact with Tat, behaving identially to DOPC, so the phosphorus position is the same as that of DOPC. A continuous $h(r)$ that satisfies the above criteria is

$$h(r) = \begin{cases} z_{\text{phos}} & \text{if} \quad 0 \le r < R_1 \\ mr + b & \text{if} \quad R_1 \le r < R_2 \\ z^0_{\text{phos}} & \text{if} \quad R_2 \le r < R_3 \end{cases} \tag{18}$$

with $m = (z_{\text{phos}} - z^0_{\text{phos}})/(R_1 - R_2)$ and $b = (z^0_{\text{phos}} R_1 - z_{\text{phos}} R_2)/(R_1 - R_2)$. Assuming that the simulation box is a cylinder, we have $R_3 = \sqrt{NA_{\text{L}}/\pi}$. z_{phos} can be measured directly from simulation trajectories. z^0_{phos} is a half of the average phosphorus-phosphorus distance in a DOPC simulation, which can be easily obtained from the SIMtoEXP program. The average height profile over the monolayer, $\langle h(r) \rangle$, can be also obtained from the program in the same manner. The only unknown is R_2, which is the relaxation length of DOPC-Tat interaction.

Let us calculate $\langle h(r)\rangle$. In the cylindrical coodinates,

$$\langle h(r)\rangle = \frac{1}{\pi R_3^2}\int_0^{2\pi} d\phi \int_0^{R_3} dr\, rh(r) \tag{19}$$

The ϕ integration is trivial. The r integration is

$$\begin{aligned}
&\int_0^{R_3} dr\, rh(r) \\
&= \int_0^{R_1} dr\, z_{\rm phos} r + \int_{R_1}^{R_2} dr(mr+b)r + \int_{R_2}^{R_3} dr\, z^0_{\rm phos} r \\
&= \frac{1}{2}\left[z_{\rm phos}R_1^2 + z^0_{\rm phos}(R_3^2 - R_2^2)\right] + \frac{1}{3}m\left(R_2^3 - R_1^3\right) + \frac{1}{2}b\left(R_2^2 - R_1^2\right) \\
&= \frac{1}{2}\left[z_{\rm phos}R_1^2 + z^0_{\rm phos}(R_3^2 - R_2^2)\right] + \frac{1}{3}\left(z^0_{\rm phos} - z_{\rm phos}\right)\left(R_2^2 + R_1R_2 + R_1^2\right) \\
&+ \frac{1}{2}\left(z_{\rm phos}R_2 - z^0_{\rm phos}R_1\right)(R_1 + R_2)
\end{aligned} \tag{20}$$

Using Eq. (20), we get

$$\langle h(r)\rangle = \frac{\left(z_{\rm phos} - z^0_{\rm phos}\right)\left(R_1^2 + R_1R_2 + R_2^2\right) + 3z^0_{\rm phos}R_3^2}{3R_3^2} \tag{21}$$

Eq. 21 is a quadratic equation in terms of R_2. Solving for R_2 gives

$$R_2 = \frac{-R_1 + \sqrt{R_1^2 + 4C}}{2} \tag{22}$$

with

$$C = \frac{3R_3^2\left(z^0_{\rm phos} - \langle h(r)\rangle\right)}{z^0_{\rm phos} - z_{\rm phos}} - R_1^2 \tag{23}$$

The model quantities calculated from the MD simulations are shown in Table.

5 Some Lipid Bilayer Data

We also estimated the structure by fitting the experimental form factors to a model using the SDP method with the component groups identified in Fig. (what The positions of these groups were free parameters and the agreement with the experimental form factors was excellent (see Fig. S.M. 5). Absolute

total electron density profiles and the Tat profiles are shown for many samples in Fig. 6 (A-C). It must be emphasized, however, that, while the total EDP is well determined by this fitting procedure. Indeed, there are local minima in the fitting landscape, including one with Tat closer to the center of the bilayer as shown in Fig. S.M. 5. The simulations help to discard that result. For the results shown in Fig. 6, a consistent trend is that Tat moves away from the bilayer center as concentration increases.

Figure shows that A_L as defined by $(V_L - V_{HL})/D_c$ decreases as the amount of DOPE increases for systems without Tat. This is consistent with the previous studies (or predictions?) and attributed to the small size of PE head group. Because DOPE has smaller head group than DOPC, lipids in DOPC/DOPE bilayers pack more compactly than DOPC bilayers do, leading to a smaller A_L. Consequently, bilayers composed of DOPC and DOPE tend to have a higher order parameter than DOPC alone. (NO THEY DON'T. WHAT'S GOING ON HERE?). Similarly, the thickness of bilayers is larger at higher PE content.

Figure shows that Tat is located further out from the bilayer center with higher content of PE lipids. This is also consistent with MD simulation PMF, which showed that arginine insertion cost more energy for PE membrane than PC membrane, the result of which was attributed to more possible hydrogen bonding between PE group and arginines.

More structural detail from the modeling and from the simulations is shown in Fig. 7. The bilayer thickness can be described as DHH, which is the distance between the maxima in the electron density profile, or as DPP, which is the distance between the phosphocholines on the opposing monolayers. Figs. 7A and 7B show that both these quantities decrease with increasing Tat mole fraction (P/(L+P)), showing that Tat thins membranes, increasingly so as its concentration is increased, even though both simulation and modeling suggest that Tat moves further from the membrane center with increasing concentration as shown in Fig. 7D. Fig. 7C shows that the area per lipid AL usually increases with increasing mole fraction of Tat, as would be expected from consideration of conservation of lipid volume. Interestingly, the bilayer thickness did not increase for DOPC/DOPE (3:1) bilayers with x less than 0.03.

Figure 8 shows that the Sxray orientational order parameters generally decreases with increasing concentration of Tat for most of the membrane mimics studied. These decreases in membrane chain order are compatible with the increase in softening of membranes by Tat observed by a decrease

in the bending energy, KC, in Fig. 2.

References

[1] Norbert Kuerka, John F. Nagle, Jonathan N. Sachs, Scott E. Feller, Jeremy Pencer, Andrew Jackson, and John Katsaras. Lipid bilayer structure determined by the simultaneous analysis of neutron and x-ray scattering data. *Biophysical Journal*, 95(5):2356 – 2367, 2008.

[2] Anthony R. Braun, Jonathan N. Sachs, and John F. Nagle. Comparing simulations of lipid bilayers to scattering data: The gromos 43a1-s3 force field. *The Journal of Physical Chemistry B*, 117(17):5065–5072, 2013.

[3] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[4] Joakim P. M. Jmbeck and Alexander P. Lyubartsev. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *The Journal of Physical Chemistry B*, 116(10):3164–3179, 2012.

[5] Joakim P. M. Jmbeck and Alexander P. Lyubartsev. An extension and further validation of an all-atomistic force field for biological membranes. *Journal of Chemical Theory and Computation*, 8(8):2938–2948, 2012.

[6] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.