

Finite system and periodicity effects in free energy simulations of membrane proteins

Turgut Baştuğ, Swarna M. Patra, Serdar Kuyucak *

School of Physics, University of Sydney, Sydney, NSW 2006, Australia

Received 23 March 2006

Available online 16 May 2006

Abstract

Periodic boundary conditions are routinely used in molecular dynamics simulations of biological systems, although they are inherently nonperiodic. Simulation artifacts that may arise from such artificial periodicity have not been well studied for membrane proteins. Here we investigate the finite size and periodicity effects in the gramicidin A peptide embedded in lipid bilayer, which offers a simple and well established test system. Comparison of free energies of translocating a potassium ion from bulk to the gramicidin A center in systems with different sizes indicate that there is no discernible effect on the free energy calculations due to finite system size and periodicity.

© 2006 Elsevier B.V. All rights reserved.

One of the limitations of molecular dynamics (MD) simulations is finite system size. Even with the latest technology, the system size rarely goes beyond $(100 \text{ \AA})^3$. The most common solution to having vacuum outside the simulation box is to employ periodic boundary conditions, where the central box is duplicated in all directions filling the space. Various truncation schemes and lattice-sum methods have been devised to deal with the long-range Coulomb interactions that ensue from a periodic system [1,2]. At present the lattice-sum methods based on various versions of Ewald sum are thought to be superior and routinely used in MD simulations of biomolecules [3]. While these methods avoid vacuum outside and allow accurate calculation of the long-range interactions, the artificial periodicity imposed on the system could induce its own artifacts. Effects of artificial periodicity and finite size have so far been investigated, for example, on transport of polymers and water [4,5], solvation and interaction of ions [6,7], conformations of globular proteins [8–10] and DNA [11,12], and lipid bilayers [13,14]. These investigations indicate that 3–4 layers of water around a solute molecule should be sufficient to avoid periodicity artifacts in simulations of neutral biomolecules.

Due to lack of molecular structure, membrane proteins have rarely been considered in MD simulations before 1998. Determination of the crystal structure of the KcsA potassium channel [15], followed by many others, have generated a great deal of interest in MD simulations of ion channels and membrane proteins (for recent reviews see [16–18]). Because the system size in simulations of membrane proteins is already quite large, it is difficult to carry out a direct investigation of the periodicity effects within the MD framework. Probably the simplest system where such an investigation is still feasible is the gramicidin A (gA) channel, formed from the dimer of right-handed $\beta^{6,3}$ helices with 15 residues [19,20]. Recently the gA channel has been the focus of several studies testing the accuracy of current MD force fields [21–24]. In all cases, the energy barrier faced by a K^+ ion at the center of the channel was calculated as ~ 20 kT, which is too large compared to the value of 5 kT obtained from the inversion of permeation data [25]. A 15 kT discrepancy in the barrier height would lead to suppression of the channel conductance by 6 orders of magnitude compared the experimental value. Thus quantifying simulation artifacts for this system to see if they help to reduce this discrepancy is also important for testing of the force fields.

The finite size correction to the free energy of translocating a K^+ ion from bulk to the center of the gA channel was

* Corresponding author. Fax: +61 2 9351 7726.

E-mail address: serdar@physics.usyd.edu.au (S. Kuyucak).

estimated from continuum electrostatics as -2.7 kT for a system containing 20 lipid and 1080 water molecules [22]. The energy barrier faced by ions is lowered by the same amount, which helps to reduce the 15 kT discrepancy noted above by 20%. Here we examine the finite size and periodicity effects in the gA channel within the MD framework by increasing the system size systematically and comparing the calculated free energy differences in each case. This is done in two stages: we first study the effect of the bilayer area and establish the minimal number of lipid molecules required to obtain accurate free energies. Then the number of water molecules in this system are doubled and quadrupled – thus increasing the distance between the neighboring bilayers in the same proportion – to investigate periodicity artifacts in the direction of the bilayer normal.

To prepare the simulation systems, we first construct a bilayer of DMPC molecules with 128 lipid molecules per layer and embed the 1JNO structure of gA [26] in the middle of this system. The lipids at the periphery of this system are gradually removed while preserving a rectangular shape, so as to obtain four systems with 64, 32, 16 and 10 lipids per layer. The smallest system is similar to that in Ref. [22], and included here in order to facilitate comparison with the estimate of finite size corrections in that work. During the initial phase, the gA atoms are constrained to their initial positions. Each system is hydrated with about 40 water molecules per lipid molecule and placed in an orthorhombic periodic box. Following an energy minimization of 10000 steps, they are equilibrated with surface-tension coupling until the surface area converged to the experimental lipid density of 60 \AA^2 per lipid [27]. At this stage, a number of water molecules are replaced with K^+ and Cl^- ions so that the solution has the physiological concentration of about 150 mM. In the remaining simulations, the periodic box is fixed in the x and y directions and a pressure coupling of 1 atm is applied in the z -direction. Each system is further equilibrated for 1 ns (the z size is obtained from its average during this equilibration period). The restraints applied to the gA atoms are then gradually relaxed in MD simulations lasting 3 ns. These systems are then employed in free energy calculations. For the second stage, we prepare two more systems by doubling and quadrupling the number of water molecules and ions in the system with 16 lipids. The simulation parameters for the six systems thus obtained are listed in Table 1. The average thickness of the bilayer is about 36 \AA . Thus in the first four systems, the thickness of the water layer (or the distance between the neighboring bilayers) is about 35 \AA , which is increased to 70 and 144 \AA in the last two systems.

MD simulations are carried out using the NAMD code, version 2.5 [28] with the PARAM27 version of the CHARMM force field [29], which provides a complete set of parameters for all the atoms in the system. An NpT ensemble is used with periodic boundary conditions. Pressure is kept at 1 atm using the Langevin piston method with a damping coefficient of 5 ps^{-1} . Similarly, temperature is maintained at 298 K through Langevin damping with a coefficient of

Table 1
System parameters used in different MD simulations

Lipid # (per layer)	Water #	Ion #	Size (x, y, z) (\AA)
10	1156	2	29, 29, 72
16	1532	4	38, 32, 71
32	3050	8	52, 44, 72
64	5210	16	64, 64, 70
16	3002	8	38, 32, 106
16	5992	16	38, 32, 180

Systems in the first four rows are used to study the effect of the bilayer area, and the second together with the last two are used to study the effect of the bilayer distance on free energy results. Note that gA covers an area of about 250 \AA^2 , roughly equal to that of four lipid molecules.

5 ps^{-1} . Electrostatic interactions are computed using the particle-mesh Ewald algorithm. The list of nonbonded interactions is truncated at 13.5 \AA , and a switching cut off distance of 10 \AA is used for the Lennard–Jones interactions. A time-step of 2 fs is employed for all simulations. Trajectory data is written at 1 ps intervals during both equilibration and production runs.

We calculate the free energy difference for translocating a K^+ ion from bulk water to the channel center using the thermodynamic integration method [30]

$$\Delta G = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda, \quad (1)$$

where $H(\lambda) = (1 - \lambda)H_0 + \lambda H_1$, with H_0 and H_1 representing the Hamiltonians of the initial and final states, respectively (e.g. if the initial state is an ion in the channel and a water molecule in bulk, in the final state these two are interchanged). The integral in Eq. (1) is performed using a Gaussian quadrature. We have experimented with various numbers of quadrature points (e.g., 3, 5, 7 and 12 points), and found that 7 point-quadrature provides sufficient accuracy for our purposes. This value is used in all subsequent calculations. Each system is equilibrated for at least 200 ps (longer in some cases) before production runs. In all cases, the integrals are evaluated from 700 ps of production runs.

When a K^+ ion is in the channel center, the dipoles of water molecules in the pore point away from the ion, whereas when there are only water molecules in the channel they all point in the same direction. Thus an alchemical transformation of a K^+ ion in the channel center to a water molecule disrupts the orientation of half of the water molecules. The resulting fluctuations in the free energy calculations can be reduced by performing the transformation via an intermediate state with no charge, which we choose as a water molecule with the partial charges set to zero (denoted as W_0). Thus we perform two calculations, $\Delta G(\text{K}^+ \rightarrow W_0)$ and $\Delta G(W_0 \rightarrow W)$, whose sum gives the desired free energy change for the $\text{K}^+ \rightarrow W$ transformation. Note that a similar transformation, $W \rightarrow W_0 \rightarrow \text{K}^+$ is carried out in bulk simultaneously to find the ionic free energy difference between the channel and bulk water. We find that the free energy difference for the $W_0 \rightarrow W$ leg is negligibly small (smaller than the ~ 1 kT statistical error) regardless of the system size. Therefore in the following, we present only

the results for the free energy difference of the first leg, i.e. $\Delta G(K^+ \rightarrow W_0)$, although we will keep referring to it as ΔG for the whole process.

The free energy difference results are summarized in Table 2. We first consider the dependence of the free energies on the bilayer area. In the system with 10 lipids, we observe a substantial difference between the forward and reverse free energies, which is much larger than the 1 kT statistical fluctuations. A large difference between the forward and reverse results often points to sampling problems and the simulation results are not reliable. The situation improves with increasing number of lipids. For example, the system with 16 lipids is a borderline case because the average ΔG encompasses both the forward and reverse values but they are just within the statistical fluctuations. In the systems with 32 and 64 lipids, the forward and reverse ΔG values are very close to each other, well within the statistical fluctuations.

To make the differences among the four systems clearer and also demonstrate the convergence of the free energies, we show the running averages of the free energy differences for the forward (ΔG_+) and backward ($-\Delta G_-$) directions in Fig. 1. We see that unlike the other three cases, the relatively large hysteresis effect for the system with 10 lipids persists throughout the 700 ps simulation. This suggests that using only 10 lipids per layer may not be sufficient to provide an adequate membrane environment for the peptide when it moves freely. The hysteresis effect in all other cases remains within the statistical fluctuations after the initial 300 ps. Thus a minimal number of 16 lipids appears to be necessary for obtaining reliable free energies from MD simulations of the gA system. Of course, if simulation time is not a problem, use of 32 lipids will provide an even more accurate simulation environment. As the results in Table 2 indicate, there is not much point in going beyond 32 lipids. Our results are consistent with a similar study of lipid bilayers, where periodicity artifacts are investigated by increasing the number of lipid molecules up to 64 per layer [14]. It is found that 36 lipids per layer is suf-

Table 2

Free energy differences for translocating a K^+ ion from bulk to the channel center (ΔG_+), negative of the reverse transfer ($-\Delta G_-$), and their average (ΔG_{av})

	ΔG_+	$-\Delta G_-$	ΔG_{av}
<i>Lipid #</i>			
10	11.1 ± 1.0	15.9 ± 0.9	13.5 ± 1.0
16	11.4 ± 0.9	13.3 ± 0.9	12.4 ± 0.9
32	13.5 ± 0.8	13.5 ± 1.0	13.5 ± 0.9
64	13.5 ± 0.9	13.9 ± 1.0	13.7 ± 1.0
<i>Water #</i>			
1532	11.4 ± 0.9	13.3 ± 0.9	12.4 ± 0.9
3002	12.8 ± 0.9	11.7 ± 0.8	12.3 ± 0.9
5992	13.1 ± 0.8	11.7 ± 1.0	12.4 ± 0.9

The top part shows the free energy differences obtained from systems where the number of lipid molecules is increased while keeping the bilayer distance fixed. The bottom part shows the same for a fixed bilayer area (16 lipids per layer) but with an increasing bilayer distance. All free energies are given in units of kT.

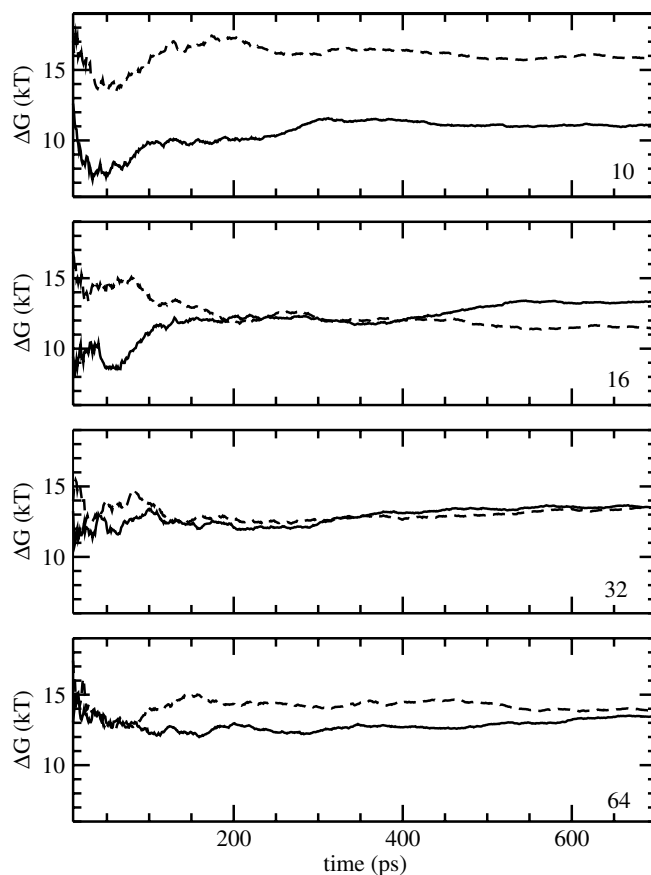


Fig. 1. Running averages of the free energy differences ΔG_+ (solid lines) and $-\Delta G_-$ (dashed lines) for transferring a K^+ ion from bulk to the channel center and the reverse process, respectively. The number of lipids in the simulation system is increased progressively as indicated in each panel. Here we show only the $K^+ \rightarrow W_0$ legs of the transformation. The $W_0 \rightarrow W$ legs are negligibly small in all cases, and therefore are not shown. The values of ΔG_{av} in Table 2 are obtained from the average of the two endpoints in the curves.

ficient to provide an accurate description of the lipid properties and a reasonable description can already be obtained with 16 lipids but using 9 lipids is not enough.

With regard to the effect of the bilayer area on the free energy differences, we see no significant corrections (i.e., greater than the statistical fluctuations) to the stabilization energy of the ion at the center of gA. If we leave the 10 lipid case aside because it is unreliable, the stabilization energy increases by about 1 kT with the number of lipid molecules. For comparison, this correction was estimated in a similar system from continuum electrostatics as -2.7 kT [22], which is in the opposite direction and larger than the estimate obtained here directly from MD simulations.

We next investigate effects of periodicity in the direction of the bilayer normal. Because we have limited computer time, we use the minimal 16 lipids per layer system for this purpose. The free energy differences for the three systems, where the bilayer distance between the neighboring simulation boxes is increased from 35 to 70 and 144 Å, are presented at the bottom of Table 2, and their running averages are shown in Fig. 2. The average ΔG values in

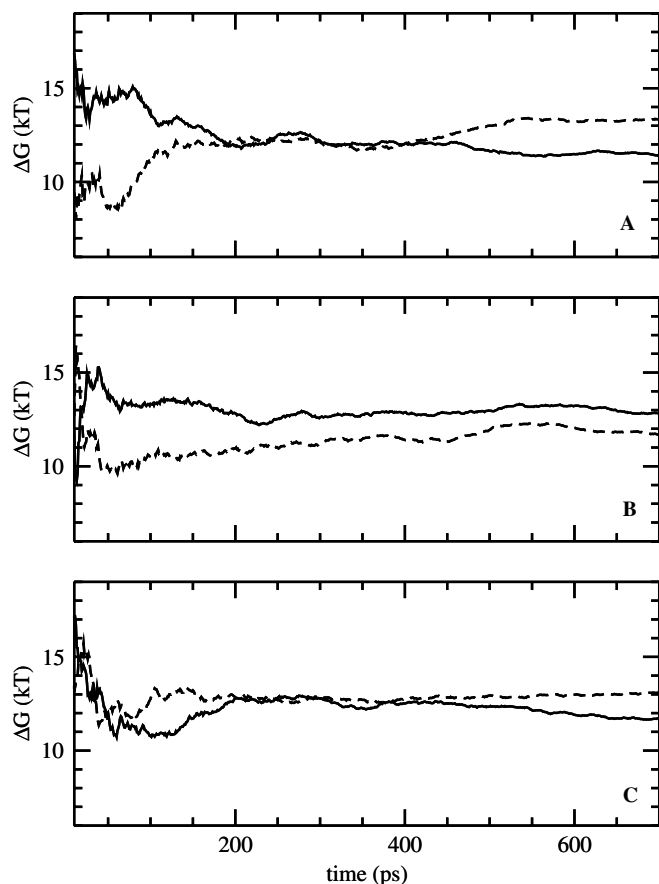


Fig. 2. Similar to Fig. 1 but for systems with increasing number of water molecules and fixed number of lipids (16 per layer). The thickness of the water layer in the simulation box (and hence the bilayer distance) is 35 Å in (A), 70 Å in (B), and 144 Å in (C).

Table 2 exhibit no discernible variation with increasing bilayer distance. The running averages in Fig. 2 are also very similar – the only effect of increasing the number of water molecules in the system is to reduce the hysteresis effect slightly. Thus our results indicate no periodicity artifacts due to replication of bilayers as long as they are well separated by several layers of water. In fact, the 5–6 layers of water employed between the lipid bilayer and the simulation box in our smallest system appears to be more than enough. In most MD simulations, 3–4 layers of water around a solute are deemed sufficient.

In conclusion, we have presented a systematic study of finite size and periodicity effects in MD simulations of the gramicidin A channel. Our results indicate that there are no significant periodicity artifacts in the calculated ionic free energies as long as one employs a minimal number 16 lipids per layer and 5–6 layers of water between the lipid and the simulation box. These conditions are satisfied in most MD simulations of the gA channel, hence their results do not suffer from any significant periodicity artifacts. The present results also have important ramifications for the tests of the MD force fields. Because periodicity artifacts do not help at all in reducing the large energy barrier found in MD simulations using the current nonpolarizable force

fields, the solution must come from elsewhere. The most obvious candidate is the polarization interaction that is neglected in the current force fields. In this regard, semimicroscopic models of the gA channel which include the polarization interaction provide valuable clues as they do not suffer from such a high energy barrier problem [31,32].

Acknowledgements

This work was supported by grants from the Australian Research Council. Calculations were carried out using the Barossa cluster at the Australian Center for Advanced Computing and Communications, and the SGI Altix cluster at the Supercomputer Facility of the Australian National University.

References

- [1] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, London, 1987.
- [2] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, second edn., Academic Press, San Diego, 2002.
- [3] C. Sagui, T.A. Darden, *Ann. Rev. Biophys. Biomol. Struct.* 28 (1999) 155.
- [4] B. Dunweg, K. Kremer, *J. Chem. Phys.* 99 (1993) 6983.
- [5] I.C. Yeh, G. Hummer, *J. Phys. Chem. B* 108 (2004) 15873.
- [6] P.H. Hunenberger, J.A. McCammon, *Biophys. Chem.* 78 (1999) 69.
- [7] X. Rozanska, C. Chipot, *J. Chem. Phys.* 112 (2000) 9691.
- [8] P.H. Hunenberger, J.A. McCammon, *J. Chem. Phys.* 110 (1999) 1856.
- [9] W. Weber, P.H. Hunenberger, J.A. McCammon, *J. Phys. Chem. B* 104 (2000) 3668.
- [10] M.A. Villarreal, G.G. Montich, *J. Biomol. Struct. Dyn.* 23 (2005) 135.
- [11] O.N. de Souza, R.L. Ornstein, *Biophys. J.* 72 (1997) 2395.
- [12] J. Norberg, L. Nilsson, *Biophys. J.* 79 (2000) 1537.
- [13] J. Wohler, O. Edholm, *Biophys. J.* 87 (2004) 2433.
- [14] A.H. de Vries, I. Chandrasekhar, W.F. van Gunsteren, P.H. Hunenberger, *J. Phys. Chem. B* 109 (2005) 11643.
- [15] D.A. Doyle, J.M. Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon, *Science* 280 (1998) 69.
- [16] S. Kuyucak, O.S. Andersen, S.H. Chung, *Rep. Prog. Phys.* 64 (2001) 1427.
- [17] B. Roux, T. Allen, S. Bernèche, W. Im, *Q. Rev. Biophys.* 37 (2004) 15.
- [18] W.L. Ash, M.R. Zlomislic, E.O. Oloo, D.P. Tieleman, *Biochim. Biophys. Acta* 1666 (2004) 158.
- [19] D.D. Busath, *Ann. Rev. Physiol.* 55 (1993) 473.
- [20] B.A. Wallace, *J. Struct. Biol.* 121 (1998) 123.
- [21] T.W. Allen, T. Bastug, S. Kuyucak, S.H. Chung, *Biophys. J.* 84 (2003) 2159.
- [22] T.W. Allen, O.S. Andersen, B. Roux, *Proc. Natl. Acad. Sci. USA* 101 (2004) 117.
- [23] T. Bastug, A. Gray-Weale, S.M. Patra, S. Kuyucak, *Biophys. J.* 90 (2006) 2285.
- [24] T. Bastug, S. Kuyucak, *Biophys. J.* 90 (2006) 3941.
- [25] S. Edwards, B. Corry, S. Kuyucak, S.H. Chung, *Biophys. J.* 83 (2002) 1348.
- [26] L.E. Townsley, W.A. Tucker, S. Sham, J.F. Hinton, *Biochemistry* 40 (2001) 11676.
- [27] J.F. Nagle, S. Tristram-Nagle, *Biochim. Biophys. Acta* 1469 (2000) 159.
- [28] L. Kale et al., *J. Comp. Phys.* 151 (1999) 283.
- [29] A.D. MacKerell Jr. et al., *J. Phys. Chem. B.* 102 (1998) 3586.
- [30] D.L. Beveridge, F.M. DiCapua, *Ann. Rev. Biophys. Biophys. Chem.* 18 (1989) 431.
- [31] J. Åqvist, A. Warshel, *Biophys. J.* 56 (1989) 171.
- [32] V.L. Dorman, P.C. Jordan, *Biophys. J.* 86 (2004) 3529.