Interaction of Tryptophan Analogs with POPC Lipid Bilayers Investigated by Molecular Dynamics Calculations

Alan Grossfield[†,‡] and Thomas B. Woolf*[,†,§]

Department of Biophysics and Biophysical Chemistry and Department of Physiology, Johns Hopkins Medical School, 725 North Wolfe Street, Baltimore, Maryland 21205

Received May 2, 2001. In Final Form: August 1, 2001

Several experimental structures of integral membrane proteins have demonstrated that tryptophan side chains are found preferentially within the membrane interface, but no clear physical explanation for this preference has been demonstrated. The present work contains a series of molecular dynamics simulations designed to provide insight into this phenomenon. Specifically, we performed conventional *unbiased* simulations, as well as *biased* molecular dynamics of indole and *N*-methylindole within palmitoyloleoyl phosphatidylcholine (POPC) bilayers. The *unbiased* calculations examined the behavior of the indoles in three environments—the membrane center, the headgroup region, and the water outside the membrane. The *biased* calculations were performed using a new dynamic windowing procedure, which efficiently moved the indoles through the bilayer. These calculations allowed us to explore the orientation, hydrogen bonding, and energetics of the indole, as a function of its location along the membrane normal. Furthermore, free energy profiles were constructed from the *biased* simulations, facilitating comparison to experiment and a possible explanation of indole's partitioning behavior.

1. Introduction

Structural knowledge of membrane proteins lags far behind that for soluble proteins. While there are more than 13000 tertiary structures in the Protein Data Bank, there are only about 20 high-resolution membrane protein structures.[1,2] This is true in large part because it is difficult to apply standard X-ray crystallographic and liquid-state NMR techniques to membrane proteins and their native solvent environment—the lipid bilayer. As a result, one must look elsewhere for molecular knowledge on membrane protein structure and the role of the membrane environment. Computer simulations can be used to fill this gap.

One approach for simulating membranes and membrane proteins is to represent the system with full atomic detail and to perform sampling using molecular dynamics. A variety of groups have run molecular dynamics simulations of explicit lipid bilayers.[3−9] Moreover, several considered the behavior of various membrane permeants, such as peptides[10] or proteins.[11−19] Strong interactions between tryptophan residues and the surrounding bilayer have been noted in several of them.[20,21]

However, these simulations are technically complex and computationally demanding. Because of the computational expense, simulations can only follow a small number of molecules (typically 20−100 lipids) for a short period of time (0.5−10 ns of trajectory time). This must be compared with relaxation time scales of lipid bilayers; acyl chain isomerization takes place on the 10−100 ps time scale, while lipid "wobbling" occurs on the nanosecond time scale.[22] Larger lipid displacements, or "jumps", require tens of nanoseconds, and collective, many-molecule motions such as bilayer undulations can take still longer.[9,22,23] Clearly, molecular dynamics simulations do not sample the longer time scale variations. This does not mean that such simulations should not be performed, only that one must be careful about the choice of systems and initial conditions. Moreover, analysis should be restricted to those

* To whom correspondence may be addressed. Phone: 410-614-2643. Fax: 410-614-4436. E-mail: woolf@groucho.med.jhmi.edu.
† Department of Biophysics and Biophysical Chemistry.
‡ Present address: Department of Biochemistry and Molecular Biophysics Washington University School of Medicine, St. Louis, MO 63110.
§ Department of Physiology.

(1) Preusch, P. C.; Norvell, J. C.; Cassatt, J. C.; Caussman, M. *Nat. Struct. Biol.* **1998**, *5*, 12−14.
(2) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242. http://www.rcsb.org/pdb.
(3) Heller, H.; Schaefer, M.; Schulten, K. *J. Phys. Chem.* **1993**, *97*, 8343−8360.
(4) Egberts, E.; Marrink, S. J.; Berendsen, H. J. C. *Eur. Biophys. J.* **1994**, *22*, 423−436.
(5) Pastor, R. W. *Curr. Opin. Struct. Biol.* **1994**, *4*, 486-492.
(6) Chiu, S. W.; Clark, M.; Balaji, V.; Subramaniam, S.; Scott, H. L.; Jakobsson, E. *Biophys. J.* **1995**, *69*, 1230−1245.
(7) Feller, S. E.; Zhang, Y.; Pastor, R. W. *J. Chem. Phys.* **1995**, *103*, 10267−10276.
(8) Marrink, S. J.; Sok, R. M.; Berendsen, H. J. C. *J. Chem. Phys.* **1996**, *104*, 9090−9099.
(9) Lindahl, E.; Edholm, O. *Biophys. J.* **2000**, *79*, 426−433.
(10) Damodaran, K. V.; Merz, K. M., Jr.; Bager, B. P. *Biophys. J.* **1995**, *69*, 1299−1308.
(11) Woolf, T. B.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11631−11635.
(12) Woolf, T. B. *Biophys. J.* **1997**, *73*, 2376−2392.
(13) Petrache, H. I.; Grossfield, A.; MacKenzie, K. R.; Engelman, D. M.; Woolf, T. B. *J. Mol. Biol.* **2000**, *302*, 727−746.
(14) Tieleman, D. P.; Berendsen, H. J. C.; Sansom, M. S. P. *Biophys. J.* **1999**, *76*, 3186−3191.
(15) Forrest, L. R.; Tieleman, D. P.; Sansom, M. S. P. *Biophys. J.* **1999**, *76*, 1886−1896.
(16) Tieleman, D. P.; Sansom, M. S. P.; Berendsen, H. J. C. *Biophys. J.* **1999**, *76*, 1757−1769.
(17) Tieleman, D. P.; Sansom, M. S. P.; Berendsen, H. J. C. *Biophys. J.* **1999**, *76*, 40−49.
(18) Forrest, L. R.; Kukol, A.; Arkin, I. T.; Tieleman, D. P.; Sansom, M. S. P. *Biophys. J.* **2000**, 78.
(19) Mihailescu, D.; Smith, J. C. *Biophys. J.* **2000**, *79*, 1718−1730.
(20) Woolf, T. B. *Biophys. J.* **1998**, *74*, 115−131.
(21) Petrache, H. I.; Zuckerman, D. M.; Sachs, J. N.; Killian, J. A.; Koeppe, R. E.; Woolf, T. B. *Langmuir*, in press.
(22) Pastor, R. W.; Feller, S. E. In *Biological Membranes*: *A molecular perspective from Computation and Experiment*; Merz, K. M., Jr., Roux, B., Eds.; Birkhauser: Basel, 1996; pp 3−30.
(23) Feller, S. E.; Pastor, R. W. In *Pacific Symposium on Biocomputing '97*; Altman, R. B., Dunker, A. K., Hunter, L., Klein, T. E., Eds.; World Scientific: Singapore, 1996; pp 142−150.

10.1021/la0106485 CCC: $22.00
Published on Web 12/11/2001

properties that are well defined on the size and time scales available in the simulations.

One interesting, but puzzling, feature of most membrane proteins solved to date is the pronounced tendency for aromatic residues, especially tryptophan, to be found in the membrane−water interface. This is a general phenomenon, also seen for example in a statistical study of membrane protein sequences.[24] The reason for this preference is not clear. Partitioning of side chain analogues into bulk hydrophobic solvents indicates that tryptophan is very hydrophobic,[25,26] so one might expect that tryptophan residues would tend to be found in the membrane interior. However, this view is predicated on a simplistic model of the membrane as a region of bulk hydrocarbon, surrounded by water. It ignores the "tumultuous chemical heterogeneity"[27] of the membrane interface, a broad region containing significant populations of water, headgroup atoms, and methylene groups. This complex environment resembles neither the membrane interior nor bulk water, obscuring the mechanism for stability of aromatic residues.

There is experimental evidence that has been interpreted to indicate tryptophan and other aromatic residues partition favorably into the membrane interface. For example, host−guest studies on tri- and pentapeptides demonstrated that, of all amino acid residues, tryptophan most favors partitioning into the membrane−water interface.[27−29] The initial intepretation of the peptide data was that interfacial partitioning was a compromise allowing the peptide backbone to form hydrogen bonds with water while partially burying hydrophobic groups, such as the tryptophan side chain, in the membrane interior.[27] However, more recent solid-state NMR experiments on isolated indole compounds (analogues to tryptophan side chains *in the absence of the peptide backbone*) seem to indicate that they partition not into the acyl core, as one might have expected, but rather into the broad interfacial region.[30]

The present work describes a series of molecular dynamics simulations of indole (IND) and *N*-methylindole (NMI) bound to fully solvated POPC bilayers. The goal of the simulations is to rationalize the behavior of indole and *N*-methylindole when bound to lipid bilayers, which in turn has implications for the behavior of tryptophans in membrane proteins. Indole is chosen because chemically it is analogous to a tryptophan sidechain, with the β-carbon replaced with a hydrogen. *N*-methylindole is meant to assess the importance of hydrogen bonding, by replacing the imino hydrogen with a methyl group. A series of conventional molecular dynamics simulations, starting the IND or NMI in different locations relative to the membrane center (hydrophobic core, membrane−water interface, and water), allow the direct analysis of a variety of properties, such hydrogen bonding between indole imino group and the water and lipid found in the interface. However, the simulation time scale is not long enough to allow the indoles to explore the entire range of membrane environments. Accordingly, a series of *biased* simulations were performed, where an "umbrella" potential was added to the system to move the indoles from the membrane center through the interface to the water. The *biased* simulations allow us to analyze various indole properties, including molecular volume, interaction energies with the environment, and hydrogen bonding, in far greater detail as a function of location in the membrane. They also allow calculation of the relative free energy profile along the membrane normal; this in turn facilitates the calculation of the average location of the indoles when bound to the membrane and can be compared to experimental binding coefficients.

2. Methods

Six simulations were run using *unbiased* molecular dynamics, three with indole and three with *N*-methylindole. Additionally, two simulations were run using umbrella sampling to drag the indole (or *N*-methylindole) along the membrane normal from the center of the bilayer to the water. For the purposes of this paper, IND will refer to the indole or the indole simulations and NMI to *N*-methylindole or the *N*-methylindole simulations. The word "indoles" will be used to describe both indole and *N*-methylindole when common behavior is discussed.

Construction of Indole−Membrane Systems. Palmitoyloleoyl phosphatidylcholine (POPC) bilayers were constructed by cutting 18 lipid/leaflet sections (36 lipid molecules total) from 10 larger POPC bilayers, taken from a neat POPC simulation by Armen et al.[31] Smaller systems were used to reduce computational cost; the focus of the simulations was to examine the behavior of the indoles, making a large simulation box an unnecessary expense. We then overlaid each system with water from preequilibrated boxes and removed all new waters overlapping heavy atoms from the system or in the membrane interior. This created fully solvated membrane systems with approximately 8900 atoms (roughly 38 waters/lipid, or 47% water by weight).

Constrained minimization was used to eliminate any unfavorable contacts created when the waters were introduced. A 10 kcal/mol-Å^2 harmonic restraint was applied to the glycerol C2 carbon atoms of each lipid to maintain the position along the *z* axis (membrane normal). With this restraint, a series of steepest descent minimizations was performed, gradually raising the nonbonded cutoff distance. We then ran 25 ps of Langevin dynamics at 300 K to further relax the system.

At this point, we inserted a single indole (IND) molecule into each system, in three separate locations (membrane center, membrane−water interface, bulk water). The center and bulk water structures were created by blindly inserting the indole at the origin or 5 Å above the bilayer and performing a series of minimizations, starting with the van der Waals and electrostatics parameters scaled to $^1/_{10}$th their normal values, and then gradually scaling them back up. The structure with the most favorable interaction energy between indole and the rest of the system was used for future dynamics.

Sites for interfacial insertion were selected by visual inspection. We used the same insertion protocol as for the other systems, except that a restraint was applied to position the indole to form a hydrogen bond with a phosphate oxygen from the POPC. Again, the most favorably interacting structure was chosen for further study. Each system came from a different initial lipid structure.

The *N*-methylindole (NMI) systems were constructed in the same way, except that no hydrogen bond restraint

(24) Landolt-Marticorena, C.; Williams, K. A.; Deber, C. M.; Reithmeier, R. A. F. *J. Mol. Biol.* **1993**, *229*, 602−608.
(25) Fauchere, J.-L.; Pliska, V. *Eur. J. Med. Chem.* **1983**, *18*, 369−375.
(26) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664−1670.
(27) Jacobs, R. E.; White, S. H. *Biochemistry* **1989**, *28*, 3421−3437.
(28) Brown, J. W.; Huestus, W. H. *J. Phys. Chem.* **1993**, *97*, 2967−2973.
(29) Wimley, W. C.; White, S. H. *Nat. Struct. Biol.* **1996**, *3*, 842−848.
(30) Yau, W. M.; Wimley, W. C.; Gawrisch, K.; White, S. H. *Biochemistry* **1998**, *37*, 14713−14718.
(31) Armen, R. S.; Uitto, O. D.; Feller, S. E. *Biophys. J.* **1998**, *75*, 734−744.

was applied. To facilitate comparison between simulations, the same starting membrane structures were used as for the equivalent indole systems.

Equilibration and Unbiased Molecular Dynamics Methods. After construction, we equilibrated each system using restrained Langevin dynamics at 300 K. We ran a sequence of 1 ps trajectories, with harmonic restraints applied to all indole and POPC heavy atoms. The restraints were reduced from 100 to 10 kcal/mol-Å^2, stepping by tens, then from 10 to 1 kcal/mol-Å^2 stepping by ones. This was followed by 50 ps of unrestrained dynamics, rescaling velocities every 3 ps for temperature control. During this time, the periodic boundary was set to 34.5 × 34.5 Å in the membrane plane, by 78.0 Å along the membrane normal, with a 10 Å nonbonded cutoff.

Next, the systems were run for 75 ps of dynamics with constant surface tension and temperature. At this point, we replaced the electrostatic cutoffs with particle mesh Ewald sums. Each of the six systems was then run for roughly 1.2 ns. In all cases, the first 200 ps were treated as equilibration and not used for analysis. The NMI simulations were run in the same manner, except that constant surface tension was used throughout. The 50 ps of velocity scaling dynamics was replaced with constant surface tension (35 dyn/cm) and temperature (300 K) dynamics as well. This allowed us to avoid some stability problems early in the equilibration.

All system construction and dynamics was performed with CHARMM, using the lipid−protein parameter set of Schlenkrich et al.[32] Parameters for *N*-methylindole were created by modifying the indole parameters.

Biased Dynamics Methods. This work represents, to our knowledge, the first application of umbrella sampling methods to the binding of permeants to lipid bilayers. The protocol used for the *biased* simulations was in most ways the same as that used for the *unbiased* molecular dynamics systems. Snapshots from the simulations of IND and NMI in the membrane center were used as starting points. The configurations were chosen such that both IND and NMI were roughly 5 Å from the center of the bilayer. The electrostatic interactions were calculated using particle−mesh Ewald summation, as implemented in CHARMM version 26.[33] The position of the indole center of mass along the membrane normal (the *z* direction) was restrained using a 5 kcal/mol-Å^2 harmonic potential. Constant surface area was maintained in each dynamics window.

Each dynamics window began with 500 steps of adopted basis set Newton−Raphson minimization, followed by 100 steps of steepest descent minimization. This was followed by 25 ps of equilibration. Finally, 50 ps of production dynamics was run, recording coordinates every 50 fs. The final coordinate set for the window was used as the initial structure for the next one. Each simulation window required roughly 36 h using two 400 MHz Pentium II processors.

The location of the biasing potential in consecutive simulation windows was chosen according to a dynamic windowing scheme. This method, described in detail in the Appendix, can be summarized as an attempt to use information from the current simulation window to determine the optimal location for the next window. The probability distribution function along the reaction coordinate is calculated, and the next window is chosen such that its center significantly overlaps the current distribution. This allows us to guarantee a sufficient number of data points at all locations along the reaction coordinate without oversampling the regions where the simulation relaxes quickly. The end result was a total of 93 windows for IND and 128 windows for NMI. The *biased* IND and NMI simulations contained a total of 7.0 and 9.6 ns of dynamics, requiring several months each to run.

The free energy profiles for the indoles were constructed from the *biased* simulations using WHAM, the weighted histogram analysis method.[34−36] Specifically, a time course for the position of the indole relative to the lipid center of mass was extracted from each window. The final 25 ps from each window was read into a locally written Python program implementing the WHAM procedure.

The WHAM procedure is only reliable when there are a significant number of data points in each bin, so that the estimate of the probability density can be trusted. Moreover, since WHAM performs a global fit, insufficient sampling at the edges of the range can in principle distort the entire free energy curve. Accordingly, we first calculated a single histogram containing all data points from each simulation, to indicate the range of the reaction coordinate where sufficient sampling was defined. We found that there was sufficient data to calculate free energy curves over the range 7−30 Å from the membrane center for IND, and 4−30 Å for NMI. This decision was verified by comparing the calculated free energies to curves with narrower ranges. Calculations using broader ranges either did not converge in 10000 iterations or produced unphysical free energy curves.

It is common practice in the literature to restrict the analysis of *biased* simulations to the calculation of free energy curves. However, there is a great deal of additional information available in these simulations. To extract this information, we calculated the averages and fluctuations of a variety of quantities over the course of each window independently. However, given the large number of windows and their uneven spacing, the resulting plots are difficult to read. Accordingly, we have grouped the results according to the location of the biasing potentials.

All other analyses of the simulations were performed using CHARMM version 27[33] and locally written scripts in Perl and Python. See Appendix B for further discussion of the technical issues raised by these simulations.

3. Results

Molecular dynamics simulations allow us to examine the structure and dynamics of biological systems with atomic resolution. Here we present *unbiased* and umbrella *biased* simulations of indole and *N*-methylindole in POPC bilayers. We first consider the structure of the bilayer and the perturbations induced by the indoles. We then examine the structural and dynamic characteristics of the indole interactions with their local environment. Finally, our estimates for the overall energetics of indole partitioning will be discussed.

Bilayer Structure. An obvious first question is whether the presence of an indole molecule sufficiently perturbs the overall bilayer structure. The simplest way to do this is to compare the two lipid *monolayers*, one with an indole and the one without.

(32) Schlenkrich, M.; Brickmann, J.; MacKerell, A. D. J.; Karplus, M. In *Biological membranes*: *A molecular perspective from computation and experiment*; Merz, K. M., Jr., Roux, B., Eds.; Birkhauser Press: Basel, 1996; pp 31−81.

(33) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(34) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(35) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339−1350.

(36) Roux, B. *Comput. Phys. Comm.* **1995**, *91*, 275−282.

First, we looked at the electron density profiles of the bilayers, using the technique of Petrache et al.[37] By integrating the electron density for the leaflet containing the indole over the course of the *unbiased* simulations, we showed that the indole constitutes roughly 0.5% of the total electron density. We then divided the electron density into 0.1 Å bins; the indoles never accounted for more than 3% of the electron density in any bin. Visual inspection of the electron density curves for the lipid leaflets with and without the indole electron density shows no difference visible above the noise in the density function. This is not unexpected; Damodaran et al.[10] simulated a tripeptide inserted into the interface of a DMPC bilayer without noting significant disruption of the membrane structure—an indole should be an even smaller perturbation. Moreover, experimental work using far higher concentrations of indole revealed only minor structural changes to the bilayer.[30]

Another simple metric for examining overall bilayer structure is its thickness, most easily quantified by the distance between the average positions of phosphorus atoms in the two monolayers. When we calculated this property for each window of the *biased* simulations, we found that there was little variation among windows. The peak to peak separation is approximately 38 Å in both simulations, while the width of the peaks (≈5 Å) serves as a reminder of the significant structural disorder present in a fluid lipid bilayer. The upper leaflet of each bilayer, the one containing the indole, is slightly narrower than the lower one, as indicated by the location of the phosphate peaks (−19.5 and 18.5 Å for IND, −19.0 and 18.5 Å for NMI). This is consistent with solid-state NMR results showing that indole and *N*-methylindole reduce the lipid order parameters near the ends of the palmitoyl chains.[30] However, the difference may not be significant: the binning was done in 0.5 Å increments, and there was little variation as the indoles were moved through the system. As such, it is possible that this difference is an artifact of the system's initial construction or the method of measurement, rather than its response to the presence of the indoles.

While our simulations do not accurately capture long time scale lipid motions, they are long enough to look at isomerization of acyl chain dihedrals. By classifying dihedral angles as either *gauche* +, *gauche* −, or *trans*, we can probe local bilayer structure. Specifically, we examined the fraction of *trans* states for all acyl chain dihedrals. The first step in this analysis is confirming that this property converges on the time scale of our simulations. We combined the data from all of the biased windows from the IND and NMI simulations and compared the two lower leaflets, the leaflets not containing an indole, as shown in Figure 1. We calculated the frequency of trans states for each dihedral, averaged over all lipids in the *monolayer* (18 lipid molecules). Note, however, that this three-state division is artificial for the cis/trans state of the double bond and that we did not see any true changes in isomerization about the double bond in the simulations. The results from the IND and NMI simulations are very similar, and the fluctuations in the per monolayer average seem to indicate that there was sufficient time for a large number of states to be sampled.

Given this apparent convergence, we then proceeded to assess the effect of the indoles on the fraction of trans states. Figure 2 compares the unperturbed leaflets to the

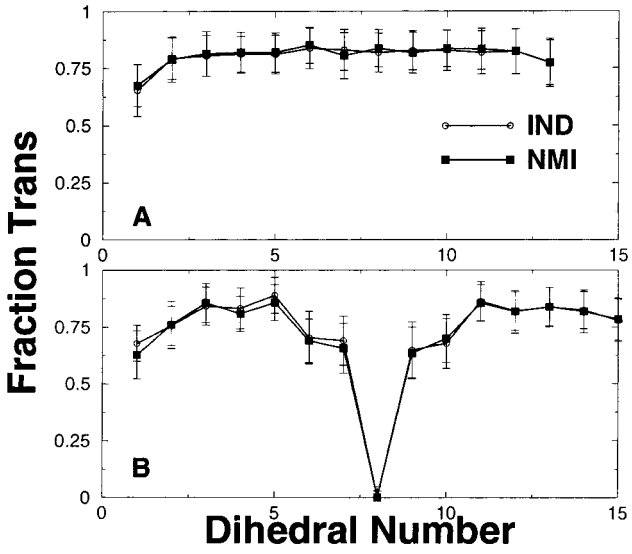

Figure 1. Frequency of trans dihedrals in unperturbed leaflets. This figure shows the fraction of trans states for a given dihedral, averaged over the entire monolayer and all windows of the *biased* simulations. The error bars indicate the fluctuations in this average quantity. By convention, dihedral 1 involves the carbon closest to the lipid headgroup. Part A compares the results for the palmitoyl chain from IND and NMI, while part B shows the oleoyl chain. Dihedral 8 of the oleoyl chain is the cis double bond. The two unperturbed bilayers are extremely similar.

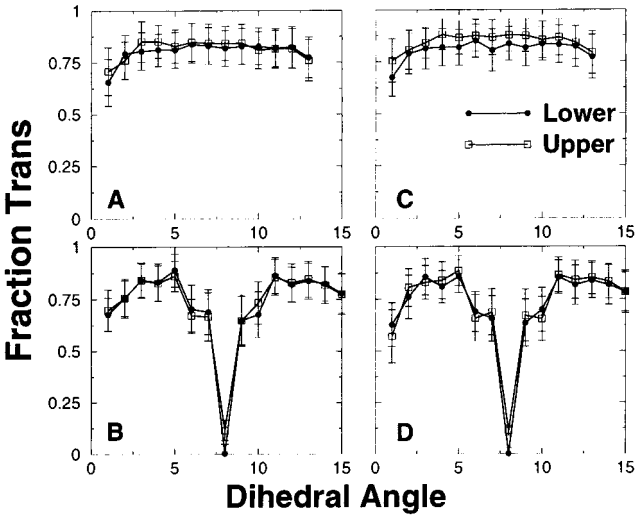

Figure 2. Frequency of trans dihedrals in the biased simulations. This figure shows the same quantity as Figure 2, comparing the two leaflets. The upper leaflet (empty squares) contains the indole, while the lower leaflet (filled circles) does not. Parts A and B show the palmitoyl and oleoyl results from the *biased* IND simulation, respectively. Parts C and D show the same quantities, from the *biased* NMI simulation.

leaflets with indoles for IND and NMI. Parts A and B of Figure 2 further suggest that IND makes no gross distortions to the computed bilayer structure. The palmitoyl chain has a higher trans frequency near the top of the chain, perhaps indicating ordering effects of the indole. This effect is more pronounced in part C of Figure 2, where all palmitoyl dihedrals have a higher trans frequency in the leaflet containing *N*-methylindole. This is consistent with recent solid-state NMR experiments, showing that IND and NMI increase the lipid order parameters near the top of the palmitoyl chain while lowering order parameters further down the chain.[30] There is no clear

(37) Petrache, H. I.; Feller, S. E.; Nagle, J. F. *Biophys. J.* **1997**, *72*, 2237−2242.

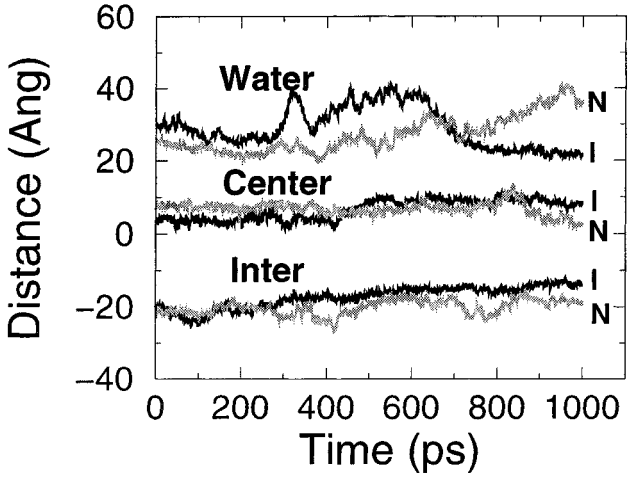

Figure 3. Time course of indole position. Time courses for indole position along the membrane normal from the *unbiased* IND and NMI simulations. "Center", "Inter", and "Water" refer to the initial environment of the indole. The NMI simulation data are gray and marked with an "N", while the IND simulation data are black and marked with an "I".

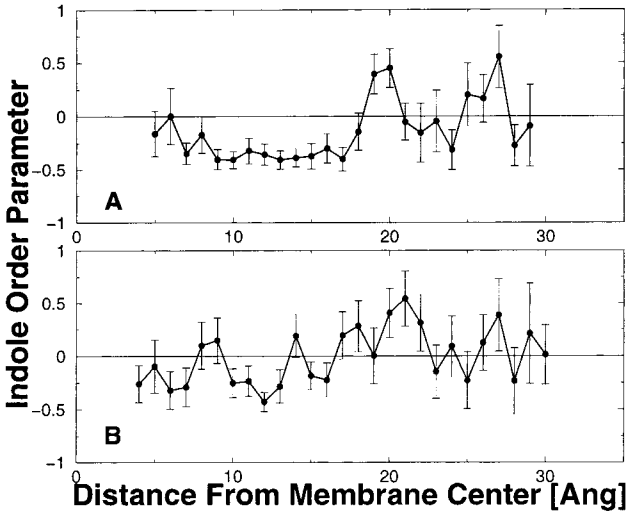

Figure 4. Indole order parameter. The order parameter $S = \langle 3\cos^2\theta - 1\rangle/2$, where θ is the angle between the indole normal and the membrane normal, calculated from the *biased* simulations. Part A shows the IND data, and part B shows the NMI data. Averages and fluctuations are calculated for each window and then grouped according to the location of the minimum of the biasing potential.

pattern to the trans frequency differences seen for the oleoyl chain in either simulation.

Ideally, the simulations could be used to determine the local molecular origins of these differences, by tracking the behavior of the lipids nearest the indole. Unfortunately, lipids are large molecules with many degrees of freedom and long relaxation times, while the indoles are quite mobile, and interact with many different lipids over the course of the trajectories. As a result, any given lipid molecule may not have time to relax in the presence of the indole (a process that could easily take 10−100 ps or more) before the indole moves away. Individual lipid molecules in our simulations have a wide range of self-energies (≈ 15 kcal/mol) and conformations; there is no obvious way to separate these large fluctuations from perturbations due to the indole at the level of a single molecule.

The long time scales of lipid relaxation make comparison difficult between computed lipid order parameters and those from experimental solid state NMR. This is because a small lipid patch with a permeant, simulated for less than tens of nanoseconds, is unlikely to reproduce experimental order parameters for each lipid (ergodic versus spatial averaging). Because other simulations, partly through their initial construction process and spatial averaging process, have demonstrated that the present parameters and methodologies are capable of reproducing these quantities,[38,39] and since our simulation cells were extracted from just this type of well-equilibrated simulation, we made no further attempt to select our snapshots to quantitatively reproduce the experimental order parameters. Unfortunately, since our simulations were not long enough to adequately sample lipid tilting motions (which would require tens of nanoseconds[22]), the *unbiased* simulations produce qualitative but not quantitative agreement with experiment (data not shown). This does not indicate that the lipid structures are physically poor but rather that lipid order parameters are a weak measure for the quality of a small simulation cell, simulated on a nanosecond time scale.

Indole Behavior and Local Environment. Figure 3 shows the position along the membrane normal (z) of IND or NMI over the course of the six *unbiased* simulations, where $z = 0$ is the membrane center. In the center simulations, both IND and NMI remain entirely in the hydrophobic core of the membrane, although they were clearly not trapped in a single location. In the interfacial simulations, both permeants for the most part remain associated with the membrane (they were inserted into the lower leaflet), although NMI did occasionally dissociate from the membrane briefly. When NMI was started in the water region, it briefly approached but never actually associated with the membrane. By contrast, IND had clearly made contact with the membrane by the end of the simulation. This is consistent with the experimental results indicating that indoles should bind to the membrane interface. It is clear that the indoles did not have enough time to migrate between the different environments during 1 ns of dynamics, further demonstrating the need for the *biased* simulations.

It is interesting to consider the overall orientation of the indoles, specifically the orientation of the indole ring, as this indicates how the indoles pack in different environments. We calculated the orientational order parameter

$$S = {}^1/_2\langle 3\cos^2\theta - 1\rangle \qquad (1)$$

for the indole as a whole, where θ is the angle between the normal to the indole ring and the membrane normal, using the *biased* simulations. Figure 4 shows the order parameter plotted as a function of indole position in the membrane. This order parameter is analogous to quantities available experimentally. Indeed, order parameters for the individual C−H bonds on the indole rings have been measured.[30] These quantities converge very slowly in molecular dynamics simulations, because they depend not only on the ring normal but on the precise orientation of the molecule, with the result that their direct calculation from a 1 ns simulation of a single indole is implausible. Figure 4 shows that the indoles tend to pack in certain orientations, largely because the rigid indole ring can only be accommodated when roughly aligned with the nearby lipids. However, the bond order parameters also contain

(38) Feller, S. E.; Venable, R. M.; Pastor, R. W. *Langmuir* **1997**, *13*, 6555−6561.
(39) Feller, S. E.; Pastor, R. W. *J. Chem. Phys.* **1999**, *111*, 1281−1287.

information about rotation about the ring normal; such rotations do not significantly alter overall packing, so the membrane does not immediately drive the indole to a preferred orientation, and the simulations are not long enough for thermal tumbling alone to vary quantities sufficiently for statistical convergence.

The most interesting feature of Figure 4 is the existence of a plateau in the IND curve, between 8 and 17 Å from the membrane center. The order parameter there corresponds to an angle of 70° with the membrane normal, in good agreement with NMR work suggesting the indole plane is generally parallel to the membrane normal.[30] As the indole moves further toward the membrane center, it becomes significantly less ordered; the larger fluctuations in the region indicate IND is becoming more mobile, rather than being strongly oriented parallel to the membrane. This decrease in order mirrors the changes in lipid order parameters in the same region; it seems likely that indole orientation and mobility are being governed by the orientation of the surrounding lipid. The indole also becomes less ordered when far from the membrane center. This makes sense given that the water has less anisotropy than the lipid environment.

The NMI order parameter curves have fewer direct trends. Generally, they are more disordered than IND, with an average angle of roughly 50° and larger fluctuations. Like IND, NMI is only weakly oriented in the water region.

IND and NMI both have a peak in their order parameters roughly 20 Å from the membrane center. The effect is more pronounced in IND, where it seems to correspond with the formation of hydrogen bonds to the lipid rather than water (e.g., see Figure 7). However, it is present in NMI as well; it may correspond to a partial resolvation by lipid, indicated by the drop in indole−lipid interaction energy that will be shown in Figure 11.

One critical aspect of the interaction of IND with its environment is the presence of hydrogen bonding. We can examine this in a molecular dynamics trajectory by calculating the number of oxygen atoms within 2.5 Å of the imino hydrogen and computing a radial distribution function. Figure 5 shows these results, as calculated from the *unbiased* simulations; we performed the analogous calculation using the methyl carbon from NMI for comparison purposes. Part A of Figure 5 demonstrates the presence of hydrogen bonding; specifically, the presence of a well-defined minimum after the hard-sphere exclusion peak indicates a favorable hydrogen-bonding interaction. This minimum is present in the two IND curves; not surprisingly, there is no corresponding minimum in the oxygen distribution about the NMI methyl group.

The curves in part B of Figure 5 represent the radial distribution for lipid oxygens only. Only the interfacial IND simulation shows any significant hydrogen bonding to the lipid, accounting for roughly one-fourth of the total hydrogen bonding seen in Figure 5A. The peak is also slightly nearer the hydrogen, around 1.8 Å, possibly indicating a tendency to form stronger hydrogen bonds.

This impression is confirmed by examination of hydrogen bonding lifetimes in the *unbiased* IND interface and water simulations. The trajectories were sampled every picosecond, and all oxygens within 2.5 Å of the imino hydrogen were recorded. The cutoff value of 2.5 Å was chosen to coincide with the minima from the radial distribution functions. The lifetime of a given hydrogen bond was estimated from the number of calculated consecutive structures. One difficulty with this estimate is its bias toward very short lifetimes, due to "flickering"

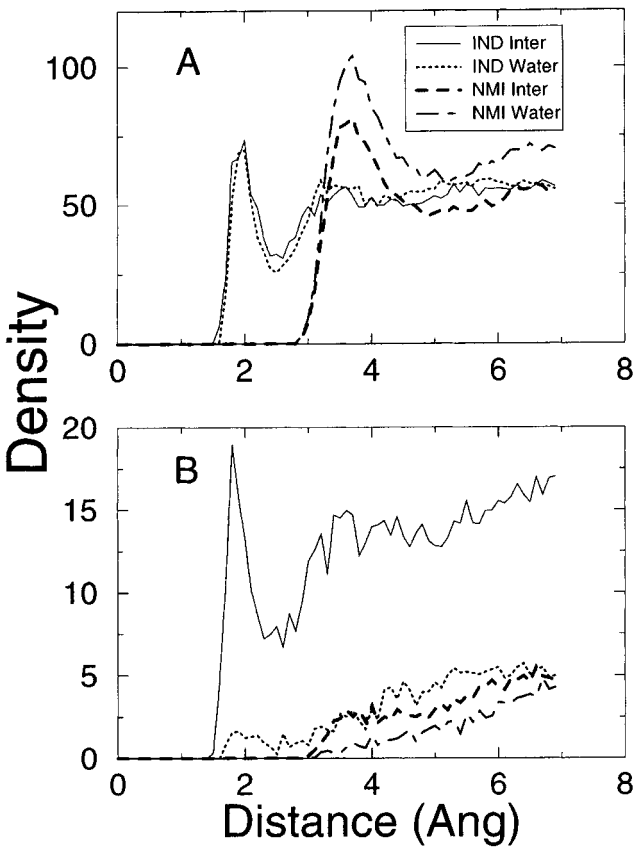

Figure 5. Radial distribution functions for oxygen about the imino hydrogen and *N*-methyl carbon from the *unbiased* IND and NMI simulations, respectively. Part A is the distribution for all oxygens, while part B includes only the lipid oxygens.

of hydrogen bonds near the cutoff length. A correction for this effect could be defined, but would probably require a full paper of its own to defend the approach; thus the simplest possible method was used. Since our main concern is comparison across the simulations, a consistent method was our main concern. Because the simulations in both environments contained hydrogen bonds to both lipid and water, their results were combined.

The *unbiased* simulations indicate that some form of hydrogen bond is present about 79% of the time in the interface simulations and for about 70% of the water simulation. The average lifetime for a hydrogen bond to lipid was 4.5 ps, while hydrogen bonds to water lasted only 2 ps on average. Figure 6 shows the probability distribution for the lifetimes of hydrogen bonds to lipids and to water. Unsurprisingly, both distributions are dominated by many very short hydrogen bonds, in part because of the simplistic hydrogen bond criterion. The log-scale plot of the probability distribution of the lifetimes of hydrogen bonds to water is nearly linear. The inverse of the slope (0.9 ps) is the decay time for these hydrogen bonds, if we assume that the lifetimes of hydrogen bonds fall in a Poisson distribution. The distribution of lipid hydrogen bond lifetimes, however, is not linear, with longer lifetime interactions, compared to water. Indeed, there are several events of hydrogen bonds lasting more than 25 ps (data not shown)—these are not seen for water despite the significantly larger number of events. This could imply that hydrogen bonds to lipid are stronger than those to water, consistent with the shift seen in the radial distribution functions. This is consistent with the long (approximately 50 ps) decay time seen for hydrogen bonding between water and lipid, in other molecular

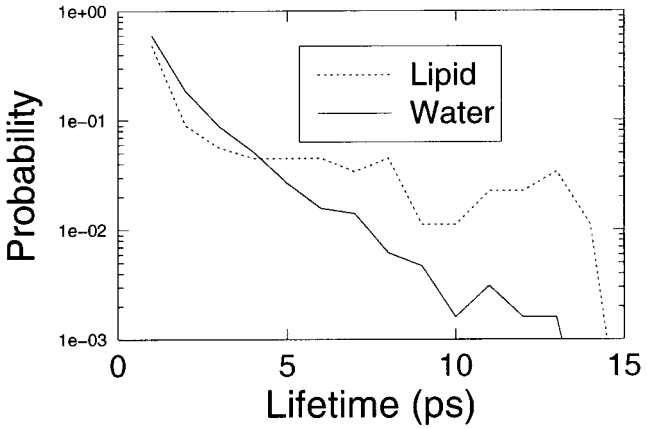

Figure 6. Distribution functions for lifetimes of hydrogen bonds to lipid and water, from the *unbiased* IND interface and water simulations.

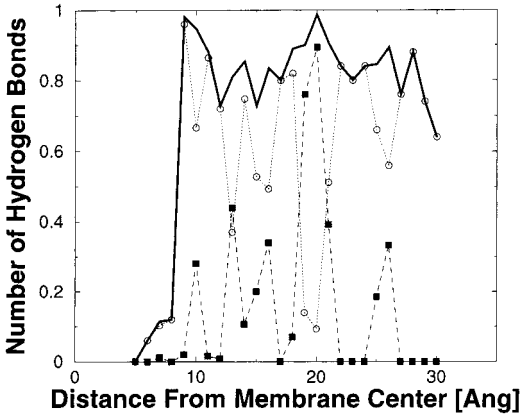

Figure 7. Average number of hydrogen bonds made by IND, as a function of location in the membrane, calculated from the *biased* simulation windows. The thick solid curve counts all hydrogen bonds. The empty circles connected with a dotted line indicate hydrogen bonds to water, and the filled squares connected with a dashed line indicate hydrogen bonds to lipid. Averages and fluctuations are grouped as in Figure 5.

dynamics simulations.[40] However, it is also possible that this is simply the result of insufficient sampling due to the relatively small number of lipid hydrogen bonds observed. Alternatively, it could simply reflect the lower mobility of lipid molecules relative to water.

We can use the *biased* simulations to further characterize the propensity to form hydrogen bonds, as a function of indole location. Figure 7 shows the number of hydrogen bonds present, averaged over the windows and then grouped by bias location. These results are consistent with *unbiased* simulation results seen in Figure 5: an imino hydrogen bond is present roughly 80% of the time, most of the hydrogen bonding is to water rather than lipid, and the average number of hydrogen bonds drops slightly when the indole is in the water. The most startling difference is the presence of occasional hydrogen bonding when the indole is only 6 Å from the membrane center, an indication of the thermal disorder present in lipid bilayers. No such interaction was visible in the *unbiased* dynamics, perhaps in part because there was insufficient data to parse the results as a function of location.

Finally, we can characterize the local environment surrounding the indoles by examining their molecular

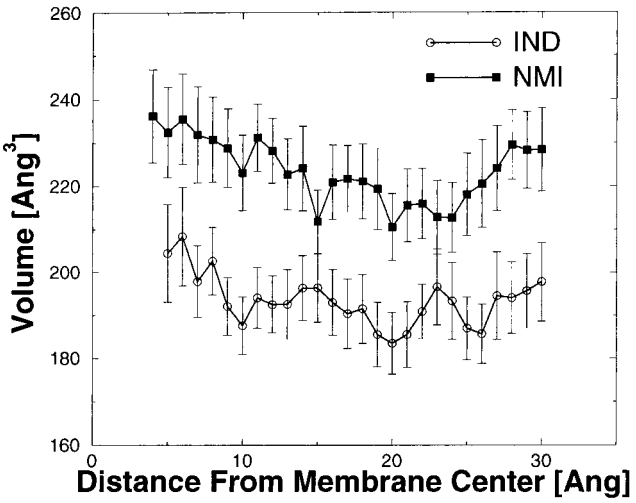

Figure 8. Indole volumes. Total molecular volume for IND and NMI, calculated from the *biased* simulations using the Voronoi procedure. The averages and fluctuations are calculated as in Figure 5.

volume. Specifically, we use an implementation of the Voronoi algorithm by Gerstein et al.,[41] calculating, for a given set of coordinates, the volumes associated with each atom. It should be noted that these volumes will be different (and larger) than simple hard sphere volumes, because all of space is allocated to atom volumes. Figure 8 shows the results for IND and NMI from the *biased* simulations. The most obvious characteristic of these curves is that NMI is larger, by about 30 Å^3. This is consistent with the volume previously calculated for a lipid methylene group.[37]

For the most part, IND and NMI volumes have the same dependence on membrane location. The largest volumes occur in the center of the membrane. This makes sense, given the lower overall density of the membrane interior[42] and high disorder of the lipid chains measured by NMR.[43,30] The volumes drop as the indoles approach the membrane–water interface and then rise as they leave the membrane entirely. The volumes seen 30 Å from the center are the same as those calculated from simulations of indole and *N*-methylindole in pure water (data not shown). The presence of a minimum at 20 Å in both curves is indicative of the high density of the interface and the tight packing of the indole with the lipid headgroups.

Energetics and Thermodynamics. One of the major advantages of molecular dynamics simulations is the ability to dissect molecular interactions in ways unavailable to experiment. Of course, interaction energies by themselves do not give the full thermodynamic picture, especially in an environment as complex as a membrane. However, they do help to directly characterize the indole behavior. Furthermore, the interaction energies can be compared to the free energies discussed in the next section.

Figure 9 shows the average interaction energies from the windows, calculated from the *biased* simulations and grouped as in Figure 4. Both IND and NMI have their most favorable interactions when located in the headgroup region, roughly 20 Å from the membrane center. The presence of a volume minimum at precisely the same location is a further indication that the indoles are capable of efficient packing in the headgroup region. It is interest-

(40) Pasenkiewicz, M.; Takaoka, Y.; Miyagawa, H.; Kitamura, K.; Kusumi, A. *J. Phys. Chem. A* **1997**, *101*, 3677–3691.

(41) Gerstein, M.; Tsai, J.; Levitt, M. *J. Mol. Biol.* **1995**, *249*, 955–966.

(42) Wiener, M. C.; White, S. H. *Biophys. J.* **1992**, *61*, 428–433.

(43) Seelig, A.; Seelig, J. *Biochim. Biophys. Acta* **1975**, *406*, 1–5.

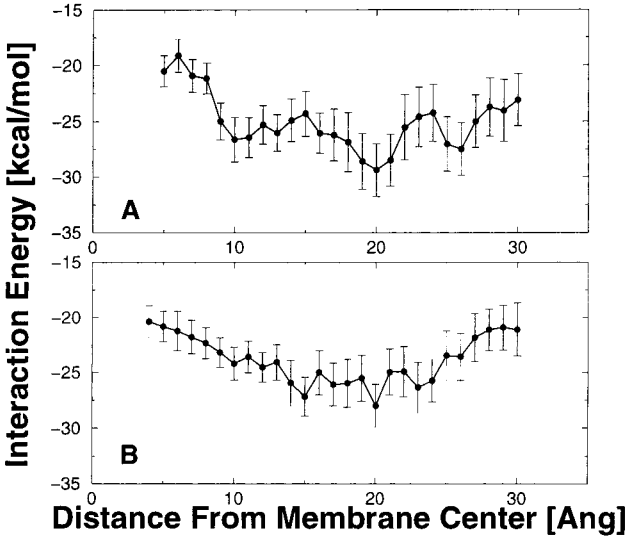

Figure 9. Total interaction energy between indoles and the environment, from the *biased* simulations. Part A is IND. Part B is NMI. The averages and fluctuations are grouped as in Figure 5.

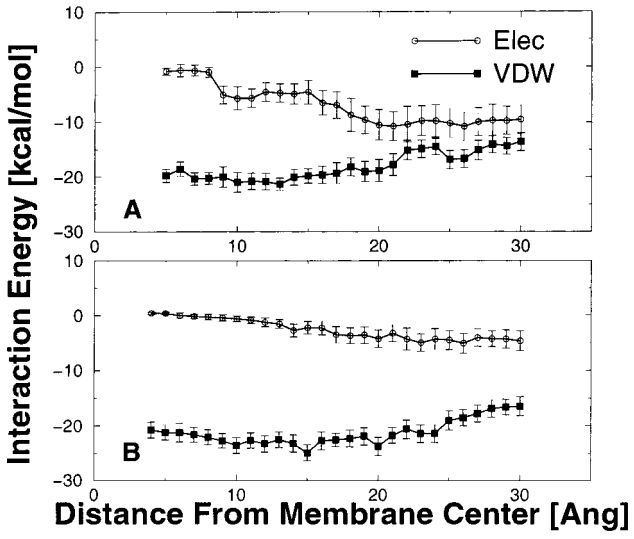

Figure 10. Electrostatic and van der Waals components of indole interaction with the environment, from the *biased* simulations. Part A is IND. Part B is NMI. The averages and fluctuations are grouped as in Figure 5.

ing to note that the IND interactions are relatively flat over the region 10−18 Å from the membrane center, the same region where its order parameter was approximately constant. Furthermore, the interactions are of roughly the same magnitude as in the water region (28−30 Å). The dip at 25 Å appears to be due to shared hydrogen bonds with both lipid and water; hydrogen bonds to lipid appear to be stronger than those to water, as shown in Figures 5 and 6. The NMI curve is smoother than the one for IND; the headgroup minimum is less pronounced and seems to be part of a broad well, perhaps because it has no hydrogen bonding moiety. However, as with IND, its interaction energy with the environment is almost the same in the membrane interior as it is in the water. It is interesting to note that the most favorable interaction energies occur in precisely the same environment where the indole molecular volumes are minimized. This, plus the presence of favorable van der Waals energies, indicates that the indoles are capable of packing quite efficiently with the lipid headgroups.

Figure 10 shows the dependence of the electrostatic and van der Waals (VDW) components of the interaction energy on membrane position: in the membrane center, the VDW interactions are dominant, with the electrostatic interactions gradually becoming more significant as the indole moves outward. NMI electrostatic interactions remain weak (≤5 kcal/mol), while IND interactions plateau around 10 kcal/mol. Note that exchanging water for lipid as a hydrogen bonding partner at 20 Å does not induce a significant decrease in the overall electrostatic interaction energy.

Still more insight can be gained by examining the indole interactions with lipid and water separately. This is shown in Figure 11. The main features of the curves are readily seen: the indoles interact mostly with the lipids when deeply buried and gradually acquire interactions with the water as they move outward. IND starts interacting with water quite early: the hydrogen bonds discussed in Figure 7 appear to account for approximately 5 kcal/mol of favorable interaction. Note that while the electrostatic profile is relatively smooth, there is a peak in the water interaction 25−27 Å from the membrane center and a corresponding dip in interaction with the lipid. This is

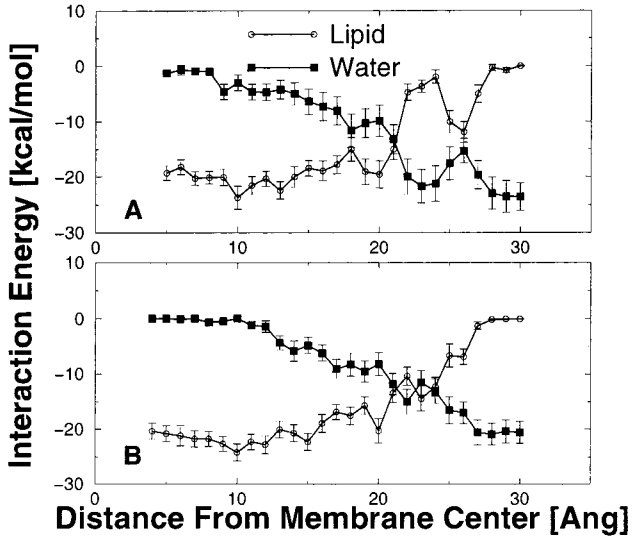

Figure 11. Indole interaction energy with lipid and water, from the *biased* simulations. Part A is IND. Part B is NMI. The averages and fluctuations are grouped as in Figure 5.

another indication of the late-forming indole−lipid hydrogen bonds discussed above and shown in Figure 7.

Interaction energies alone do not tell the entire thermodynamic story, rather they emphasize the enthalpic component and ignore the entropic. The more complete quantity is the free energy, calculated as a function of indole position. The *biased* simulations were used to construct the potential of mean force curves for IND and NMI, using WHAM,[34−36] as described above. The resulting free energy curves are shown in Figure 12.

The free energy curves do not parallel the interaction energies, as one might naively expect. Neither IND nor NMI has a minimum in the interface: IND has a weak minimum approximately 11 Å from the membrane center, while NMI simply plateaus at about 7 Å. Moreover, the free energy curves show no clear barrier to leaving the membrane. Rather, the free energy rises smoothly until about 25 Å from the membrane center and then either plateaus (in the case of IND) or tails slightly (in the case

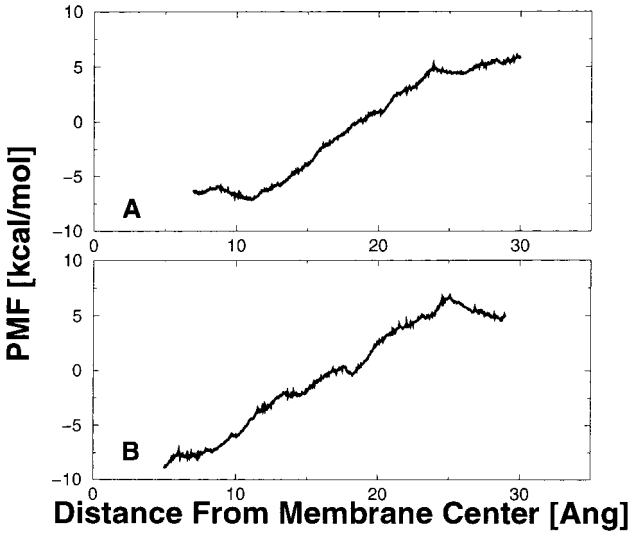

Figure 12. Free energy curves. Free energy curves as a function of position along the membrane normal. Part A is IND, and Part B is NMI. Since WHAM produces relative (as opposed to absolute) free energies, the midpoint of each curve was arbitrarily offset to 0 kcal/mol.

of NMI). While it is possible to interpret these results physically, as we do below, it is important to acknowledge that this result could also be an artifact of insufficient relaxation and sample time; too little equilibration time per window would likely cause the simulation to overestimate the free energy increase in the direction of sampling, in this case moving from the membrane center outward to the water.

Our free energy curves indicate that the Boltzmann averaged position for bound IND is about 11 Å from the membrane center, while NMI is lower, at about 7.4 Å. This region contains significant amounts of acyl chain and carbonyl and phosphate groups, in addition to a small amount of water.[44] This is consistent with the solid-state NMR experiments of Yau et al.[30] The average location for NMI is approximately 3.5 Å closer to the membrane center, perhaps because it is larger, more hydrophobic, and lacks a hydrogen bonding moiety to favor interaction with water. It should be noted, however, that the absence of free energy values for the region between 0 and 5 Å from the membrane center may cause us to significantly overestimate the average distance from the membrane center.

4. Discussion

There are two primary sources of experimental data on the binding of indoles to lipid bilayers: the binding thermodynamics measured by Wimley and White[45] and the solid-state and magic-angle spinning NMR measurements of indole location and orientation by Yau et al.[30]

Binding Thermodynamics. Wimley and White[45] measured the binding affinity of tryptophan analogs to POPC large unilamellar vesicles, using equilibrium dialysis methods. They found that IND and NMI both favor binding to the membranes, by 7.6 and 8.1 kcal/mol, respectively. These values can be compared to the free energy curves shown in Figure 12, once certain corrections have been applied. For example, we must define what in our simulation constitutes the bound and unbound states, since the experiments operate under the two-state assumption. Moreover, we must also compensate for the different concentrations of indole and lipid in the experiments versus the simulations.

Once the system's conformations are partitioned into bound and unbound macrostates, the microstates within each macrostate can then be Boltzmann averaged, allowing us to calculate the free energy difference between them. However, as discussed above and shown in Figure 12, the free energy curves themselves show no significant barriers as the indoles leave the membrane. As a result, we need to consider other ways to define the bound state. Beyond 20 Å from the membrane center, IND and NMI are both consistently surrounded more by water than lipid (data not shown). Figure 7 shows that hydrogen bonding between IND and lipid peaks falls sharply past 20 Å. Figure 8 shows minimum volumes around 20 Å, and Figure 9 shows the most favorable interaction energies take place around 20 Å. These factors, taken together, seem to indicate that 20 Å can be used as a reasonable dividing line. Using this definition of the membrane bound state, we can calculate the free energy of binding from our simulations as 9.2 kcal/mol for IND and 11.2 kcal/mol for NMI. It should be noted that while this definition of the bound state is arbitrary (i.e., indole binding to bilayers is not a simple two-state process), it is necessary for comparison to the experimental results. Further, the resulting free energy differences are not sensitive to the choice of dividing line.

The simulation values are still not directly comparable to experiment, because the experiments were performed using concentrations of lipid and indole different from those in the simulation. We estimate the effect of lipid concentration using a simple mixing entropy term. The average volume of our simulation cell was 85 622 ± 675 Å^3 for IND, and 85 512 ± 555 Å^3 for NMI. Counting the region from −20 to 20 Å as "indole binding", the volume fraction of bound states in our simulations χ_{sim} is 0.56 for IND and 0.57 for NMI.

Estimating the volume fraction of "indole binding" states under experimental conditions is more complicated. If we assume that the "indole binding" volume is equivalent to the total lipid volume, estimated by Wimley and White as 0.6 L/M,[45] their given concentration range of 2−4 mM produces volume fractions χ_{exp} in the range 0.12−0.24. The free energy change due to mixing entropy is approximately

$$\Delta G_{mix} = k_B T \ln \frac{\chi_{sim}}{\chi_{exp}} \qquad (2)$$

yielding a correction of roughly 0.5−0.9 kcal/mol for each simulation.

Similarly, we must also correct for the different concentrations of indole in simulations and experiments, since this will effect the chemical potential of the indole once it reaches bulk solution in water. The simplest way to estimate this free energy change is to use Raoult's law

$$\Delta G_R = k_B T \ln \frac{x_{sim}}{x_{exp}} \qquad (3)$$

where x_j is the mole fraction of indole in system j. If we assume all waters contribute to the bulk solvation of the indole, its mole fraction is $x_{sim} = 1/(1 + N_w) = 0.00074$. The bulk indole concentration in the experiments was 20−40 μM;[45] given the concentration of water in water is 56 M, $x_{exp} \approx 3.6 \times 10^{-7}$ to 7.2×10^{-7}. Substituting these values in the eq 2 yields $\Delta G_R \approx 4.1-4.5$ kcal/mol. Of course, indole almost certainly does not behave as an ideal solute in

(44) Wiener, M. C.; White, S. H. *Biophys. J.* **1992**, *61*, 434−447.
(45) Wimley, W. C.; White, S. H. *Biochemistry* **1993**, *32*, 6307−6312.

Table 1. Free Energy of Binding[a]

analogue	ΔG_{sim}	$\Delta G_{[indole]}$	$\Delta G_{[POPC]}$	ΔG_{tot}	ΔG_{exp}
IND	9.2	−4.1	−0.5	4.6	7.6
NMI	11.2	−4.1	−0.5	6.6	8.1

[a] Free energy of indole binding to POPC bilayers compared to experimental work of Wimley and White.[45] ΔG_{Sim} is the free energy difference calculated directly from simulation, $\Delta G_{[indole]}$ and $\Delta G_{[POPC]}$ are correction factors adjusting for different concentrations of indole and lipid in experiment and simulation, ΔG_{tot} is the sum of these terms, and ΔG_{exp} is the experimental value. All units are kcal/mol. For simplicity, the $\Delta G_{[indole]}$ and $\Delta G_{[POPC]}$ are the smallest of the possible range. See text for details.

water, so this correction is probably not particularly accurate. We include it here mainly to indicate that a correction for the different concentrations could be important and to reemphasize the subtleties involved in comparing microscopic simulations to macroscopic thermodynamic measurements. Table 1 summarizes these results.

Our free energy curves indicate that indole and *N*-methylindole both have their lowest free energies in the hydrophobic core, with Boltzmann averaged positions 7−11 Å from the membrane center. This is closer to the membrane center than was previously inferred from the analysis of the tryptophan locations in the sequences of integral membrane proteins.[24] However, these analyses assume that all transmembrane helices are oriented parallel to the membrane normal and that the bilayer width is not distorted by the helices. In fact, this is not the case; the recent simulations of Petrache et al.[21] indicate that the average tilt of isolated helices depends sensitively on the location of the tryptophans and that the lipids immediately surrounding the peptide adjust to place the tryptophans in the interfacial region. Moreover, this average location is also consistent with the NMR measurements of Yau et al.[30]

Both IND and NMI have their most favorable overall interaction energies with their environment when 20 Å from the membrane center, in the middle of the headgroup region, and far from their free energy minimum. These states may be thermodynamically unfavorable because they disrupt the headgroup region, perhaps by forcing partial desolvation of the lipid headgroups. Previous simulations have indicated that "bridging" waters often form hydrogen bonds to two lipids, forming local clusters of two to seven lipids.[40] It seems quite reasonable that indole would partially disrupt these structures, given the frequency with which it forms hydrogen bonds. If this is indeed the case, then the origin of tryptophan's interfacial location becomes clearer. The heterogeneous chemical nature of the headgroup region allows indole to optimize its interaction energies. However, without a protein attachment, it cannot make enough strong interactions to compensate for its disruptive properties and so partitions further into the membrane. On the other hand, when an α-helix is inserted into a lipid bilayer, the cost of disrupting the membrane has already been paid by the backbone, and the possibility of strong interaction energies with the environment make it favorable to place tryptophan residues in the interface. Unfortunately, the present work cannot confirm or disprove this hypothesis. Such analysis will require simulations with a variety of aromatics, small peptides, and helices bound to bilayers (such as the work of Petrache et al.[21]).

Alternatively, as stated above, it is not entirely certain that the free energy results are free of sampling artifacts. Significant care was taken to ensure adequate sampling; the dynamic windowing procedure was intended to guarantee that the distributions from consecutive windows had sufficient overlap. The individual distributions were also examined visually; for the most part, they were reasonably smooth and Gaussian-like and free from obvious sampling artifacts. Furthermore, the time courses along the membrane normal were examined visually, to ensure that systematic motion in response to the new restraint was not included as part of the equilibrium sample. It was for this purpose that only the final 25 ps of each window was analyzed. Other properties, such as interaction energy, volume, orientation, and hydrogen bond frequency did not appear to require as long to equilibrate but rather appeared free of systematic change after the initial 25 ps period. However, for consistency we used the same data points for all analyses.

The only way to eliminate the uncertainty due to undersampling is to greatly increase the amount of sampling performed. For example, the *biased* simulations could be run in the opposite direction, dragging the indole from the water to the membrane center. Additionally, it would be instructive to greatly increase the dynamics time in each window, by as much as a factor of 10 or more. However, such calculations are beyond our present resources; as it is, each of the *biased* simulations required several months of computer time to complete.

NMR Experiments. There has been a major effort to use NMR methods to better understand the structural characteristics of tryptophan analogs bound to a POPC bilayer.[30,46,47] Much of these data can be compared, directly or indirectly, with quantities calculated in our molecular dynamics simulations.

For example, Yau et al.[30] used solid-state deuterium NMR to measure the effect of indole and *N*-methylindole on the lipid palmitoyl chain order parameters. They found that both indoles increased the order in the upper parts of the chains while decreasing the order farther down. Direct comparison with the experimental order parameters is difficult, because the lipid order parameters depend sensitively on properties such as overall lipid tilt and thus relax on a time scale longer than the current simulations.[22] However, Figure 2 shows that the indoles appear to increase the fraction of trans states, especially in the upper chains. It is not surprising that we do not see a decrease in trans states further down the chains; the bond orientations near the ends of the acyl chains are mediated not just by local chemical geometry but by the distribution of isomers above them.

Several kinds of experiments have been performed in an attempt to describe the average location and orientation of the indoles in the bilayer. Neutron diffraction was used to locate the tryptophan side chain of a blocked tripeptide.[27] These data, also presented in ref 30, clearly show the tryptophan residues overlapping the headgoups, far from the membrane interior. However, it must be recalled that the peptide is much larger and contains a polar backbone which resists desolvation. Yau et al.[30] analyzed the chemical shifts induced by the indole ring currents. The largest induced shifts were on the uppermost carbons of the acyl chains and in the headgroups, with relatively small shifts along the rest of the acyl chains and in the glycerol region. While these experiments do not give the indole location to high resolution, they do seem to indicate that the indoles are probably found mostly in the upper acyl chain and headgroup regions. This is not inconsistent

(46) Yau, W. M.; Steinbach, P. J.; Wimley, W. C.; White, S. H.; Gawrisch, K. *Biophys. J.* **1998**, *74*, A303.
(47) Yau, W. M.; Gawrisch, K. *Biophys. J.* **1999**, *76*, A355.

with the results presented with the results in this paper. It is true that the experimental "structure" of a liquid crystalline bilayer of Wiener and White[44] indicated that the interface begins somewhat farther from the membrane center. However, those experiments were performed at very low hydration, where the surface area per lipid is approximately 60 Å^2, while in the present simulations, performed at full hydration, the surface area is approximately 67 Å^2.

Moreover, in interpreting the experimental data, one must recall that the induced chemical shift on a given group due to indole ring currents is affected by the overall probability of having indole in close proximity to that group, the length of time that proximity is maintained, and the orientation of the indole relative to both the group and the magnetic field. Figure 4 indicates that indole orientation (and mobility) depends significantly on location, so that the induced shifts cannot be interpreted on the basis of proximity alone. Specifically, the high mobility and relatively flat orientation observed for IND from 5 to 9 Å from the membrane center would indicate that it should not induce large chemical shifts in the nearby protons. Moreover, IND's highly ordered plateau region from 10 to 17 Å corresponds to the upper acyl chains; indole molecules located in this region would have strong ring current effects. Similarly, hydrogen bonds to lipid (see Figure 7) would produce the sort of long-lived, strongly oriented configurations needed to produce the relatively large induced chemical shifts seen experimentally. The result would be an effective overrepresentation of signal from indoles located in this region.

Yau et al.[30] also argued that indole is usually oriented with the plane of the ring parallel to the membrane normal, on the basis of induced chemical shift and deuterium quadrupolar splitting data. This orientational preference is also seen in the simulations (especially IND), as demonstrated by Figure 4.

5. Conclusions

The present work examined the behavior of indole and *N*-methylindole, bound to lipid bilayers. The fundamental question we wished to answer was, why do tryptophan residues appear frequently in the interface when bulk partitioning studies say they are hydrophobic? While our simulations do not entirely resolve this issue, they are a step in that direction. The present results indicate that the membrane interior is not entirely hydrophobic; IND began to form hydrogen bonds when only 9 Å from the membrane center. Moreover, these hydrogen bonds may well be stronger than those found in water. IND and NMI both make their most favorable energetic interactions with their surroundings when embedded in the headgroup region, although their free energy minima and average locations are somewhat closer to the membrane center, in an environment containing acyl chain, headgroup, and water. This disparity can best be accounted for by assuming that indole disrupts headgroup−water interactions, such that on average they are lower in the interface despite the energetic favorability of the headgroup region. This in turn helps explain the propensity for tryptophan residues to appear at the ends of transmembrane helices: the helix intrinsically disrupts the local structure of the bilayer, paying the price necessary to gain the favorable membrane−indole interactions observed here. Further simulations, of tryptophan-containing helices bound to lipid bilayers,[21] should help to prove or disprove this hypothesis.

Acknowledgment. We wish to thank Dr. Klaus Gawrisch and Dr. Stephen White for many helpful discussions, especially relating to the interpretation of the experimental results. Dr. Scott Feller gave us the snapshots from his POPC simulations we used to construct our systems. Horia Petrache helped us perform the volume and electron density calculations. He and Dan Zuckerman also made many helpful comments on this manuscript. This work was funded by the NIH (RR12600), the AHA (96011480), and the Department of Physiology.

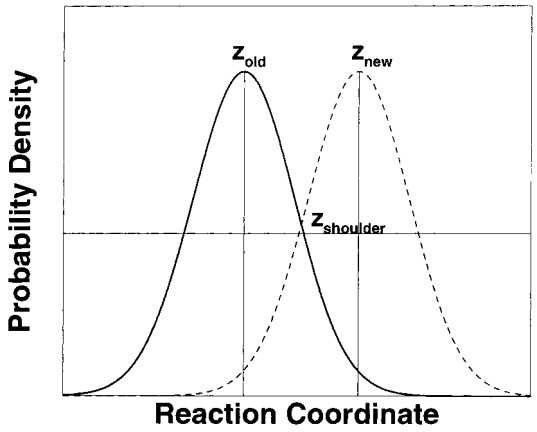

Figure 13. Schematic of the dynamic windowing procedure. z_{old} is the location of the maximum in the probability distribution from the current window (shown as a solid line). For simplicity, this figure assumes that z_{old} is also the location of the bias in the current window. The half-maximum for the distribution is drawn as a solid horizontal line. The position where the distribution intersects the half-maximum is marked by $z_{shoulder}$. z_{new} is the location for the bias in the next window. The dotted line shows an estimate of the probability distribution that would be produced by the next window. Note that the two distributions overlap significantly, such that no region is left unsampled.

Appendix A: Dynamic Windowing in Biased Simulations

Typically, the trajectory length for each window and the separation between windows (the distance between the minima in the biasing potentials of consecutive windows) are predetermined at the beginning of the simulation. As is often the case with molecular dynamics simulations, this choice reflects a compromise between computational cost and the need for adequate sampling.[48] Uniformly spacing the locations for the biasing potentials is intuitively appealing, since the goal of *biased* molecular dynamics is even sampling over the entirety of the reaction coordinate. This assumption is probably reasonable in many protein systems, where the system's relaxation times do not depend strongly on the reaction coordinate. By contrast, the relaxation times for the indoles vary with their locations in the membrane. Longer relaxation times imply that more trajectory time is needed for adequate sampling, since consecutive time points are strongly correlated. One could simply use window lengths and bias spacings appropriate to the longest relaxation times present anywhere in the system, but this would clearly be prohibitively expensive for membrane systems.[22,23] We instead chose to fix the simulation time per window and use information from each window to indicate the appropriate location for the next window, in a procedure we call dynamic windowing (see Figure 13 for a schematic).

To perform our windowing procedure, we used the following approach: after completing the dynamics for each window, we constructed a histogram of the position

(48) Kollman, P. A. *Chem. Rev.* **1993**, *93* (*7*), 2395−2416.

of the indole center of mass along the membrane normal. The biasing potential generally caused these histograms to be unimodal and roughly Gaussian. We located the maximum value for the histogram, then found the position of the half-maximum shoulder on the $+z$ side (because the simulation dragged the indole from bilayer center, $z \approx 5$ Å, to the water, $z \approx 30$ Å). The location for the restraint in the next window was calculated as

$$z_{new} = z_{old} + 2(z_{old} - z_{shoulder}) \quad (4)$$

subject to the constraint that

$$z_{old} + 0.1 \leq z_{new} \leq z_{old} + 0.5 \quad (5)$$

For convenience, z_{new} was always rounded to one decimal place. This restraint forces the system to continue moving in the correct direction along the reaction coordinate, without allowing it to move too quickly. More importantly, as long as the width of the distributions do not vary too much, this approach guarantees even sampling of the reaction coordinate.

Appendix B: Methodological Issues

Molecular dynamics simulation of lipid membranes is still a young field. The present work is, to the best of our knowledge, the first attempt to use umbrella sampling methods to explore the behavior of small molecules embedded in bilayers, and we have found that there are some methodological issues involved in *biased* simulations, particularly of membranes, which require discussion.

One issue has to do with the appropriate use of constant surface tension algorithms versus constant surface area calculations. There is extensive literature on this issue, and the choice of surface tension appropriate for a relatively small membrane patch has not been definitively resolved.[49−51] At issue is the relatively small size of current membrane simulations, such that the periodic boundary conditions abolish the undulations characteristic of flaccid lipid bilayers. Thus, while a macroscopic bilayer could properly be said to have zero surface tension, the surface tension of a flat microscopic patch is unclear. Furthermore, while it is true thermodynamically that the choice of ensemble has no influence on equilibrium average properties, it does influence properties based on fluctuations. More important for the present circumstances, in the context of a small simulation, the various ensembles equilibrate at different rates. A recent set of simulations of α-helices in explicit bilayers discussed the fact that constant surface tension simulations relax more slowly than constant surface area calculation and that the latter are sufficiently accurate to allow calculation of the membrane compressibility.[21]

The presence of membrane permeants and biasing potentials further complicates the situation. The current calculations originally consisted of a series of *unbiased* and *biased* simulations, both using the NPγT ensemble with a surface tension of 35 dyn/cm. This protocol produced stable bilayers when used with *unbiased* dynamics. However, the virial calculation is complicated by the presence of external forces (due to the biasing potential), with the result that *biased* simulations experience a rapid increase in the surface area/lipid. This is a known difficulty with constant surface tension simulations.[52] For this reason, the *biased* constant surface tensions were discarded, and the simulations were rerun with the surface area fixed at the value found in the starting configurations (67.8 Å^2/lipid for IND and 66.2 Å^2/lipid for NMI). We should note that while this difference in surface area may appear significant, it results from a difference in cell dimension less than 0.5 Å, significantly less than the fluctuations seen in *unbiased* constant surface tension simulations.

Another issue for the *biased* simulations concerns the appropriate form for the biasing potential. In particular, we initially applied harmonic restraints to the indole along the z axis, centered about a fixed plane in space, using the MMFP module of CHARMM. However, the true reaction coordinate was not the absolute z position, but rather the distance from the membrane center. The approximation initially seemed reasonable, since the dynamics algorithm prevented center of mass translation of the whole system.[53] However, it turned out that this was not the case; during minimization, the lipid center of mass moved in response to the restraint. We resolved this difficulty by applying an additional restraint on the lipid center of mass, effectively keeping it from moving. However, this is still not an ideal approach. While the "rolling" motion of the system did not cause any forces within the simulation cell, our solution does. Moreover, applying this force to only a subset of the atoms could in principle perturb the statistical distributions of the system. It does not appear likely that this perturbation is significant, since the restrained degree of freedom (the lipid center of mass) should not fluctuate very much under reasonable circumstances, but a nonperturbing method would have been preferred. A better solution may have been to directly restrain the distance of the indole from the lipid center of mass. The exact location of the minimum biasing energy would thus fluctuate with small motions in the lipid center of mass, but would be constant with respect to the reaction coordinate.

Finally, we consider the sampling issues unique to *biased* simulations of lipid bilayer. That is, the simulator must choose the length and location of the simulation windows for *biased* simulations with an eye toward both adequate sampling at acceptable computational cost, in several disparate environments. The usual approach involves choosing a specific window length and number of windows prior to the simulations. While this approach is reasonable in many circumstances, where the environment is relatively homogeneous and the system's relaxation times do not vary with location along the reaction coordinate, it seems likely that it is not the most efficient method for a heterogeneous system where relaxation times vary strongly with location. For the present simulations, we implemented a new, more flexible method, the dynamic windowing procedure. Its basic premise is to let the simulation decide how quickly to move the system along the reaction coordinate. Specifically, each window's probability distribution is used to indicate where the next window should be placed. We believe that this approach can efficiently sample reaction coordinates with a heterogeneous distribution of relaxation times.

Dynamic windowing is not the first attempt to use preliminary simulation results to guide the choice of biasing potential. In an analogous effort to dynamically

(49) Roux, B. *Biophys. J.* **1996**, *71*, 1346−1347.
(50) Jahnig, F. *Biophys. J.* **1996**, *71*, 1348−1349.
(51) Chiu, S. W.; Clark, M.; Balaji, V.; Subramaniam, S.; Jakobsson, E. *Biophys. J.* **1995**, *69*, 1230−1245.
(52) Cheatham, T. E.; Brooks, B. R. *Theor. Chem. Acc.* **1998**, *99*, 279−288.
(53) Harvey, S. C.; Tan, R. K. Z.; Cheatham, T. E., III *J. Comput. Chem.* **1998**, *19*, 726−740.

adjust the biasing potential to produce optimal sampling, Bartels and Karplus developed adaptive umbrella sampling.[54,55] In this approach the umbrella potential is iteratively adjusted to try and create a flat surface for sampling during the dynamics. This contrasts with the approach used here, where the location of the minima of the biasing potential for sequential windows are chosen based on the previous window but the functional form remains unchanged.

LA0106485

(54) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450−1462.

(55) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865−880.