Supplementary Information for 'Sequence-Dependent Interfacial Adsorption and Permeation of Dipeptides across Phospholipid Membranes'

Chenyu Wei[1,2*] *and Andrew Pohorille*[1,2*]

[1] NASA Ames Research Center, Mail Stop 239-4, Moffett Field, CA 94035, USA

[2] Department of Pharmaceutical Chemistry, University of California San Francisco

* Corresponding authors:

chenyu.wei@nasa.gov; andrew.pohorille@nasa.gov

Table S1. The coordination number ⟨CN⟩ from the nearest neighboring water and the oxygen atoms on lipid headgroup around a selected atom on dipeptide at the POPC membrane interface.

⟨CN⟩ (from water)	Ace-Ser-Leu-NMe	Ace-Phe-Leu-NMe
O (Ser or Phe backbone)	1.0	1.4
O (Leu backbone)	1.1	1.0
N (Ser or Phe backbone)	0.6	0.7
N (Leu backbone)	0.6	0.3
OG (Ser side chain)	1.8	/

⟨CN⟩ (from lipid headgroup)	Ace-Ser-Leu-NMe	Ace-Phe-Leu-NMe
N (Ser or Phe backbone)	0.6	0.3
N (Leu backbone)	0.4	0.1
OG (on Ser side chain)	0.5	/

Table S2. Position dependent diffusion coefficient for the three dipetides inside POPC membrane at various *z* position, which is calculated from autocorrelation function of random forces.[1]

D(x 10^{-6} cm^2/s)	0 Å	-6 Å	-12 Å
Ace-Ser-Ser-NMe	0.45	0.47	0.36
Ace-Ser-Leu-NMe	0.41	0.42	0.22
Ace-Phe-Leu-NMe	0.40	0.60	0.14

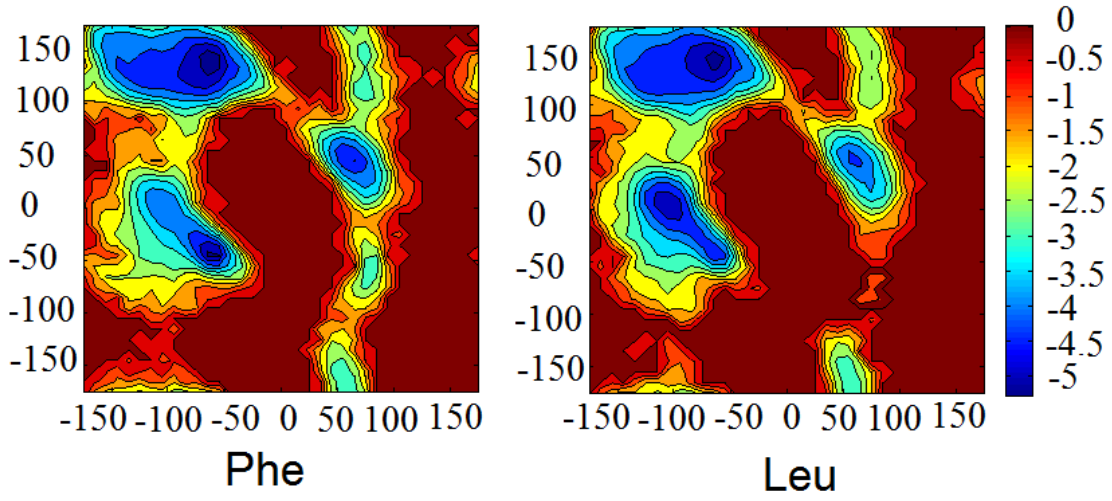


Figure S1. Ramachandran map for the backbone conformation of Ace-Phe-Leu-NMe in aqueous solution.

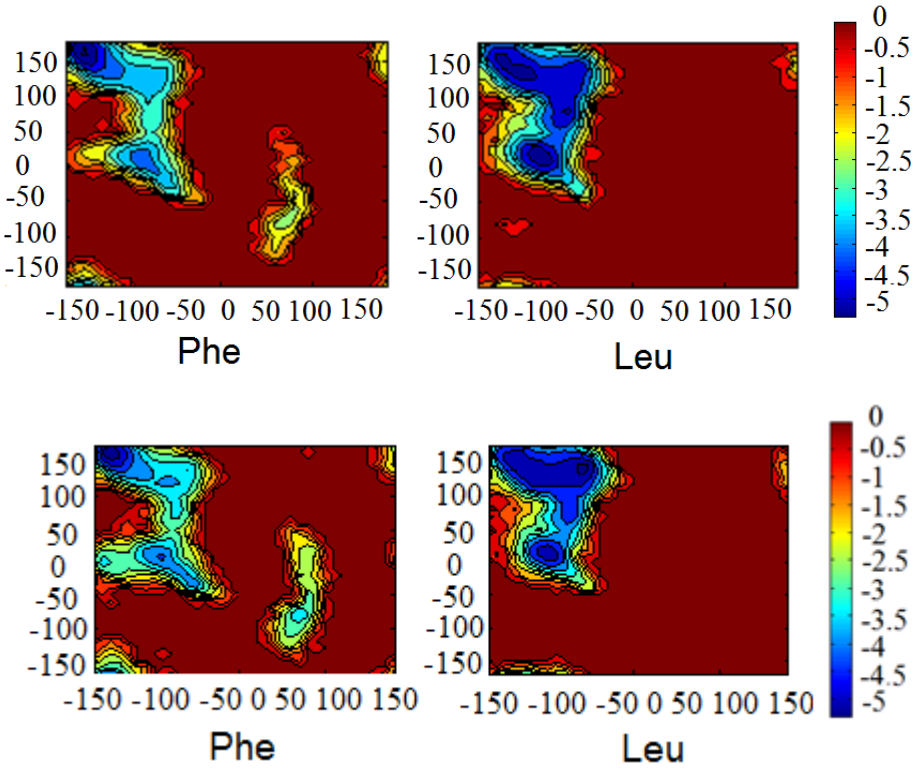


Figure S2. Ramachandran map for the backbone conformation of Ace-Phe-Leu-NMe in decane solution (upper panel) and the center region of POPC membrane (lower panel).

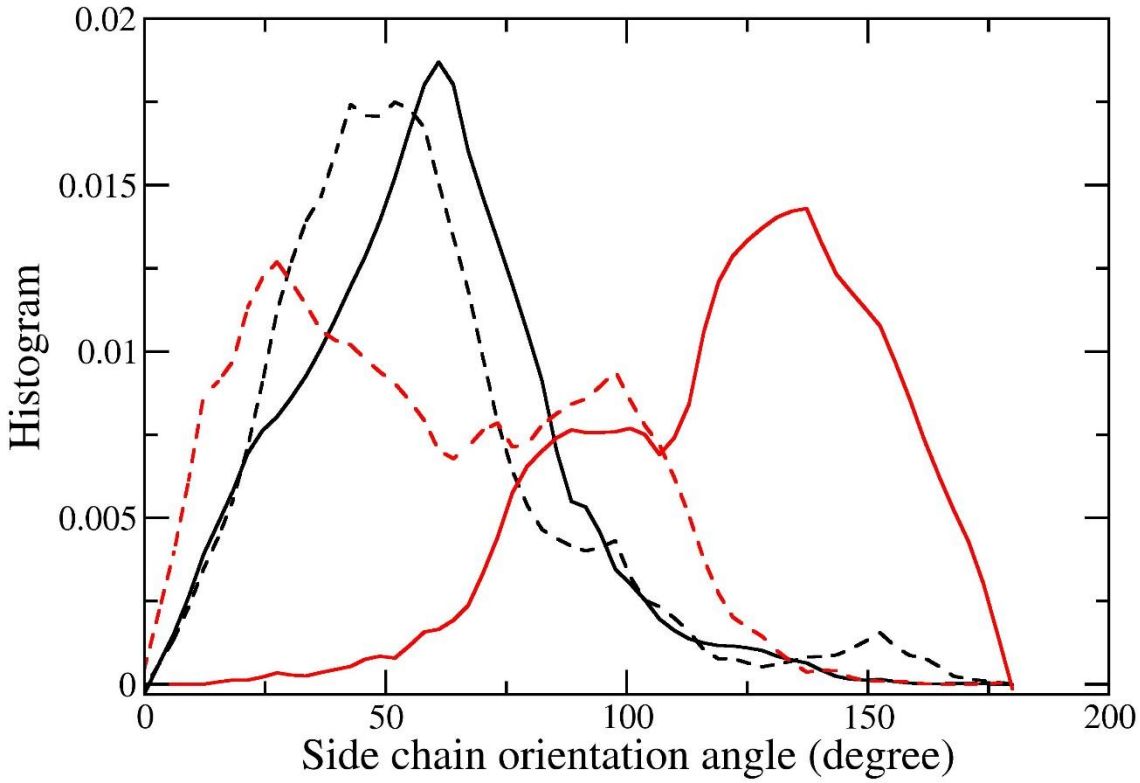
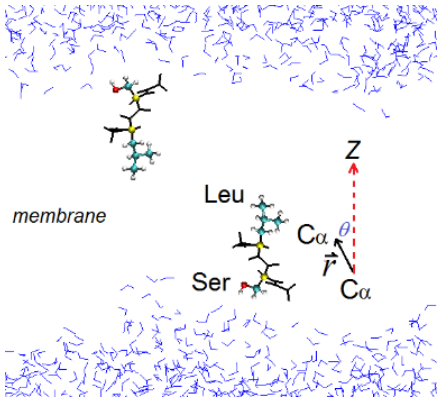
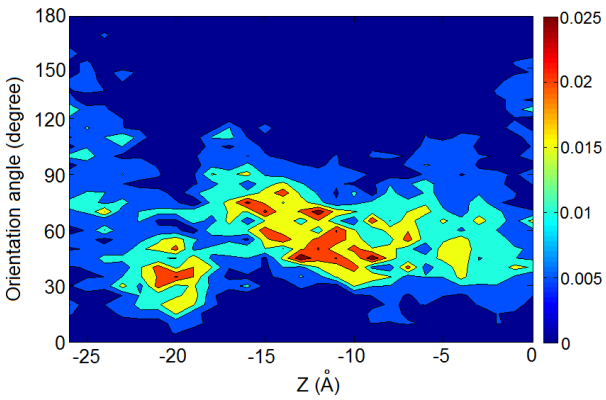


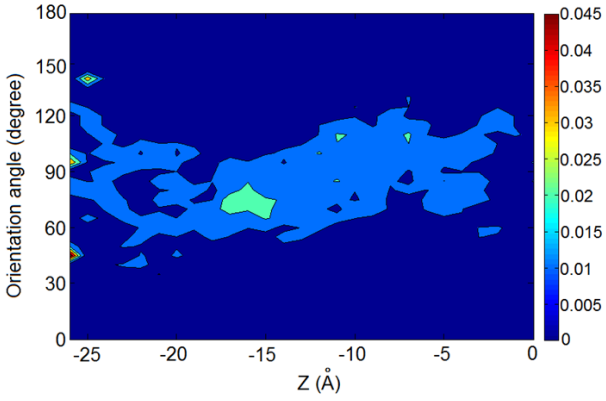
Figure S3. The distribution for the orientation of the side chains on Ace-Phe-Leu-Nme (black) and Ace-Ser-Leu-Nme (red) dipeptide. The solid and dashed black curve is for Phe and Leu residue on Ace-Phe-Leu-Nme, respectively; and the solid and dashed red curve is for Ser residue at N-terminal and C-terminal on Ace-Ser-Ser-Nme, respectively.



(a)



(b)



(c)

Figure S4. (a) The illustration of the orientation angle θ between the vector $\vec{r}_{mol}^{\,Ser-Leu}$ and $\boldsymbol{Z}$ axis. Schematic plot of Ace-Ser-Leu-Nme inside POPC membrane, with Ser side chain in closer promixity to the interface than Leu side chain is also shown (side chains are represented in ball-and-stick model and C_α atom is marked in yellow). (b) 2D distribution of Ace-Ser-Leu-NMe dipeptide orientation as a funtion of the distance z from the membrane center and the orientation angle θ; (c) 2D distribution of Ace-Phe-Leu-NMe dipeptide orientation as a funtion of the distance z from the membrane center and the orientaiton angle θ.

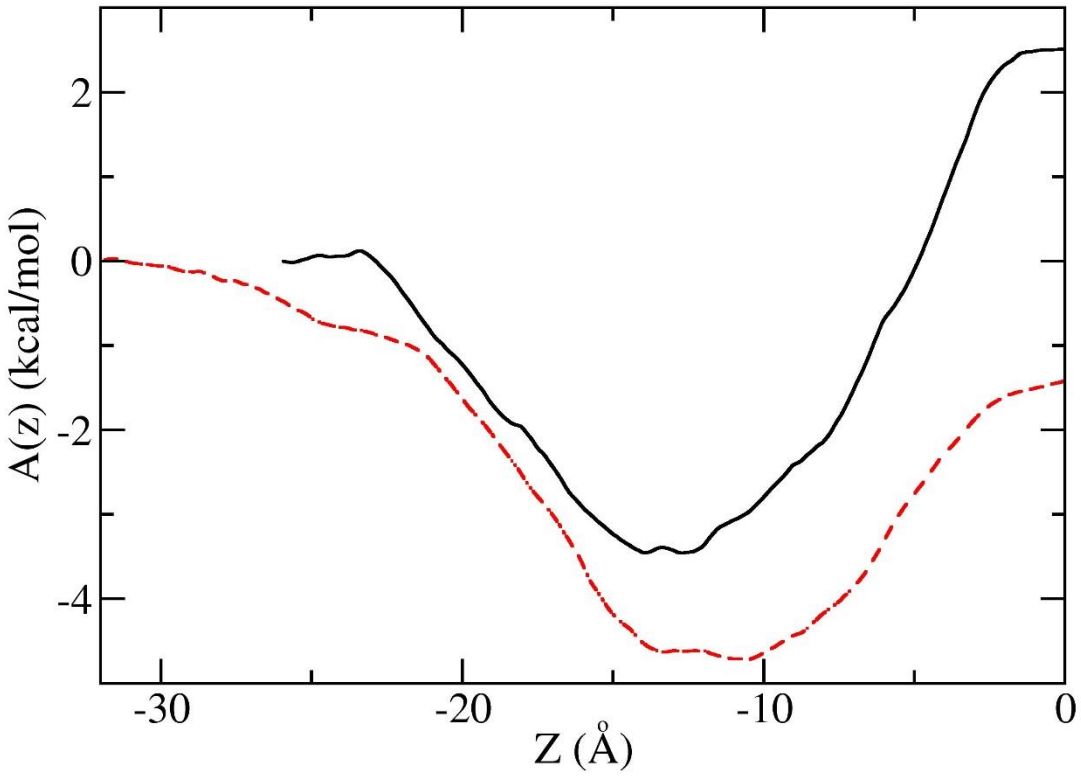


Figure S5. Free energy profile for Ace-Phe-NMe (solid black curve) and 3-methyl indole (red dashed curve) molecule as a function of the distance from the POPC membrane center ($z = 0$)

Bayesian estimate of populations of different states of Ace-Ser-Ser-NMe dipeptide near the middle of the membrane.

As argued in the main text, the probabilities, $p(s_0, z)$, $p(s_+, z)$, and $p(s_-, z)$ of observing the dipeptide in states s_0, s_+, and s_- at z near the center of the bilayer are:

$$p(s_0, z) = \frac{\exp(-\beta \Delta A_0)}{2\cosh(\beta fz) + \exp(-\beta \Delta A_0)} \quad \text{(S1)}$$

$$p(s_+, z) = \frac{\exp(\beta fz)}{2\cosh(\beta fz) + \exp(-\beta \Delta A_0)} \quad \text{(S2)}$$

$$p(s_-, z) = \frac{\exp(-\beta fz)}{2\cosh(\beta fz) + \exp(-\beta \Delta A_0)}, \quad \text{(S3)}$$

where f is the absolute value of the average force acting on the dipeptide in states s_+ and s_- in this range of z, and $\beta = 1/k_B T$. Here and ΔA_0 is a parameter that can be interpreted as the free energy difference between s_0 and s_+ or s_- state at $z = 0$. The free energies of s_+ and s_- are equal at this plane by symmetry.

We wish to find the probability distribution of different values of ΔA_0, $P(\Delta A_0 \mid S)$, given the observed populations, s, of states s_0, s_+, and s_- at n different value of z. To do so, we follow a standard Bayesian approach[2]. The posterior probability, $P(\Delta A_0 \mid M, S)$, where M is the probability model defined in Eq. (S1), Eq. (S2) and Eq. (S3), is expressed as:

$$P(\Delta A_0 \mid M, S) = \frac{P(S \mid \Delta A_0, M) P(\Delta A_0 \mid M)}{P(S \mid M)}. \quad \text{(S4)}$$

If the prior, $P(\Delta A_0 | M)$ is uniform, the posterior becomes proportional to the likelihood function.

$$P(\Delta A_0 \mid M, S) \propto P(S \mid \Delta A_0, M). \quad \text{(S5)}$$

Our goal is to estimate the probability, $P(S \mid \Delta A_0, M)$, of generating data with the observed populations of the three states of interest at the given n values of z

$$P(S \mid \Delta A_0, M) = \prod_{i=1}^{n} P(S_i^0 \mid \Delta A_0, M) P(S_i^+ \mid \Delta A_0, M) P(S_i^- \mid \Delta A_0, M), \quad \text{(S6)}$$

where $P(S_i^0 \mid \Delta A_0, M)$, $P(S_i^+ \mid \Delta A_0, M)$ and $P(S_i^- \mid \Delta A_0, M)$ are the probabilities of generating observed populations of states s_0, s_+ and s_-, respectively, at position z_i given a specific value of ΔA_0. We wish to maximize this function with respect to ΔA_0. Taking advantage of the monotonicity of logarithm, we maximize $g(S, \Delta A_0) = \ln P(S \mid \Delta A_0, M)$ instead.

$$g(S \mid \Delta A_0) = \sum_{i=1}^{n} \left[\ln P(S_i^0 \mid \Delta A_0, M) + \ln P(S_i^+ \mid \Delta A_0, M) + \ln P(S_i^- \mid \Delta A_0, M) \right]. \quad \text{(S7)}$$

Taking advantage of Eq. (S1), Eq. (S2) and Eq. (S3), $g(S, \Delta A_0)$ takes the form:

$$\begin{aligned} g(s, \Delta A_0) = & \sum_{i=1}^{n} \ln \left[\frac{\exp(fz / k_B T)}{2\cosh(fz / k_B T) + \exp(-\Delta A_0 / k_B T)} \right]^{\alpha_i^+} \\ & + \sum_{i=1}^{n} \ln \left[\frac{\exp(-fz / k_B T)}{2\cosh(fz / k_B T) + \exp(-\Delta A_0 / k_B T)} \right]^{\alpha_i^-} \quad \text{(S8)} \\ & + \sum_{i=1}^{n} \ln \left[\frac{\exp(-\Delta A_0 / k_B T)}{2\cosh(fz / k_B T) + \exp(-\Delta A_0 / k_B T)} \right]^{\alpha_i^0}, \end{aligned}$$

where α_i^0, α_i^+ and α_i^- are proportional to the populations of states s_0, s_+ and s_- at z_i. Therefore, $\alpha_i^0 + \alpha_i^+ + \alpha_i^- = C$, where C is a constant independent of z_i. Then

$$g(s,\Delta A_0) = -\left[\sum_{i=1}^{n}\alpha_i^0\right]\beta\Delta A_0 - \sum_{i=1}^{n}\ln\left[2\cosh(\beta f z_i) + \exp(-\beta\Delta A_0)\right] + \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)\beta f z_i. \quad \text{(S9)}$$

In this equation, C was taken as 1, since its specific value has no influence on the shape of g as a function of ΔA_0. The last term in Eq. (S9) is independent of ΔA_0 and, therefore, will contribute only to the normalization constant. This reflects the fact that ΔA_0 in Eq. (S1), Eq. (S2) and Eq. (S3) changes only the proportion of S_0 to $S_+ + S_-$, but not the proportion between S_+ and S_-. The latter is defined by the value of f and the symmetry requirement with respect to the plane at

$z = 0$. Thus, the function that we need to consider is

$$h(S,\Delta A_0) = -\left[\sum_{i=1}^{n}\alpha_i^0\right]\beta\Delta A_0 - \sum_{i=1}^{n}\ln\left[2\cosh(\beta f z_i) + \exp(-\beta\Delta A_0)\right]. \quad \text{(S10)}$$

Then, the posterior probability can be expressed as,

$$P(\Delta A_0 \mid M, S) = \exp\left[h(S,\Delta A_0)\right] / N. \quad \text{(S11)}$$

where N is the normalization constant obtained from integrating $\exp\left[h(S,\Delta A_0)\right]$ over ΔA_0. In this particular instance, the observed populations of S_0 were 0.31, 0.10, 0.38, 0.11, 0.09, 0.12 0.10 and 0.51 for simulations at z equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8 Å, respectively. The length of molecular dynamics trajectory for each value of z was approximately 400 ns. This yielded $P(\Delta A_0 \mid M, S)$ shown in Figure S6. As we can see that this distribution is fairly broad and Gaussian-like, although the positive- ΔA_0 tail decays slower that the negative-ΔA_0 tail. The maximum likelihood ΔA_0 is 0.05 kcal/mol, and the average ΔA_0 is 0.18 kcal/mol with the standard deviation of 0.56 kcal/mol.

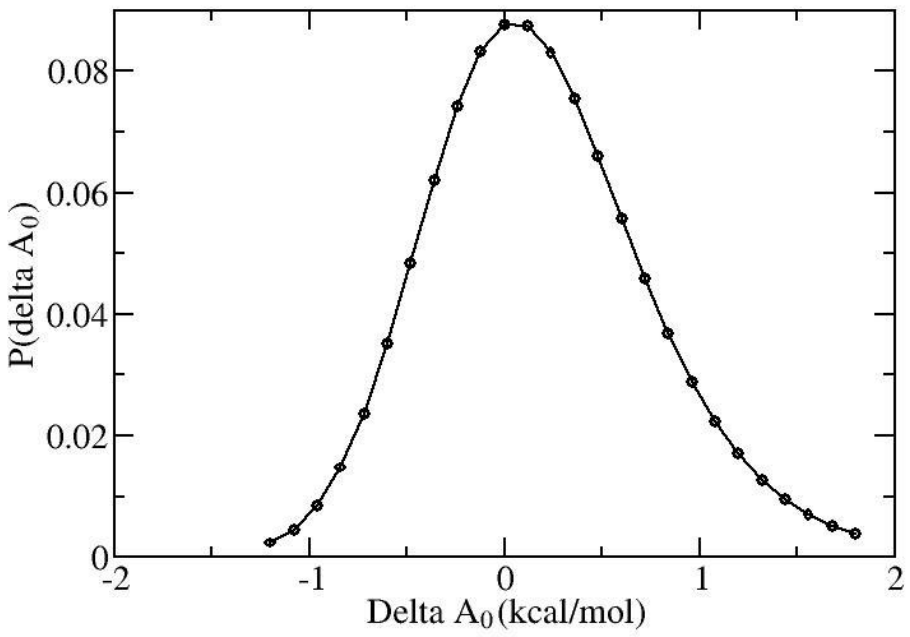


Figure S6. Normalized posterior probability distribution $P(\Delta A_0 \mid M, S)$, as a function of ΔA_0.

References.

(1.) Kubo, R. The Fluctuation-Dissipation Theorem. *Rep. Prog. Phys.* **1966**, *29*, 255–284.

(2.) Sivia, D. S. *Data Analysis. A Bayesian Tutorial.* Claredon Press, Oxford, 1996.