Challenges in analysing and visualizing large-scale molecular dynamics simulations: domain and defect formation in lung surfactant monolayers

E Mendez-Villuendas, S Baoukina, D P Tieleman
Department of Biological Sciences and Institute of Biocomplexity and Informatics, University of Calgary
E-mail: tieleman@ucalgary.ca

Abstract. Molecular dynamics simulations have rapidly grown in size and complexity, as computers have become more powerful and molecular dynamics software more efficient. Using coarse-grained models like MARTINI system sizes of the order of 50 nm x 50 nm x 50 nm can be simulated on commodity clusters on microsecond time scales. For simulations of biological membranes and monolayers mimicking lung surfactant this enables large-scale transformation and complex mixtures of lipids and proteins. Here we use a simulation of a monolayer with three phospholipid components, cholesterol, lung surfactant proteins, water, and ions on a ten microsecond time scale to illustrate some current challenges in analysis. In the simulation, phase separation occurs followed by formation of a bilayer fold in which lipids and lung surfactant protein form a highly curved structure in the aqueous phase. We use Voronoi analysis to obtain detailed physical properties of the different components and phases, and calculate local mean and Gaussian curvatures of the bilayer fold.

1. Introduction
Computer simulations of models of biological membranes and monolayers have a long history [1, 2]. They have evolved from simulations of a few dozen simplified lipids in a clearly defined lamellar geometry to very complex systems including a variety of lipids and proteins in geometries that can be quite complex [2, 3]. This increase in size and complexity is driven by increases in computer power and more efficient simulation software, but received a major boost with the advent of coarse-grained models that are up to several orders of magnitude faster than atomistic simulations. The popular MARTINI model used in this paper allows simulations of a scope 2-3 orders of magnitude larger than atomistic simulations for the same computational cost, at the obvious expense of a loss of detail [4, 5]. In practice, simulations with a size of the order of 50 x 50 x 50 nm can now be run relatively easily on a few hundred CPU cores for simulation times of the order of 10 microseconds. At these time and length scale, bilayer and monolayers can show major changes in organization. Combined with increasingly complex compositions this poses challenges for analysis, as properties of interest vary throughout the inhomogeneous system and cannot be calculated by simple averaging. Instead, methods to identify regions of interest first, followed by identification of the molecules in that region, are required to do further analyses.

In addition, there is a growing disparity between the size of a typical system and the smallest interactions of interest. In a typical bilayer simulation of 15 years ago, the system contained 128 lipids plus water, in a box that was ca. 6 x 6 nm laterally, while an individual lipid has a surface area in the
membrane of ca. 0.6 nm². To zoom from a view of the whole system to a view of individual lipids requires about a 10-50-fold ‘magnification’, or, in an alternative point of view, zooming from a view that shows 15,000 atoms to a view that shows 100. Current systems can be of the order of 100 x 100 x 100 nm, with millions of atoms, with a much larger difference in scale between the whole system and molecular details, while both are still of interest.

In this report we examine a simulation of a scope and complexity we think will become typical in the next few years, even if it is currently at the edge of what is feasible. In this simulation, as in several previously published papers, we are inspired by questions related to the function of lung surfactant. Lung surfactant is a monomolecular film that covers the interface of the alveoli and air. The alveoli are small structures in the lung where the actual gas exchange takes place. During the breathing cycle the area of this interface changes substantially. Upon inhalation the alveoli expand, increasing the area of the interface, while during exhalation the opposite occurs. At most interfaces, for instance the air/water interface, changing the area requires a significant amount of work, expressed quantitatively by the surface tension at the interface. The main role of lung surfactant is to reduce the surface tension to very low values, so that breathing does not require significant work [6]. Lung surfactant has several components but consists primarily of the zwitter-ionic lipid dipalmitoyl-phosphatidyl choline (DPPC), and contains anionic lipids, unsaturated lipids and cholesterol. In addition, there are a number of proteins associated with lung surfactant, two of which are thought to play a critical role in its primary function: surfactant-associated proteins B and C (SP-B, SP-C) [7].

In previous simulation studies of lung surfactant, we focused on a number of model systems. We studied monolayers formed by DPPC [8], a more realistic lipid mixture of DPPC, palmitoyl-oleoyl-phosphatidylglycerol (POPG), and cholesterol and surfactant proteins to test two different models of lung surfactant reservoirs [9], the general mechanism of single-phase monolayer breakdown [10]; stability conditions for monolayers [11], and the ability of SP-B to facilitate fusion and lipid flow between a monolayer and bilayers/vesicles [12, 13]. However, these simulations, by design, ignored the possibility that lung surfactant monolayers may contain coexisting liquid-disordered and liquid-ordered phases. We also did not analyze in any detail the distribution of different lipids near irregular features such as the connection points of vesicles to a monolayer or fusion intermediates. Here we describe methods to analyze a mixture mimicking lung surfactant under conditions of surface tension and temperature where there is clear evidence of domain formation, protein sorting, and eventually monolayer breakdown or defect formation.

2. Methods
The simulation here uses the same methods previously described [13]. We refer to that paper for details of the MARTINI force field and simulation parameters. The system setup consisted of a water slab in vacuum with two symmetric monolayers at the two water/vacuum interfaces. The monolayers consist of dipalmitoyl-phosphatidylcholine (DPPC), dioleoyl-phosphatidylcholine (DOPC), palmitoyloleoyl-phosphatidylglycerol (POPG), and cholesterol in ratio 2:1:1:1, respectively. Two hydrophobic surfactant proteins SP-B and SP-C were added to the monolayers. Each monolayer contains 2304 lipids, and 6 SP-B and 8 SP-C proteins. The water slab contains 124,000 water particles, of which 398 were substituted for Ca²⁺ ions to neutralize the unbalanced charge resulting from anionic POPG lipids and net positively charged proteins.

The lipids were randomly mixed in the starting configuration; SP-B proteins were placed in water and SP-C proteins in vacuum in the vicinity of each monolayer. Initially the monolayers were equilibrated for 1 μs at a surface tension of 30 mN/m and a temperature of 290 K, at which the monolayers remained roughly homogeneous and stable at the interface. After equilibration, the surface tension was reduced to 10 mN/m, which led to lipid segregation and monolayer collapse from the interface via formation of a bilayer fold induced by SP-B proteins.
To analyze the frames from the trajectory, we developed a Matlab-based (v.R2011b) analysis program to perform quantitative analysis of complex mixtures based on Voronoi tessellation [Mendez-Villuendas et al, in preparation]. In the figures this is based on 2D tessellation using specific lipid particles as centers of Voronoi polygons: the R1 site in cholesterol (near the OH headgroup), and the C1A and C1B sites in the phospholipids, the first beads of the hydrocarbon chain (two Voronoi cells per lipid).

During the simulation, two distinct phases appear. We assigned each Voronoi cell to a specific phase using a cluster analysis based on a connectivity matrix with a cutoff parameter. The cutoff corresponded roughly to the square root of the average area of the Voronoi polygons in the more ordered phase. In addition, lipids belonging to a more ordered phase were required to have an average chain order parameter, Sz, larger than a selected value. Sz is a measure of the orientation of a lipid chain – a high value corresponds to a gel or liquid-ordered domain, while a lower value corresponds to a liquid expanded or liquid-disordered domain.

The monolayer curvature was characterized using a Matlab-based (v.R2011b) program (Mendez-Villuendas et al, in preparation). The phosphate groups of lipids (PO4 sites) were fitted to a surface using a binomial filter in 2D with a 6 nm characteristic length to remove the noise. This convoluted surface was then converted to an equally-spaced grid (0.6 nm spacing), for which partial derivatives were calculated to find the principal curvatures, c1 and c2, the mean curvature H = ½ (c1+c2) and the Gaussian curvature K = c1c2

3. Results and Discussion

Figure 1 shows a number of snapshots of the simulation. The snapshots 1A, B, C show a view from the water phase normal to the monolayer. At t=0, the system is in a metastable state, induced from a stable state by lowering the surface tension below the equilibrium. Figure 1B shows that after ca. 7 microseconds, there is clear evidence of phase separation and domain formation. The surfactant
proteins preferentially interact with DOPC/POPG components forming a liquid-disordered (Ld) phase, while DPPC and cholesterol separate out into a liquid-ordered (Lo) domain with occasional inclusions of POPG, preferentially at the boundary. Surfactant proteins induce local perturbation of the monolayer.

The monolayer at the interface is still metastable, and a significant defect forms after 8 microseconds (Figure 1C), which can be clearly seen in the side views (along two different axes) of this monolayer in Figures 1D and 1E. The defect represents a bilayer fold of a disc-like shape, connected to a monolayer. The bilayer fold is induced by an SP-B dimer via local bending of the monolayer. We previously observed formation of a bilayer fold from a monolayer below the equilibrium surface tension induced by SP-Bs in the absence of phase separation [9]. Here it is clear even without numerical analysis that the fold nucleates in the Ld phase and becomes highly enriched in POPG/DOPC/protein, with little cholesterol or DPPC.

Figure 2 illustrates how the Voronoi analysis is helpful in analyzing this complex mixture. We considered (Figure 2A) a flat monolayer with Ld and Lo domains (at 7 μs) to characterize in-plane distribution of components. Figure 2B shows 2D Voronoi tessellation based on selected lipid sites only (see Methods). Once lipids have been assigned to a Voronoi cell, additional analysis becomes possible on subsets of cells that share common features. We can use the Voronoi polygons to calculate areas of individual lipids and to group lipids to domains of distinct phases. Figure 2C superimposes the protein structures on the Voronoi cells. It becomes apparent that lipids near the proteins have larger areas. We can readily identify lipids that are near proteins, determined by a distance criterion or other means, and analyze these as group.

Figure 3 shows variation of lipid properties as function of in-plane radial distance from the proteins based on Figure 2C. The areas per lipid (Figure 3A) increase noticeably in the protein vicinity due to partitioning of apolar protein residues in the monolayer. Interestingly, this increase is followed by a decrease in the area per DOPC lipid compared to its average area in the monolayer.
Figure 3. (A) Area per individual lipid and (B) lipid composition as function of the in-plane radial distance from surfactant proteins SP-B and SP-C. Lipid distribution around proteins was calculated by averaging over concentric shells of 1 nm increment. The points at 30 nm converge to the average values.

This is possibly due to an increase of its local concentration in the protein vicinity, which can be seen in the composition profile (Figure 3B). The distribution of lipid components also shows a moderate increase in POPG concentration near proteins, and a local depletion of saturated DPPC lipids. These differences in lipid concentration near proteins promote lipid segregation between the ordered and disordered phases. At large distances, the composition of the monolayer is reaches its average value, as expected.

A large number of other analyses are possible. We can trivially analyze averages over all DOPC lipids in the system, because we know the nature of all lipids in the simulation. However, DOPC lipids behave differently in different phases, and for more meaningful analyses we need to distinguish between DOPC lipids in Lo and Ld phases. From a detailed examination of the images it becomes obvious that lipid components do not fully segregate between the phases; the Ld phase, while highly enriched in DOPC, also contains POPG and a small fraction of DPPC and cholesterol. We can group lipids into Lo/Ld domains by building the connectivity matrix based on selected physical properties. These properties include lipid packing (which can be expressed by a distance cutoff or a number of neighbors), or chain order parameter. By analyzing the edges of the Voronoi polygons, we can also characterize the boundary between the phases/domains (Figure 2D).

Once domains have been identified, we can track the number of individual domains over time and characterize the kinetics of domain growth. We can analyze the properties that are essential for understanding the energetics of domains, such as the area of each phase and the length of the phase boundary, and dynamics, such as diffusion of lipids within each domain. We can also obtain other properties of each phase, including lipid composition, partial areas of each lipid component and order parameters. We can follow the response of all these properties to changes in e.g. temperature, composition, and surface tension.

We are extending the software to do a 3D tessellation, but for mostly flat geometries 2D is adequate. There is an interesting opportunity to combine the Voronoi analysis with other grid-based analyses. One key example is given by the Gaussian and mean curvatures that can be calculated from the principal curvatures. To calculate these curvatures we need to overlay the molecular system, which has a substantial roughness/noise at any given time, with a filter to smooth the surface, and then discretize it. From this smoothed surface on a mesh grid we can numerically calculate the partial derivatives that define the local curvature.
Figure 4A again shows a snapshot of the monolayer system described in more detail in Figure 1. Figure 4B and C show the calculated mean and Gaussian curvature for the bilayer fold connected to the monolayer (at 8.65 μs), corresponding to Figure 1C-E, using a filter with a 6 nm resolution. The resolution of the filter is limited by the validity of the concept of curvature as a physically meaningful quantity for membranes on the low end and by the desired resolution of the analysis at the high end. Specifically, if the filter is too small curvature can be calculated for individual lipid molecules but this has no useful physical meaning, while if the filter is very large important features on smaller scales will be ignored. Thus there is significant room for further development as well as combining this curvature grid with the Voronoi grid.

4. Outlook
We have illustrated some current challenges in analyzing state of the art simulations of biological systems, representing complex lipid-protein 3D assemblies. The Voronoi analyses allow us to numerically distinguish different environments for the same molecules, so that properties can be calculated for a specific type of environment only, or the interfacial length between two phases, quantities that could be used to study the factors explaining the energy gain in one molecular mechanism over another. For example, proteins may play a role in reducing the free energy of formation of a new nucleation site via an energetic preference for a particular lipid, leading to the formation of a new phase. Further, models describing domain energetics may be linked to mesoscopic theories that have the overarching goal of explaining phenomena at the biological scale. Curvature plays a key role in this class of theories; the ability to link molecular simulation results through curvature to mesoscopic Helfrich-type models will open a new realm of applicability of molecular simulation and strengthen both large-scale lipid simulations and the use of continuum models.

Acknowledgements
This work was supported by the Canadian Institutes of Health Research (EMV) and the Natural Sciences and Engineering Research Council of Canada (SB). DPT is an Alberta Innovates Health Solutions Scientist and Alberta Innovates Technology Futures Strategic Chair in (Bio)Molecular Simulation. Simulations were carried out on WestGrid/Compute Canada facilities. This research was supported in part by the Project of Knowledge Innovation Program (PKIP) of Chinese Academy of Sciences, Grant No. KJCX2.YW.W10
Supplementary information: An animation of the simulation described here is available with the online version (m4222sp-10.290.avi). This movie shows a ~ 9 microsecond trajectory at a surface tension of 10 mN/m and a temperature of 290 K starting from the monolayer equilibrated at a 30 mN/m as described in the Methods. The color scheme as in Figure 1.

References
[1] Tieleman D P, Marrink S J and Berendsen H J C 1997 A computer perspective of membranes: Molecular dynamics studies of lipid bilayer systems Biochim. Biophys. Acta 1331 pp 235-270
[2] Marrink S J, de Vries A H, and Tieleman D P 2009 Lipids on the move: Simulations of membrane pores, domains, stalks and curves. Biochim. Biophys. Acta Biomembranes 1788 pp 149-168
[3] Marrink S J and Tieleman D P 2002 Molecular dynamics simulation of spontaneous membrane fusion during a cubic-hexagonal phase transition Biophys. J. 83 pp 2386-2392
[4] Marrink S J et al 2007 The MARTINI force field: coarse grained model for biomolecular simulations J. Phys. Chem. B 111 pp 7812-7824
[5] Monticelli L et al 2008 The MARTINI coarse-grained force field: extension to proteins J. Chem. Theo. Comp. 4 pp 819-834
[6] Rugonyi S, Biswas S C, and Hall S B 2008 The biophysical function of pulmonary surfactant Respiratory Physiology & Neurobiology 163 pp 244-255
[7] Zuo Y Y et al 2008 Current perspectives in pulmonary surfactant - Inhibition, enhancement and evaluation Biochim. Biophys. Acta Biomembranes 1778 pp 1947-1977
[8] Baoukina S et al 2007 Pressure-area isotherm of a lipid monolayer from molecular dynamics simulations Langmuir 23 pp 12617-12623
[9] Baoukina S et al 2007 The molecular mechanism of monolayer-bilayer transformations of lung surfactant from molecular dynamics simulations Biophys. J. 93 pp 3775-3782
[10] Baoukina S et al 2008 The molecular mechanism of lipid monolayer collapse Proc. Natl. Acad. Sci. USA 105 pp 10803-10808
[11] Baoukina S, Marrink S J, and Tieleman D P 2010 Lateral pressure profiles in lipid monolayers Faraday Discussions 144 pp 393-409
[12] Baoukina S and Tieleman D P 2010 Direct Simulation of Protein-Mediated Vesicle Fusion: Lung Surfactant Protein B Biophys. J. 99 pp 2134-2142
[13] Baoukina S and Tieleman D P 2011 Lung surfactant protein SP-B promotes formation of bilayer reservoirs from monolayer and lipid transfer between the interface and subphase Biophys. J. 100 pp 1678-1687