Statistical Convergence of Equilibrium Properties in Simulations of Molecular Solutes Embedded in Lipid Bilayers

Chris Neale,[†,‡] W. F. Drew Bennett,[§] D. Peter Tieleman,[§] and Régis Pomès*[,†,‡]

[†]Molecular Structure and Function, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8

[‡]Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

[§]Department of Biological Sciences, Institute for Biocomplexity and Informatics, University of Calgary, Calgary, Alberta, Canada

ABSTRACT: In recent years, atomistic molecular simulations have become a method of choice for studying the interaction of small molecules, peptides, and proteins with biological membranes. Here, we critically examine the statistical convergence of equilibrium properties in molecular simulations of two amino acid side-chain analogs, leucine and arginine, in the presence of a hydrated phospholipid bilayer. To this end, the convergence of the standard binding free energy for the reversible insertion of the solutes in the bilayer is systematically assessed by evaluating dozens of separate sets of umbrella sampling calculations for a total simulation time exceeding 400 μs. We identify rare and abrupt transitions in bilayer structure as a function of solute insertion depth. These transitions correspond to the slow reorganization of ionic interactions involving zwitterionic phospholipid headgroups when the solutes penetrate the lipid–water interface and when arginine is forced through the bilayer center. These rare events are shown to constitute hidden sampling barriers that limit the rate of convergence of equilibrium properties and result in systematic sampling errors. Our analysis demonstrates that the difficulty of attaining convergence for lipid bilayer-embedded solutes has, in general, been drastically underestimated. This information will assist future studies in improving accuracy by selecting a more appropriate reaction coordinate or by focusing computational resources on those regions of the reaction coordinate that exhibit slow convergence of equilibrium properties.

INTRODUCTION

Biological membranes enable life by maintaining different environments within and between cells, protecting cellular machinery from harsh and dilute external environments,[1] and compartmentalizing eukaryotic cells.[2] The defining component of a biological membrane is the lipid bilayer, a bilamellar sheet of oriented amphipathic lipids that collectively sandwich a hydrophobic interior between two hydrophilic surfaces, thus creating a barrier to the passage of materials.[3] While biological membranes are significantly more complex than neat lipid bilayers,[4,5] understanding the properties of neat lipid bilayers and their interactions with simple molecular solutes is a key step to understanding the properties of biological membranes.

The structure and bulk properties of lipid bilayers have been investigated with theoretical methods including molecular dynamics (MD) simulations.[6,7] In recent years, atomistic MD simulations have been used to compute bilayer properties[8,9] and equilibrium distributions of solutes across lipid bilayers.[10−17] While simulations are useful tools for determining the equilibrium properties of complex systems, all simulation studies are susceptible to sampling errors,[18] particularly when simulation time scales are less than or similar to the autocorrelation times of the degrees of freedom (DOFs) on which observed values depend.[19] In this case, statistical measures such as the mean, μ, and standard deviation, σ, of values observed in simulations change significantly with increased sampling. This time dependence is a hallmark of insufficient sampling. Two types of sampling errors may arise from insufficient sampling: statistical and systematic errors. Statistical sampling errors primarily affect a value's precision, causing oscillations about an average value as the simulation time is increased. Conversely, systematic sampling errors directly affect a value's accuracy. When systematic sampling errors arise from insufficient sampling, statistical values drift unidirectionally toward the correct values as the equilibration time is increased and the systematic sampling errors become smaller.

In some cases, a separation of time scales exists whereby one DOF relaxes much more slowly than all of the other DOFs. One may then conduct a series of simulations restrained to different positions along this slowly relaxing DOF or reaction coordinate, thus reducing the amount of simulation time required for statistical properties to converge. One technique that makes use of this procedure is known as umbrella sampling (US).[20,21] However, the convergence and accuracy of average properties computed from US simulations can also suffer from systematic sampling errors introduced by slow relaxation in DOFs orthogonal to the selected reaction coordinate.[22] The barriers that underlie long autocorrelation times in orthogonal DOFs are referred to as hidden barriers because, although they retard convergence, they are not apparent along the reaction coordinate. Conversely, an explicit barrier lies along the reaction coordinate. A schematic representation of a hidden barrier is presented in Figure 1. In principle, the system outlined in Figure 1 could be sampled much more efficiently if the DOF with the largest autocorrelation time was chosen as the reaction coordinate for US. Unfortunately, optimal reaction coordinates are rarely known *a priori*, and reaction coordinates are thus often selected simply because they are physically intuitive. To compute
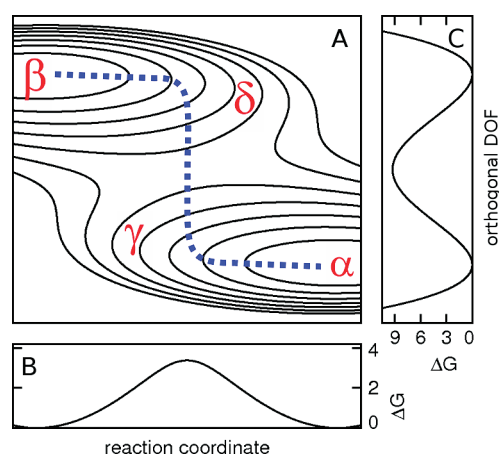
Figure 1. Schematic illustration of a free energy sampling barrier in a degree of freedom (DOF) orthogonal to the reaction coordinate. (A) Isoenergy contour representation of an example two-dimensional free energy surface. Two local minima designated α and β are connected by a dotted line that, as it progresses along the designated reaction coordinate (x axis), follows the path of lowest free energy in an orthogonal DOF (y axis). Two states that do not lie along this path are designated γ and δ. (B) Free energy profile along the reaction coordinate. (C) Free energy profile along the orthogonal DOF.

the equilibrium distribution of molecular solutes in lipid bilayers, one intuitive reaction coordinate is the distance between the solute and the bilayer center along the bilayer normal.

In this study, we examine the statistical convergence of the standard binding free energy for the immersion of two amino acid side-chain analogs in a lipid bilayer. First, we examine the partitioning of a chemically simple hydrophobic solute, methylpropane, the side chain analog of leucine. Second, we consider an amphipathic cationic solute, *n*-propylguanidinium, the side chain analog of arginine. These solutes are small and have few internal DOFs. This simplicity allows us to identify hidden barriers to solute insertion systematically. Moreover, these solutes are biologically relevant moieties whose solvation in a lipid bilayer may be relevant in the context of larger solutes, such as proteins.

For each solute, we use US to calculate the free energy profile or potential of mean force (PMF) governing the solute's axial probability distribution from bulk water to the center of a 1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine (DOPC) bilayer. These simulations extend two earlier studies of the distribution of hexane[13] and the side chain analogs methylpropane and *n*-propylguanidinium.[23] We increase the respective sampling of methylpropane and *n*-propylguanidinium by factors of 80 and 40 and compute the PMF many times using different initial conditions. The increased simulation time and the systematic evaluation of the dependence of the binding free energy on initial conditions, both made possible by increases in available computer power, make it possible to quantify systematic sampling errors. Importantly, our in-depth analysis of statistical sampling convergence in simulations of lipid bilayers identifies the immersion depths at which these side chain analogs are particularly susceptible to systematic sampling errors. We elucidate the molecular underpinnings of these sampling errors and present a comprehensive view of the structural and thermodynamic bases of lipid solvation of these two biologically relevant solutes.

■ THEORY AND METHODS

Simulation Protocol. The simulation systems consisted of methylpropane or *n*-propylguanidinium in hydrated DOPC bilayers. MD simulations were conducted with version 4.0.7 of the GROMACS simulation package.[24] The water model was TIP3P.[25] Methylpropane and *n*-propylguanidinium were modeled by the OPLS-AA/L parameters[26,27] for the side chains of leucine and arginine, respectively, where the α-carbon was replaced by a hydrogen atom and the charge on the β-carbon was adjusted to yield an integral molecular charge.[23] DOPC was modeled by the Berger parameters.[28] For combination with OPLS-AA/L solutes, the Coulombic 1−4 intramolecular interactions of DOPC were reduced to half magnitude in spite of the fact that self-consistent combination of the Berger and OPLS-AA/L parameter sets is now possible using the half-ε double-pairlist method.[29] Lennard-Jones interactions were evaluated using a group-based cutoff and truncated at 1 nm without a smoothing function. Coulomb interactions were calculated using the smooth particle-mesh Ewald method[30,31] with a real-space cutoff of 1 nm and a Fourier grid spacing of 0.12 nm. Simulation in the NpT ensemble was achieved by semi-isotropic coupling to Berendsen barostats[32] at 1 bar with coupling constants of 1 ps and coupling the water, lipids, and solute to three separate Berendsen thermostats[32] at 298 K with coupling constants of 0.1 ps as in the previous study.[23] Bonds involving hydrogen atoms were constrained with SETTLE[33] and P-LINCS[34] for water and other molecules, respectively. The integration time step was 2 fs. The nonbonded pairlist was updated every 20 fs. Coordinates were saved every 10 ps.

System Setup and Umbrella Sampling. A configuration containing a DOPC lipid bilayer with 32 lipids per leaflet was obtained from the methylpropane-in-DOPC simulations of MacCallum et al.[23] To increase the spatial separation of the lipid bilayers across the periodic boundary, the simulation box was elongated along the bilayer normal in the Cartesian **z** dimension. Excess water was added to this newly-created cavity, resulting in a total of 4555 water molecules in the entire system. This DOPC bilayer was simulated for 108 ns in the absence of any solute, during which the area per lipid (APL; see Table 1 for acronyms and symbols used throughout this article) did not drift systematically (data not shown). Over these 108 ns, the APL was 0.644 ($\sigma = 0.01$) nm^2 and the spatial extent of the box along **z** was 10.8 ($\sigma = 0.2$) nm. (Throughout this article, standard deviations of the sample, $\sigma = [1/(N-1)\sum_{i=1}^{N}(v_i - \mu)^2]^{1/2}$, for N values of the sample, v, with mean μ are shown in parentheses while the $\pm$ symbol is reserved for standard deviations of the mean, $\sigma_M = [1/(M-1)\sum_{j=1}^{M}(\mu_j - \bar{\mu})^2]^{1/2}$, for M estimates of the mean, μ, with overall mean $\bar{\mu}$. We have chosen to display the standard deviation of the mean instead of the standard error of the mean because the former reflects the range of values that can be expected for the mean from a single set of US simulations, which is currently the most common application of US.) Coordinates were extracted after 8, 48, and 108 ns for use in US simulations. A solute was embedded in each of these bilayer conformations using the inflategro routine.[35] This solute insertion process was repeated 65 times while varying the axial position of the center of mass (COM) of the solute relative to that of the lipid bilayer from $z = -3.2$ nm to +3.2 nm in 0.1 nm increments, where z represents the solute immersion depth in the bilayer. This entire protocol was then repeated, for each of the three bilayer conformations and each of 65 axial coordinates, after translating the bilayer by

B

dx.doi.org/10.1021/ct200316w |*J. Chem. Theory Comput.* XXXX, XXX, 000–000

Table 1. Selected Acronyms and Symbols Used Throughout This Article

symbol	definition
APL	area per lipid
COM	center of mass
SDF	spatial distribution function
σ_M	standard deviation of the mean
t_{eq}	simulation time discarded as equilibration
$\mathbf{z}$	the Cartesian z axis
z	distance along $\mathbf{z}$ from the bilayer COM to the solute COM (solute immersion depth)
z_i^0	center of the harmonic restraining potential along $\mathbf{z}$ for umbrella i
ΔG_z	free energy as a function of z
ΔG_{bind}°	standard binding free energy
ρ_{H_2O}	water density profile across a neat bilayer
ρ_P	lipid phosphorus density profile across a neat bilayer
N_{wat}	number of water oxygen atoms in the solute's first solvation shell
N_P	number of lipid phosphorus atoms in the solute's first solvation shell
r_P	distance between lipid phosphorus atom and the solute COM in the $\mathbf{xy}$ plane
z_P^{upper}	lipid phosphorus atom height deviation in upper leaflet (not necessarily the proximal leaflet) from their mean position in a neat bilayer
ΔS_{acyl}^{upper}	deviation in acyl chain order parameters in the upper leaflet (not necessarily the proximal leaflet) from their mean values in a neat bilayer
θ	angle (used to report the solute orientation)

half a box length in the bilayer plane. This translation doubled the number of uncorrelated bilayer surfaces used to determine the PMF of solute insertion. In the methylpropane simulations, the solute was embedded in two different initial orientations, with the solute $C_\gamma H_\gamma$ vector aligned to the $\mathbf{z}$ axis or perpendicular to it. Note that solute atomic nomenclature is based on the cognate amino acid side chain. In the n-propylguanidinium simulations, the all-*trans* solute was embedded in three different initial orientations, with the $C_\delta C_\zeta$ vector of the guanidino group pointing either up or down along the bilayer normal or along the bilayer plane. Considering that each bilayer has two leaflets, the methylpropane and n-propylguanidinium simulations yield, respectively, a total of 24 and 36 distinct sets of US simulations for separate evaluations of the PMF from bulk water to the center of the bilayer.

The primary objective of these simulations is to evaluate the rate at which free energies attain convergence in US simulations of solute insertion along the lipid bilayer normal. Thus, we have not evaluated the effect of salt concentration on the mechanism of solute insertion. In addition, we have deliberately avoided a potential source of quasi-nonergodic sampling, namely, the slow equilibration of the distribution of a single counterion such as Cl^-. To this end, we did not add a counterion to the n-propylguanidinium systems, which retain a +1 net charge. Because the n-propylguanidinium PMF is determined not only by strong Coulombic interactions between charged lipid headgroups and the solute but also, indirectly, by the network of ionic interactions between the headgroups themselves, it is unlikely that the addition of a single counterion would have a significant effect on the outcome of this study (see the Results). Although there has been much debate about the protonation state of arginine within a lipid bilayer[36,37] and it is possible to compute pK_a profiles for charged residues from molecular simulations,[38−40] we assume that n-propylguanidinium remains protonated, and we focus on the statistical sampling convergence of the bilayer interacting with n-propylguanidinium in its cationic state.

US simulations were conducted for 205 ns for each initial conformation and under the influence of each restraining potential (umbrella). During these simulations, the solute insertion depth, z, was harmonically restrained to a specified value, z_i^0, in each umbrella i, with a force constant of 3000 kJ/mol/nm², and was stored every 1 ps. While we did not systematically evaluate the effect of varying the spacing or the force constant of our umbrella potentials on the PMF, we did ensure that there was sufficient overlap between adjacent histograms along z to permit the computation of PMFs (data not shown).

Finally, the methylpropane and n-propylguanidinium US simulations of MacCallum et al.[23] were extended to 205 ns per umbrella using pre-existing restart files and GROMACS 4.0.7.

Methodological Comparison to Previous Simulations. The simulations reported in this study were conducted under similar conditions to those of MacCallum et al.,[23] with seven differences: (i) We used GROMACS version 4.0.7[24] in place of version 3.3.1.[41] We also used (ii) a greater number of water molecules, (iii) different starting conformations, and (iv) a different method to embed solute molecules at specified bilayer depths. (v) Our systems contained a single solute molecule rather than the two molecules separated by 3.7 nm along the bilayer normal as employed earlier.[23] (vi) Our n-propylguanidinium simulations did not contain a counterion. Finally, (vii) our simulations were considerably longer. We simulated methylpropane for a total of 160 μs and n-propylguanidinium for a total of 240 μs compared to earlier totals of 2 and 6 μs, respectively.[23] The water model was TIP3P in both studies in spite of the fact that MacCallum et al. indicated using the SPC water model[42] in their publication.[23]

Free Energies and Standard States. The values of z sampled by each solute in the US simulations were converted to PMFs using Alan Grossfield's implementation[43] of the weighted histogram analysis method (WHAM).[44] To this end, recorded values of the solute insertion depth in the range $-3.25 \leq z \leq 3.25$ nm were distributed among 2600 histogram bins, and the WHAM calculation was performed with a tolerance of 1×10^{-5}. This was done separately for each set of US simulations. Each resulting PMF describes the free energy as a function of solute immersion depth, ΔG_z, from bulk water ($z = -3.25$ nm) across the bilayer center ($z = 0$ nm) to bulk water ($z = 3.25$ nm). Exploiting the

symmetry of the system with respect to the $z = 0$ nm plane, we present the PMF for the absolute value of z. Each PMF was then shifted such that the average value of ΔG_z in the range $3 < z \leq 3.25$ nm equaled zero. Finally, the standard binding free energy, $\Delta G^{\circ}_{\mathrm{bind}}$, was determined by trapezoid integration of this PMF according to

$$e^{-\beta \Delta G^{\circ}_{\mathrm{bind}}} = \frac{\int_{z\mathrm{min, bound}}^{z\mathrm{max, bound}} e^{-\beta \Delta G_z}\, \mathrm{d}z}{\int_{z\mathrm{min, unbound}}^{z\mathrm{max, unbound}} e^{-\beta \Delta G_z}\, \mathrm{d}z} \tag{1}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$, k_{B} is the Boltzmann constant, and T is the absolute temperature. Additionally, $z$min,bound and $z$max, bound define the bound state and, likewise, $z$min,unbound and $z$max,unbound define the unbound state. We assign a constant value of 0 kcal/mol to the PMF in the unbound state. Equation 1 is integrated over the range Δz:

$$\begin{aligned} \Delta z &= z\mathrm{max, bound} - z\mathrm{min, bound} \\ &= z\mathrm{max, unbound} - z\mathrm{min, unbound} \end{aligned} \tag{2}$$

which ensures that the available volume is equal in the bound and unbound states,[45] thus calculating $\Delta G^{\circ}_{\mathrm{bind}}$ on the basis of a volume-fraction partition coefficient.[46] Integrating over a single leaflet of the bilayer, we set $z$min,bound = 0 nm and $z$max,bound = 3 nm, which is beyond the point at which the solute becomes fully hydrated and the mean force acting on it vanishes (see the Results). Equation 2 ensures that eq 1 provides a standard-state evaluation of the binding free energy, since the relative bound and unbound probabilities depend on the volume of space available to the solute in these respective states. Thus, doubling the thickness of the water layer in the integral appearing in the denominator of eq 1 would double the population of the unbound state. Note that since we are integrating over z, this volume dependence reduces to a length dependence.

As shown above, the PMF allows definition of the bound state in a rational manner, by placing the boundary at the point where the mean force becomes zero. Nevertheless, sampling errors can lead to different definitions of the bound state in repeat simulations. While an increase in Δz has the same additive effect on the probabilities of the bound and unbound states, the already large probability of the bound state (see the Results) dictates that enlargement of the bound and unbound state definitions affects the unbound state integral (denominator in eq 1) much more dramatically than the bound state integral (numerator in eq 1). Considering the range of possible definitions of $z$max,bound based on the point where the mean force becomes zero (3.0 nm in this work and 3.7 nm in MacCallum et al.[23]), $\Delta G^{\circ}_{\mathrm{bind}}$ has an additional uncertainty component of $\beta^{-1} \ln(3.7/3.0) = 0.12$ kcal/mol. Confining the bound state to a hydrophobic length of 2.0 nm per leaflet, as some authors have done,[47] increases this uncertainty to $\beta^{-1} \ln(3.7/2.0) = 0.36$ kcal/mol, and even larger uncertainties are possible with larger solutes for which the mean force becomes zero at very large values of $z$max,bound.

Unrestrained Simulations. To assess the binding mechanism directly, we conducted 10 unrestrained simulations of methylpropane in the presence of a DOPC bilayer. To this end, initial coordinates of the hydrated bilayer were extracted from the 108-ns simulation of a neat DOPC bilayer, and 10 methylpropane molecules were placed in bulk water. After energy minimization, 10 205-ns MD trajectories with different initial

velocities were generated in the absence of restraints. Note that these simulations were not used to evaluate the binding free energy.

Additional US Simulations at the Lipid−Water Interface. To probe for the presence of a hidden sampling barrier at the lipid−water interface, we conducted 20 additional 180-ns US simulations of methylpropane with an umbrella at $z_i^0 = 2.0$ nm. Ten distinct starting conformations were drawn from unrestrained simulations in which a methylpropane molecule located near $z = 2.0$ nm subsequently progressed deeply into the hydrophobic interior. An additional 10 starting conformations in which methylpropane molecules located near $z = 2.0$ nm did not subsequently bind the bilayer but instead diffused back into bulk water were also selected.

Data Analysis. The number of water molecules in the first solvation shell of the solutes, $N_{\mathrm{H_2O}}$, was calculated on the basis of a heavy-atom cutoff distance of 0.435 nm. This distance corresponds to the first minimum in the radial distribution function (RDF) of water oxygen atoms around methylpropane heavy atoms in aqueous solution (data not shown). The cutoff for determining the number of phosphorus atoms interacting closely with the solutes, N_{P}, was 0.585 nm. Because the corresponding RDF is complex (data not shown), we chose to simply increase the aqueous cutoff by 0.15 nm, approximately the length of the P−O bond in the lipid phosphate group. Axial density profiles for water oxygen atoms, $\rho_{\mathrm{H_2O}}$, and for lipid headgroup phosphorus atoms, ρ_{P}, were calculated from the 108-ns simulation of a neat DOPC bilayer with the GROMACS g_density tool (correcting for fluctuations in the length of the box along z by modifying the g_density algorithm to bin the data outward from the bilayer center) after centering the bilayer along z using the GROMACS trjconv tool (note that the centering algorithm was modified to center the COM rather than the mean of the maximum and minimum values). For ease of comparison, density profiles were scaled such that their maximum value is numerically similar to the maximum $N_{\mathrm{H_2O}}$ for that solute. The distances in the **xy** plane between the solute COM and each headgroup phosphorus atom, r_{P}, were measured with the GROMACS g_dist tool. The distances along **z** between the bilayer COM and the headgroup phosphorus atom of lipids in the upper leaflet were measured with the GROMACS g_dist tool, and deviations of the measured values from the mean distance in a neat DOPC bilayer (2.01 nm) are represented by $z_{\mathrm{P}}^{\mathrm{upper}}$, which, for some figures, was computed only for lipids with $r_{\mathrm{P}} \leq 1$ nm. Order parameters were computed for all saturated nonterminal acyl-chain carbon atoms by reconstructing hydrogen atom positions assuming tetrahedral geometry using the GROMACS g_order tool, as outlined previously,[7] for those lipids with $r_{\mathrm{P}} \leq 1$ nm. These order parameters were averaged for each chain position in each simulation, and the deviations of these mean values from reference values obtained from simulations of a neat bilayer were computed for each chain position and then grouped together as $\Delta S_{\mathrm{acyl}}^{\mathrm{upper}}$.

As we did for PMFs, most data are shown as a function of the absolute value of z in order to enhance statistics. The only exception is the solute orientation, which was evaluated as the angle, θ, between the positive bilayer normal and the $C_\gamma H_\gamma$ vector of methylpropane or the $C_\delta C_\zeta$ vector of n-propylguanidinium, and was processed as a function of the signed value of z to provide a measure of convergence.

Spatial distribution functions (SDFs) were created with the GROMACS g_spatial tool with a bin width of 0.05 nm after centering the bilayer COM at **z** = 0 nm and the solute COM
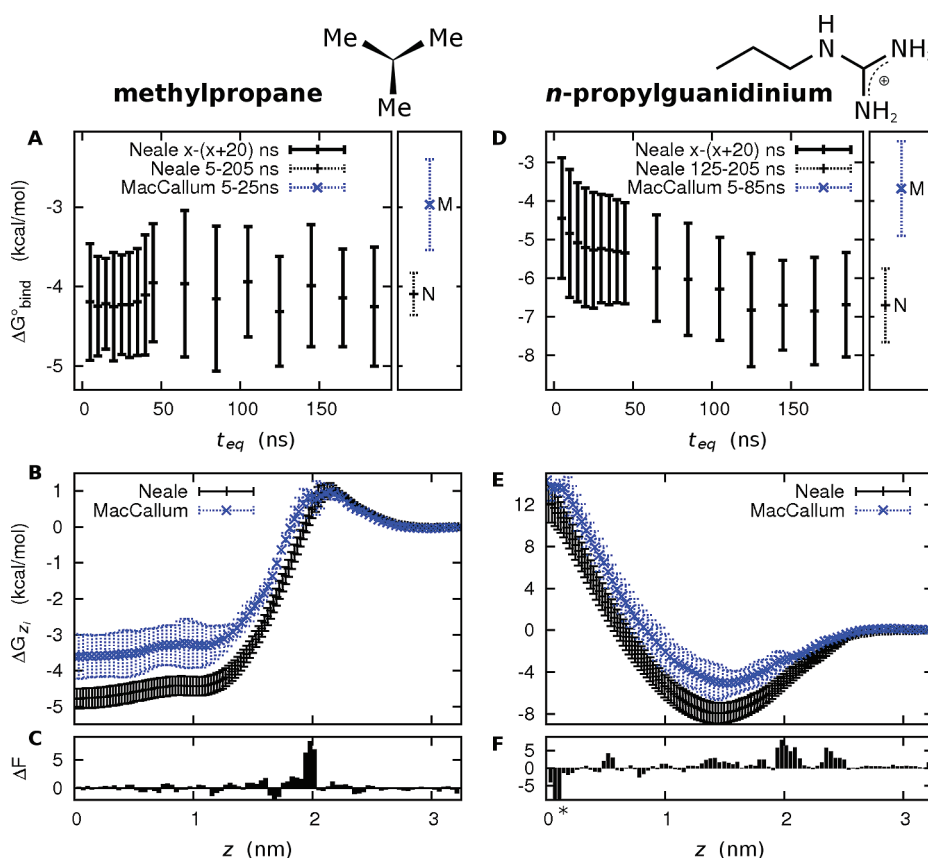
D

dx.doi.org/10.1021/ct200316w |*J. Chem. Theory Comput.* XXXX, XXX, 000–000

Figure 2. Free energies of (left) methylpropane and (right) n-propylguanidinium interacting with a DOPC bilayer. Chemical structures of each solute are displayed beside the titles. (A and D) ΔG°_{bind} and the standard deviation of the mean, σ_M, based on a 20 ns sample per umbrella after discarding an increasing amount of simulation time, t_{eq}, as equilibration. The two values with dotted error bars represent ΔG°_{bind} from full-production sampling in (black "+" labeled "N") this work and (blue "x" labeled "M") MacCallum et al.[23] (refer to Theory and Methods for the definition of full-production sampling). (B and E) The PMF and its σ_M for the solute from bulk water to the bilayer center from full-production sampling in (black "+" with solid error bars) this work and (blue "x" with broken error bars) MacCallum et al.[23] (C and F) The slope of the PMF from MacCallum et al.[23] is subtracted from the slope of the PMF calculated in this work to yield ΔF in kcal/mol/nm. (∗) In part F, ΔF values extend to -20 kcal/mol/nm at $z = 0.1$ nm.

at $x = y = 0$ nm and then, to enhance resolution, repetitively concatenating the simulation trajectory data rotated about z from 10° through 360° in 10° increments. Molecular visualizations were prepared with VMD.[48] In figure captions, the term "full-production sampling" denotes the following sample per umbrella: for methylpropane, 5−205 ns in this work and 5−25 ns in MacCallum et al.,[23] and, for n-propylguanidinium, 125−205 ns in this work and 5−85 ns in MacCallum et al.[23] All error bars in figures represent σ_M, the standard deviation of the mean.

■ RESULTS

Solute Distribution Across a DOPC Bilayer. In this section, we analyze the statistical convergence of the standard binding free energy, ΔG°_{bind}, for the insertion of methylpropane and n-propylguanidinium in a DOPC bilayer based on PMF profiles along the bilayer normal. The PMF profiles were computed from US simulations with 205 ns at each of 33 umbrellas from bulk water to the bilayer center, repeated 24 and 36 times for methylpropane and n-propylguanidinium, respectively. These results are compared to those of previous US simulations with 25 and 85 ns per umbrella for methylpropane and n-propylguanidinium, respectively, each performed twice.[23] ΔG°_{bind} was computed by integrating the PMF in the standard state (see Theory and Methods). To

select the sampling time range on which the PMFs were based, we used block averaging[49] to identify the duration of visible initial systematic sampling errors from our simulations and excluded that data from the computation of our PMFs. The PMFs from previous studies were taken from ref 23 without any modification.

Methylpropane. The value of ΔG°_{bind} for methylpropane binding to a DOPC bilayer is shown as a function of equilibration time, t_{eq}, in Figure 2A. The average ΔG°_{bind} computed on the basis of 20 ns per umbrella remains constant at -4.12 ± 0.14 kcal/mol with increasing t_{eq} between 5 and 185 ns (Table 2 and Figure 2A). Accordingly, the value of ΔG°_{bind} computed on the basis of 5−205 ns per umbrella, -4.10 ± 0.27 kcal/mol, is within this range (Table 2 and Figure 2A). In comparison, the average value of ΔG°_{bind} determined from the two previous simulations with less sampling and different starting conformations,[23] -2.97 ± 0.57 kcal/mol, differs from the average value obtained in this study by 1.1 kcal/mol (Table 2 and Figure 2A), and both absolute values of ΔG°_{bind} from the two previous simulations are smaller than the smallest of 24 independent evaluations of ΔG°_{bind} obtained in this work (Table 2). Importantly, the extension of the previously published simulations yields $\Delta G^{\circ}_{bind} = -3.96 \pm 0.29$ kcal/mol when computed on the basis of 5−205 ns per umbrella, comparable with the average value of ΔG°_{bind} obtained from 24 PMFs of similar length in this work (Table 2).

Table 2. Standard Binding Free Energies of Methylpropane and *n*-Propylguanidinium to a DOPC Bilayer

sample (ns)	$\Delta G^{\circ}_{\text{bind}}$ (kcal/mol)	
	$\mu \pm \sigma_M$	component values
(i) methylpropane		
$\{x \rightarrow (x+20)\}$ $x = \{5, 25, 45, ...,185\}$	-4.12 ± 0.14 [a]	not shown
$5 \rightarrow 205$	-4.10 ± 0.27	$-4.80, -4.41, -4.39, -4.37, -4.32, -4.29, -4.28, -4.22, -4.22, -4.21, -4.16, -4.16,$
		$-4.07, -4.02, -4.01, -3.99, -3.87, -3.87, -3.86, -3.83, -3.82, -3.75, -3.71, -3.70$
$5 \rightarrow 25$ [c]	-2.97 ± 0.57	$-3.37, -2.57$
$5 \rightarrow 205$ [d]	-3.96 ± 0.29	$-4.16, -3.76$
(ii) *n*-propylguanidinium		
$\{x \rightarrow (x+20)\}$ $x = \{125, 145, 165, 185\}$	-6.77 ± 0.08 [b]	not shown
$125 \rightarrow 205$	-6.71 ± 0.96	$-9.12, -9.03, -8.87, -8.56, -8.25, -8.11, -8.00, -7.98, -7.92, -7.51, -7.49, -7.47,$
		$-7.39, -7.10, -7.07, -7.04, -6.90, -6.75, -6.57, -6.47, -6.47, -6.31, -6.29, -5.96,$
		$-5.86, -5.85, -5.80, -5.66, -5.49, -5.43, -5.39, -5.33, -4.89, -4.39, -4.32, -4.00$
$5 \rightarrow 85$ [c]	-3.68 ± 1.23	$-4.56, -2.81$
$125 \rightarrow 205$ [d]	-4.69 ± 0.01	$-4.70, -4.69$

[a] Here, σ_M is computed from 10 blocks of 20 ns, each of which is first averaged over 24 PMFs. [b] Here, σ_M is computed from 4 blocks of 20 ns, each of which is first averaged over 36 PMFs. [c] MacCallum et al. [23] (2 PMFs). [d] Simulations extended from MacCallum et al. [23] (2 PMFs).

The largest difference between the PMFs for methylpropane insertion into a DOPC bilayer derived in the present and the previous study[23] occurs for solute insertion depths $1.85 \le z \le 2.05$ nm (Figure 2B,C), which correspond to the lipid−water interface (see below). At these depths, the PMF computed in this study smoothly decreases toward the global minimum located at the bilayer center, whereas the previous PMF does not.

n-Propylguanidinium. The value of $\Delta G^{\circ}_{\text{bind}}$ for *n*-propylguanidinium binding to a DOPC bilayer is shown as a function of t_{eq} in Figure 2D. The value of $\Delta G^{\circ}_{\text{bind}}$ drifts systematically as t_{eq} increases between 5 and 125 ns per umbrella. It is only after discarding the initial 125 ns of simulation that the value of $\Delta G^{\circ}_{\text{bind}}$ computed on the basis of 20 ns per umbrella converges to -6.77 ± 0.08 kcal/mol (Table 2 and Figure 2D). This estimate is consistent with the value of -6.71 ± 0.96 kcal/mol based on 125−205 ns per umbrella (Table 2 and Figure 2D). The estimate of $\Delta G^{\circ}_{\text{bind}}$ determined from the previous simulations,[23] -3.68 ± 1.23 kcal/mol, differs from the average value obtained in this study by 3 kcal/mol (Table 2 and Figure 2D). The extension of the previous simulations yields an estimate of $\Delta G^{\circ}_{\text{bind}} = -4.69 \pm 0.01$ kcal/mol based on 125−205 ns per umbrella. This value still differs by 2 kcal/mol from the average value of $\Delta G^{\circ}_{\text{bind}}$ obtained from the 36 PMFs based on new simulations of similar length (Table 2).

The global minimum of the PMF is located at insertion depths $1.4 \le z \le 1.5$ nm (Figure 2E). The largest differences with previous simulations occur near insertion depths $z = 2.4, 2.0$, and 0.1 nm (Figure 2E,F). At these depths, the slope of the new PMF profile is steeper than previously computed.[23]

To ensure that our results were not influenced by pressure artifacts resulting from simulating a periodic system with net charge,[50] we evaluated the APL from simulations in which either methylpropane or *n*-propylguanidinium was restrained in bulk water at $z_i^0 = \pm 3.2$ nm based on a time-range of 125−205 ns per umbrella. The APL values were 0.642 ± 0.006 ($\sigma = 0.014$) nm^2 and 0.641 ± 0.005 ($\sigma = 0.014$) nm^2 for methylpropane and for *n*-propylguanidinium, respectively. It appears, therefore, that the APL in these simulations was not affected by the presence of a net charge.

Bilayer Reorganization. In this section, we analyze the structural properties underlying the slow convergence of $\Delta G^{\circ}_{\text{bind}}$ identified in Figure 2. Solute insertion induces two types of structural perturbation of the lipid bilayer, which involve either (i) depression or (ii) protrusion of the bilayer surface. Depression occurs when the phosphate groups of lipids near the solute retract toward the bilayer center (Figure 3B,C,G,H) as acyl chains become more disordered (Figure 3D,I). Depression of the bilayer surface leaves the solute hydrated to a greater extent than expected on the basis of the water density profile across a neat bilayer, $\rho_{\text{H}_2\text{O}}$ (Figure 3A,F). Conversely, a protrusion of the bilayer surface occurs when the phosphate groups of lipids near the solute extend away from the bilayer center (Figure 3B,C,G,H) as acyl chains become more ordered (Figure 3D,I) and, as a result, the solute is hydrated to a lesser extent than expected on the basis of $\rho_{\text{H}_2\text{O}}$ (Figure 3A,F). These patterns of bilayer reorganization are apparent in the time-averaged spatial distribution functions (SDFs) depicted in Figure 4 and the simulation snapshots displayed in Figure 5.

To describe bilayer reorganization as a function of z, we conceptually divide a neat DOPC bilayer into four regions: (i) bulk water for $z > 2.5$ nm, (ii) the aqueous interface between water and phosphate groups spanning $2.0 < z \le 2.5$ nm, (iii) the dry transition region between phosphate groups and acyl chains spanning $1.5 < z \le 2.0$ nm, and finally, (iv) the hydrophobic region of the bilayer for $z \le 1.5$ nm. Although this model differs from previous decompositions[51] of the axial distribution of hydrated bilayers, it is pertinent to the present analysis.

Methylpropane. When methylpropane resides at the aqueous interface, the bilayer surface excludes the solute by forming a depression (Figure 3A−D). As a corollary, the number of lipid headgroup phosphorus atoms that closely interact with the solute is maximal in this region (Figure 3A). This type of lipid reorganization is depicted by the SDF in Figure 4A.

By contrast, when methylpropane is located in the dry transition region, the bilayer surface protrudes to include the solute (Figures 3A−D), as depicted by the SDF in Figure 4B. Thus, the local depression of the water−headgroup interface induced by the presence of the hydrophobic solute at the aqueous interface (Figures 4A, 5A) becomes a protrusion (Figures 4B, 5B) once the solute pierces the surface formed
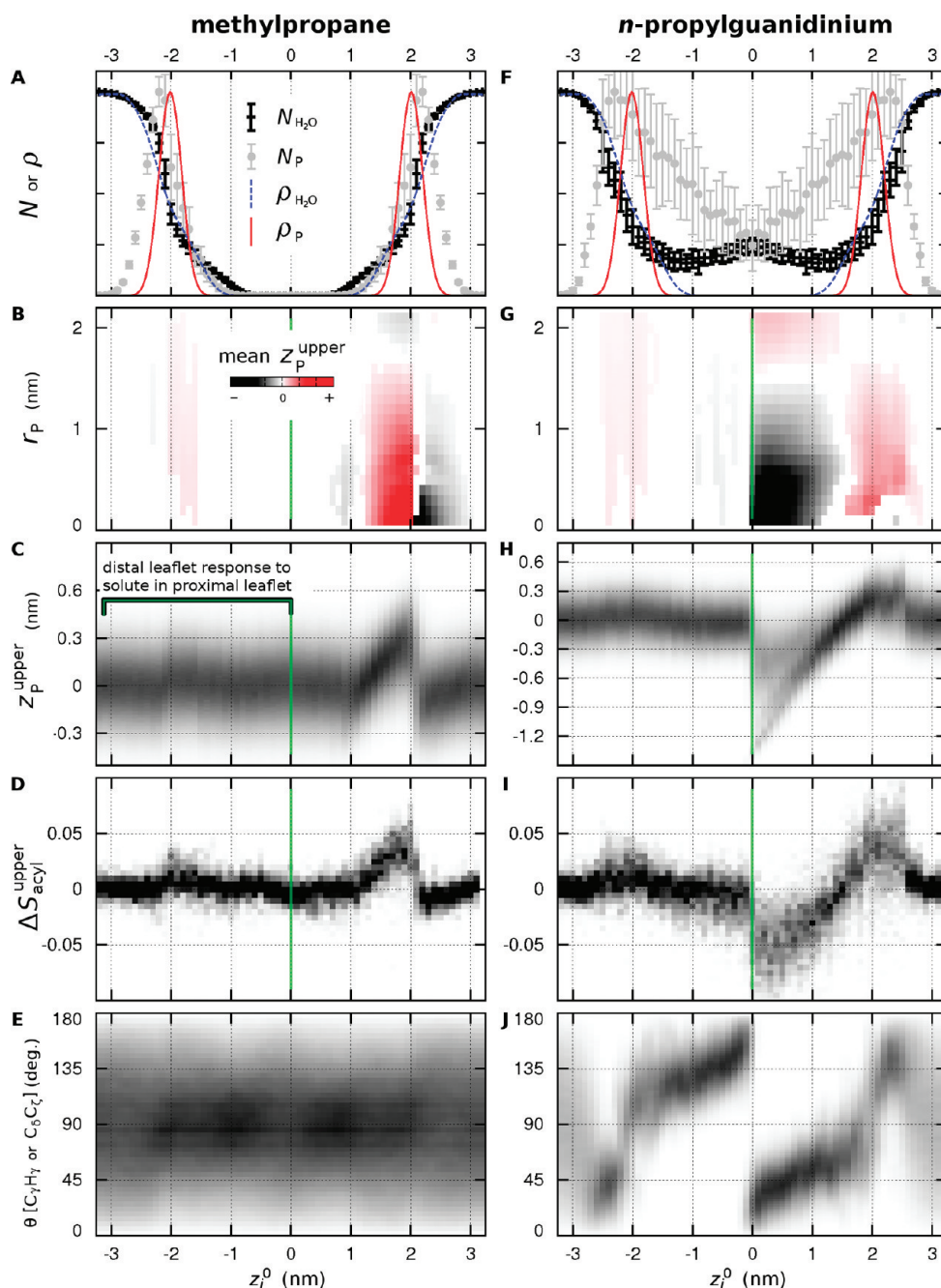
Figure 3. Structural reorganization of the bilayer as a function of restraint position, z_i^0, for (left) methylpropane and (right) n-propylguanidinium partitioning into a DOPC bilayer via US. (A and F). Solvation of the solute by (black error bars) water, N_{H_2O}, and (gray error bars) lipid phosphorus atoms, N_P. Overlaid on this plot are the density profiles in a neat DOPC bilayer for (broken blue line) water, ρ_{H_2O}, and (solid red line) lipid phosphorus atoms, ρ_P. All profiles are scaled to the same maximum value. (B and G) Mean values of lipid height deviation in the upper leaflet, z_P^{upper}, are plotted for (red) positive and (black) negative displacements against (x axis) z_i^0 and (y axis) proximity of the lipid to the solute in the bilayer plane, r_P. The values of z_P^{upper} in B and G were filtered by setting the mean value of $z_P^{upper} = 0$ whenever the standard deviation of the mean value of z_P^{upper} included zero. (C and H) Probability distributions of z_P^{upper} as a function of z_i^0 for lipids with $r_P < 1$ nm. (D and I) Distributions of mean values of the deviation in acyl chain order parameters in the upper leaflet from their mean values in a neat bilayer, ΔS_{acyl}^{upper}, as a function of z_i^0 for lipids with $r_P < 1$ nm. (E and J) Probability distributions of solute orientation, θ, as a function of z_i^0. In parts B-D and G-I, a vertical green line at $z_i^0 = 0$ nm indicates that the z axes of parts B and G and the y axes of parts C, D, H, and I represent changes that occur in a single leaflet as the solute traverses that leaflet (positive x axis values) and the opposing leaflet (negative x axis values). Histograms in parts C—E and H—J were normalized within each z_i^0 value.

by zwitterionic interactions involving phosphate (and choline) groups (Figure 3B,C).

Finally, as methylpropane enters the hydrophobic region of the bilayer, near $z = 1$ nm, it introduces a packing defect that draws water molecules to a region that is normally dehydrated in a neat DOPC bilayer (Figure 3A), resulting in a slight retraction of nearby lipids toward the bilayer center (Figure 3B,C). Methylpropane is entirely dehydrated only for $z < 0.7$ nm (Figure 3A),
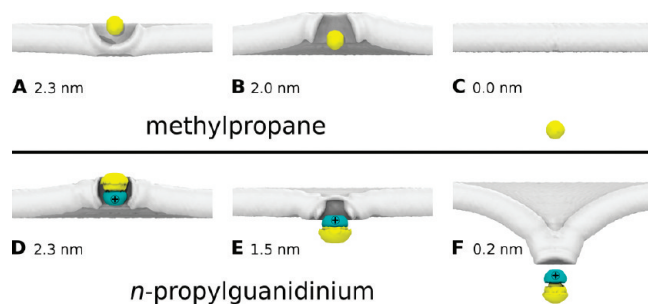
methylpropane

A 2.3 nm   **B** 2.0 nm   **C** 0.0 nm

methylpropane

D 2.3 nm   **E** 1.5 nm   **F** 0.2 nm

n-propylguanidinium

Figure 4. Spatial distribution functions (SDFs) of the solute and the proximal bilayer leaflet viewed from the side, looking along the plane of the bilayer. The SDFs depict (gray) phosphorus atoms in the proximal leaflet and (yellow) hydrophobic and (cyan) hydrophilic moieties of the solute for (A−C) methylpropane and (D−F) n-propylguanidinium. The z_i^0 value is reported alongside each SDF. Images are aligned vertically according to the bilayer COM. To facilitate solute visualization, the SDF of the phosphorus atoms depicts only the density that lies behind the COM of the solute.

where the solute does not affect the localization of the headgroups above it (Figures 3B, 3C, 4C, 5C).

Throughout the bilayer, methylpropane adopts nearly isotropic orientations whereby the angle between the positive bilayer normal and the $C_\gamma H_\gamma$ vector, θ, is sampled with a probability proportional to $\sin(\theta)$ (Figure 3E). There is a slight population shift toward orientations in which the $C_\gamma H_\gamma$ vector is parallel to the global bilayer plane when methylpropane is embedded in the bilayer, except at the bilayer center.

n-Propylguanidinium. When n-propylguanidinium resides at the aqueous interface, the bilayer surface protrudes to interact with the solute (Figure 3F−I). This reorganization is depicted by the time-averaged SDF in Figure 4D and illustrated by the snapshot in Figure 5D. In the aqueous limit of this region, at $z = 2.5$ nm, n-propylguanidinium is strongly oriented with the guanidino group pointing toward the bilayer center (Figure 3J). As the solute reaches $z = 2.0$ nm, a depth that corresponds to the peak in axial phosphate density in a neat DOPC bilayer (Figure 3F), the orientation of the solute is reversed, with the alkyl chain inserted in the nonpolar core and the guanidino group pointing away from the bilayer center (Figure 3J).

As n-propylguanidinium moves from the dry transition region to an immersion depth that corresponds to the hydrophobic core of a neat DOPC bilayer, the bilayer forms a depression that facilitates the continued interaction of the bilayer surface with the solute (Figure 3F−I). This reorganization is evident in the snapshot shown in Figure 5E and in the SDF depicted in Figure 4E, which correspond to the global minimum of the PMF, at $1.4 \leq z \leq 1.5$ nm (Figure 2E). The magnitude of the structural perturbation of the bilayer grows with deeper solute immersion (Figure 3F−I). Throughout the hydrophobic region of the bilayer, n-propylguanidinium remains partially hydrated and solvated by lipid headgroups (Figure 3F) and is strongly oriented with the guanidino group directed outward (Figure 3J). The extreme reorganization of the bilayer structure that occurs as the solute nears the bilayer center is evident in Figure 4F.

When n-propylguanidinium is restrained near the center of the bilayer, at $z = 0$ or 0.1 nm, bulk water may penetrate either leaflet (data not shown) and both polarized orientations are populated (Figure 3J).
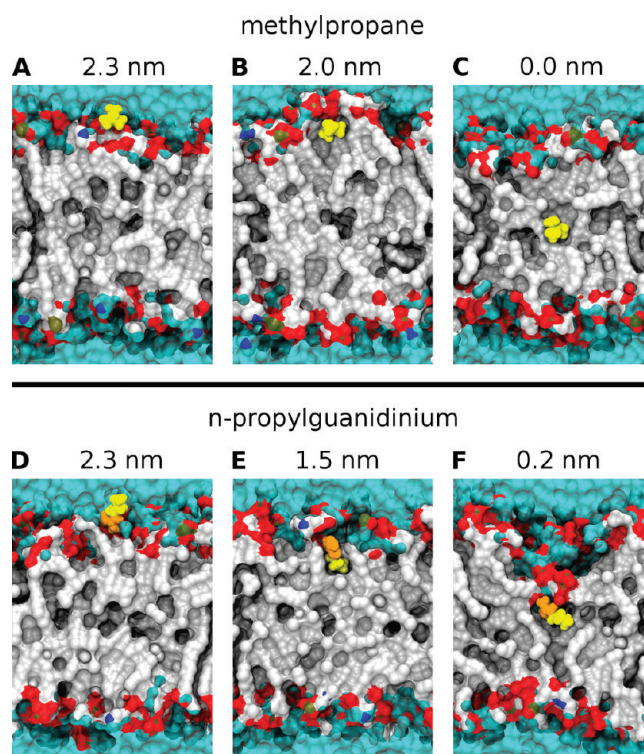

Figure 5. Snapshots of (A−C) methylpropane and (D−F) n-propylguanidinium after 200 ns of simulation. The z_i^0 value is reported alongside each snapshot. Color is added for (cyan) water, (white) lipid carbon, (red) lipid oxygen, (brown) lipid phosphorus, (blue) lipid nitrogen, (yellow) solute hydrophobic moiety, and (orange) solute hydrophilic moiety.

Structural Perturbations of the Opposing Leaflet. Significant changes in bilayer organization take place in both leaflets as the molecular solutes are embedded. We refer to the leaflet in which the solute is embedded as the proximal leaflet and to the opposing leaflet as the distal leaflet. As described above, the reorganization of the proximal leaflet predominantly occurs in the vicinity of the solute. In contrast, the reorganization of the distal leaflet is more global, with uniform displacements of the water−bilayer interface (Figure 3B,G) and uniform changes in order parameters (data not shown). As a result, the bilayer does not bend. Note that global bending modes of the bilayer may be inhibited to some extent by the presence of periodic boundary conditions in the **x** and **y** dimensions. Instead, local depressions and protrusions of the proximal interface induce respective increases in disorder and order of lipids throughout the distal leaflet (Figure 3D,I). These compensating changes in lipid ordering result in thinning and thickening of the distal leaflet, respectively, both of which uniformly displace the distal interface (Figure 3B,G). Compensating effects also extend to more distant regions of the proximal leaflet (Figure 3B,G) whenever the solute is restrained to values of z at which the PMF has a large slope (see Figure 2B,E). Throughout these processes, the change in average area per lipid remains smaller than 0.013 nm^2 (data not shown).

Effect of Restraints on the Sampling of Bilayer Perturbations. To assess the relevance of the US ensembles depicted in Figure 4 to the mechanism governing the partitioning of methylpropane into the bilayer, we conducted nonequilibrium simulations in the absence of restraints. Ten simulations, each
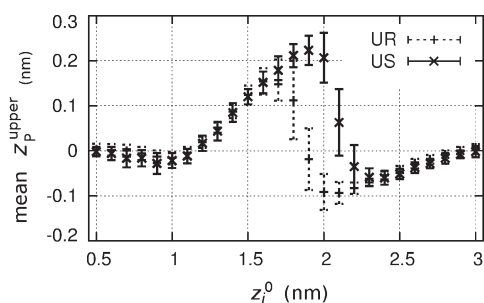
H

dx.doi.org/10.1021/ct200316w |*J. Chem. Theory Comput.* XXXX, XXX, 000−000

Figure 6. Bilayer reorganization as a function of methylpropane immersion sampled by different methods. The mean value of z_P^{upper} for lipids with $r_P < 1$ nm is shown from ("+" and broken lines denoted UR) unrestrained sampling and ("x" and solid lines) US. Error bars represent the standard deviation of the mean. The histogram bin width for unrestrained sampling is 0.1 nm.

containing 10 methylpropane molecules initially located in the aqueous phase, were run for 205 ns each. In the course of these simulations, an average of 6.4 ± 1.3 solute molecules spontaneously partitioned into the lipid bilayer, no unbinding events took place, and the distribution of methylpropane along the bilayer normal was still drifting with increasing simulation time when the simulations were stopped (data not shown). Binding events were characterized by long waiting times (90 ± 56 ns considering only the 64 binding events that occurred) followed by very rapid progression through the transition state to the bound state (0.3 ± 0.3 ns from $z = 2.5$ nm to $z = 1.5$ nm).

Importantly, there is a discrepancy in the nature and extent of bilayer perturbation sampled during nonequilibrium unrestrained vs equilibrium restrained simulations. In Figure 6, we compare the local reorganization of the proximal bilayer leaflet that occurred during unrestrained and US simulations. During unrestrained sampling, the most populated bilayer state changed from a depression to a protrusion with deeper solute immersion near $z = 1.9$ nm (Figure 6). However, in the restrained simulations, this transition occurred near $z = 2.1$ nm (Figure 6).

To evaluate the possibility that this discrepancy results from systematic sampling errors at $1.9 \leq z \leq 2.1$ nm in our restrained simulations, we conducted additional US simulations at $z_i^0 = 2.0$ nm. These simulations directly probe the probability that US simulations may be trapped in metastable bilayer protrusion states at $1.9 \leq z \leq 2.1$ nm due to hidden sampling barriers along z_P^{upper}, the deviation from the mean of the distance along **z** between the center of the bilayer and the phosphorus atom of lipids in the upper leaflet. These simulations were initiated with conformations drawn from unrestrained simulations in which a methylpropane molecule located near $z = 2.0$ nm either did or did not subsequently progress into the hydrophobic interior of the bilayer. We refer to these starting conformations as binding and nonbinding, respectively. Not only do both binding (Figure 7A) and nonbinding (Figure 7B) initial conformations drawn from unrestrained simulations all predominantly sample bilayer protrusion states in US simulations at $z_i^0 = 2.0$ nm but they also do so to the same extent as the original set of US simulations at this depth (Figure 7C). Furthermore, rare transitions between local bilayer protrusion and depression are evident for a number of US simulations in Figure 7A and B, demonstrating that the apparent barrier along z_P^{upper} does not prevent the slow
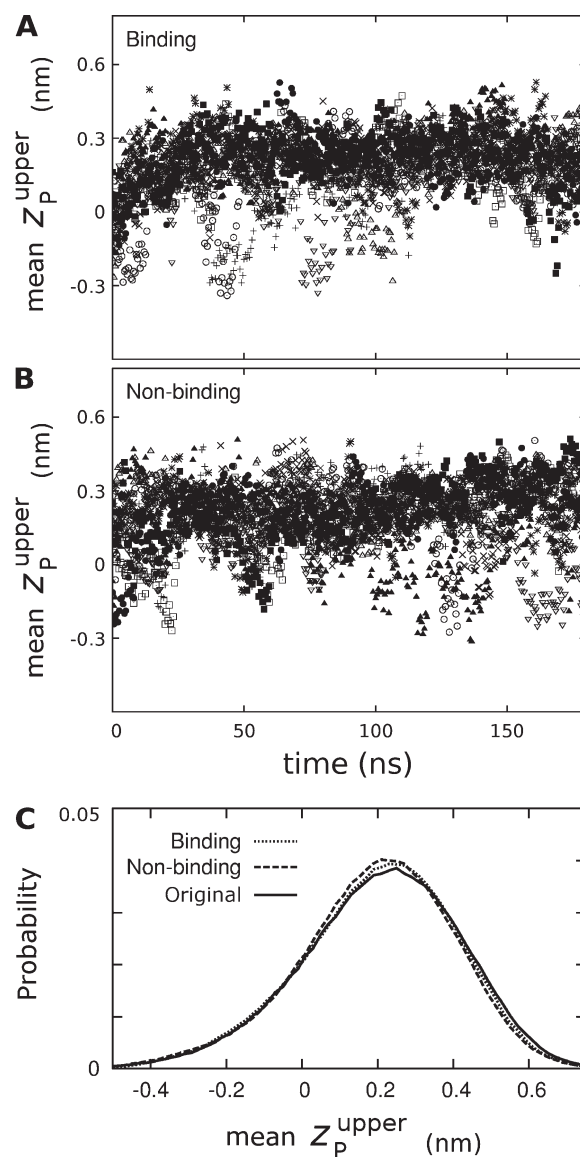

Figure 7. Bilayer reorganization as a function of simulation time from US at $z_i^0 = 2.0$ nm. (A and B) For each simulation, a different symbol represents the time series of the mean value of z_P^{upper} for lipids with $r_P < 1$ nm, block-averaged in 500 ps intervals. The reader is not expected to distinguish the different symbols throughout the sampling time. Rather, we use discrete symbols instead of depicting this data as a heat map to make it clear that excursions of sampling away from the main basin are transiently conducted by different simulations. Starting conformations were drawn from 10 snapshots obtained from unrestrained simulations in which the identified methylpropane molecule (A) subsequently bound the bilayer, progressing deeply into the hydrophobic interior, or (B) did not bind the bilayer, diffusing back into bulk water. (C) Histograms showing the probabilities at which values of z_P^{upper} were sampled during US initiated from (dotted line) productive binding snapshots, (dashed line) nonbinding snapshots, and (solid line) the 24 main US simulations of methylpropane in this work. The histogram bin width was 0.02 nm.

equilibration of protrusion and depression states at this depth on the simulation time scale. As discussed in the next section, this discrepancy questions the relevance of solute-restrained conformations to the binding mechanism.

■ DISCUSSION

Systematic Sampling Errors. The analysis of potential of mean-force free energy calculations for the insertion of small molecular solutes into a lipid bilayer shows that statistical convergence is limited by the rate of conformational reorganization of the lipid bilayer.

The accuracy of ΔG_{bind}° can be strongly affected by systematic sampling errors occurring at specific locations along the reaction coordinate. In the previous study of methylpropane,[23] systematic sampling errors in only two umbrellas caused a systematic ΔG_{bind}° error of 1 kcal/mol (Table 2). The two umbrellas that contained systematic sampling errors were centered at $z_i^0 = 1.9$ and 2.0 nm (Figure 2C). In a neat DOPC bilayer, this region corresponds to the rugged interface between the bilayer and water (Figure 3A). There, methylpropane is abruptly being dehydrated with increasing bilayer immersion, following a sharp transition from local depression to protrusion of the proximal bilayer−water interface with increasing solute depth between $z_i^0 = 2.1$ and 2.0 nm in the long simulations presented in this work (Figure 3B−D). In shorter simulations,[23] the two umbrella potentials centered at $z_i^0 = 1.9$ and 2.0 nm were trapped in a metastable state of leaflet depression similar to the one depicted in Figure 4A and did not sample the state of leaflet protrusion depicted in Figure 4B (data not shown), which is more favorable at these depths (Figure 3B−D). This finding is consistent with the significant spread in ΔG_{bind}° values obtained from 20 ns samples per umbrella in this work (Figure 2A). Accordingly, the extension of the previously published simulations presented in this study resolved these systematic sampling errors (Table 2).

Systematic sampling errors were also revealed upon increasing the amount of sampling in simulations of *n*-propylguanidinium. In this case, the systematic ΔG_{bind}° error is 3 kcal/mol (Table 2). In the previous study of *n*-propylguanidinium,[23] systematic sampling errors occurred near $z = 2.0$ and 2.4 nm (Figure 2F) and, similarly to the previous simulations of methylpropane,[23] also resulted from insufficient sampling of a state involving local lipid protrusion (Figures 3G−I). In this case, bilayer reorganization was concurrent with reorientation of the solute (Figure 3J), and the extension of the previously published simulations[23] for an additional 120 ns per umbrella did not resolve these systematic sampling errors, although the difference in ΔG_{bind}° values was reduced (Table 2). Nevertheless, the 36 independent measures of ΔG_{bind}° in this study range from −9.1 to −4.0 kcal/mol (Table 2), and it is only as an average that they converge to −6.7 ± 1.0 kcal/mol. The finding that a single set of US simulations lasting 205 ns per umbrella may be insufficient to obtain adequate convergence in the value of ΔG_{bind}° for *n*-propylguanidinium should be regarded as an important cautionary tale for the calculation of equilibrium properties from simulations of peptides in lipid bilayers.

Increased sampling also affected the free energy profile of *n*-propylguanidinium near the center of the bilayer (Figure 2E,F). Here, changes in the orientation of the guanidino group (Figure 3J) are coupled to the presence, in one leaflet or the other, of a conic lipid surface defect that is lined by lipid headgroups (Figures 3F−I, 4F, 5F) and filled with water molecules that partially solvate the guanidino moiety (Figure 3F). Conformational states in which such defects occur in the distal leaflet are only metastable.[52] The reason why *n*-propylguanidinium PMFs in this work show no indication of systematic sampling errors in the center of the bilayer for US

simulations with increasing t_{eq} is that the orientational auto-correlation time exceeds the simulation time scale (data not shown). That is, systematic sampling errors only become apparent in PMFs after a transition leading to a state of lower free energy occurs for the first time. Sampling errors in this region do not affect the computed binding free energy significantly because solute residence at the bilayer center is the least favorable state in the entire PMF (Figure 2E). However, the large kinetic barrier to convergence at the middle of the bilayer does affect the symmetry of the solute orientation about the bilayer center (Figure 3J) and the computed free energy barrier to bilayer traversal.[52] At $z_i^0 = 0$ nm, both polarized orientations of the solute should be equally likely. Here, the presence of US restraining potentials, by preventing significant axial mobility of the solute, forces the system to cross pre-existing barriers that could otherwise be circumvented. This source of quasi-nonergodicity may be alleviated by the addition of equilibrium exchange.[22]

Importantly, our results suggest that systematic sampling errors may also be present in other US studies of similar duration in which solutes interact with a lipid bilayer. Free energy profiles for the insertion of molecular solutes along the bilayer normal have been obtained using US and similar techniques for a long list of increasingly complex solutes including, among others, the small molecules hexane,[13] oxygen, and ammonia;[14] the steroids cholesterol[53] and cortisone;[54] the drugs valproic acid,[11] various adamantanes,[12] and the nonsteroidal anti-inflammatories acetylsalicylic acid and ibuprofen;[10] a panel of amino acid side chain analogs;[23] and, finally, peptides including a model hexapeptide,[55] a transmembrane helix,[52] the 18-residue cationic antimicrobial peptide protegrin-1,[16] a 34-residue Kv channel gating-modifier toxin,[56] and a 42-residue fragment of the amyloid-β peptide.[57] In the aforementioned simulations, either ΔG_{bind}° drifted systematically with increasing t_{eq} (see ref 55 and C.H. Davis, personal communication regarding ref 57), it depended significantly on the starting conformation,[56] or free-energy convergence measures based on t_{eq} were not conducted.[10,11,13,14,16,23,52−54] From this list, only the study of adamantanes evaluated the systematic drift in the value of ΔG_{bind}° based on t_{eq} and, using a statistical test, concluded that the simulations had converged after 15 ns per umbrella with $t_{eq} = 2$ ns.[12] In light of the present study, it remains possible that these simulations of adamantanes, amphipathic cationic molecules larger than *n*-propylguanidinium, contained systematic sampling errors but simply did not yet show it. Furthermore, it is important to keep in mind that some of the 36 PMFs that we constructed, each based on 205 ns of equilibration and sampling per umbrella, still contain unresolved systematic sampling errors. Finally, the simulations presented in this work may also be subject to systematic sampling errors that are not evident on this time scale.

Interestingly, the application of additional restraints on lipid headgroups appears to avoid the slow relaxation of z_P^{upper} and speed up convergence of PMF calculations from US simulations of peptide insertion along the bilayer normal.[56] Note, however, that such restraints preclude large deformations of the lipid bilayer, which may be mechanistically important, such as those depicted in Figures 4 and 5.

Effect of Simulation Box Size. The simulations in this work employed a relatively small patch of bilayer, with only 32 lipids per leaflet. For the small solutes investigated in this study, systematic sampling errors persisted for up to 125 ns per umbrella (Figure 2D), and even longer equilibration times may be necessary for larger solutes. Thus, we chose to keep the

simulation box small in order to attain long simulation times. Importantly, at certain values of z_i^0, the bilayer reorganization induced by the insertion of the solute involved the entire proximal leaflet (Figures 3B,G and 4B,F), suggesting that the finite size of the simulation box may result in a systematic bias of the PMF. This possible bias appears unlikely to dramatically affect the value of ΔG_{bind}° for n-propylguanidinium. This is because bilayer reorganization remained predominantly local for $z_i^0 \geq 1.5$ nm (Figure 3G), which encompasses the region traversed by the solute from bulk water to the global minimum of the PMF at $1.4 \leq z \leq 1.5$ nm (Figure 2E). For methylpropane, however, bilayer reorganization induced by solute migration through the lipid−water interface involved most of the proximal leaflet for $2.1 \geq z_i^0 \geq 1.5$ nm (Figure 3B), a region of the PMF located between bulk water and the global minimum of the PMF at $z_i^0 = 0.0$ nm (Figure 2B), so that the box size may have a significant effect on the estimate of ΔG_{bind}°. A systematic evaluation of the dependence of ΔG_{bind}° on bilayer size is an interesting avenue for future research. However, PMF calculations based on a conventional definition of solute insertion depth will be subjected to an increasingly degenerate reaction coordinate as the size of the bilayer is increased. This is because larger bilayers can experience larger undulations. When using very large bilayer patches, it may thus become possible for the COM of the solute to coincide with the COM of the bilayer along the bilayer normal even as the solute remains in bulk water.

Comparison to Experimental Transfer Free Energies. Our results indicate that previously published estimates of methyl-propane and n-propylguanidinium binding free energies[23] contain systematic sampling errors. Nevertheless, those results[23] are in good agreement with the experimental water-to-cyclohexane transfer free energies reported by Radzicka and Wolfenden[58] for methylpropane (-4.9 kcal/mol) and n-propylguanidinium ($+14.9$ kcal/mol). This may be because the range of experimentally determined values for each side chain analog remains large in comparison to the theoretical uncertainties. For methylpropane, MacCallum et al. obtained a transfer free energy of -3.6 ± 0.4 kcal/mol,[23] using the value of the PMF at the bilayer center, whereas we obtain -4.8 ± 0.3 kcal/mol (Figure 2B). While the free energy value that we report appears to have improved accuracy, it should be noted that the experimental free energy quoted is actually the transfer free energy for n-butane.[58] In a separate simulation study, MacCallum and Tieleman used a thermodynamic cycle to calculate the free energy for the transfer of amino acid side chain analogs from water to cyclohexane.[59] The binding free energy value obtained for methylpropane using the OPLS-AA/L force field and a different treatment of long-range nonbonded interactions was, depending on the water model, -5.4 to -5.7 kcal/mol.[59] Alchemical pathways can be advantageously used to determine ΔG_{bind}° in cases where the bound state is known. For methylpropane, an alchemical pathway can circumvent the hidden sampling barriers at the lipid−water interface. For n-propylguanidinium, however, and in the general case, knowledge of the PMF is required to define the bound state.

Moreover, it is difficult to make experimental comparisons for the value of ΔG_{bind}° in the case of n-propylguanidinium because water-to-cyclohexane or -octanol transfer free energies are expected to differ from the free energy of transfer to the center or the interface of a DOPC bilayer, where the solute remains solvated by water molecules and lipid headgroups (Figure 3F) in a complex and highly anisotropic chemical environment (Figures 4D−F, 5D−F). In addition, the drastic reorganization

of the lipid bilayer that occurs in response to even small molecular solutes and the sensitivity of these perturbations to the nature of the solute (Figures 3−5) complicate the assumption of group additivity underlying the application of the free energy of partitioning of Ace-WLLxL peptides (where x includes L and R) between water and a bilayer similar to DOPC.[60] Indeed, the limitations of group contribution approaches have also been pointed out in simulation studies of hydration free energies.[61]

Relevance of Solute-Restrained Conformations to the Binding Mechanism. The bilayer reorganization induced by the insertion of methylpropane in equilibrium restrained simulations differs from the reorganization observed in nonequilibrium unrestrained simulations (Figure 6). Because conformations sampled by restrained simulations along physical reaction coordinates are attainable by unrestrained simulations, it is tempting to interpret the path of lowest free energy along z as the unrestrained solute binding pathway.[55] Generally, however, restrained simulations may not yield mechanistically representative states in the absence of a separation of time scales whereby unrestrained molecular diffusion along the reaction coordinate is much slower than diffusion along other degrees of freedom. This is because weak ergodicity breaking[62] may occur during binding events in unrestrained simulations but not in restrained simulations over sufficiently long time scales (Figure 6). When spontaneous binding conformations near $z = 2.0$ nm are extracted from unrestrained simulations and used to initiate US simulations, the distribution of z_P^{upper} requires 10 ns to converge to a state of local lipid protrusion (Figure 7A). However, during all 64 binding events observed in our unrestrained simulation, methylpropane passed through this region in significantly less than 1 ns. In other words, the conformation of the bilayer does not reach equilibrium as methylpropane freely diffuses inward through the headgroup region. Schematically, when the reaction coordinate is not the slowest relaxing DOF, restrained simulations on sufficiently long time scales follow the path of lowest free energy in orthogonal DOFs (blue dotted line in Figure 1), whereas unrestrained sampling over the landscape presented in Figure 1 may most readily progress from α to β by way of γ if diffusion along the reaction coordinate (**x** axis) is faster than along other DOFs (**y** axis). This effect is similar to differences in experimental time averaging.[63]

As a result, it would be incorrect to conclude from the time-averaged data in Figures 3B,C,D; 4B; and 5B that, during binding events, the bilayer is likely to protrude to encapsulate methylpropane when methylpropane is at a distance of 2.0 nm from the bilayer center. Indeed, Figure 6 demonstrates that, during spontaneous binding events, the bilayer interface is only likely to protrude once methylpropane is within 1.8 nm of the bilayer center. Inversely, it is likely that fluctuations within the bound state up to and including spontaneous unbinding events, none of which occurred in our unrestrained simulations, involve even larger protrusions of the bilayer surface than observed in the equilibrium picture obtained from ensemble averaging. Because it describes a reversible (i.e., infinitely slow) reaction, the free energy surface represents an ensemble average over both binding and unbinding events, regardless of the direction of travel of the solute and its history. In such a case, the lowest free-energy pathway through the saddle point in Figure 1 represents an average between trajectories following either γ or δ, depending on direction. This analysis underscores the importance of considering time-dependent information when one desires to draw mechanistic conclusions from equilibrium sampling.

General Implications to Membrane Solvation of Peptides and Proteins. The mechanisms of bilayer reorganization uncovered in the present study are likely to be relevant in the context of larger molecules such as peptides or proteins, in which a hydrophobic or charged moiety is anchored at a particular depth. In such cases, as seen here for *n*-propylguanidinium, the bound state may involve substantial deformation of the bilayer. In particular, the deformations outlined in Figures 4 and 5 are parts of bilayer adaptation to hydrophobic mismatches that may be introduced by transmembrane proteins.[64] During such adaptations, lipid molecules in close proximity to the solute adjust their own immersion depth to facilitate the inclusion of hydrophobic moieties and the exclusion of hydrophilic moieties from the hydrophobic region of the bilayer. Importantly, these phenomena may also exist in the interaction of lipid bilayers with interfacial membrane proteins or peptides.

Interestingly, the states depicted in Figures 4 and 5 suggest that peripheral membrane proteins or interfacial peptides could induce bilayer bending by inducing changes in lipid ordering. The induction of bilayer bending is important in any cellular process that relies on membrane tension, including endocytosis, exocytosis, cell motility, and viral membrane fusion during infection.[65] Furthermore, protein-induced changes in membrane tension are capable of feeding back into the function of mechanosensitive membrane proteins.[66]

■ CONCLUSIONS

The above results suggest that the difficulty of attaining convergence in simulations of molecular solutes embedded in lipid bilayers has, in general, been drastically underestimated due to the presence of hidden sampling barriers involving the slow reorganization of the lipid—water interface in response to solute insertion. Although we focused our analysis of systematic sampling errors on a single study,[23] it seems likely that similar systematic sampling errors pervade the literature. It is unfortunate that many published simulation studies do not report measures of convergence, since this information contributes crucially to the inferred accuracy of the measured quantities.[67] The publication of convergence measures, especially those that are capable of detecting systematic drifts in values obtained with increasing equilibration time, should be systematically encouraged.

On the basis of the above analysis, we propose that simulation studies could benefit from the following suggestions:

1. Performing multiple distinct sets of simulations of the same system from different starting conformations and computing PMFs separately from each set. Degrees of freedom predicted to contribute significantly to the binding free energy should be evaluated systematically. It may also be useful to use a variety of methods to generate these starting conformations in order to vary the starting conformations in unexpected ways.
2. Evaluating a key observable(s), such as the binding free energy, as a function of both simulation time and equilibration time and reporting these convergence measures.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: pomes@sickkids.ca.

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Edidin, M. Lipids on the frontier: a century of cell-membrane bilayers. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (5), 414–418.

(2) Dyall, S. D.; Brown, M. T.; Johnson, P. J. Ancient invasions: from endosymbionts to organelles. *Science* **2004**, *304* (5668), 253–257.

(3) Deamer, D. W.; Bramhall, J. Permeability of lipid bilayers to water and ionic solutes. *Chem. Phys. Lipids* **1986**, *40* (2−4), 167–188.

(4) Engelman, D. M. Membranes are more mosaic than fluid. *Nature* **2005**, *438* (7068), 578–580.

(5) Simons, K.; Vaz, W. Model systems, lipid rafts, and cell membranes. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33* (1), 269–295.

(6) Egberts, E.; Berendsen, H. J. C. Molecular dynamics simulation of a smectic liquid crystal with atomic detail. *J. Chem. Phys.* **1988**, *89* (6), 3718–3732.

(7) Tieleman, D. P.; Marrink, S. J.; Berendsen, H. J. C. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *BBA—Rev. Biomembranes* **1997**, *1331* (3), 235–270.

(8) Takaoka, Y.; Pasenkiewicz-Gierula, M.; Miyagawa, H.; Kitamura, K.; Tamura, Y.; Kusumi, A. Molecular dynamics generation of nonarbitrary membrane models reveals lipid orientational correlations. *Biophys. J.* **2000**, *79* (6), 3118–3138.

(9) Poger, D.; Mark, A. E. On the validation of molecular dynamics simulations of saturated and *cis*-monounsaturated phosphatidylcholine lipid bilayers: a comparison with experiment. *J. Chem. Theory Comput.* **2010**, *6* (1), 325–336.

(10) Boggara, M. B.; Krishnamoorti, R. Partitioning of nonsteroidal antiinflammatory drugs in lipid membranes: a molecular dynamics simulation study. *Biophys. J.* **2010**, *98* (4), 586–595.

(11) Ulander, J.; Haymet, A. D. J. Permeation across hydrated DPPC lipid bilayers: simulation of the titrable amphiphilic drug valproic acid. *Biophys. J.* **2003**, *85* (6), 3475–3484.

(12) Chew, C. F.; Guy, A.; Biggin, P. C. Distribution and dynamics of adamantanes in a lipid bilayer. *Biophys. J.* **2008**, *95* (12), 5627–5636.

(13) MacCallum, J. L.; Tieleman, D. P. Computer simulation of the distribution of hexane in a lipid bilayer: spatially resolved free energy, entropy, and enthalpy profiles. *J. Am. Chem. Soc.* **2006**, *128* (1), 125–130.

(14) Marrink, S. J.; Berendsen, H. J. C. Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *J. Phys. Chem.* **1996**, *100* (41), 16729–16738.

(15) Bemporad, D.; Luttmann, C.; Essex, J. W. Computer simulation of small molecule permeation across a lipid bilayer: dependence on bilayer properties and solute volume, size, and cross-sectional area. *Biophys. J.* **2004**, *87* (1), 1–13.

(16) Vivcharuk, V.; Kaznessis, Y. Free energy profile of the interaction between a monomer or a dimer of protegrin-1 in a specific binding orientation and a model lipid bilayer. *J. Phys. Chem. B* **2010**, *114* (8), 2790–2797.

(17) Yeh, I.-C.; Olson, M. A.; Lee, M. S.; Wallqvist, A. Free-energy profiles of membrane insertion of the M2 transmembrane peptide from influenza A virus. *Biophys. J.* **2008**, *95* (11), 5021–5029.

(18) van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 149–153.

(19) Zuckerman, D. M. Equilibrium sampling in biomolecular simulations. *Ann. Rev. Biophys.* **2011**, *40* (1), 41–62.

(20) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.

(21) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91* (1−3), 275–282.

(22) Neale, C.; Rodinger, T.; Pomès, R. Equilibrium exchange enhances the convergence rate of umbrella sampling. *Chem. Phys. Lett.* **2008**, *460* (1−3), 375–381.

(23) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. Distribution of amino acids in a lipid bilayer from computer simulations. *Biophys. J.* **2008**, *94* (9), 3393–3404.

(24) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.

(25) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(26) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.

(27) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105* (28), 6474–6487.

(28) Berger, O.; Edholm, O.; Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **1997**, *72* (5), 2002–2013.

(29) Chakrabarti, N.; Neale, C.; Payandeh, J.; Pai, E. F.; Pomès, R. An iris-like mechanism of pore dilation in the CorA magnesium transport system. *Biophys. J.* **2010**, *98* (5), 784–792.

(30) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(31) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.

(32) Berendsen, H. J. C.; Postma, J. P. M.; vanGunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

(33) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13* (8), 952–962.

(34) Hess, B. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (1), 116–122.

(35) Kandt, C.; Ash, W. L.; Tieleman, D. P. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41* (4), 475–488.

(36) Hristova, K.; Wimley, W. C. A look at arginine in membranes. *J. Membr. Biol.* **2011**, *239* (1−2), 49–56.

(37) Roux, B. Lonely arginine seeks friendly environment. *J. Gen. Physiol.* **2007**, *130* (2), 233–236.

(38) Yoo, J.; Cui, Q. Does arginine remain protonated in the lipid membrane? Insights from microscopic pKa calculations. *Biophys. J.* **2008**, *94* (8), L61–L63.

(39) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. *J. Gen. Physiol.* **2007**, *129* (5), 371–377.

(40) Li, L.; Vorobyov, I.; MacKerell, A. D.; Allen, T. W. Is arginine charged in a membrane?. *Biophys. J.* **2008**, *94* (2), L11–L13.

(41) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7* (8), 306–317.

(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1981; pp 331−342.

(43) Grossfield, A. WHAM: the weighted histogram analysis method. http://membrane.urmc.rochester.edu/content/wham (accessed August 3, 2011).

(44) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.

(45) Ben-Tal, N.; Ben-Shaul, A.; Nicholls, A.; Honig, B. Free-energy determinants of alpha-helix insertion into lipid bilayers. *Biophys. J.* **1996**, *70* (4), 1803–1812.

(46) General, I. J. A note on the standard state's binding free energy. *J. Chem. Theory Comput.* **2010**, *6* (8), 2520–2524.

(47) Grossfield, A.; Woolf, T. B. Interaction of tryptophan analogs with POPC lipid bilayers investigated by molecular dynamics calculations. *Langmuir* **2002**, *18* (1), 198–210.

(48) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38.

(49) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91* (1), 461–466.

(50) Bogusz, S.; Cheatham, T. E.; Brooks, B. R. Removal of pressure and free energy artifacts in charged periodic systems via net charge corrections to the Ewald potential. *J. Chem. Phys.* **1998**, *108* (17), 7070–7084.

(51) Berendsen, H. J. C.; Marrink, S.-J. Molecular dynamics of water transport through membranes: water from solvent to solute. *Pure Appl. Chem.* **1993**, *65* (12), 2513–2520.

(52) Li, L.; Vorobyov, I.; Allen, T. Potential of mean force and pKa profile calculation for a lipid membrane-exposed arginine side chain. *J. Phys. Chem. B* **2008**, *112* (32), 9574–9587.

(53) Zhang, Z.; Lu, L.; Berkowitz, M. L. Energetics of cholesterol transfer between lipid bilayers. *J. Phys. Chem. B* **2008**, *112* (12), 3807–3811.

(54) Vijayan, R.; Biggin, P. C. A steroid in a lipid bilayer: localization, orientation, and energetics. *Biophys. J.* **2008**, *95* (7), L45–L47.

(55) Babakhani, A.; Gorfe, A.; Kim, J. E.; McCammon, J. A. Thermodynamics of peptide insertion and aggregation in a lipid bilayer. *J. Phys. Chem. B* **2008**, *112* (34), 10528–10534.

(56) Wee, C. L.; Gavaghan, D.; Sansom, M. S. P. Lipid bilayer deformation and the free energy of interaction of a Kv channel gating-modifier toxin. *Biophys. J.* **2008**, *95* (8), 3816–3826.

(57) Davis, C. H.; Berkowitz, M. L. Interaction between amyloid-β (1−42) peptide and phospholipid bilayers: a molecular dynamics study. *Biophys. J.* **2009**, *96* (3), 785–797.

(58) Radzicka, A.; Wolfenden, R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **1988**, *27* (5), 1664–1670.

(59) MacCallum, J. L.; Tieleman, D. P. Calculation of the water−cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field. *J. Comput. Chem.* **2003**, *24* (15), 1930–1935.

(60) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **1996**, *3* (10), 842–848.

(61) König, G.; Boresch, S. Hydration free energies of amino acids: why side chain analog data are not enough. *J. Phys. Chem. B* **2009**, *113* (26), 8967–8974.

(62) Bouchaud, J. P. Weak ergodicity breaking and aging in disordered systems. *J. Phys. I France* **1992**, *2* (9), 1705–1713.

(63) Paula, L. C.; Wang, J.; Leite, V. B. P. Statistics and kinetics of single-molecule electron transfer dynamics in complex environments: a simulation model study. *J. Chem. Phys.* **2008**, *129* (22), 224504.

(64) Kandasamy, S. K.; Larson, R. G. Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: a systematic investigation of hydrophobic mismatch. *Biophys. J.* **2006**, *90* (7), 2326–2343.

(65) Nambiar, R.; McConnell, R. E.; Tyska, M. J. Control of cell membrane tension by myosin-I. *P. Natl. Acad. Sci. U. S. A.* **2009**, *106* (29), 11972–11977.

(66) Phillips, R.; Ursell, T.; Wiggins, P.; Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature* **2009**, *459* (7245), 379–385.

(67) Lyman, E.; Zuckerman, D. M. Ensemble-based convergence analysis of biomolecular trajectories. *Biophys. J.* **2006**, *91* (1), 164–172.

(68) Loken, C.; Gruner, D.; Groer, L.; Peltier, R.; Bunn, N.; Craig, M.; Henriques, T.; Dempsey, J.; Yu, C.-H.; Chen, J.; Dursi, L. J.; Chong, J.; Northrup, S.; Pinto, J.; Knecht, N.; Van Zon, R. SciNet: lessons learned from building a power-efficient top-20 system and data centre. *J. Phys.: Conf. Ser.* **2010**, *256* (1), 012026.