

Measurement of the Most Important Physics Quantity Ever Discovered

by

Kiyotaka Akabori

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
Department of Physics
Pittsburgh, Pennsylvania

Advised by Professor John F. Nagle and Stephanie Tristram-Nagle

June 14, 2014

Contents

1	Introduction	1
1.1	HIV life cycle	1
1.2	Overview of Tat	2
2	Materials and Methods	5
2.1	X-ray optics	5
2.2	Hydration Chamber	5
2.3	Sample Preparation	5
2.3.1	Stock Solutions	5
2.3.2	Thin Film Samples	6
2.4	CCD detector	7
3	Structural Perturbation on Lipid Bilayers Due to Tat Peptide	8
3.1	Introduction	9
3.2	Materials and Methods	12
3.2.1	Volume Measurement	12
3.2.2	Analysis of Diffuse Scattering	13
3.2.3	Modeling the Bilayer Structure	15
3.2.4	Molecular Dynamics Simulation	23
3.3	Analysis of Molecular Dynamics Simulation Data	24
3.3.1	SIMtoEXP program	24
3.3.2	Local Thinning of Membranes	24
3.3.3	Lateral Decay Length of Membrane Thinning	26
3.4	Results	28
3.4.1	Bending and Bulk Modulus	28
3.4.2	Volume results	28

3.4.3	Electron Density Profile Modeling	29
3.4.4	Hard Wall Constraint Fits	38
3.4.5	Summary of Electron Density Profile Modeling	38
3.4.6	Molecular Dynamics Simulations	41
3.5	Discussion	49
3.6	Conclusion	53
4	Ripple Phase	46
4.1	Introduction	46
4.2	Materials and Methods	47
4.2.1	Sample Preparation	48
4.2.2	Low Angle X-ray Scattering Experiment	49
4.2.3	Near Grazing Incident Wide Angle X-ray Scattering Experiment	49
4.2.4	Transmission Wide Angle X-ray Scattering Experiment	50
4.3	Some Theories	50
4.3.1	Lattice Structure	50
4.3.2	Sample q -space	51
4.3.3	Geometric (Lorentz) Correction	52
4.3.4	Absorption Correction for LAXS	56
4.3.5	Absorption Correction for WAXS	59
4.4	Model	59
4.4.1	Contour Part of the Form Factor	59
4.4.2	Transbilayer Part of the Form Factor	60
4.5	Results	61
4.5.1	Data	61
4.5.2	Electron Density Profile	62
4.5.3	Near Grazing Incident Wide Angle X-ray Scattering (NGIWAXS)	63
4.5.4	Transmission WAXS	63
4.6	Discussion	63
4.7	Conclusion	63
Appendices		65
A Tat		62
A.1	Fixed Angle Analysis of NFIT	62

A.2	Proper Incorporation of Mosaic Spread to NFIT analysis	62
A.3	Domain Size Distribution: Gaussian and Exponential	62
A.4	Hard Wall Constraints on SDP	62
A.5	Nonsymmetrized Profiles of MD	62
B	Ripple Phase	63
B.1	Rotation of a Two-Dimensional Function	63
B.2	Derivation of the transbilayer part of the form factor in the 2G hybrid model	64
B.3	Derivation of the contour part of the form factor	66
B.4	Correction due to refractive index	68

List of Tables

3.1	Some Amino Acids Data	12
3.2	DOPC basic structural parameters. n_i^e and ρ_i are the number of electrons and average electron density per component, respectively.	20
3.3	DOPE basic structural parameters. The notations are the same as in Table 3.2.	20
3.4	DOPC:DOPE (3:1) basic structural parameters. The notations are the same as in Table 3.2.	20
3.5	Tat basic structural parameters. The notations are the same as in Table 3.2.	21
3.6	Volumetric constraints. A and B refer to two different models described in the text.	21
3.7	Important Quantities for Tat Peptide	29
3.8	Measured Quantities in	29
3.9	Fitting Results for DOPC membranes for the THG model. $\Delta z_1 = z_{PC} - z_{CG}$ and $\Delta z_2 = z_{CG} - z_{HC}$	32
3.10	Fitting Results for DOPC:DOPE (3:1) membranes for the THG model. $\Delta z_1 = z_{PC} - z_{CG}$ and $\Delta z_2 = z_{CG} - z_{HC}$	33
3.11	Fitting Results for DOPC:DOPE (1:1) membranes for the THG model. $\Delta z_1 = z_{PC} - z_{CG}$ and $\Delta z_2 = z_{CG} - z_{HC}$. (Need to work)	36
3.12	Fitting Results of the bound THG model for DOPC membranes. $\Delta z_1 = z_{PC} - z_{CG}$ and $\Delta z_2 = z_{CG} - z_{HC}$	40
3.13	Comparison of the simulated form factors to the experimental form factors.	46

3.14	Summary of simulation results. $\langle D_{\text{PP}} \rangle$, phosphorus-phosphorus distance averaged over all lipids; D_{PP} , Tat-perturbed phosphorus atoms; x , thickness away from Tat; Δt , $\langle D_{\text{PP}}^{\text{DOPC}} \rangle - D_{\text{PP}}$; H_{Tat} , Tat height; R_{Tat} , radius of Tat cylinder; R_2 , radius of the calculated in-plane Tat-perturbed region; R_3 , effective radius of the simulation box.	47
3.15	Summary of weighted average results. The caption is the same as Table 3.14.	47
4.1	Definitions of Z_{CH_2} and Z_{W}	61
4.2	Lattice constants	62

List of Figures

3.1	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, phosphate-choline and carbonyl-glycerol. The hydrocarbon chain region is also divided into two components, methelence+methine and terminal methyl groups.	16
3.2	A model electron density profile for DOPC with Tat.	17
3.3	test	25
3.4	test	26
3.5	The best fits to DOPC form factors (left) and the corresponding electron density profiles (right) with $x_{\text{Tat}} = 0, 0.016, 0.034,$ and 0.059 (from top to bottom).	31
3.6	The best fits to DOPC:DOPE (3:1) form factors (left) and the corresponding electron density profiles (right) with $x_{\text{Tat}} = 0, 0.016, 0.034,$ and 0.059 (from top to bottom).	34
3.7	The best fits to DOPC:DOPE (1:1) form factors (left) and the corresponding electron density profiles (right) with $x_{\text{Tat}} = 0, 0.016, 0.034,$ and 0.059 (from top to bottom).	35
3.8	χ^2 as a function of z_{Tat} for DOPC, DOPC:DOPE (3:1), and DOPC:DOPE (1:1) (from left to right) with $x_{\text{Tat}} = 0.016, 0.034,$ and 0.059 (from top to bottom). $\sigma_{\text{Tat}} = 3.0$. The THG model (black squares) and the THC model (red circles).	37
3.9	χ^2 as a function of z_{Tat} for DOPC with $x_{\text{Tat}} = 0.016, 0.034,$ and 0.059 (from top to bottom). $\sigma_{\text{Tat}} = 3.0$. The bound THG model was used. .	39
3.10	DPP graph	41
3.11	DHH graph	41
3.12	AL graph	41

3.13 zTat graph	41
3.14 MD simulated form factors for DOPC at $A_L = 68 \text{ \AA}^2$ (blue solid line), 70 \AA^2 (red solid line), and 72 \AA^2 (green solid line) compared to the experimental form factor (open circles) scaled vertically to best match the form factor for 70 \AA^2	42
3.15 The simulated, symmetrized electron density profile for DOPC at $A_L =$ 70 \AA^2 as a function of the distance away from the bilayer center. Each component profile is labeled with its name: PC (phosphate-choline), CG (carbonyl-glycerol), CH ₂ +CH (methylene-methine combination), CH ₃ (terminal methyl). The sum of all the components is labeled as total.	43
3.16 MD simulated form factors for DOPC with $x_{\text{Tat}} = 0.015$ at $A_L =$ 72 \AA^2 (top) and 74 \AA^2 (bottom), with $z_{\text{Tat}} = 18 \text{ \AA}$ (red solid lines), 16 \AA (green solid lines), and 14 \AA (blue solid lines) compared to the experimental form factor (open circles) scaled vertically to best match the form factor for $z_{\text{Tat}} = 18 \text{ \AA}$	44
3.17 MD simulated form factors for DOPC with $x_{\text{Tat}} = 0.030$ at $A_L =$ 74 \AA^2 (top) and 76 \AA^2 (bottom), with $z_{\text{Tat}} = 18 \text{ \AA}$ (red solid lines), 16 \AA (green solid lines), and 14 \AA (blue solid lines) compared to the experimental form factor (open circles) scaled vertically to best match the form factor for $z_{\text{Tat}} = 18 \text{ \AA}$	45
3.18 Electron density profiles of guanidinium groups from the four best matched simulations for DOPC with $x_{\text{Tat}} = 0.015$ (one Tat on each leaflet). Tat on the lower and upper leaflets are shown on the left and right plots, respectively.	48
4.1 caption goes here	53

4.2	q -space representations of Bragg peaks and Bragg rings for $h = 1$ and $k = 0, 1$, and 2 in q_{hk}^z planes. The shaded rectangles show cross sections of the rotating Ewald sphere along q_{hk}^z plane. The intersection between the Ewald sphere and a Bragg peak/ring is indicated in red. The observed intensity for the $k \neq 0$ orders is proportional to the fraction of the length of red arcs in the circumference. This fraction is equal to one for a $k = 0$ order. Because the orders are not in the same q_z plane, the range of q_y integration indicated by the height of the rectangle is different for different orders. The magnitude of curvature of arcs is exaggerated.	54
4.3	The path of X-rays within the sample. The incident angle is ω and the total scattering angle is 2θ . An X-ray with a penetration depth of z is shown. The total thickness of the sample is t	57

Chapter 3

Structural Perturbation on Lipid Bilayers Due to Tat Peptide

As discussed in chapter 2, the two main techniques employed in this thesis were molecular dynamics (MD) simulation and low angle X-ray scattering (LAXS). First, we discuss the results and analysis of diffuse X-ray scattering. The general protocol was the following; LAXS data were fitted to a model X-ray scattering pattern from a stack of fluctuating membranes via NFIT program, the analysis of which yielded the bending modulus, K_C , and the bulk modulus, B . Dividing the experimental data by the model, then, gave the absolute X-ray form factor, $|F(q_z)|$, which is the Fourier transform of bilayer electron density profile along the bilayer normal direction, z . We fitted $|F(q_z)|$ to a model density profile using the scattering density profile (SDP) program. The SDP program allows us to model a bilayer density according to volumetric spacing constraint. The advantage of this program is that we can see fine details of bilayer structure such as an individual head group, terminal methyl, and so on. The model requires many parameters that are not so well determined. We then constrain many parameters from the past experimental data and MD simulations. This is discussed in section ?.

The second main method is MD simulation. From simulation trajectory, we calculated the so called simulated X-ray form factor using the SIMtoEXP program (ref). The best matching simulation result was chosen as the best prediction of the bilayer structure. We, then, calculated many structural details from the trajectory that were not accessible experimentally.

Section X discusses the implication of the results obtained in the proceeding sec-

tions. While this study does not probe dynamics of Tat translocation, it supports Tat's ability to interact with neutral membranes. This finding is compared with recent studies on a single arginine molecule.

3.1 Introduction

The name cell-penetrating peptide (CPP) connotes a peptide that easily penetrates cell membranes (for Reviews see [1-3]).

This thesis focuses on the transactivator of translation, Tat, from the HIV-1 virus, which plays a role in AIDS progression. Earlier work showed that the HIV-Tat protein (86 amino acids) was efficiently taken up by cells, and concentrations as low as 1 nM were sufficient to transactivate a reporter gene expressed from the HIV-1 promoter [4, 5]. It has been reported that Tat protein uptake does not require ATP [6]. Studies using inhibitors of different types of endocytosis, including clathrin and caveolae-mediated, or receptor-independent macropinocytosis reached the same conclusion that ATP mediated endocytosis is not involved in Tat protein permeation [7-10]. However, this issue is controversial, as other studies found evidence for endocytosis in Tat protein import [11-19]. Still other studies have concluded that an ATP requirement for Tat protein entry depends on the size of the cargo attached to Tat protein, or on the specific cell type [20-22]. The part of the Tat protein responsible for cellular uptake was assigned to a short region Tat (48-60), G48RKRRQRRRPPQ60, which is particularly rich in basic amino acids [6]. Deletion of three out of eight positive charges in this region caused loss of its ability to translocate [6]. In this thesis, short basic regions will be called Tat, while the entire 86- amino acid protein will be called Tat protein. Tat was shown to be responsible for the Tat proteins permeation into the cell nucleus and the nucleoli [6], and this was confirmed using live cell fluorescence in SVGA cells [23]. Tat (48-60) was shown to have little toxicity on HeLa cells at 100 M concentration [6], but the longer Tat protein (2-86) was toxic to rat brain glioma cells at 1-10 M [24]. Interestingly, no hemolytic activity was found when human erythrocytes were incubated with a highly neurotoxic concentration (40 M) of Tat (2-86) [24]. These results prompt the question, what is the mechanism of Tats translocation through membranes? To address this question, many biophysical studies have used simple models of biological membranes composed of a small number of lipid types. These studies are valuable because there is no possibility

for ATP-dependent translocation, thus ruling out endocytosis if translocation occurs. For example, Mishra et al. reported that the rate of entry into giant unilamellar vesicles (GUVs) composed of PS/PC (1:4 mole ratio) lipids of rhodamine-tagged Tat is immeasurably slow, but it crosses a GUV composed of PS/PC/PE (1:2:1) lipids within 30 seconds [25]. This study suggests that negative curvature induced by the inclusion of PE facilitates translocation. In a subsequent study using much smaller unilamellar vesicles (LUVs), Tat did not release an encapsulated fluorescent probe in LUVs composed of lipids modeling the outer plasma membrane, PC/PE/SM/Chol (1:1:1:1.5), but did release the probe in LUVs composed of BMP/PC/PE (77:19:4) [26]; BMP (bis(monoacylglycero)-phosphate) is an anionic lipid specific to late endosomes. In that study [26], the inclusion of PE did not suffice to cause leaky fusion in LUVs in the absence of a negatively charged lipid. The contrasting results in these two experiments may also be due to the use of LUVs instead of GUVs since it was reported that Tat does not translocate across LUVS of PC/PG (3:2) but does translocate across GUVs of the same lipid composition [27]. In a similar experiment, Tat did not translocate into egg PC LUVs [28]. In another experiment confirming these results, Tat did not translocate into GUVS containing only PC with 20 mol% cholesterol, but when PS or PE was included with PC, then rapid translocation of Tat was observed [29]. These experiments demonstrate that the choice of lipids and model systems influences Tat translocation.

Is a pore formed during Tat translocation? Although direct conductance measurements of Tat and lipid membranes have not been carried out, two studies measured conductance with the somewhat similar CPP oligoarginine R9C peptide. Using single-channel conductance of gramicidin A in planar lipid membranes consisting of anionic, neutral or positively charged lipids, R9C did not increase conductance, even in anionic lipid membranes [30]. By contrast, in a similar experiment using planar lipid membranes, a current was induced by R9C in PC/PG (3:1) membranes, with increasing destabilization over time [31]. Thus questions remain about pore formation of Tat in membranes. In the GUV experiment with Tat mentioned above [29], Ciobanasu et al., using size exclusion methods, suggested a pore in the nanometer range, which could only be passed by small dye tracer molecules. Thus, if a true pore forms, it is likely to be small and transitory.

The secondary structure of Tat have been characterized by many researchers. Ref.[27] carried out Circular dichroism (CD) spectroscopy on a variation of Tat where

the penultimate proline on Tat (48-60) was replaced by a tryptophan [27]. Their study found a random coil secondary structure in aqueous solution as well as when Tat was mixed with PC/PG/PE (65:35:5) LUVs. Ziegler et al.[10] obtained the same result using CD in PC/PG (3:1) vesicles. In addition, solid state NMR has identified a random coil structure of Tat in DMPC/DMPG (8:7 mole ratio) multibilayers [32]. In the larger Tat-(1-72)-protein NMR measurements at pH 4 have determined there is no secondary structure, with a dynamical basic region [33]. Similarly, NMR was used to study the full Tat protein and found a highly flexible basic region [34]. These previous studies indicate that an alpha helix is not required for Tats translocation ability.

Regarding the mechanism of translocation of this randomly structured, short basic peptide, many models have been proposed based on the conflicting results listed above. Molecular dynamics simulations offer some insight into the molecular details of translocation. Herce and Garcia simulated the translocation of Tat (Y47GRKKRRQRRR57) across DOPC at various lipid:peptide molar ratios [35]. Their simulations indicated that Tat binds to the phosphate headgroups, with 1 Tat binding with 14 lipids, each positive charge on Tat associated with nearly 2 phosphate groups [35]. Translocation involved a localized thinning, and snorkeling of arginine side chains through the hydrophobic layer to interact with phosphates on the other side of the membrane. This allowed some water molecules to penetrate the membrane along with Tat, forming a pore [35]. In this simulation, performed without inclusion of counterions, pore formation was only observed at high ratios of peptide:lipid (1:18) or at elevated temperature. However, a subsequent Gromacs simulation with counterions found no thinning and no pore formation when Tat was added to DOPC membranes [36]. Instead it found a membrane invagination associated with a cluster of Tat peptides. From their findings, the authors suggested that micropinocytosis could be the model for Tat translocation across membranes [36].

In this work we combine experimental low-angle X-ray scattering (LAXS) data with MD simulations to obtain the structure of fully hydrated, oriented lipid bilayers with Tat (47-57) added at several mole ratios. The lipid systems were DOPC, DOPC/DOPE (3:1 mole ratio), DOPC/DOPS (3:1), DOPC/DOPE (1:1) and a mimic of the nuclear membrane (POPC/POPE/POPS/SoyPI/Chol, 69:15:2:4:11). Accessory techniques, densitometry, wideangle X-ray scattering (WAXS), neutron scattering, CD spectroscopy were also applied to further characterize Tat/membrane inter-

actions.

3.2 Materials and Methods

3.2.1 Volume Measurement

Multilamellar vesicles (MLVs) were prepared by mixing dried lipid mixtures with MilliQ water to a final concentration of 2-5 wt% in nalgene vials and cycling three times between 20 °C and 60 °C for ten minutes at each temperature with vortexing. Pure Tat was dissolved in water at 0.4 wt%.

Volumes of lipid mixtures with and without peptides in fully hydrated multilamellar vesicles (MLV) were determined at 37 ± 0.01 °C using an Anton-Paar USA DMA5000M (Ashland, VA) vibrating tube densimeter [?]. This instrument measures the average density of a solution.

The Tat peptide sequence used in X-ray experiments and MD simulations was Y₄₇GRKKRRQRRR₅₇. Table 3.1 lists the chemical formulas and molecular weights of these amino acids for convenience. The molecular weight of this sequence is $181.2 + 75.1 + 146.1 + 2 \times 146.2 + 6 \times 174.2 - 10 \times 18 = 1560$. The Tat peptides were

Code	Amino acid	Chemical Formula	Molecular weight (g/mol)
K	Lysine	C ₆ H ₁₄ N ₂ O ₂	146.2
R	Arginine	C ₆ H ₁₄ N ₄ O ₂	174.2
G	Glycine	C ₂ H ₅ NO ₂	75.1
Y	Tyrosine	C ₉ H ₁₁ NO ₃	181.2
Q	Glutamine	C ₅ H ₁₀ N ₂ O ₃	146.1

Table 3.1: Some Amino Acids Data

synthesized in trifluoroacetic acid, which has the chemical formula CF₃CO₂H, and made into a powder form by the freeze-dry method. Therefore, each positively charged amino acid such as an arginine and lysine was counter-balanced by a trifluoroacetate (TFA) (C₂F₃O₂). Since Tat has six arginines and two lysines, it came with eight trifluoroacetates. This complex has a molecular weight of $1560 + 113 \times 8 = 2464$. We used the molecular weight of this complex in order to calculate the molarity of a Tat in water solution correctly.

The Tat volume V_{Tat} was calculated from the measured average density of a Tat-water solution in the following way. Assuming that Tat molecules in water does not

change the volume of water molecules, the density of Tat-water solution is equal to the mass of Tat-water solution divided by the sum of volumes of water and Tat,

$$\rho_{\text{sol}} = \frac{m_w + m_c}{V_w + V_c N_c}, \quad (3.1)$$

where m_w and m_c are the total mass of water and Tat-TFA complex, V_w is the total volume of water, V_c is the molecular volume of a Tat-TFA complex, and N_c is the total number of this complex in the solution. Denoting $V_w = m_w / \rho_w$ and $N_c = N_A m_c / W_c$, where W_c is the molecular weight of the complex, N_A is the Avogadro's number, and ρ_w is the density of water, we have

$$V_c = \frac{W_c}{\rho_{\text{sol}} N_A} \left(1 + \frac{m_w}{m_c} \left(1 - \frac{\rho_{\text{sol}}}{\rho_w} \right) \right), \quad (3.2)$$

which allows us to calculate the molecular volume of a Tat-TFA complex from the experimentally measured quantities. Assuming that the molecular volume scales with the molecular weight gives the volume of Tat, $V_{\text{Tat}} = 1560/2464 \times V_c \text{ \AA}^3$.

3.2.2 Analysis of Diffuse Scattering

During an X-ray exposure, the sample was continuously rotated. A different approach to diffuse scattering analysis is proposed in appendix A.1. An advantage of the alternative analysis is our experimental ability to collect very precise background scattering data by tilting the sample at the negative angle. While the results look promising, this method was not used for getting Tat perturbed bilayer structure.

Another proposed improvement in the analysis is to include mosaic spread properly. This is discussed in appendix A.2.

The analysis of diffuse X-ray scattering pattern begins with separating $|F(q_z)|$ from $S(\mathbf{q})$. To this end, we used an analysis program called NFIT. The derivation is described in Yufeng Liu's thesis in detail. In this section, we describe the theoretical model for $S(\mathbf{q})$ to outline the theory.

We assume that a stack of bilayers can be accurately described by the smectic liquid crystal theory, so that the free energy of the system is

$$F = \quad (3.3)$$

where K_C and B are the bending and bulk modulus, respectively. Writing the membrane height profile in terms of the Fourier modes, $u = \sum \exp(i\mathbf{q} \cdot \mathbf{r})$,

$$F \quad (3.4)$$

3.2.3 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 3.1. The phosphate/choline (PC) and carbonyl/glycerol (CG) components together make up the lipid headgroup whereas the hydrocarbon chain region is divided into two components, the methylene (CH_2) and methine (CH) group 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.

Functional forms

Our model for electron density profile (EDP) of Tat/lipid bilayer system consists of five structural subgroups: PC, CG, CH_2+CH , CH_3 , and Tat. (Fig. ??). The volume probability distributions of components PC, CG, CH_3 , and Tat are described by Gaussian functions,

$$P_i(z) = \frac{c_i}{\sqrt{2\pi}} \left(\exp\left\{-\frac{(z+z_i)^2}{2\sigma_i^2}\right\} + \exp\left\{-\frac{(z-z_i)^2}{2\sigma_i^2}\right\} \right), \quad (3.5)$$

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

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

$$P_{\text{HC}}(z) = \frac{1}{2} [\text{erf}(z, -z_{\text{HC}}, \sigma_{\text{HC}}) - \text{erf}(z, z_{\text{HC}}, \sigma_{\text{HC}})], \quad (3.6)$$

where

$$\text{erf}(z, z_i, \sigma_i) = \frac{2}{\sqrt{\pi}} \int_0^{\frac{z-z_i}{\sqrt{2}\sigma}} dx e^{-x^2}. \quad (3.7)$$

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

$$P_{\text{CH}_2+\text{CH}}(z) = P_{\text{HC}}(z) - P_{\text{CH}_3}(z). \quad (3.8)$$

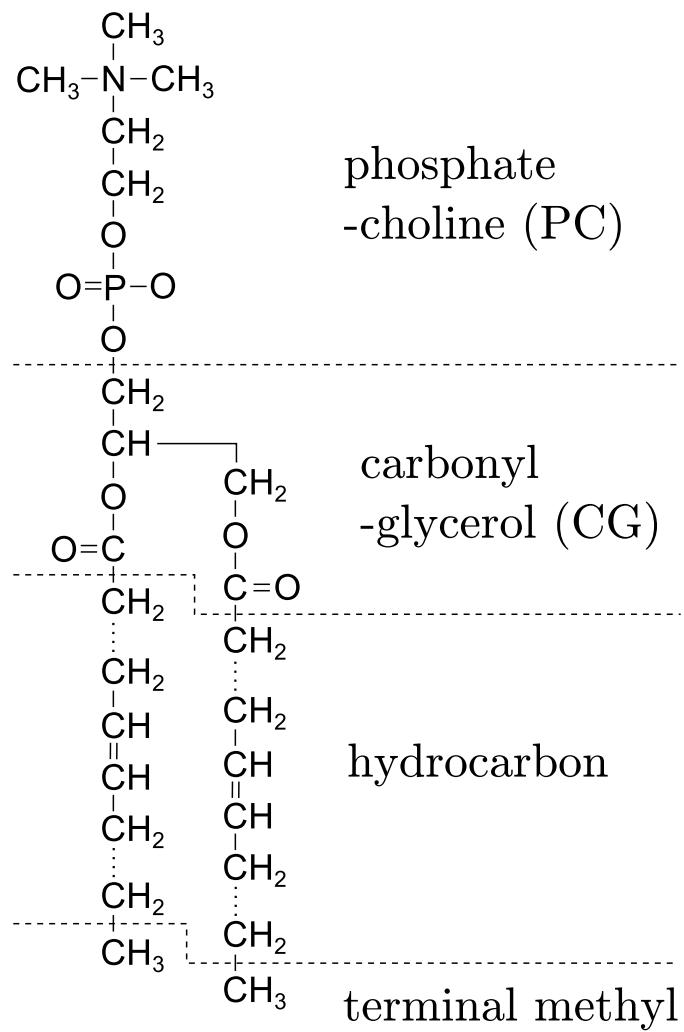


Figure 3.1: Schematic of DOPC showing each lipid component. The dash lines show where the lipid is divided into different components. The lipid headgroup is divided into two components, phosphate-choline and carbonyl-glycerol. The hydrocarbon chain region is also divided into two components, methylene+methine and terminal methyl groups.

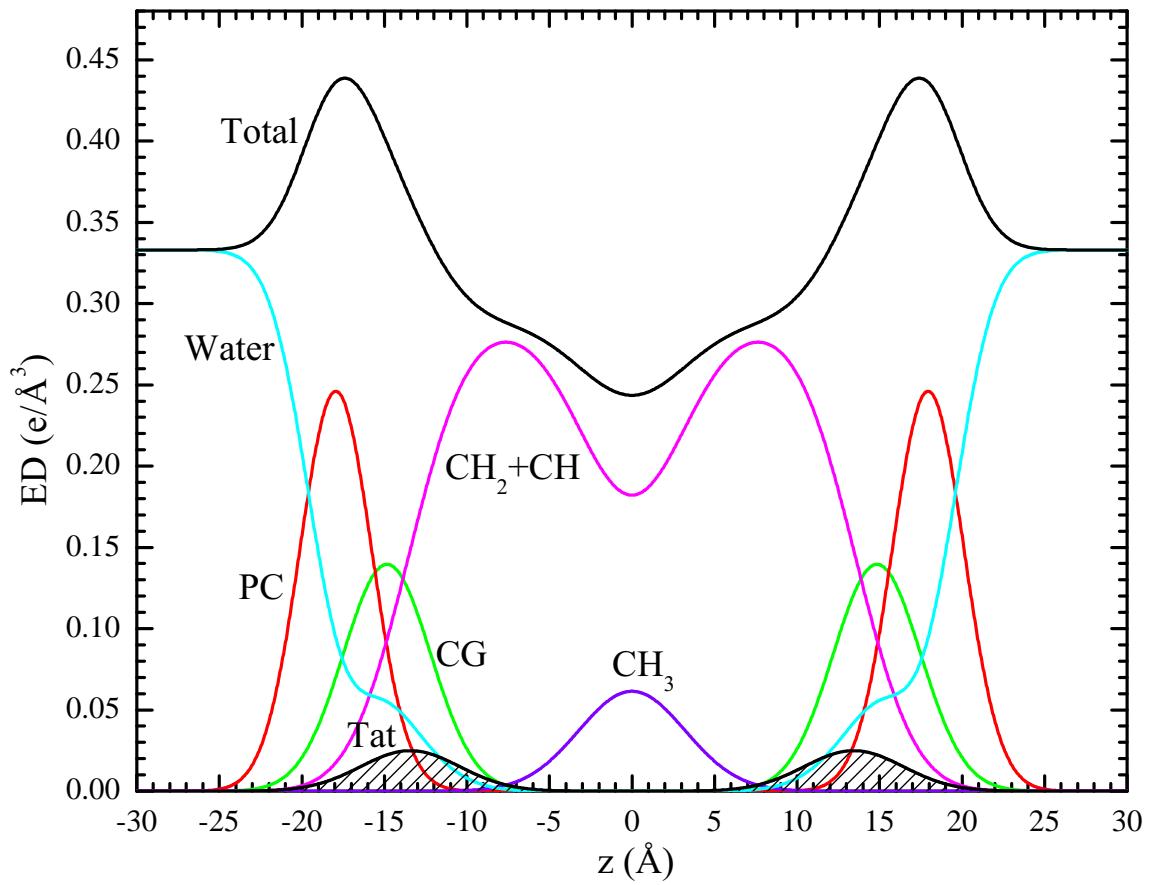


Figure 3.2: A model electron density profile for DOPC with Tat.

This definition enforces the total probability P_{HC} in the hydrocarbon chain region to equal one, which in turn means that placement of Tat in the chain region is prohibited. We call the model defined by Eq. (3.8) model A (better name?). To allow Tat to be placed inside the hydrocarbon chain region, we also consider an alternative definition,

$$P_{\text{CH}_2+\text{CH}}(z) = P_{\text{HC}}(z) - P_{\text{CH}_3}(z) - P_{\text{Tat}}(z), \quad (3.9)$$

where the volume probability of CH_2+CH combined component is reduced by the Tat volume probability distribution. We call this model B. The spatial conservation requires the water volume probability distribution to be

$$P_{\text{W}}(z) = 1 - P_{\text{PC}}(z) - P_{\text{CG}}(z) - P_{\text{Tat}}(z) - P_{\text{HC}}(z) \quad (3.10)$$

for model A and

$$P_{\text{W}}(z) = 1 - P_{\text{PC}}(z) - P_{\text{CG}}(z) - P_{\text{HC}}(z) \quad (3.11)$$

for model B.

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,

$$F(q_z) = 2 \int_0^{\frac{D}{2}} dz \left(\sum_i (\rho_i - \rho_{\text{W}}) P_i(z) \right) \cos(q_z z), \quad (3.12)$$

where $i = \text{PC}, \text{CG}, \text{Tat}, \text{CH}+\text{CH}_2$, and CH_3 .

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 $z_{\text{CH}_3} = 0$ 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 \AA^3 for gel phase phosphatidylcholine bilayers [2], and we assume the same volume for the fluid phase phosphatidylcholine bilayers. The volumes of PC and CG components satisfy

$$V_{\text{PC}} + V_{\text{CG}} = V_{\text{HL}}, \quad (3.13)$$

and the volumes of CH_3 and CH_2+CH components satisfy

$$2(16V_{\text{CH}_2+\text{CH}} + V_{\text{CH}_3}) = V_L - V_{\text{HL}}. \quad (3.14)$$

These component volumes constrain the height of the Gaussians as

$$c_{\text{PC}} = \frac{V_{\text{PC}}}{A_L \sigma_{\text{PC}}} \quad (3.15)$$

$$c_{\text{CG}} = \frac{V_{\text{CG}}}{A_L \sigma_{\text{CG}}} \quad (3.16)$$

$$c_{\text{CH}_3} = \frac{2V_{\text{CH}_3}}{A_L \sigma_{\text{CH}_3}} \quad (3.17)$$

$$c_{\text{Tat}} = \frac{V_{\text{Tat}}}{A_L \sigma_{\text{Tat}}} \quad (3.18)$$

where A_L is area per lipid.

The ratio of the carbonyl/glycerol volume to the headgroup volume V_{HL} was reported to be 0.41 [3], so we constrain the CG component volume to 135.7 \AA^3 and the PC component volume to 195.3 \AA^3 .

The most detailed structural study on DOPC to date was published by Braun *et al.* [3], and many of constraints on our model parameters can be derived from their study. However, in that work, the authors used the SDP model [1], which is specifically tailored for combined analysis of neutron and X-ray form factors. Therefore, we need to convert their structural results to the corresponding parameters in our simpler model. For example, from the reported values of the ratio of the volumes of the chain terminal methyl (CH_3) to the chain methylenes (CH_2) and the ratio of the volumes of the chain methines (CH) to the chain methylenes, we can calculate the ratio r_{CH_3} of the volumes of CH_3 to the CH_2 and CH combined component. Furthermore, the study by Braun *et al.* was at 30°C while our study was at 37°C , so our measured volume of DOPC was slightly higher.

At 30°C , the volume of DOPC was reported to be 1303 \AA^3 , so the volume of hydrocarbon chain region at the same temperature is $1303 - 331 = 972 \text{ \AA}^3$. The ratio r of the volumes of the chain terminal methyl (CH_3) to the chain methylenes (CH_2) was reported to be 1.95, and the ratio r_{12} of the volumes of the chain methines (CH) to the chain methylenes 0.91 at 30°C . Because there are 14 CH_2 groups, 2 CH groups, and 1 CH_3 group in each DOPC hydrocarbon chain, we have $2 \times (14V_{\text{CH}_2} +$

$2V_{\text{CH}} + V_{\text{CH}_3}) = 972 \text{ \AA}^3$. Using $V_{\text{CH}_3}/V_{\text{CH}_2} = 1.95$ and $V_{\text{CH}}/V_{\text{CH}_2} = 0.91$, we get $V_{\text{CH}_2} = 27.3 \text{ \AA}^3$, $V_{\text{CH}} = 24.9 \text{ \AA}^3$, and $V_{\text{CH}_3} = 53.3 \text{ \AA}^3$. These calculated volumes lead to $V_{\text{CH}_3}/V_{\text{CH}_2+\text{CH}} = 1.97$ for 30°C .

At 37°C , the volume of DOPC was measured to be 1313.5 \AA^3 , so we have $2 \times (16V_{\text{CH}_2+\text{CH}} + V_{\text{CH}_3}) = 1313.5 - 331$. Assuming that the ratio $V_{\text{CH}_3}/V_{\text{CH}_2+\text{CH}}$ at 37°C is the same as that at 30°C gives $V_{\text{CH}_2+\text{CH}} = 27.3 \text{ \AA}^3$ and $V_{\text{CH}_3} = 53.9 \text{ \AA}^3$. We constrain the components for the hydrocarbon chain region in our model to these calculated values.

		component	n_i^e	$V_i (\text{\AA}^3)$	$\rho_i (\text{e}/\text{\AA}^3)$
number of e/lipid	434	PC	97	195.3	0.497
volume/lipid (\AA^3)	1313.5	CG	67	135.7	0.494
		CH_2+CH	7.875	27.3	0.288
		CH_3	9	53.9	0.167

Table 3.2: DOPC basic structural parameters. n_i^e and ρ_i are the number of electrons and average electron density per component, respectively.

		component	n_i^e	$V_i (\text{\AA}^3)$	$\rho_i (\text{e}/\text{\AA}^3)$
number of e/lipid	410	PE	73	94.1	0.776
volume/lipid (\AA^3)	1212.3	CG	67	135.7	0.494
		CH_2+CH	7.875	27.3	0.288
		CH_3	9	53.9	0.167

Table 3.3: DOPE basic structural parameters. The notations are the same as in Table 3.2.

		component	n_i^e	$V_i (\text{\AA}^3)$	$\rho_i (\text{e}/\text{\AA}^3)$
number of e/lipid	428	PC/PE	91	170	0.535
volume/lipid (\AA^3)	1288.2	CG	67	135.7	0.494
		CH_2+CH	7.875	27.3	0.288
		CH_3	9	53.9	0.167

Table 3.4: DOPC:DOPE (3:1) basic structural parameters. The notations are the same as in Table 3.2.

		ratio	n_{Tat}^e	V_{Tat} (\AA^3)
number of e/Tat	838	62:1	13.6	30.5
volume/Tat (\AA^3)	1877	28:1	29.5	66.1
ρ_{Tat} (e/ \AA^3)	0.446	16:1	53.0	118.8

Table 3.5: Tat basic structural parameters. The notations are the same as in Table 3.2.

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

Table 3.6: Volumetric constraints. A and B refer to two different models described in the text.

Fits with Lower Bounds

Non-linear least squared fits with upper and lower bounds for the model parameters are implemented using an internal-external parameter transformation method. This method is described in MINUIT User's Guide, section 1.3 [4]. This section briefly describes the method. The details can be found in the MINUIT website [5].

Basically, instead of a model parameter, which is also called the external variable, the minimization procedure varies a related variable called the internal variable. This internal variable can take any values between $-\infty$ to $+\infty$. At every χ^2 calculation, the internal variable is transformed to the external variable, which can take values only between the lower and upper bounds (a and b). This non-linear transformation allows an existing minimization algorithm that was developed for fits with no bounds to work for fits with bounds. This point was important because it allowed us to implement bound fits in the model fitting program called the SDP program, fully developed by Dr. Norbert Kucerka, without too many additional changes. Downsides of the transformation method include turning a linear problem into a non-linear one and some computational overhead, neither of which is particularly problematic in this study.

For variables with both lower and upper bounds (a and b , respectively), the transformation between the internal and external variables is

$$P_{\text{int}} = \arcsin\left(2\frac{P_{\text{ext}} - a}{b - a} - 1\right) \quad (3.19)$$

$$P_{\text{ext}} = a + \frac{b - a}{2}(\sin P_{\text{int}} + 1). \quad (3.20)$$

For variables with a lower bound a only, the transformation is

$$P_{\text{int}} = \sqrt{(P_{\text{ext}} - a + 1)^2 - 1} \quad (3.21)$$

$$P_{\text{ext}} = a - 1 + \sqrt{P_{\text{int}}^2 + 1}, \quad (3.22)$$

and for variables with an upper bound b only,

$$P_{\text{int}} = \sqrt{(b - P_{\text{ext}} + 1)^2 - 1} \quad (3.23)$$

$$P_{\text{ext}} = b + 1 - \sqrt{P_{\text{int}}^2 + 1}. \quad (3.24)$$

3.2.4 Molecular Dynamics Simulation

This section describes the MD simulations performed by Dr. Kun Huang. Systems with different DOPC/Tat mole ratios (128:0, 128:2 and 128:4, corresponding to 0, 0.015 and 0.030 mole fractions) were simulated atomistically using the Gromacs 4.6.1 package [6]. DOPC was modeled by the Slipid force field [7, 8] and HIV Tat was modeled by Amber 99SB [9]. Tip3p water was used [53]. The number of Tats was divided equally on each side of the bilayer to mimic experimental conditions. All systems were simulated at 310 K with a constant area in the x - y plane and 1 atm constant pressure in the z direction. Each system was simulated for 100 ns and the last 50 ns was used as the production run. At each DOPC/Tat mole ratio, we studied systems with three different area/lipid (A_L). For the DOPC system, we fixed $A_L = 68, 70, 72 \text{ \AA}^2$; DOPC/Tat (128:2), we fixed the $A_L = 72, 74, 76 \text{ \AA}^2$; DOPC/Tat (128:4), we fixed the $A_L = 72, 74, 76 \text{ \AA}^2$. For each DOPC/Tat system at fixed A_L , we then conducted seven independent simulations with the center of mass (COM) of each Tat constrained at different bilayer depths from the bilayer center (18, 16, 14, 12, 10, 8 and 5 \AA). In total, 45 independent simulations were conducted. The goal of constrained simulations is to find the best match between experimental and MD simulation form factors. Comparison to the X-ray form factors was performed using the SIMtoEXP software [54]. Additional details concerning the MD simulations are in Supplementary Data 6.

The center of mass (COM) distance between each peptide and the bilayer was constrained by an umbrella potential with a force constant k of 3000 kJ/mol/nm². Essentially, this potential acts as a spring, where its potential energy depends on the deviation of the distance between the center of mass of Tat and DOPC from a preferred value, z_0 ,

$$U(z_1^{\text{Tat}}, \dots, z_1^{\text{DOPC}}, \dots) = -\frac{1}{2}k(z_{\text{cm}}^{\text{Tat}} - z_{\text{cm}}^{\text{DOPC}} - z_0)^2.$$

Then, $-\partial U / \partial z_i$ is 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. With this COM constraint, each Tat was allowed to move laterally and rotate, changing its configuration during simulations.

3.3 Analysis of Molecular Dynamics Simulation Data

3.3.1 SIMtoEXP program

This section briefly describes the SIMtoEXP program developed by Dr. Norbert Kucerka [10]. Essentially, for each snapshot, positional distribution of each atom averaged over the xy plane is calculated. Then, the distribution is averaged over snapshots. The product of this distribution and the average electron density gives the electron density profile of the atom. The sum of all the atoms provide the total electron density profile. This total electron density profile minus the average electron density of water is Fourier transformed to provide the X-ray form factor.

$$F^{\text{sim}}(q_z) = \int_0^{\infty} dz (\rho(z) - \rho_W) \cos(q_z z). \quad (3.25)$$

Electron density profiles were symmetrized and then X-ray form factors were calculated with $\rho_W = 0.326$. Because $\rho(z)$ is equal to ρ_W outside the bilayer, the upper integration limit takes on a finite value. Since the total electron density profile is symmetric with respect to the bilayer center, the simulated form factor F_{sim} is real.

Because the experimental form factor is on arbitrary units, it is scaled by a single constant a to produce the best fit to the simulated form factor through a linear least squared fit,

$$\chi^2 = \sum_i \left(\frac{1}{\sigma_i} (a|F_i^{\text{exp}}| - |F^{\text{sim}}(q_{z,i})|) \right)^2 \quad (3.26)$$

where σ is the input experimental uncertainty and F_{exp} is the experimental form factor. Because the SIMtoEXP program does not scale the input uncertainties, the relative errors ($\sigma / |F|$) depend on the overall scaling factor by which the experimental form factor is multiplied to produce the best fit of an experimental form factor to a simulated one. Consequently, The χ^2 values calculated by the program were multiplied by $1 / a^2$. These corrected χ^2 are reported in this chapter.

3.3.2 Local Thinning of Membranes

The SIMtoEXP program only gives the average quantities for each leaflet. While our x-ray data are sensitive to the bilayer average 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 around the Tat peptides.

The presence of Tat may result in compression of lipid bilayer along z -direction. If so, the phosphorus-phosphorus distance D'_{PP} of the bilayer near Tat may be different from the distance D_{PP} away from Tat. For small Tat concentration, D_{PP} would be the same as that of pure DOPC if the distance from all Tats is large enough. 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, and D'_{PP} is what we want to measure.

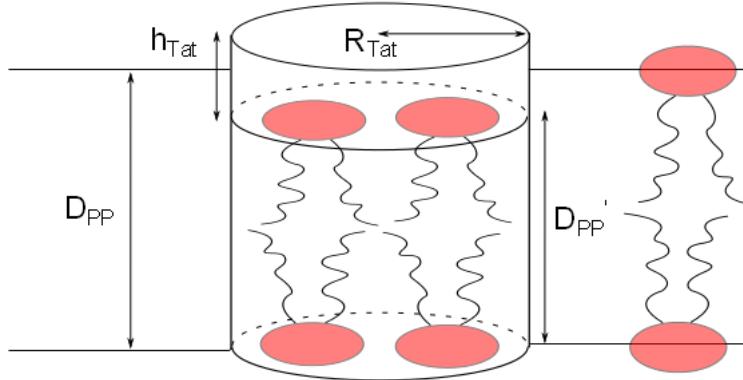


Figure 3.3: test

First, let us define what we mean by lipids close to Tat. As in Fig. 3.3, 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 $R_{\text{Tat}} = 9 \text{ \AA}$ comes from the volume of Tat = 1876 \AA^3 and $H_{\text{Tat}} = 7.6 \text{ \AA}$ 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 the middle of the amino acid sequence. Then let us define $D_{\text{PP}'}$ using only those lipids whose phosphorus atoms lie within these 9 \AA cylinders around the Tats. Then $D_{\text{PP}} = z_{\text{phos}}^+ - z_{\text{phos}}^-$ 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, 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. We averaged over many snapshots to gain better statistics.

3.3.3 Lateral Decay Length of Membrane Thinning

This section describes a method to measure the lateral decay length of membrane thinning due to Tat-lipid interactions. As in the previous section, Tat is modeled here as a cylinder with its radius equal to R_1 , height H_{Tat} , and volume V_{Tat} such that $R_1 = \sqrt{V_{\text{Tat}}/(\pi H_{\text{Tat}})}$. Let $h(r)$ represent the phosphorus height profile of a leaflet. The two leaflets are assumed to be decoupled. In our model, lipids are separated into three regions: suppressed, boundary, and unperturbed region . The suppressed region

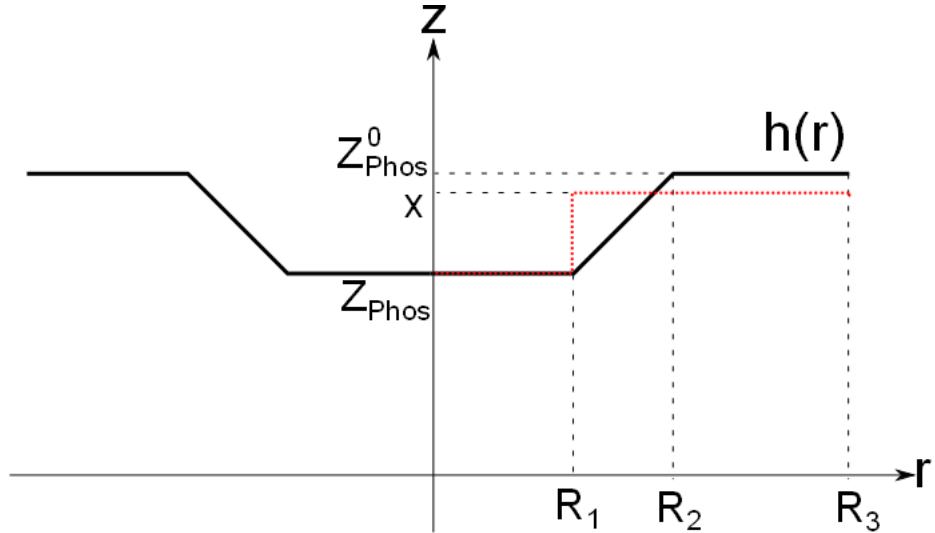


Figure 3.4: test

extends from $r = 0$ to R_1 and is directly beneath (above) Tat in the top (bottom) leaflet. In this region, lipids are uniformly compressed by Tat toward the center of the bilayer, so that $h(r)$ is a constant equal to z_{phos} . From $r = R_1$ to R_2 is the boundary region, where $h(r)$ is assumed to linearly increase with the lateral distance r . The lateral decay length of membrane thinning is given by $R_2 - R_1$. In the unperturbed

region ($r > R_3$), lipids do not interact with Tat, behaving identically to DOPC, so the phosphorus position is the same as that of DOPC. A continuous $h(r)$ that satisfies the above criteria is

$$h(r) = \begin{cases} z_{\text{phos}} & \text{if } 0 \leq r < R_1 \\ mr + b & \text{if } R_1 \leq r < R_2 \\ z_{\text{phos}}^0 & \text{if } R_2 \leq r < R_3 \end{cases} \quad (3.27)$$

with $m = (z_{\text{phos}} - z_{\text{phos}}^0)/(R_1 - R_2)$ and $b = (z_{\text{phos}}^0 R_1 - z_{\text{phos}} R_2)/(R_1 - R_2)$. Assuming that the simulation box is a cylinder gives $R_3 = \sqrt{NA_L/\pi}$. z_{phos} can be measured directly from simulation trajectories. z_{phos}^0 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 .

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

$$\langle h(r) \rangle = \frac{1}{\pi R_3^2} \int_0^{2\pi} d\phi \int_0^{R_3} dr r h(r) \quad (3.28)$$

The ϕ integration is trivial. The r integration is

$$\begin{aligned} & \int_0^{R_3} dr r h(r) \\ &= \int_0^{R_1} dr z_{\text{phos}} r + \int_{R_1}^{R_2} dr (mr + b)r + \int_{R_2}^{R_3} dr z_{\text{phos}}^0 r \\ &= \frac{1}{2} [z_{\text{phos}} R_1^2 + z_{\text{phos}}^0 (R_3^2 - R_2^2)] + \frac{1}{3} m (R_2^3 - R_1^3) + \frac{1}{2} b (R_2^2 - R_1^2) \\ &= \frac{1}{2} [z_{\text{phos}} R_1^2 + z_{\text{phos}}^0 (R_3^2 - R_2^2)] + \frac{1}{3} (z_{\text{phos}}^0 - z_{\text{phos}}) (R_2^2 + R_1 R_2 + R_1^2) \\ &\quad + \frac{1}{2} (z_{\text{phos}} R_2 - z_{\text{phos}}^0 R_1) (R_1 + R_2) \end{aligned} \quad (3.29)$$

Using Eq. (3.29), we get

$$\langle h(r) \rangle = \frac{(z_{\text{phos}} - z_{\text{phos}}^0) (R_1^2 + R_1 R_2 + R_2^2) + 3z_{\text{phos}}^0 R_3^2}{3R_3^2} \quad (3.30)$$

Eq. 3.30 is a quadratic equation in terms of R_2 . Solving for R_2 gives

$$R_2 = \frac{-R_1 + \sqrt{R_1^2 + 4C}}{2} \quad (3.31)$$

with

$$C = \frac{3R_3^2(z_{\text{phos}}^0 - \langle h(r) \rangle)}{z_{\text{phos}}^0 - z_{\text{phos}}} - R_1^2 \quad (3.32)$$

3.4 Results

3.4.1 Bending and Bulk Modulus

Show X-ray data. Show fitting boxes. Show the K_C values. Also, show the resultant form factors, which qualitatively show the membrane thinning. Fig. XX shows the scattering intensity pattern from DOPC/DOPE (1:1) with mole fraction $x=0.034$ Tat. The diffuse lobes are due to equilibrium fluctuations that occur in these fully hydrated, oriented lipid/peptide samples. The intensity $I(q)$ in the diffuse patterns provide the absolute values of the form factors $F(q_z)$, which are the Fourier transforms of the electron density profile, through the relation $I(\mathbf{q}) = S(\mathbf{q})|F(q_z)|2/q_z$, where $\mathbf{q} = (q_r, q_z)$, $S(\mathbf{q})$ is the structure interference factor, and q_z^1 is the usual LAXS approximation to the Lorentz factor [39, 55, 56]. The first step in the analysis takes advantage of the q_r dependence of the scattering to obtain the bending modulus K_C with results shown in Fig. 2. As positively charged Tat concentration was increased, the lamellar repeat spacing D generally increased in neutral lipid bilayers and decreased in negatively charged bilayers, consistent with changes in electrostatic repulsive interactions. With few exceptions, the water space between bilayers exceeded 20 Å.

The analysis that obtains K_C also obtains the structure factor $S(\mathbf{q})$ and then the unsigned form factors $|F(q_z)|$ are obtained from the intensity $I(q)$ by division. Results for five different membrane mimics are shown in Fig. 3. Vertical lines indicate the zero position between the lobes of diffuse data where $F(q_z)$ change sign. In every sample, the zero positions shift to larger q_z , indicating a thinning of the membranes.

3.4.2 Volume results

(Should probably include this info in appendix) First, the mass of Tat and water were measured to be 3.7 and 1212.6 mg via a digital balance. The density of water and

Tat-water solution were measured to be 0.993325 and 0.99418 g/cm³, respectively. The measured values of these quantities are shown in Table 3.8. This value is in a

Molecule	Molecular Weight	Volume
Tat (YGRKKRRQRRR)	1560	1876
Tat + TFA	2464	2964

Table 3.7: Important Quantities for Tat Peptide

ρ_{sol}	0.994180 g/cm ³
ρ_w	0.993325 g/cm ³
m_w	1212.6 mg
m_T	3.73 mg

Table 3.8: Measured Quantities in

good agreement with the value calculated from a peptide calculator website [].

Experimental and simulated volumes are given in Table 2. The simulated volume was obtained using the volume app in the SIMtoEXP program. The experimental Tat volume was calculated from the measured density assuming that the lipid volume was the same as with no Tat. In general, there may be an interaction volume between the peptide and the lipid membrane as we found previously for bacteriorhodopsin [57]. As lipid was present in excess to Tat, the partial molecular volume of the lipid should be the same as with no Tat, so this way of calculating includes all the interaction volume in VTat. Comparison of VTat in water with the result for 5:1 Lipid:Tat suggests that the interaction volume may be negative, consistent with a net attractive interaction with lipid. Understandably, values of VTat were unreliable for small mole ratios of Tat:Lipid. Therefore we used simple additivity for those mimics not shown in Table 2 for the volumes used in the SDP program. All volumes obtained from the Gromacs MD simulations were somewhat smaller than the measured volumes, but it supports the Tat volume being closer to 1822 3 than the outlying values obtained experimentally at small Tat concentrations.

3.4.3 Electron Density Profile Modeling

Using the model described in section 3.2.3, we fitted our measured X-ray form factors. In all fits, the positions of component groups were free parameters, but we assumed

that the lipid headgroup is somewhat rigid so that it cannot compress or expand. This assumption led to fixing the distance $z_{\text{PC}} - z_{\text{CG}}$ between the PC and CG components as well as the distance $z_{\text{CG}} - z_{\text{HC}}$ between the CG component and the Gibbs dividing surface for the hydrocarbon chains. We also constrained the width of Tat Gaussian. 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) We constrained the Tat width because we found that this parameter tended to become very small when it was free. This tendency to a unphysically small value was due to lack of higher q_z data points. A very narrow feature in an electron density profile led to large magnitude of the form factor at larger q_z . Because data points at larger q_z were not available, this narrow feature did not get penalized. In wide angle X-ray scattering, which probes much larger q_z than LAXS does, we did not observe much diffuse scattering for $q_z > 1.0 \text{ \AA}^{-1}$ (data not shown). Also, from a stereochemical point of view, a peptide width cannot be too small. These arguments allowed us to disregard the best fits with a too small value of σ_{Tat} and led to fixing the Tat width.

Figure 3.5 shows the results for DOPC with Tat. As shown, the fits were generally very good. Table 3.9 shows the best fit parameters for DOPC bilayers. It is clear that the area per lipid A_L increased as the Tat concentration was increased. An increase in A_L implies thinning of the bilayer which is an incompressible membrane. The results for DOPC:DOPE (3:1) are shown in Fig. 3.6 and Table 3.10, and the results for DOPC:DOPE (1:1) in Fig. 3.7 and Table 3.11.

As shown in Fig. 3.5, the membrane thickness can be defined as the distance D_{PP} between the PC components in the opposing leaflets or the distance D_{HH} between the maxima in the opposing leaflets. D_{HH} is more reliable than D_{PP} because it is a property of the total electron density of a bilayer and, therefore, does not depend strongly on the specific model employed for fitting the data. Indeed, the total electron density profile can be determined independently of a bilayer model by writing the electron density profile in terms of Fourier series, Fourier transforming the profile, and fitting the resulting model independent form factor to the data. On the other hand, D_{PP} is a property that depends on lipid components, which are influenced by how the lipid is parsed and what assumptions and constraints go into the specific model. A disadvantage of using D_{HH} as a measure of the membrane thickness is that D_{HH} is influenced by the electron density of Tat because the total electron density

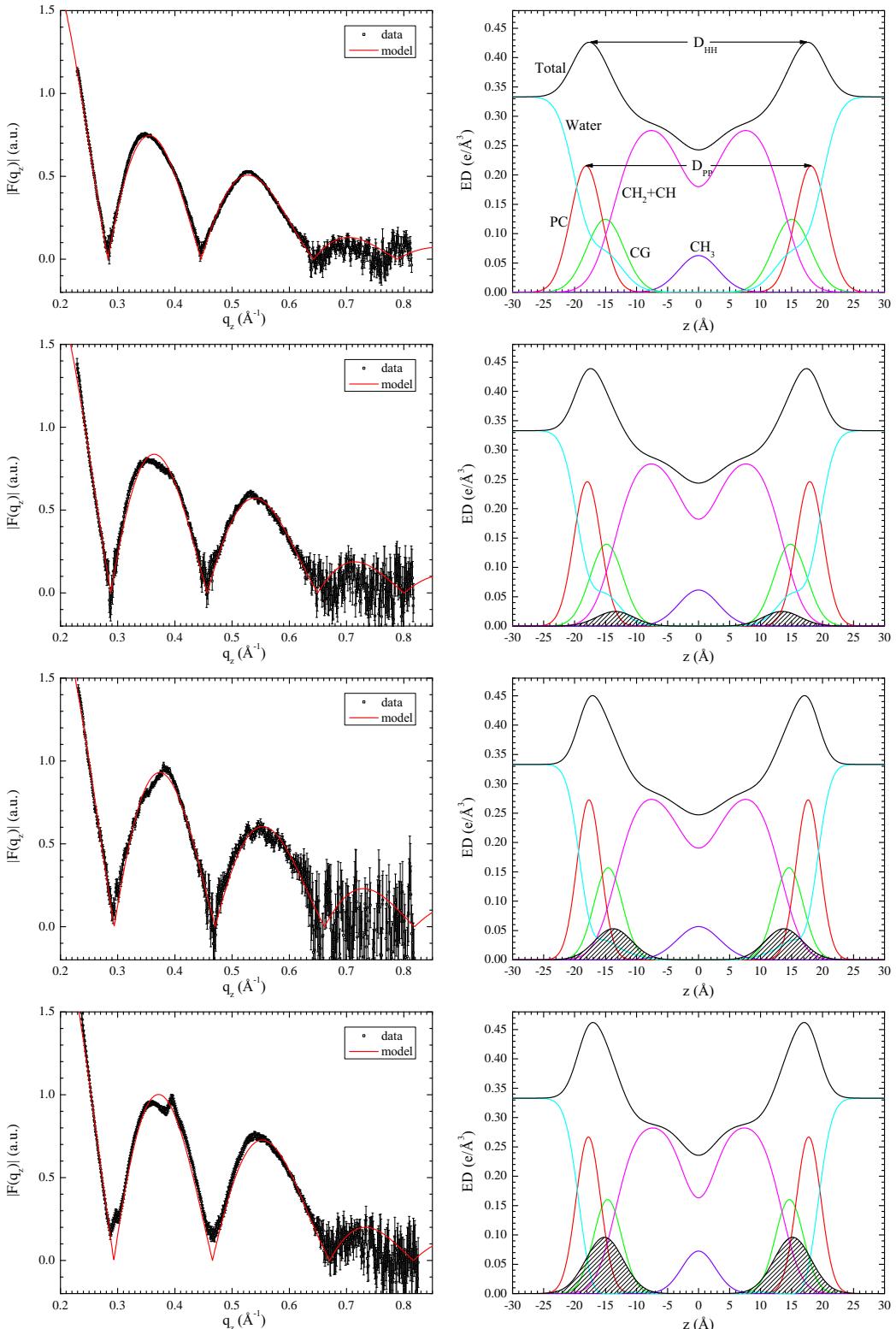


Figure 3.5: The best fits to DOPC form factors (left) and the corresponding electron density profiles (right) with $x_{\text{Tat}} = 0, 0.016, 0.034$, and 0.059 (from top to bottom).

x_{Tat}	0	1/63			1/29			1/17								
χ^2	2961	1554	1570	1581	1563	1587	1607	2342	2338	2363						
χ^2_{red}	5.307	2.785	2.813	2.833	2.807	2.849	2.885	4.197	4.189	4.235						
z_{PC}	18.1	18.0	17.9	17.9	17.8	17.7	17.6	17.8	17.8	17.7						
σ_{PC}	2.52	2.14	2.17	2.18	1.86	1.92	1.93	2.02	1.97	1.93						
z_{CG}	15.0	14.9	14.8	14.8	14.7	14.6	14.5	14.7	14.7	14.6						
σ_{CG}	3.00	2.62	2.64	2.66	2.22	2.30	2.31	2.58	2.27	2.14						
z_{HC}	13.7	13.6	13.5	13.5	13.4	13.3	13.2	13.4	13.4	13.3						
σ_{HC}	3.00	2.69	2.84	2.95	2.65	2.82	3.01	2.47	2.58	2.83						
σ_{CH_3}	3.20	3.19	3.22	3.24	3.37	3.43	3.47	2.70	2.70	2.74						
z_{Tat}	NA	12.9	13.4	14.2	13.1	13.8	14.4	15.2	15.2	15.7						
σ_{Tat}	NA	2.5	3.0	3.5	2.5	3.0	3.5	2.5	3.0	3.5						
V_L	1314	1344			1380			1432								
V_{HL}	331	362			397			450								
V_{Tat}	0	30.5			66.1			118.8								
R_{PC}	0.59	0.54			0.49			0.43								
R_{CG}	0.41	0.38			0.034			0.30								
Δz_1						3.1										
Δz_2						1.3										
r_{CH_3}						1.97										
r_{Tat}						0										
A_L	71.5	72.4	72.5	72.7	73.6	74.0	74.4	73.6	73.5	73.9						

Table 3.9: Fitting Results for DOPC membranes for the THG model. $\Delta z_1 = z_{\text{PC}} - z_{\text{CG}}$ and $\Delta z_2 = z_{\text{CG}} - z_{\text{HC}}$.

profile has a contribution from the electron density of Tat. Especially when the mole fraction of Tat in a system becomes large, the Tat electron density contributes significantly to the total electron density profile. If the Tat resided slightly outside of the PC component, the apparent membrane thickness measured by D_{HH} would be larger than D_{PP} . Then, even if the actual bilayer thickness defined by D_{PP} was reduced by the presence of Tat, the effect of thinning might not be obvious. With the above caveat in mind, we report both quantities in what follows since they can be easily calculated from the model.

x_{Tat}	0			1/63			1/29			1/17		
χ^2	924.5	4972	4985	4994	6758	6826	6863	2293	2280	2296		
z_{PC}	18.3	18.5	18.5	18.4	18.5	18.4	18.3	18.2	18.2	18.1		
σ_{PC}	2.66	2.23	2.26	2.27	2.25	2.31	2.34	2.31	2.19	2.11		
z_{CG}	15.2	15.4	15.4	15.3	15.4	15.3	15.2	15.1	15.1	15.0		
σ_{CG}	2.92	2.63	2.65	2.69	2.52	2.58	2.63	2.40	2.20	2.01		
z_{HC}	13.9	14.1	14.1	14.0	14.1	14.0	13.9	13.8	13.8	13.7		
σ_{HC}	2.73	2.70	2.83	2.91	2.86	2.79	2.84	2.25	2.38	2.60		
σ_{CH_3}	3.24	2.94	2.97	2.98	2.87	2.90	2.91	2.63	2.61	2.65		
z_{Tat}	NA	13.5	14.0	15.0	14.3	14.9	16.0	16.3	16.4	16.9		
σ_{Tat}	NA	2.5	3.0	3.5	2.5	3.0	3.5	2.5	3.0	3.5		
V_L	1288		1319			1354			1407			
V_{HL}	306		336			372			425			
V_{Tat}	0		30.5			66.1			118.8			
R_{PC}	0.59		0.54			0.49			0.43			
R_{CG}	0.41		0.38			0.034			0.30			
Δz_1					3.1							
Δz_2					1.3							
r_{CH_3}					1.97							
r_{Tat}					0							
A_L	70.9	69.8	69.9	70.1	69.5	70.0	70.6	71.3	71.4	71.7		

Table 3.10: Fitting Results for DOPC:DOPE (3:1) membranes for the THG model. $\Delta z_1 = z_{\text{PC}} - z_{\text{CG}}$ and $\Delta z_2 = z_{\text{CG}} - z_{\text{HC}}$.

We also investigated how the goodness of fits varied as the position of the Tat Gaussian was varied. Figure 3.8 plots χ^2 as a function of the fixed Tat position z_{Tat} . We found that the two models, THG and THC, resulted in similar electron density profiles, yielding similar χ^2 values when Tat was placed near the hydrocarbon-water interface region. In the THC model, the error function representing the hydrocarbon chain region became wider as Tat was placed near the interface region but further

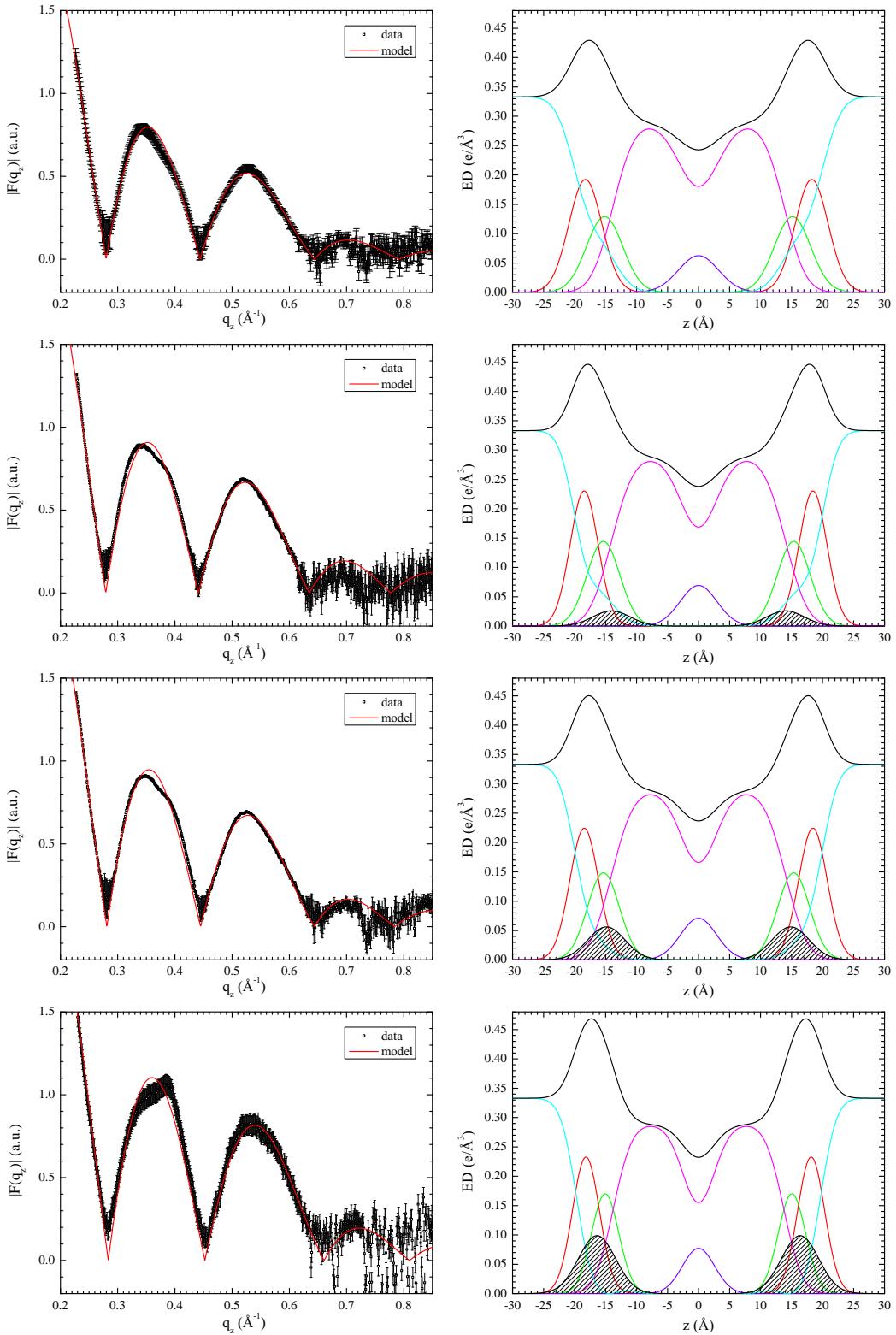


Figure 3.6: The best fits to DOPC:DOPE (3:1) form factors (left) and the corresponding electron density profiles (right) with $x_{\text{Tat}} = 0, 0.016, 0.034$, and 0.059 (from top to bottom).

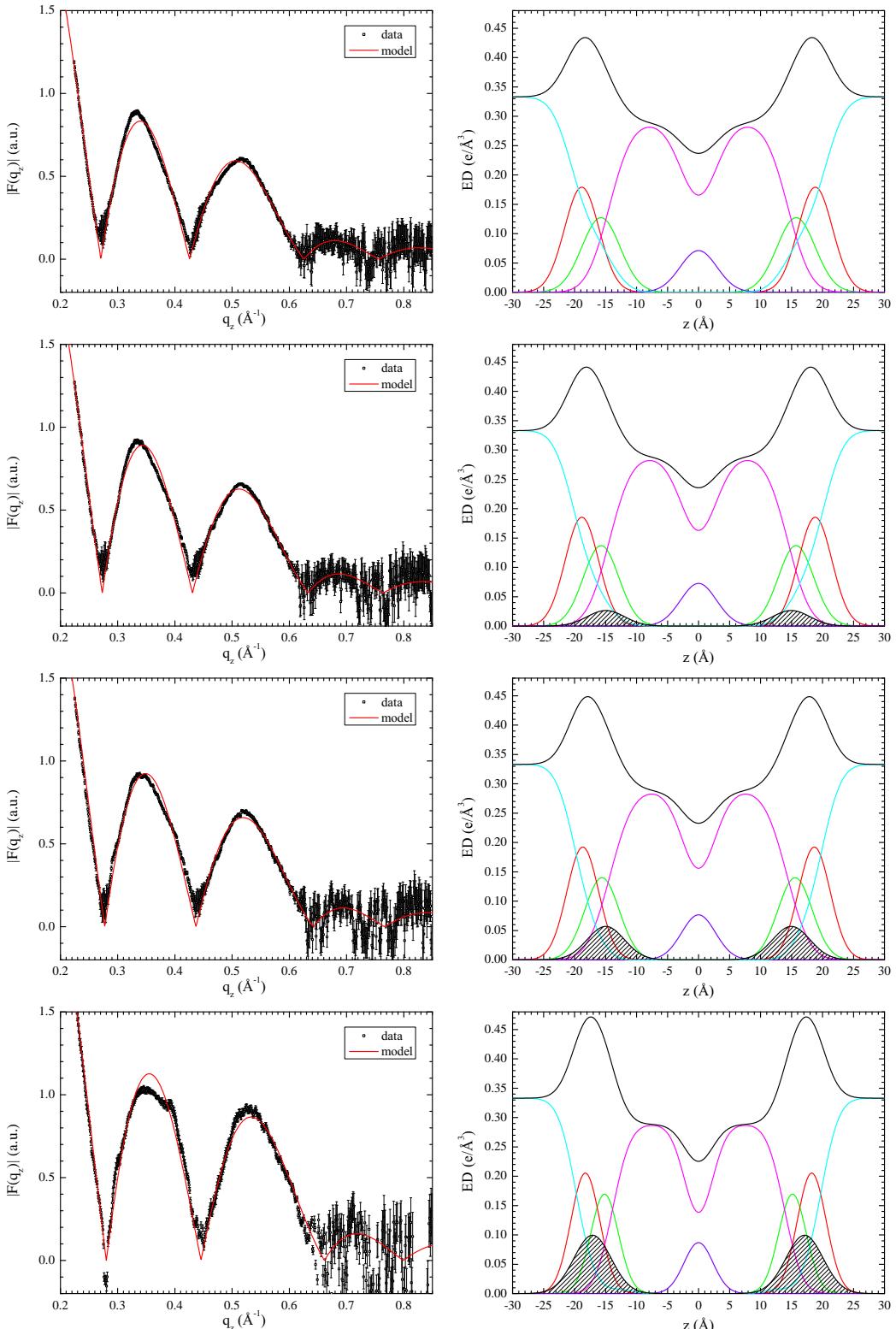


Figure 3.7: The best fits to DOPC:DOPE (1:1) form factors (left) and the corresponding electron density profiles (right) with $x_{\text{Tat}} = 0, 0.016, 0.034$, and 0.059 (from top to bottom).

x_{Tat}	0	1/63			1/29			1/17								
χ^2	2961	1554	1570	1581	1563	1587	1607	2342	2338	2363						
χ^2_{red}	5.307	2.785	2.813	2.833	2.807	2.849	2.885	4.197	4.189	4.235						
z_{PC}	18.1	18.0	17.9	17.9	17.8	17.7	17.6	17.8	17.8	17.7						
σ_{PC}	2.52	2.14	2.17	2.18	1.86	1.92	1.93	2.02	1.97	1.93						
z_{CG}	15.0	14.9	14.8	14.8	14.7	14.6	14.5	14.7	14.7	14.6						
σ_{CG}	3.00	2.62	2.64	2.66	2.22	2.30	2.31	2.58	2.27	2.14						
z_{HC}	13.7	13.6	13.5	13.5	13.4	13.3	13.2	13.4	13.4	13.3						
σ_{HC}	3.00	2.69	2.84	2.95	2.65	2.82	3.01	2.47	2.58	2.83						
σ_{CH_3}	3.20	3.19	3.22	3.24	3.37	3.43	3.47	2.70	2.70	2.74						
z_{Tat}	NA	12.9	13.4	14.2	13.1	13.8	14.4	15.2	15.2	15.7						
σ_{Tat}	NA	2.5	3.0	3.5	2.5	3.0	3.5	2.5	3.0	3.5						
V_L	1314	1344			1380			1432								
V_{HL}	331	362			397			450								
V_{Tat}	0	30.5			66.1			118.8								
R_{PC}	0.59	0.54			0.49			0.43								
R_{CG}	0.41	0.38			0.034			0.30								
Δz_1						3.1										
Δz_2						1.3										
r_{CH_3}						1.97										
r_{Tat}						0										
A_L	71.5	72.4	72.5	72.7	73.6	74.0	74.4	73.6	73.5	73.9						

Table 3.11: Fitting Results for DOPC:DOPE (1:1) membranes for the THG model. $\Delta z_1 = z_{\text{PC}} - z_{\text{CG}}$ and $\Delta z_2 = z_{\text{CG}} - z_{\text{HC}}$. (Need to work)

from the bilayer center. The subtraction of the Tat component from the hydrocarbon chain error function resulted in a smooth error function-like profile with a smaller value of σ such that the total profile calculated from the THC model was very similar to that calculated from the THG model.

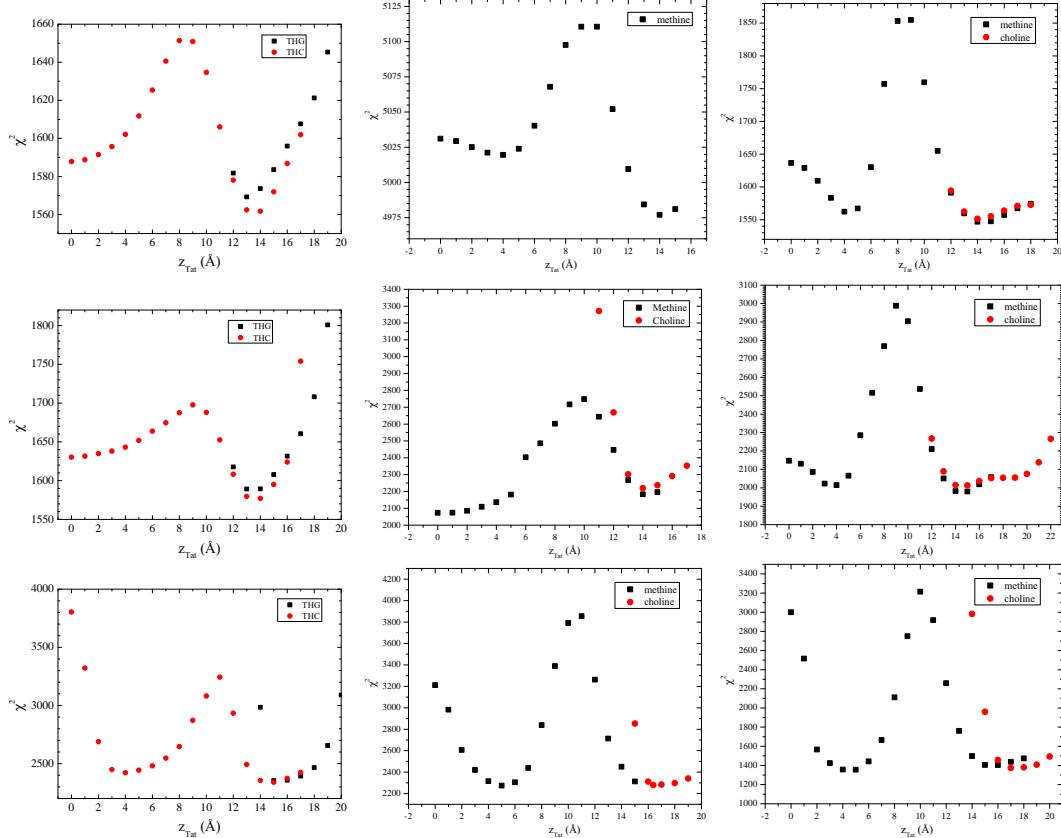


Figure 3.8: χ^2 as a function of z_{Tat} for DOPC, DOPC:DOPE (3:1), and DOPC:DOPE (1:1) (from left to right) with $x_{\text{Tat}} = 0.016, 0.034$, and 0.059 (from top to bottom). $\sigma_{\text{Tat}} = 3.0$. The THG model (black squares) and the THC model (red circles).

In general, while the total electron density profile is well determined by our modeling procedures, the values of the parameters for the components are not as well determined as the agreement of the fit to the data may suggest. in many cases, we found multiple local minima in the fitting landscape, including one with Tat closer to the center of the bilayer as shown in Fig. 3.8. χ^2 calculated at these local minima tended to be smaller for larger concentration of Tat. We also found that χ^2 with z_{Tat} in the hydrocarbon chain region and headgroup region was almost equal for the smallest value of x_{Tat} for DOPC:DOPE (1:1) bilayer. While Fig. 3.8 shows trends

for the χ^2 minima with Tat in the hydrocarbon chain region, this position seemed energetically unfavorable as Tat is a hydrophilic molecule. Also, if Tat favored to be inserted deep in a membrane, it would be difficult for Tat to leave the membrane. This difficulty seems inefficient in terms of the HIV virus infection because Tat passing through the nuclear membrane and binding to the viral integrated DNA is crucial for proliferation of HIV infected cells as discussed in section 3.1. These considerations suggested that the local minima in the chain region was artifact of our models. With a help of MD simulations performed by Dr. Kun Huang, we were able to discard the interior position as an artifact of our modeling. The simulation results are discussed in section 3.4.6.

Electron density profiles for DOPC/DOPS (3:1) and the nuclear membrane mimic were not successful, due to loss of diffuse scattering by Tats charge neutralization of these negatively charged membranes as described in section 3.4.1.

3.4.4 Hard Wall Constraint Fits

As seen from Table 3.9, 3.10, and 3.11, the widths of the headgroup components became smaller as Tat concentration increased in all membranes. These decreases seemed somewhat unreasonable; if Tat causes a bilayer to locally become thinner near where it is bound, we would expect that the headgroup components to become wider. Therefore, we also fitted the model with hard wall constrains on these headgroup widths. Namely, the minimum values of the widths of the headgroup components, PC and CG, were limited to the corresponding values for pure bilayers without Tat.

Table ?? shows results from fitting the data with lower bounds on the widths of the headgroup components for DOPC:Tat. In all cases, both headgroup widths, σ_{PC} and σ_{CG} , resulted in the same value as the value of their corresponding lower bounds. Figure ?? shows χ^2 landscape as a function of z_{Tat} similarly to Fig. 3.8. The χ^2 minima observed for $z_{\text{Tat}} > i 25 \text{ \AA}$ were artifact; Tat are essentially in the water region while the bilayer structure was significantly perturbed. This action-at-distance seemed unreasonable, so these minima were considered as artifact of the model. Indeed, when we fixed lipid component parameters in these fits to be identical to those of the DOPC model, we did not observe any minima with Tat in the water region. Although we did not note in the previous section, we observed similar minima in the unbound model as well.

x_{Tat}	0	0.016	0.016	0.16	0.034	0.034	0.034
χ^2	2961	1853	1979	2118	2398	2893	3414
z_{PC}	18.1	17.8	17.8	17.8	17.4	17.4	17.4
σ_{PC}	2.5	2.5	2.5	2.5	2.5	2.5	2.5
z_{CG}	15.0	14.7	14.7	14.7	14.3	14.3	14.3
σ_{CG}	3.0	3.0	3.0	3.0	3.0	3.0	3.0
z_{HC}	13.7	13.4	13.4	13.4	13.0	13.0	13.0
σ_{HC}	3.0	2.7	2.7	2.7	2.7	2.7	2.7
σ_{CH_3}	3.2	3.1	3.1	3.1	3.6	3.6	3.7
z_{Tat}		16.9	16.8	17.0	16.4	16.5	16.7
σ_{Tat}		2.5	3.0	3.5	2.5	3.0	3.5
Δz_1	3.1	3.1	3.1	3.1	3.1	3.1	3.1
Δz_2	1.3	1.3	1.3	1.3	1.3	1.3	1.3
A_{L}	71.5	73.5	73.5	73.5			
x_{Tat}	0.059	0.059	0.059				
χ^2	3160	4298	5539				
z_{PC}	17.5	17.4	17.3				
σ_{PC}	2.5	2.5	2.5				
z_{CG}	14.4	14.4	14.3				
σ_{CG}	3.0	3.0	3.0				
z_{HC}	13.1	13.0	12.9				
σ_{HC}	2.7	2.7	2.7				
σ_{CH_3}	2.6	2.6	2.5				
z_{Tat}	16.3	16.6	17.1				
σ_{Tat}	2.5	3.0	3.5				
Δz_1	3.1	3.1	3.1				
Δz_2	1.3	1.3	1.3				
A_{L}							

Table 3.12: Fitting Results of the bound THG model for DOPC membranes. $\Delta z_1 = z_{\text{PC}} - z_{\text{CG}}$ and $\Delta z_2 = z_{\text{CG}} - z_{\text{HC}}$.

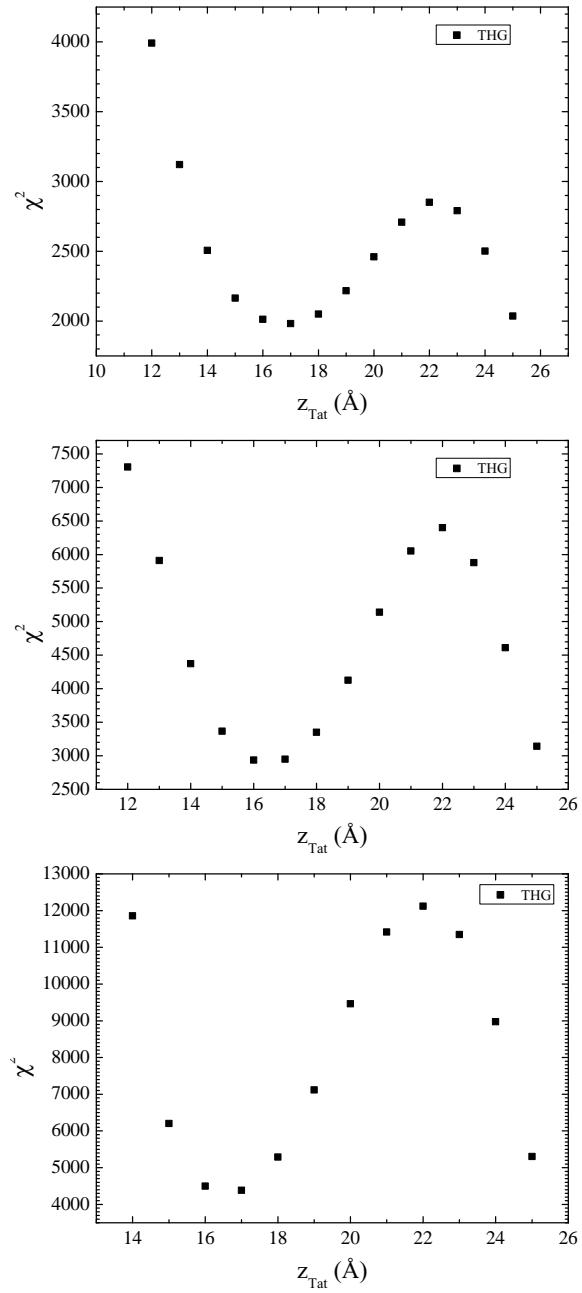


Figure 3.9: χ^2 as a function of z_{Tat} for DOPC with $x_{\text{Tat}} = 0.016, 0.034$, and 0.059 (from top to bottom). $\sigma_{\text{Tat}} = 3.0$. The bound THG model was used.

Figure 3.10: DPP graph

Figure 3.11: DHH graph

Figure 3.12: AL graph

3.4.5 Summary of Electron Density Profile Modeling

As discussed in the previous section, the bilayer thickness can be described as D_{HH} which is the distance between the maxima in the electron density profile, or as D_{PP} , which is the distance between the phosphocholines on the opposing monolayers. Figs. 3.10 and 3.11 show that the thickness defined by D_{PP} or D_{HH} tends to decrease with increasing Tat mole fraction x_{Tat} , 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. 3.13. Figure 3.12 shows that the area per lipid A_{L} 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 3.12 shows that the area per lipid A_{L} as defined by $(V_{\text{L}} - V_{\text{HL}})/D_{\text{C}}$ decreases as the mole fraction of DOPE in DOPC:DOPE membranes increases. This decrease of A_{L} is qualitatively consistent with previous studies and attributed to the small size of PE head group (references?). Because DOPE has smaller head group than DOPC, lipids in DOPC/DOPE bilayers pack more compactly than in DOPC bilayers. Then, more compact packing of lipids leads to a smaller A_{L} .

Figure 3.13 shows that Tat is located further out from the bilayer center with higher content of PE lipids. This trend is consistent with a potential mean force calculated from MD simulations, which has shown that arginine insertion costs more energy in a PE membrane than PC membrane. The higher energy cost of arginine insertion is due to more possible hydrogen bonding between PE group and arginines. (After I wrote this paragraph, I realized that the argument presented does not make much sense. z_{Tat} could be larger for PE simply because DOPE membranes are thicker than DOPC membrane. z_{Tat} must be measured with respect to, say, the hydrocarbon interface.)

Figure 3.13: zTat graph

3.4.6 Molecular Dynamics Simulations

Due to the slow relaxation in lipid bilayers and limited accuracy of the force field, a good agreement between experimental and MD simulation calculated form factors may be difficult to reach. Consequently, we carried out several constrained simulations at various A_L and z_{Tat} as described in Materials and Methods. We then compared the simulated form factor $F(q_z)$ with the experimentally measured one. Figure 3.14 shows such comparison for a DOPC bilayer. As discussed earlier, the simulated form factor shifted to larger q_z as the area per lipid was increased. From this comparison, we found the simulation at $A_L = 70 \text{ \AA}^2$ to be the best match with the experimental form factor, yielding the lowest χ^2 . However, the form factor for $A_L = 72 \text{ \AA}^2$ matched the experiment better than that for 70 \AA^2 near $q_z = 0.3 \text{ \AA}^{-1}$, which suggests that a better match might lie between 70 and 72 \AA^2 . This case was not investigated further. The electron density profile for the best fit is shown in Fig. 3.15. The comparison for DOPC with $x_{\text{Tat}} = 0.015$ where there is one Tat in each monolayer is shown in Fig. 3.16. The same comparison for DOPC with $x_{\text{Tat}} = 0.03$ is shown in Fig. 3.17.

The best match for DOPC/Tat (128:4) was found when the Tats were constrained at 18 \AA away from the bilayer center (Fig. 4.A,B). The other best fit results were: DOPC $A_L = 70 \text{ \AA}^2$ and DOPC/Tat(128:2) $A_L = 72 \text{ \AA}^2$, $z_{\text{Tat}} = 18 \text{ \AA}$. It clearly indicates that with increasing Tat concentration, A_L increases. The agreement worsened as Tat was constrained to be closer to the center of the bilayer. When Tats were constrained at 5 \AA away from the bilayer center, we observed a spontaneous formation of water pores in the MD simulation. However, as shown in Fig. 4.C the corresponding form factor calculated from MD simulations does not match well with experiments.

We summarize our results for how Tat affects the lipid bilayer in Fig. 9. The height of Tat, $H_{\text{Tat}} = 8.7 \text{ \AA}$, was the full width at half maximum of the Tat electron density profiles obtained from simulations and the cylindrical radius, $R_{\text{Tat}} = 8.3 \text{ \AA}$, was calculated to give the measured volume. The Z distances from the center of the bilayer were derived from weighted averages of four MD simulations of Tat:DOPC 2:128. The χ^2 obtained by comparison to experiment indicated that the best Z_{Tat} lay between the simulated values of 16 \AA and 18 \AA and the best area/lipid A_L lay between the simulated values of 72 \AA^2 and 74 \AA^2 , so averages were obtained from these four combinations of Z_{Tat} and A_L , weighted inversely with their χ^2 . The average positions, Z'Phos, of phosphates situated underneath the Tats were calculated by averaging over

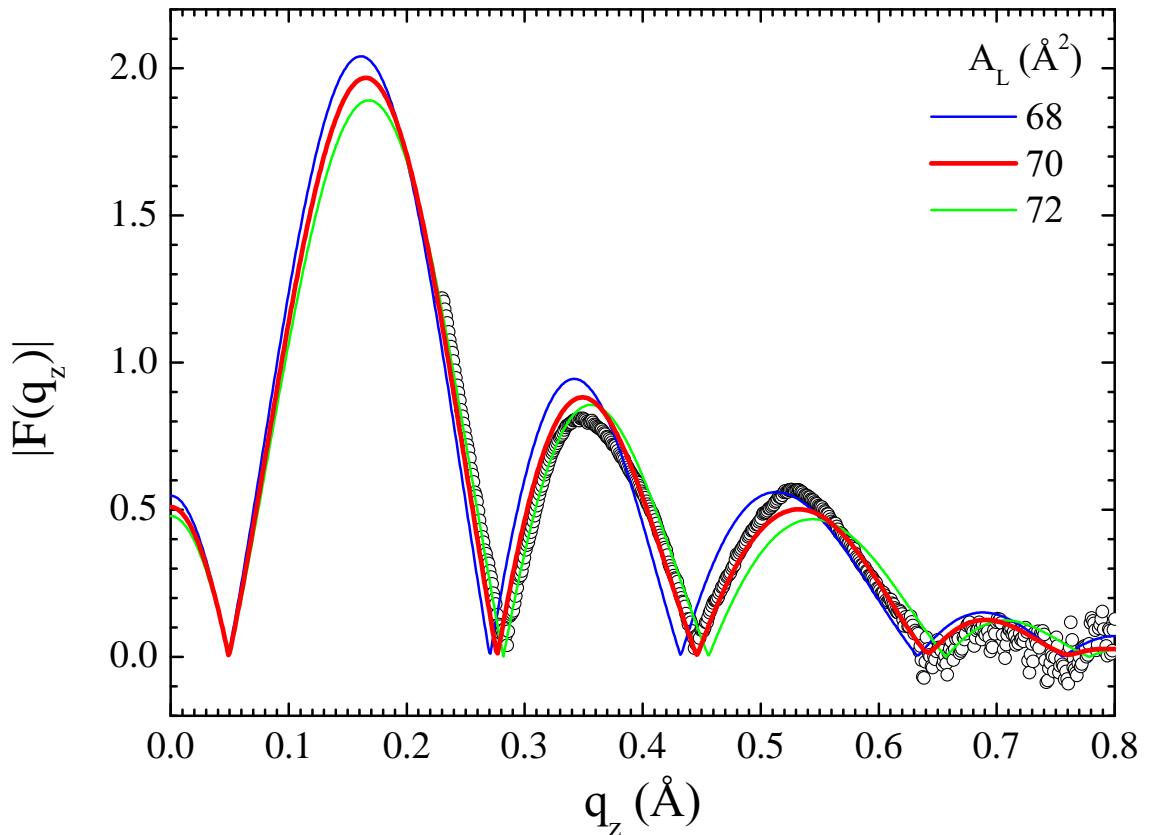


Figure 3.14: MD simulated form factors for DOPC at $A_L = 68 \text{ \AA}^2$ (blue solid line), 70 \AA^2 (red solid line), and 72 \AA^2 (green solid line) compared to the experimental form factor (open circles) scaled vertically to best match the form factor for 70 \AA^2 .

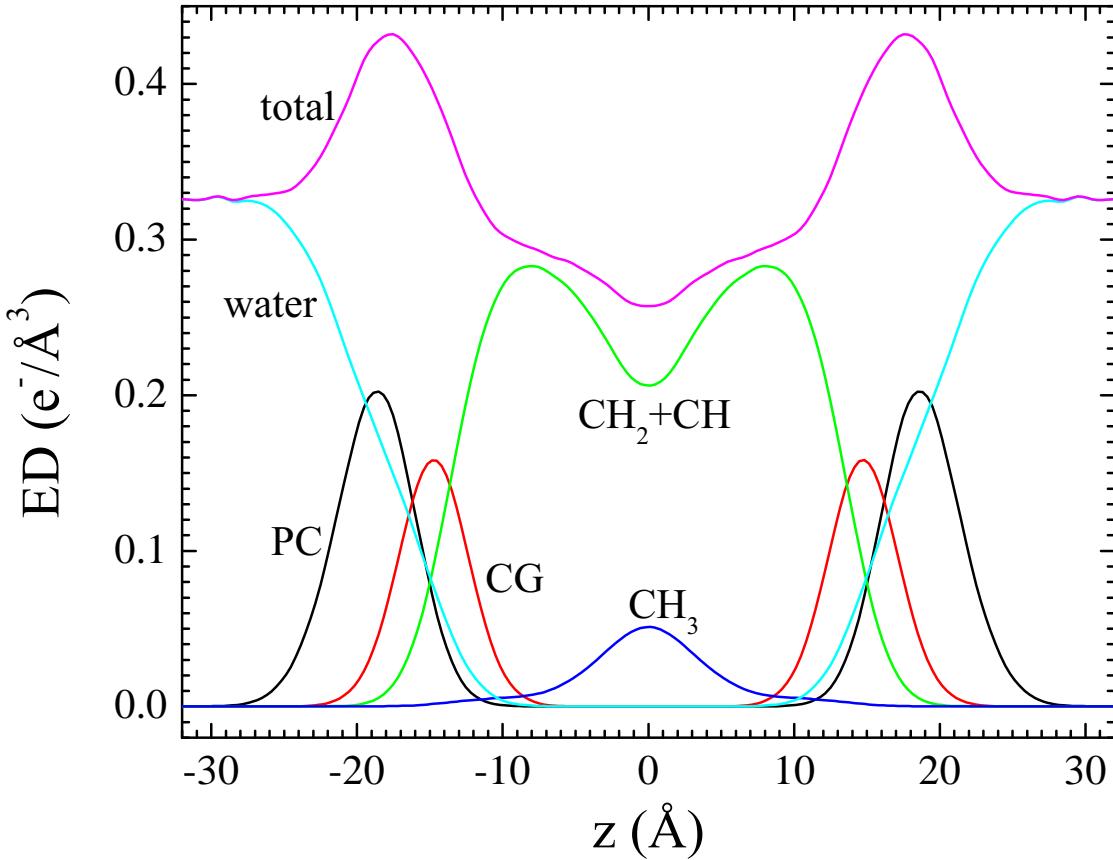


Figure 3.15: The simulated, symmetrized electron density profile for DOPC at $A_L = 70 \text{ \AA}^2$ as a function of the distance away from the bilayer center. Each component profile is labeled with its name: PC (phosphate-choline), CG (carbonyl-glycerol), CH_2+CH (methylene-methine combination), CH_3 (terminal methyl). The sum of all the components is labeled as total.

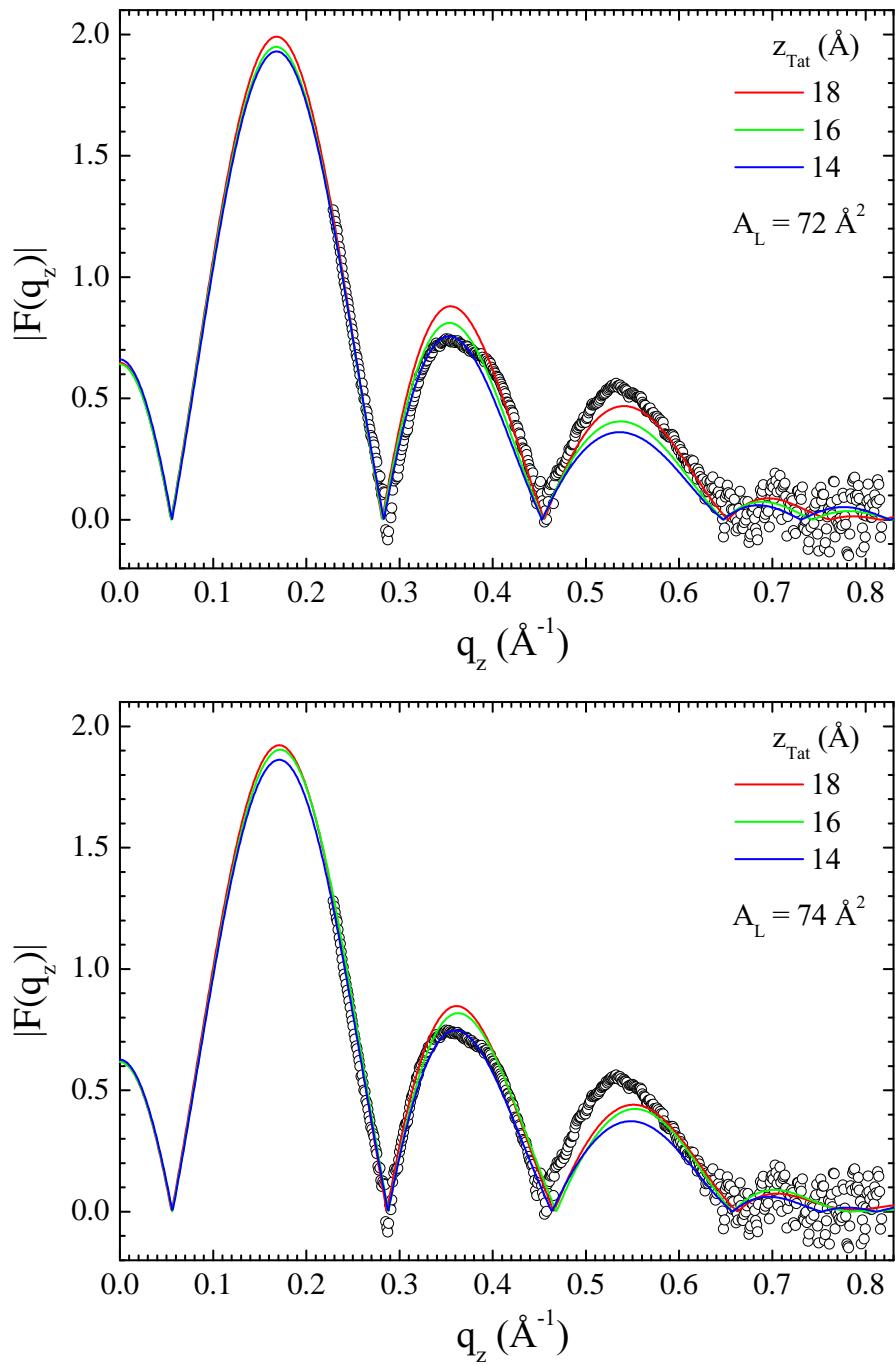


Figure 3.16: MD simulated form factors for DOPC with $x_{\text{Tat}} = 0.015$ at $A_L = 72 \text{ \AA}^2$ (top) and 74 \AA^2 (bottom), with $z_{\text{Tat}} = 18 \text{ \AA}$ (red solid lines), 16 \AA (green solid lines), and 14 \AA (blue solid lines) compared to the experimental form factor (open circles) scaled vertically to best match the form factor for $z_{\text{Tat}} = 18 \text{ \AA}$.

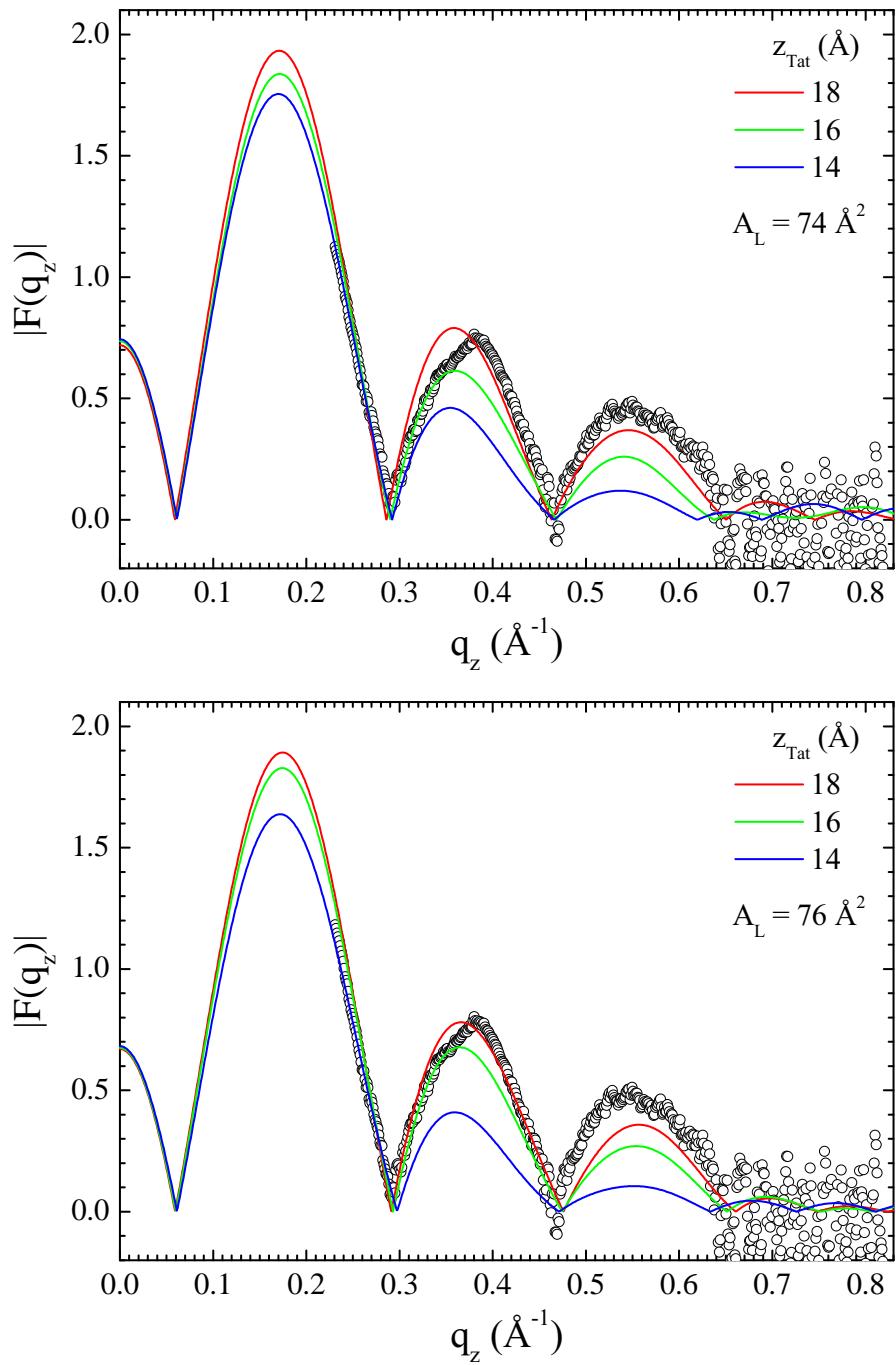


Figure 3.17: MD simulated form factors for DOPC with $x_{\text{Tat}} = 0.030$ at $A_L = 74 \text{ \AA}^2$ (top) and 76 \AA^2 (bottom), with $z_{\text{Tat}} = 18 \text{ \AA}$ (red solid lines), 16 \AA (green solid lines), and 14 \AA (blue solid lines) compared to the experimental form factor (open circles) scaled vertically to best match the form factor for $z_{\text{Tat}} = 18 \text{ \AA}$.

$x_{\text{Tat}} = 0.015$				$x_{\text{Tat}} = 0.030$			
$A_L (\text{\AA}^2)$	$z_{\text{Tat}} (\text{\AA})$	a	χ^2	$A_L (\text{\AA}^2)$	$z_{\text{Tat}} (\text{\AA})$	a	χ^2
70	18	0.621	60.1	70	18	0.621	60.1
70	16	0.568	69.1	70	16	0.568	69.1
70	14	0.439	131	70	14	0.439	131
70	12	0.285	391	70	12	0.285	391
70	10	0.199	440	70	10	0.199	440
70	8	0.196	374	70	8	0.196	374
70	5	0.159	527	70	5	0.159	527
72	18	0.72	18.0	72	18	0.72	18.0
72	16	0.65	24.9	72	16	0.65	24.9
72	14	0.6	31.4	72	14	0.6	31.4
72	12	0.426	104	72	12	0.426	104
72	10	0.219	443	72	10	0.219	443
72	8	0.205	336	72	8	0.205	336
72	5	0.165	448	72	5	0.165	448
74	18	0.722	21.3	74	18	0.722	21.3
74	16	0.704	25.9	74	16	0.704	25.9
74	14	0.631	24.7	74	14	0.631	24.7
74	12	0.412	81.9	74	12	0.412	81.9
74	10	0.312	194	74	10	0.312	194
74	8	0.246	351	74	8	0.246	351
74	5	0.177	427	74	5	0.177	427

Table 3.13: Comparison of the simulated form factors to the experimental form factors.

x_{Tat}	A_L	z_{Tat}	$\langle D_{\text{PP}} \rangle$	D_{PP}	x	Δt	H_{Tat}	R_{Tat}	R_2	z_{phos}	z_{guan}	χ^2
0	70		36.3									
0.015	72	18	35.6	32.8	35.8	3.5	9.2	8.1	15.0	14.7	15.5	18
0.015	72	16	36.1	33.0	36.3	3.3	9.4	8.0	9.0	14.9	14.5	24.9
0.015	74	18	35.0	33.0	35.1	3.3	8.6	8.3	23.9	14.9	16.5	21.3
0.015	74	16	35.0	32.1	35.2	4.2	7.6	8.9	20.4	14.0	13.5	25.9
0.030	74	18	35.3	32.6	NA	3.7	7.6	8.9	NA	14.5	15.5	24.3
0.030	74	16	35.3	31.2	NA	5.1	7.7	8.8	NA	13.1	13.5	40.1
0.030	76	18	34.2	32.0	NA	4.3	7.6	8.9	NA	13.9	16.5	14.8
0.030	76	16	34.9	31.4	NA	4.9	7.8	8.7	NA	13.3	14.5	30.4

Table 3.14: Summary of simulation results. $\langle D_{\text{PP}} \rangle$, phosphorus-phosphorus distance averaged over all lipids; D_{PP} , Tat-perturbed phosphorus atoms; x , thickness away from Tat; Δt , $\langle D_{\text{PP}}^{\text{DOPC}} \rangle - D_{\text{PP}}$; H_{Tat} , Tat height; R_{Tat} , radius of Tat cylinder; R_2 , radius of the calculated in-plane Tat-perturbed region; R_3 , effective radius of the simulation box.

x_{Tat}	A_L	z_{Tat}	$\langle D_{\text{PP}} \rangle$	D_{PP}	Δt	H_{Tat}	R_{Tat}	R_2	z_{phos}	z_{guan}
0.015	72.9	17.1	35.4	32.7	3.6	8.7	8.3	17.1	14.6	15.1
0.030	75.2	17.3	34.8	31.9	4.4	7.7	8.8	NA	13.8	15.4

Table 3.15: Summary of weighted average results. The caption is the same as Table 3.14.

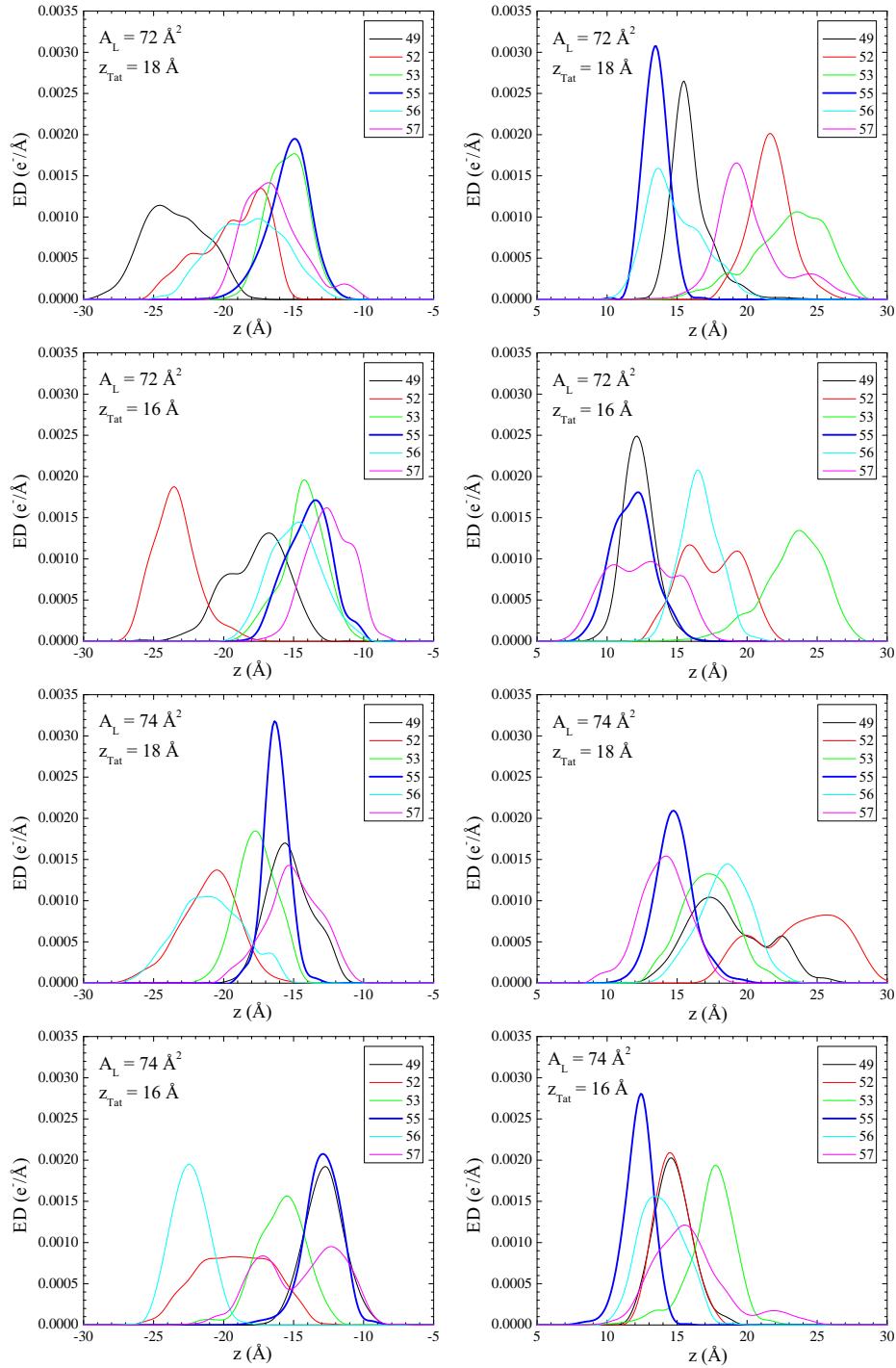


Figure 3.18: Electron density profiles of guanidinium groups from the four best matched simulations for DOPC with $x_{\text{Tat}} = 0.015$ (one Tat on each leaflet). Tat on the lower and upper leaflets are shown on the left and right plots, respectively.

the phosphates whose in-plane distance, R , from the center of Tat is smaller than $RTat$. The simulation cell extended to 38 Å, far enough to ensure that ZPhos for most of the lipids is the same as for DOPC. Assuming a simple linear ramp in ZPhos, Fig. 9 then indicates a ring of boundary lipids that extends twice as far in R as Tat itself. Although the guanidinium electron density profile was broad (Fig. S8), indicating that some were pointing away from the bilayer relative to the center of Tat, more were pointing towards the bilayer center as indicated in Fig. 9. Numerical values are given in Table S1.

3.5 Discussion

Given that 8 of the 11 amino acids in Tat (47-57) are arginines and lysines, one would have suggested 20 years ago that highly charged Tat would partition strongly into solution rather than being associated with lipid bilayers. By contrast, but in agreement with more recent perspectives on arginine partitioning into the interfacial region [58], we find that Tat interacts with lipid bilayers, even with neutral DOPC and DOPC/DOPE mixtures, as well as with negatively charged DOPC/DOPS and nuclear membrane mimic lipid mixtures. This paper presents multiple lines of evidence for a Tat/membrane interaction. Fig. 2 shows that Tat decreases the bending modulus. Although one could argue that such a decrease is only apparent and could instead be due to local changes in membrane spontaneous curvature [59], either interpretation supports a Tat-bilayer interaction. The changes with increasing Tat concentration in the X-ray membrane form factors in Fig. 3 prove that Tat affects membrane structure, and the shift of the zero positions to higher qz suggests thinning. Thinning is substantiated by quantitative analysis of the X-ray data and by MD simulations. Fig. 7A shows that the average membrane thickness, as measured by the distance DPP between phosphocholines on opposite surfaces, decreases with increasing Tat concentration. Similar thinning is shown in Fig. 7B for the distance DHH between the maxima in the electron density profiles of opposite surfaces. Compared to DPP, DHH is pulled towards both the carbonyl/glycerol groups and Tat because both have electron densities (0.4 e/3) greater than water (0.33 e/3) or hydrocarbon (0.3 e/3). Although the thinning shown in Figs. 7A and 7B is not large, it obviously requires interaction of Tat with the bilayers. Fig. 7C shows that AL increases with increasing Tat concentration, by both model fitting and MD simulations.

It is of considerable interest to learn where Tat resides, on average, in the membrane, as this would establish a base position from which translocation would be initiated. We have combined our two main methods, MD simulations and X-ray scattering, to address this question. In general, Tats locate at the bilayer/water interface as indicated in Section 3.2, and they are close to the phosphocholine headgroup region by comparing the simulated 2ZTat in Fig. 7.D with 7.A. Although the SDP modeling of the X-ray data obtains excellent fits to the experimental form factors for a model with Tat deep in the hydrocarbon interior (see Fig. S5), the corresponding MD simulation (shown in Fig. 4.C) eliminates this spurious result. Fig. 7D also shows that modeling gives smaller values for ZTat than the simulation. The modeling result is supportive of the original simulation result of Herce and Garcia that Tat resides closer to the bilayer center than do the phosphocholine groups [60]. That is a base position that would be a possibly important precursor to translocation, as would the larger AL.

Several groups have carried out calculations and MD simulations showing that the cost of moving an arginine group from water to the bilayer center is 12-26 kcal/mol [58, 61-63] or 6-7 kcal/mol if side-chain snorkeling to the surface is taken into account [64]. This is not inconsistent with our result that Tat interacts with the membrane because, as is well known, the bilayer is not just a hydrocarbon slab, but has interfacial headgroup regions where Tat can reside. It has been suggested that the free energy cost for charged amino acids entering the headgroup region is similar to that for partitioning into octanol, about an order of magnitude smaller free energy cost than partitioning into cyclohexane [65-67]. Simulations suggest that the free energy is smaller for an arginine residing in the interfacial region than in water, roughly by 3 kcal/mole, depending upon the lipid [58, 67]. Our results therefore appear energetically reasonable.

One concern with diffraction experiments on samples consisting of adjacent bilayers in a stack or in a multilamellar vesicle is that the samples have to be partially dried to obtain conventional diffraction data. But then there is no pure water layer between adjacent bilayers, so a hydrophilic peptide is forced into the interfacial, partially hydrophilic region of the lipid bilayer. In contrast, by using diffuse scattering, we obtained structure from experimental samples that had a range of lamellar D spacings (see Fig. 2 caption) that were considerably larger than the thickness of the bilayer in Fig. 7A, thereby providing an ample pure water space, typically greater

than 20. The result that 2ZTat shown in Fig. 7D is so much smaller than our repeat spacings shows that Tat preferentially associates with the membrane rather than dissociating into water.

Consistent with Tat softening the bilayers (Fig. 2), it also disorders them as indicated by Sxray decreasing with Tat concentration shown in Fig. 8. Tat also increases the mosaic spread observed by X-ray and neutron scattering as shown in Figs. S1-3; this is a much larger scale disordering of the stack of bilayers. As shown in Table 1 and in Fig. S7, Tat assumed slightly β 50membrane. Our results were determined using the DichroWEB program, which compares the mean residue ellipticity with that from standard globular proteins, with details given in Supplementary data near Fig. S7. These structures include approximately equal amounts of regular strands and turns, with half that amount of distorted strands. The next most prevalent structure was random coil (37Introduction) report a primarily random structure, determined using either CD or NMR. This difference could be due to different sample preparations, or due to a different interpretation of the CD spectra. Ref. [68] reported that CD spectra of unordered polypeptides are similar to that of the poly(Pro)II helix, and a significant fraction of the unordered conformation in globular proteins consists of poly(Pro)II helix plus distorted strands.

In an effort to better determine the secondary structure of Tat, our collaborator, Dr. Rieko Ishima, performed 1D and 2D-NMR of Tat in solution at 10, 20 and 30oC. Her results showed no evidence for backbone hydrogen bond formation, indicating that the peptide does not have a stable conformation, at least on the time scale of the NMR measurement. Additionally, we analyzed the secondary structures of Tats from MD simulations using the Define Secondary Structure of Proteins (DSSP) program [69]. Data from the MD simulation which has the best fit to experimental X-ray form factors show that Tat contains neither α nor β -helix structures. Therefore, both our solution NMR and MD simulation results find primarily random coil, with no significant structure, which contrasts with our CD findings of β 50the interpretation of CD spectra as α , P2 helix or coil is controversial, what is clear is that the membrane does not influence the conformation of solubilized Tat. In addition, no studies including our own, have implicated Tat forming an α -helix, either in solution or in the membrane.

Given our structural and elastic moduli results, we now compare to other experiments in the literature. In 2008, the Wong group implicated Tats ability to

induce saddle-splay curvature with a potential role of bidentate hydrogen bonding as key [70]. Rhodamine-tagged Tat only entered GUVs when the PE headgroup was included with PS and PC lipids (PS/PC/PE, 20:40:40), indicating that hydrogen-bonding, and/or curvature-promoting lipids are required for Tat translocation. In PS/PE (20:80) lipids, they found Tat caused a highly curved cubic phase using X-ray diffraction [70]. In our experiments, there was little effect of adding DOPE to DOPC at either a 3:1 or 1:1 mole ratio on decrease in the bending modulus, bilayer thinning, Tats outward movement with increasing concentration or disordering of chains (Sxray). Our two results are not inconsistent, however, since curvature-promotion appears not to be required for Tats ability to lower the energy required to bend, nor to locate Tat in the bilayer, nor to disorder chains, all of which may be important for Tat translocation. Yet Tat does translocate across membranes in their experiments only with PE in the membrane, so the ability to induce saddle- splay curvature may also be required for Tats translocation. Another study by Melikov et al. [26] found that Tats main mechanism of action is to induce lipid mixing and membrane leakage with lipids of late endosomes. This result is consistent with our results that Tat induced a reversible, hydration-induced increase in mosaic spread (Figs. S1-3) and a disordering of chains (Fig. 8). Both of these could induce lipid mixing and perhaps, membrane leakage. An X-ray, neutron and AFM study reported thickening upon initial Tat binding, in contradiction to our result in Fig. 7B that shows thinning [71]. We suggest that this difference was caused by their using stiff gel phase DPPC lipid that did not allow bound Tat to perturb the bilayer. Using a variety of techniques, including high sensitivity isothermal titration calorimetry and ^{2}H - and ^{31}P - NMR, Seelig et al. [72] presented evidence that the lipid bilayer remains intact upon Tat binding and our results confirm this. Finally, we compare our structural results to those obtained by solid state NMR, although at a lower hydration level than in our sample. Hong et al. [32] found that Tat lies parallel to the bilayer surface in the headgroup region of DMPC/DMPG (8:7) bilayers, similar to our cartoon in Fig. 9.

3.6 Conclusion

Although a recent MD simulation using umbrella sampling [73] found that the free energy required for cR9 to traverse a membrane was smaller if a water pore was present, we cannot directly test the existence of a transient water pore from our

X-ray or neutron scattering experiments. This is because, even with a water pore, the translocation process still requires crossing a free energy barrier which is a non-equilibrium process. X-ray form factors measure an equilibrium state. If the form factors obtained from water pore structures agreed well with experiments, it would indicate that the pore structure was thermodynamically stable. This may be the case for some antimicrobial peptides, but certainly not for cell-penetrating peptides. Finding a kinetically competent pathway for the interesting phenomenon of translocation of highly charged Tat through hydrophobic membranes is difficult. An energetically passive translocation likely occurs very seldom on an MD simulation time scale, and it probably happens quickly, so it would not significantly change the average structure of the membrane in which it occurs. Although our results in this paper do not reveal a kinetically competent pathway, they do show that Tat is drawn to the surface of the membrane, and is therefore ready for translocation at a region of local thinning. And they show that these interactions tend to soften (Fig. 2) and disorder (Fig. 8) the membrane and increase the AL, thereby likely reducing the energy barrier for passive translocation.

Appendices

Bibliography

- [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] Stephanie Tristram-Nagle, Yufeng Liu, Justin Legleiter, and John F. Nagle. Structure of gel phase DMPC determined by x-ray diffraction. *Biophysical Journal*, 83(6):3324 – 3335, 2002.
- [3] 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.
- [4] <http://seal.web.cern.ch/seal/documents/minuit/mnusersguide.pdf>.
- [5] <http://lcgapp.cern.ch/project/cls/work-packages/mathlibs/minuit/index.html>.
- [6] 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.
- [7] 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.
- [8] 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.

- [9] 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.
- [10] Norbert Kuerka, John Katsaras, and JohnF. Nagle. Comparing membrane simulations to scattering experiments: Introducing the simtoexp software. *Journal of Membrane Biology*, 235(1):43–50, 2010.
- [11] http://henke.lbl.gov/optical_constants.
- [12] M.C. Wiener, R.M. Suter, and J.F. Nagle. Structure of the fully hydrated gel phase of dipalmitoylphosphatidylcholine. *Biophysical Journal*, 55(2):315 – 325, 1989.