UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

School of Chemistry

The Development of a Coarse-Grain Biomembrane Model and its Use in Multiscale Simulations of Solute Permeability

by

Mario Orsi

Thesis for the degree of Doctor of Philosophy

July 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF CHEMISTRY

Doctor of Philosophy

THE DEVELOPMENT OF A COARSE-GRAIN BIOMEMBRANE MODEL AND ITS USE IN MULTISCALE SIMULATIONS OF SOLUTE PERMEABILITY

by Mario Orsi

A new simplified particle-based computer model for hydrated phospholipid bilayers is presented. In the model, each lipid molecule, in reality comprising more than one hundred atoms, is reduced to a collection of ten "coarse-grain" macrounits. Compared with available coarse-grain methods, three novel aspects are introduced. First, electrostatics are explicitly incorporated via charges and dipoles. Second, water is accurately (yet efficiently) described, on an individual level, by the soft sticky dipole model. Third, hydrocarbon tails are modelled using the anisotropic Gay-Berne potential. Simulations are conducted by rigid body molecular dynamics, using software specifically designed and implemented for this project. The technique developed proves two orders of magnitude less demanding of computational resources than traditional atomic-level methodology. The model is parameterised to reproduce the experimental area and volume per lipid, order parameters, and the self-assembly process. Self-assembled bilayers quantitatively reproduce experimental observables such as electron density, compressibility moduli, dipole potential, lipid diffusion and water permeability. The lateral pressure profile is calculated, along with the elastic curvature constants of the Helfrich expression for the membrane bending energy: results are consistent with experimental estimates and atomic-level simulation data. Several of the results presented are obtained for the first time using a coarse-grain method. The model is also directly compatible with atomic-level force-fields, allowing mixed systems to be simulated in a multiscale fashion. Efficient multiscale simulations are conducted to predict the permeability coefficient of a number of atomic-level solutes, including small organic molecules, large drugs and steroid hormones. Results prove broadly consistent with previous atomic-level calculations and available experimental data. In particular, despite discrepancies in the absolute magnitudes, the solute relative permeability coefficients, and hence the permeability ranking orders, are consistently reproduced.

Contents

List of Figures

List of Tables

Acknowledgements

I am very grateful to my supervisor Jon Essex for wise guidance, constant encouragement and contagious enthusiasm throughout the project.

I also thank my industrial collaborator Wendy Sanderson for her optimism and kindness; Johnson & Johnson - Janssen Pharmaceutica is acknowledged for funding this work.

Past and present members of the Essex group have provided me with a supportive and stimulating working environment, along with some remarkably pleasant time during various social events.

I have really enjoyed my time in Southampton, and this is mainly due to all the great people I've met, too many to properly acknowledge here; very special thanks are for Marty, Beppe, and particularly Monica.

An acknowledgement is also due to the Politecnico di Milano university, where I did my undergraduate studies in bioengineering; what I learnt there proved instrumental in dealing with the most difficult parts of this project.

Finally, I deeply thank my father Filippo, my mother Luisa, my sisters Alessandra and Margherita, and my "little" brother Francesco; I will always be grateful for their unquestioned support to every choice I've made.

It is often important to formulate simple ideas very clearly so that one can use them as models in thinking about more complex ideas.

Gödel, Escher, Bach: an Eternal Golden Braid (1980)
Douglas R. Hofstadter

Each of us builds, inside our head, a model of the world in which we find ourselves. The minimal model of the world is the model our ancestors needed in order to survive in it. The simulation software was constructed and debugged by natural selection, and it is most adept in the world familiar to our ancestors on the African savannah: a three-dimensional world of medium-sized material objects, moving at medium speeds relative to one another. As an unexpected bonus, our brains turn out to be powerful enough to accommodate a much richer world model than the mediocre utilitarian one that our ancestors needed in order to survive. Art and science are runaway manifestations of this bonus.

The God Delusion (2006)
Richard Dawkins

You can have as much junk in the guess as you like, provided that the consequences can be compared with experiment.

The Character of Physical Law (1965)
Richard P. Feynman

So she went on, wondering more and more at every step...

Through the Looking-Glass and What Alice Found There (1872)
Lewis Carroll

Chapter 1

Introduction

Biomembranes are dynamic structures in which proteins float in a sea of lipids[1] (Figure 1.1); they are complex and fascinating systems, characterised by highly heterogeneous structure and motion, and displaying an astonishingly rich and biologically relevant behaviour on a vast range of spatial and temporal scales.[2–4] Despite their liquid-like nature, biomembranes retain a remarkable degree of structural integrity, and for this reason they are universally employed in nature to define the boundary surface of cells and to act as barriers to prevent diffusion of substances in and out of the cell. Yet membranes contain protein transport systems, functioning as selective pumps and channels, that allow specific molecules to be taken up and unwanted compounds to be removed. Membranes are also central to several other functions indispensable for life, such as growth, energy storage, and information transduction via neural activity.

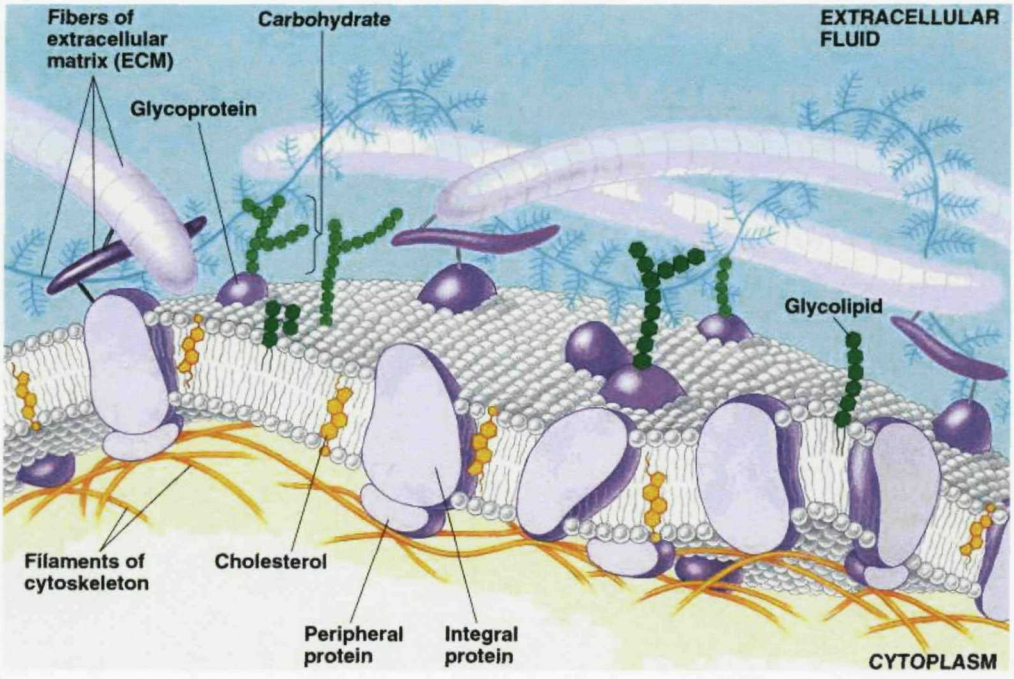


Figure 1.1: The plasma membrane. Schematic representation of a plasma membrane patch at molecular resolution; characteristic elements are highlighted.[5]

In addition to an external cell membrane, called the plasma membrane, eukaryotic cells also comprise internal membranes that enclose organelles such as mitochondria. The core of any biological membrane is the lipid bilayer. Experimental investigation of the physics of lipid bilayers allows the determination of a large body of bilayer features. The internal structure can be described via electron density profiles,[6] the ordering of the hydrocarbon tails is quantified by order parameters,[7] fluidity is studied by lateral diffusion measurements,[8] mechanical properties can be related to the measurements of elastic moduli,[9] electrostatic properties are quantified by estimating internal potentials,[10] and even the transbilayer pressure distribution can be qualitatively measured.[11] These fundamental membrane properties are central to an incredibly large number of biological mechanisms. The bilayer structure directly influences the conformation of embedded proteins, whereas lipid fluidity is crucial, for example, for membrane lateral organisation. The membrane dipole potential, and associated electric field, also play fundamental roles: for example, they are involved in the regulation of membrane proteins,[12–14] membrane fusion,[15] insertion and folding of amphiphilic peptides,[16] redox reactions,[17] permeability,[18] interaction with drugs[19,20] and signalling.[21] An even more important feature is the lateral pressure profile, which characterises the transmembrane distribution of forces. The lateral pressure profile is the most fundamental physical property of lipid bilayers:[4] it determines the interfacial area, it is at the basis of phase transitions and fusion,[22] it modulates the insertion, folding and functioning of membrane proteins,[23] and it affects permeability,[24] drug transport,[25] and anaesthesia.[26] A more detailed account of the most relevant biophysical features of phospholipid bilayers is presented in Chapter 2, along with brief descriptions of the corresponding experimental investigation techniques.

Considering the amount and variety of phenomena associated with lipid bilayers, it is clear that the development of realistic models is a delicate and challenging task. Particle-based computer simulation, typically molecular dynamics, represents a powerful tool to model biomembranes, as it can provide detailed dynamic and thermodynamic data for a broad range of systems. An overview of the molecular dynamics simulation method is presented in Chapter 3.

Atomic-level (AL) membrane models, in which every atom is explicitly simulated with a corresponding site, have been employed for decades now, and have significantly contributed to the understanding of many membrane phenomena.[27–45] However, AL methods require an enormous amount of computational resources to calculate the interactions between all atoms in the system. The large computational cost of AL models results in a number of limitations and issues. For example, collective, large-scale phenomena such as self-assembly, membrane fusion or micro-domain formation (lipid rafts) are typically intractable. In fact, apart from one exception,[28] all published AL simulations to date have been carried out on pre-assembled bilayers, due to the prohibitive computational cost of simulating self-assembly: this does not guarantee that the system is at thermodynamic equilibrium. Furthermore, the reliability of the calculation of some important properties, such as the lipid area, diffusion coefficients, or the lateral pressure profile, can be undermined by insufficient sampling. These issues can be

tackled by adopting a simplified, *coarse-grain* description.

The coarse-grain (CG) methodology generally involves grouping together selected clusters of atoms into single macro-sites, to significantly reduce the number of interactions calculated, and hence also the computational cost. The obvious downside inherent in all CG models is the loss of atomic detail. Over the past few years, the CG field has grown significantly: a large number of models have been developed, for different membrane systems, and with differing degrees of simplification. The literature background concerning coarse-grain methods has been reviewed, and a detailed account is given in Chapter 4. Here we introduce the typical features of those CG models of phospholipid membranes that retain an explicit connection with the chemical identity of the systems described, as this is the level at which our new methodology is located. Such models can be called *specific*. In the specific CG models,[46–51] a lipid molecule, which at the atomic level comprises about one hundred particles, is typically reduced to around ten CG sites. To further increase simulation efficiency, the representations of water and electrostatics are highly simplified. A popular strategy involves representing three[48–50] or four[46] water molecules as single CG units. However, in one model[51] water molecules are described individually, whereas in another[47] solvation is described implicitly. None of these models explicitly includes water electrostatics. The charges in the lipid glycerol-ester region are also not explicitly represented. As for the lipid headgroup region, electrostatics are present in most models, where, however, artificial dielectric constants are introduced to account for water screening.[46–49] Such simple models can be orders of magnitude more efficient than corresponding AL systems, and hence they can be applied to simulate phenomena at larger temporal and spatial scales, such as self-assembly,[46–48,52] phase transition[53,54] and vesicle fusion.[55] CG simulations have also been conducted to study the interaction with inclusions such as alcohols,[56] anaesthetics,[57] cholesterol[58] and even proteins.[59] These studies are interesting because they have given access to membrane processes that are difficult to investigate by any other theoretical or experimental technique. However, we have identified a number of issues in these CG models and their application. The over-simplification[46–49] or lack[50,51] of explicit electrostatics inherently precludes an accurate representation of charge-dependent properties. The membrane dipole potential, electric field, and orientational polarisation effects cannot be modelled; related phenomena might be ill-represented. The highly simplified modelling of water as a rather generic, apolar solvent is also a matter of concern. Water is a fundamental component of membrane systems: it drives the formation of lipid aggregates through the hydrophobic effect and provides the necessary electrostatic screening between charged particles (such as headgroups and ions). These features are not consistently accounted for by the referenced CG models. Water also generates a significant electric field by collectively orienting its dipoles: clearly, such a local polarisation effect cannot be captured by apolar solvent particles. Other issues affecting the specific CG models developed to date involve the membrane dynamic properties, which are reproduced with some difficulty, and the associated issue in interpreting the simulation time-scales. For example, lipid diffusion coefficients have been reported to be four[46,58] to one hundred[52] times

higher than experimental data; as already pointed out elsewhere,[60] the ad hoc rescaling of the simulation time according to these factors[46,52] is questionable, because it assumes that all dynamic events are homogeneous in time-scale, whereas, in general, dynamic processes in membrane systems are highly heterogeneous. Another issue regards the attractive possibility of mixing CG and AL representations in multiscale simulation, where selected parts of the system are described at an atomic level, whereas the surrounding environment is simulated by simplified models. It would not be straightforward to interface the available CG models with AL representations, due to the radically different description of electrostatics and solvent. An original approach to the problem involves performing a rigorous and consistent parameterisation of the mixed AL-CG interactions,[61] although charges are still absent.

In this thesis, we present a new specific CG model that addresses the issues discussed above. In common with the available models, we have significantly simplified the representation of lipid molecules to increase simulation efficiency. In contrast to the other models, we have retained an explicit description of water and the major electrostatics. Our new model is presented in Chapter 5; the construction, parameterisation and validation of the model, along with the simulation protocol, are reported in detail. In particular, the predictive power of our new CG model is demonstrated by reproducing experimental observables against which it is not directly fitted. All major physical properties of fluid-phase phospholipid bilayers are investigated; many of these properties are reproduced for the first time using a CG method.

Moreover, the CG force-field developed is shown to be directly compatible with AL representations. This allows the application of the methodology to multiscale simulation, where selected parts of the system are represented at the atomic level. In this context, solutes modelled at the atomic-level are inserted into the membrane, and permeability simulations are conducted. The results of such simulations for a set of small organic molecules are presented in Chapter 6. The methodology is then straightforwardly applied to study larger drug molecules and hormones; transmembrane permeability simulations of these solutes are described in Chapter 7.

Overall, this thesis reports work focused on both methodology development and application; important conclusions and novel results are summarised in Chapter 8, along with possible future directions. The most technical aspects of this work (such as algorithmic details of the specific software implemented and explicit formulae for forces and torques) are treated in the appendices. In summary, the main scientific aims of the project reported in this thesis are:

- develop a new simplified computer model for lipid bilayers;
- validate the model against available experimental and simulation data;
- use the model to study transmembrane solute permeability.

The results achieved will be critically discussed and contextualised within the available (alternative) methodologies.

Chapter 2

Lipids and membranes

Lipids* are the fundamental building blocks of any biological membrane; they are generally defined as substances that are soluble in organic solvent but only sparingly soluble in water.[62] As a class of molecules, lipids display a wide diversity in both structure and biological function, and they self-organise into many intriguing structures, with extraordinary material properties that have been optimised by evolutionary principles over billions of years.[4,63] Lipid membranes play crucial roles in compartmentalisation, they represent a solvent for other species (such as proteins) to function, and they contribute to the structural scaffolding of cells. Along with these rather passive functions, it has recently become evident that lipids can actively contribute to fission and fusion events, and in general to the regulation of membrane proteins. It is now accepted that lipids are as important for life as proteins, sugars, and genes.[4] In this chapter, the fundamental properties and phenomena regarding lipids and lipid bilayers are summarised, along with some of the most popular experimental techniques employed to study membrane systems.

2.1 Types of membrane lipids

There are three common classes of membrane lipids: phospholipids, glycolipids and cholesterol. Phospholipids are the major class of membrane lipids,[1] and are the main object of the modelling work presented in this thesis. Hence, in the following sections, we will focus on this particular lipid species. However, we give here a brief account on the two other common lipid types.

Glycolipids, as their name implies, are sugar-containing lipids. They are formed by association of a carbohydrate chain with lipids on the external surface of the cell membrane. Glycolipids thus extend from the plasma membrane into the aqueous environment outside the cell, acting as a recognition site for specific chemicals as well as helping to maintain the stability of the membrane and attaching cells to one another to form tissues. They also serve a role as energy stores.

*From the Greek *lipos*, meaning "fat".

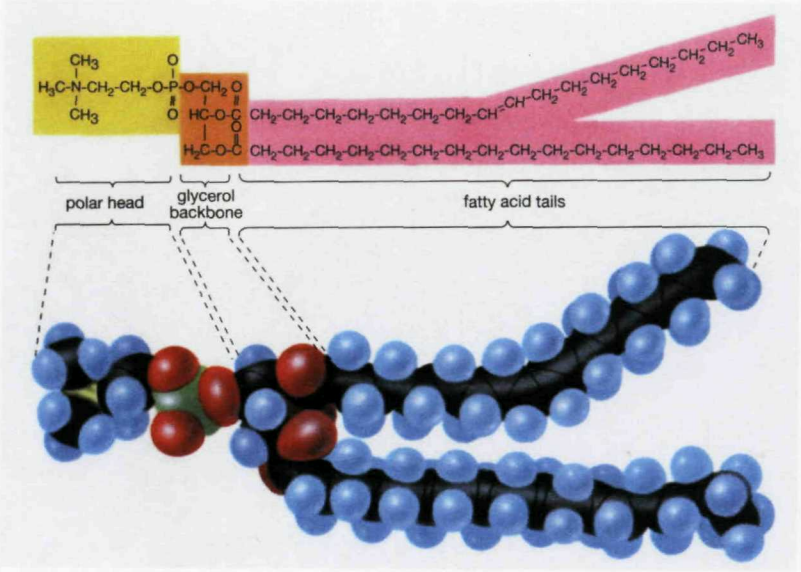
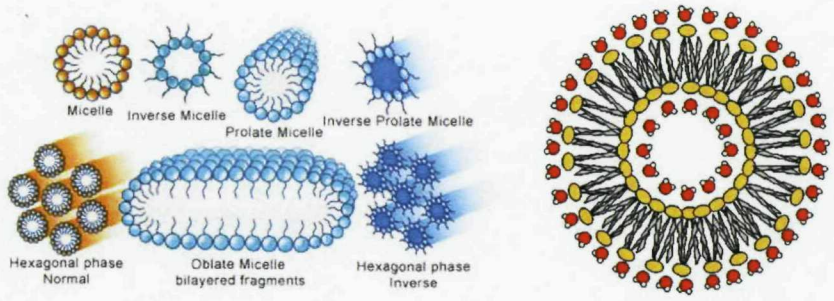


Figure 2.1: Phospholipid molecule. It is possible to identify the polar head group, the glycerol moiety, the ester groups and the hydrocarbon chains.[65]

Cholesterol is a lipid with a unique structure compared with the other lipids; it is a steroid, built from four linked hydrocarbon rings, terminated at one end by a hydrocarbon tail and on the other by a hydroxyl group. Cholesterol is universally present in the plasma membranes of all animals, in ratios of 25-50% of the total lipid content; however, it is essentially absent from some intracellular membranes, such as mitochondrial and Golgi.[4] Cholesterol is important in stabilising membranes, as its presence makes them thicker and less leaky; it is also an essential component of lipid "rafts", membrane microdomains believed to favour specific protein-protein interactions resulting in the activation of signalling cascades.[64]

2.2 Phospholipids: structure and self-assembly

The most important class of lipids in natural membranes is represented by the *phospholipids*.[1] Phospholipids are amphiphilic molecules constructed from four components: fatty acids, a "backbone" to which the fatty acids are attached, a phosphate, and an alcohol attached to the phosphate. The fatty acid components constitute a hydrophobic barrier, whereas the rest of the molecule has hydrophilic properties to enable interaction with the environment. The backbone component may be *glycerol*, a 3-carbon alcohol, in which case the phospholipid is called *phosphoglyceride*, or it may be *sphingosine*, a (more complex) amino alcohol, which constitutes *sphingolipids*. Sphingolipids comprise important lipid species such as *ceramide*, a fundamental constituent of the skin and an important molecule involved in programmed cell death.[4] However, the main lipid components of biological membranes are *phosphoglycerides* (or *glycerophospholipids*).[62] Figure 2.1 shows a typical phosphoglyceride, in particular a *phosphatidylcholine* lipid: the polar head comprises a positive group (choline $(CH_3)_3N\text{-}CH_2\text{-}CH_2$) and a negative group (phosphate $O\text{-}PO_2\text{-}O$), whereas the hydrophobic



(a) (b)

Figure 2.2: The most common aggregates formed by lipid molecules. (a) Non-bilayer phases.[66] (b) Lipid bilayer vesicle. Water molecules, hydrating the interior and exterior of the vesicle, are also schematically represented.[67]

part is made of the glycerol backbone (CH_2-CH-CH_2), two ester groups (O-CO-CH_2) and two hydrocarbon "chains" (also called "tails") of fatty acids. Hydrocarbon chains are formed by consecutive methylene (CH_2) segments terminating with a methyl (CH_3) segment. Hydrocarbon tails often comprise one or more double bonds between C atoms along the tail. In general, when dispersed in water, lipids are driven together by the *hydrophobic effect*, and self-organise in various aggregates depending on the specific lipid structure, temperature and level of hydration (Figure 2.2 a). The hydrophobic effect, that is, the tendency for oil and water to separate, is a complex phenomenon that manifests different characteristics depending on the system's temperature and pressure, the shape of the oil-like components and the size of the aggregates involved.[68,69] In the particular case of lipid/water mixtures at biological temperature and pressure, the hydrophobic effect is believed to be mainly of entropic origin.[4] Pure water systems are characterised by a tetrahedral arrangement of hydrogen-bonded molecules which maximises the system's entropy. When insoluble species (such as hydrocarbon molecules) are introduced in water, a loss in entropy is produced due to the induced local ordering of water around every insoluble molecule. To reduce such ordering, and hence to minimise entropy loss, the insoluble molecules are driven together so that the total surface area exposed to water, corresponding to the total area involved in local ordering phenomena, is decreased. This gain in entropy of the water upon assembly formation outweighs the enthalpy penalty (caused by the demixing of water and the insoluble species) and the loss of configurational entropy of the insoluble molecules due to the constraints typically imposed by the aggregate structure.[70] Owing to the fact that lipids form assemblies by self-aggregation processes that do not involve strong chemical forces, such assemblies can be categorised as *soft matter* materials. Soft matter comprises a large and ubiquitous class of systems including for example polymers, emulsions, colloids, liquid crystals and many biological materials. All these systems exist in a condensed phase which cannot be described unambiguously as either liquid or solid, as it typically possesses mixed properties. For instance, soft matter may display long-range ordering properties typical of solid mat-

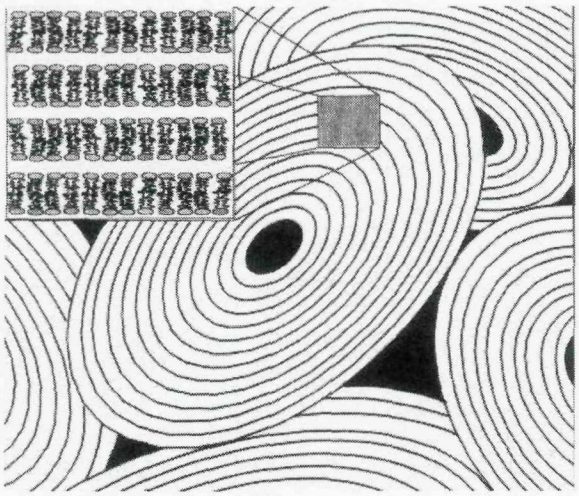


Figure 2.3: Schematic representation of multi-lamellar vesicles. The black areas represent regions of excess water. From Koenig et al.[72]

ter. However, unlike conventional solid materials, soft materials have physical properties which are largely dominated by entropy. Soft matter is highly deformable, and typically constructed in a hierarchical manner with substructures subtly interacting on several length and time scales.[4] All aggregates formed by lipids represent specific examples of soft matter structures. From a biological perspective, *bilayer* structures (Figure 2.2 b) are considered the most important category of lipid assemblies, as they form the fundamental backbone of the majority of biological membranes.

2.3 Phospholipid bilayers

Phospholipid bilayers constitute the basic material employed to encapsulate the cell and its sub-compartments. In the following sections, the main features of lipid bilayers are summarised.

2.3.1 Structure

Structural data on lipid bilayers are widely used as basic information to help understand and model biomembrane structure and the functions that take place therein.[71] Most experimental studies are performed on stacks of hydrated bilayers, especially on multi-lamellar vesicles (Figure 2.3). The internal structure is investigated by X-ray, neutron scattering, molecular-probe, and magnetic resonance techniques. The measurements obtained provide information about the membrane thickness, and, most importantly, can resolve the depth-dependent distribution of specific lipid segments across the bilayer (Figure 2.4).

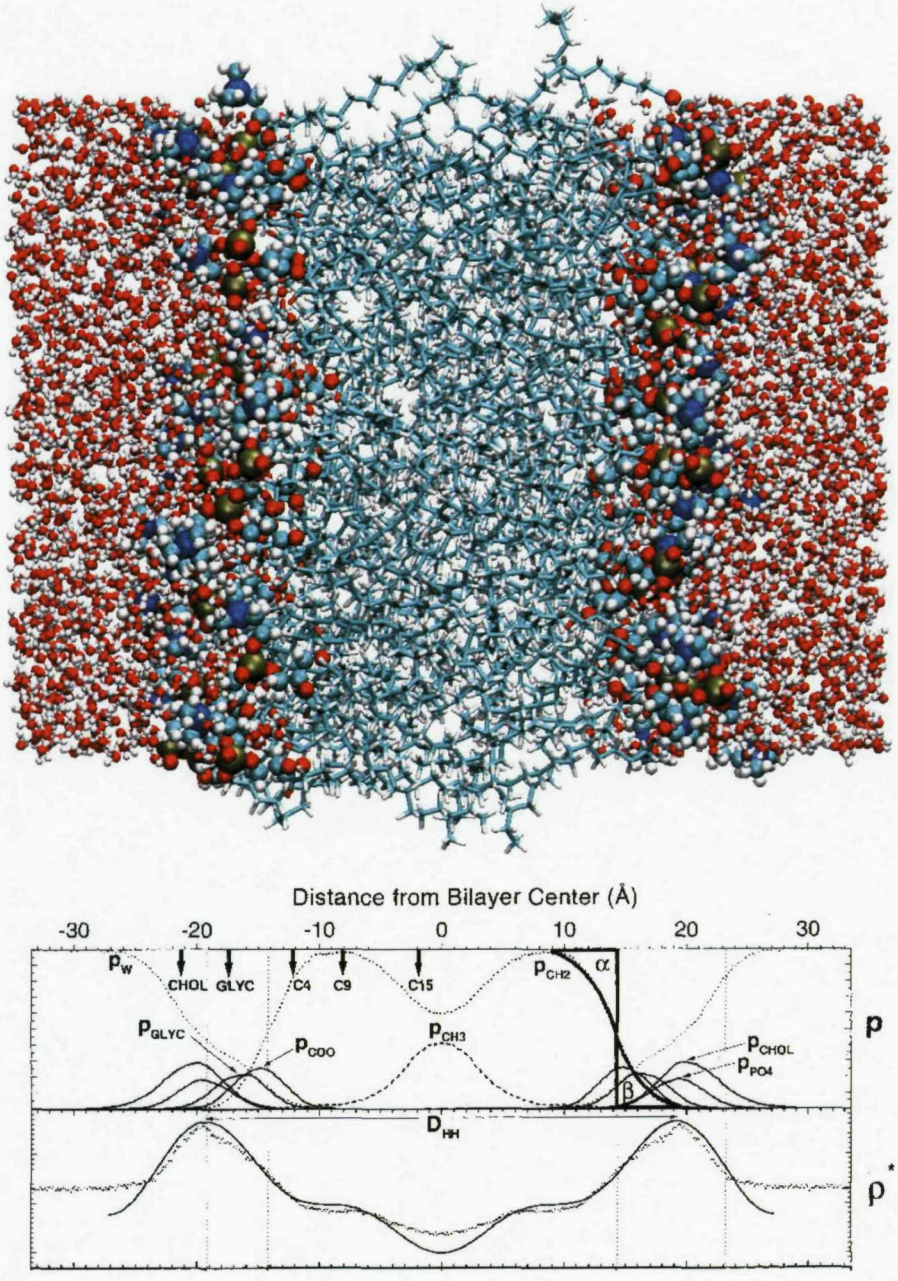


Figure 2.4: Transbilayer structure. The top panel shows a phospholipid bilayer model; hydrogens are coloured in white, oxygens in red, carbons in turquoise, nitrogens in blue and phosphoruses in dark yellow (adapted from Feller[73]). The bottom panel shows corresponding density profiles obtained from neutron-scattering and X-ray techniques; the curves give the relative probabilities of finding the different molecular segments of the phospholipid molecules (adapted from Nagle and Tristram-Nagle[71]).

Structure of the hydrocarbon region: intramolecular order parameters

The hydrocarbon tail region of a bilayer can be investigated by deuterium magnetic resonance; this technique allows tail ordering to be quantified in terms of order parameters. Order parameters generally describe the tail orientation as a function of depth. For each methylene group k along a lipid tail, the intramolecular order parameter S^k_{CD} can be defined as:[74]

$$S^k_{\mathrm{CD}} = \langle 3\cos^2\theta - 1\rangle/2 \tag{2.1}$$

with θ the instantaneous angle between the k-th C$-^2$H bond vector and the overall molecular axis. The overall molecular axis is the main axis of a lipid molecule, which, on average in the biologically-relevant fluid phase, is parallel to the "bilayer normal", that is, the direction perpendicular to the bilayer plane.[74] It is also possible to define the molecular axis of a chain segment k as the normal direction to the plane spanned by the two CH bonds of the k-th methylene group.[75] For each methyl segment k, the intramolecular *segmental* order parameter S^k_{mol} is then:

$$S^k_{\mathrm{mol}} = \langle 3\cos^2\eta - 1\rangle/2 \tag{2.2}$$

with η the instantaneous angle between the molecular axis of the k-th segment and the bilayer normal. S^k_{mol} are thus the order parameters of the segments' molecular axes with respect to the bilayer normal. Experiments can accurately determined S^k_{CD}, which can then be related to the segmental order parameters S^k_{mol} using the formulae:[75]

$$S^k_{\mathrm{mol}} = \begin{cases} -2\,S^k_{\mathrm{CD}} & \text{for the } k\text{-th } CH_2 \text{ segment} \\ -3\,S^k_{\mathrm{CD}} & \text{for the terminal } CH_3 \text{ segment} \end{cases}$$

In general, $S^k_{\mathrm{mol}} = 0$ indicates a completely random mean orientation, $S^k_{\mathrm{mol}} = 1$ indicates alignment of the segment molecular axis along the bilayer normal, whereas $S^k_{\mathrm{mol}} = -0.5$ indicates that the segment molecular axis lies in the bilayer plain (thus being perpendicular to the normal direction).

2.3.2 Phase behaviour of lipid bilayers

Lipids are able to adopt a range of phases depending on temperature and level of hydration, as already mentioned. In particular, fully hydrated bilayers composed of a single phospholipid species undergo a well-defined thermotropic phase transition in which the lipid chains change from an ordered, or gel, state to a fluid, or liquid-crystalline, state.[2] As an example, the phase diagram for the dimyristoylphosphatidylcholine (DMPC) bilayer is reported in Figure 2.5. Biologically, the most important phase is the liquid L_α phase, characterised by a high degree of disorder in the alkyl chains of the hydrophobic core.

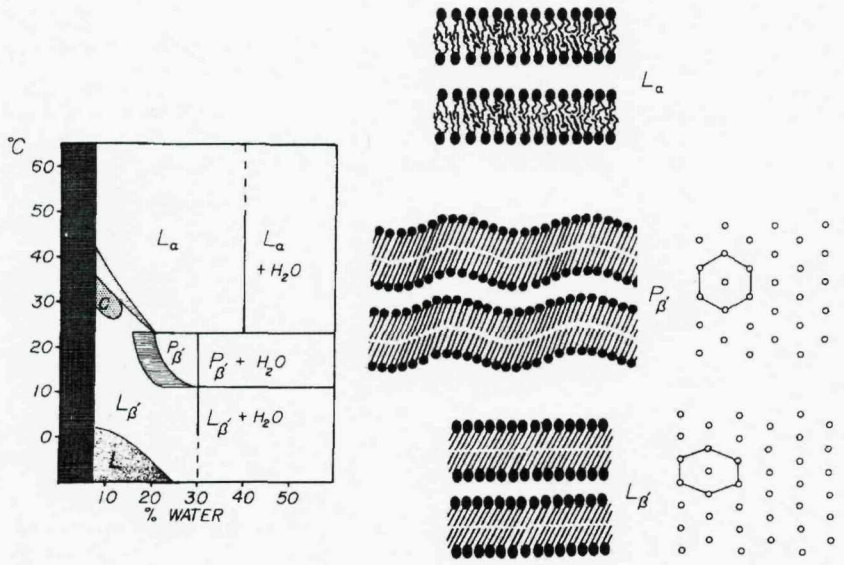


Figure 2.5: Phase diagram of hydrated DMPC bilayers, together with representations of the L_α, $P_{\beta'}$ and $L_{\beta'}$ phases. The hydrocarbon chain packing is a hexagonal array for the $P_{\beta'}$ phase and a "distorted" hexagonal lattice for the $L_{\beta'}$ phase. From Janiak et al.[76]

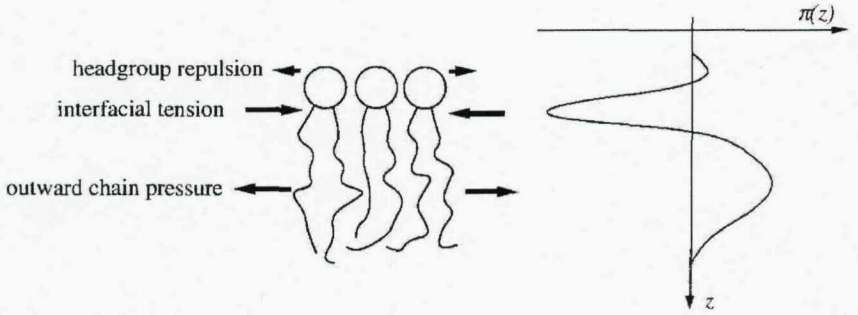


Figure 2.6: Lateral pressure profile. Proposed distribution of lateral pressure π within a flat monolayer as a function of the position along the interfacial normal z. From Templer et al.[11]

2.3.3 The lateral pressure profile

The transbilayer lateral pressure profile $\pi(z)$, where z is a spatial coordinate along the bilayer normal, is defined as the difference between the lateral and the normal pressures acting inside the bilayer: $\pi(z) = p_L(z) - p_N(z)$. The lateral pressure profile thus characterises the transmembrane distribution of forces; Figure 2.6 is an example of the proposed shape of such a distribution. In terms of magnitude, peak pressures of the order of several hundreds of atmospheres are predicted; the pressure profile results from an interplay of enormous opposing forces, that ultimately compensate each other. The lateral pressure profile changes in relation to the lipid composition (or the state of the lipid headgroups, e.g., by proton or ion binding), and as a result of the presence of cholesterol or solutes (such as drugs); the consequent depth-dependent changes in the stress distribution are predicted to affect

lipid phase behaviour and the conformation of inclusions such as proteins.[77] In fact, the lateral pressure profile controls a very large number of membrane features and phenomena: it determines the interfacial area, it is at the basis of phase transitions and fusion,[22,77–80] it affects permeability,[24] drug transport,[25] and anaesthesia,[26] it modulates the insertion and folding of membrane proteins,[23,81–84] and it is believed to directly control the functioning of several membrane proteins, such as the lipid synthesis regulatory enzyme CCT,[85,86] diacylglycerol kinase,[87] phospholipase A2[88] and C,[89] rhodopsin,[90–92] and several transbilayer channels.[23,93–96] The lateral pressure profile is also directly related to the elastic curvature constants that characterise the Helfrich expression for the bending free energy.[77,97,98] According to Helfrich's theory, the surface curvature elastic energy per unit area g is concisely expressed as:

$$g = \kappa \left(c_1 + c_2 - c_0\right)^2 / 2 + \kappa_{\mathrm{G}}\, c_1\, c_2 \tag{2.3}$$

with κ the bending rigidity, c_1 and c_2 the (local) principal curvatures, c_0 the spontaneous (or intrinsic) curvature and κ_{G} the Gaussian curvature modulus. Equivalently, the Helfrich equation can also be written:

$$g = 2\kappa \left(H - H_0\right)^2 + \kappa_{\mathrm{G}} K \tag{2.4}$$

with $H = (c_1 + c_2)/2$ the mean curvature, $H_0 = c_0/2$ the equilibrium mean curvature and $K = c_1 c_2$ the Gaussian curvature. The constants appearing in Helfrich's expressions in turn control membrane shape, and play specific roles in the mechanisms modulated by the lateral pressure profile (mathematical relations between the Helfrich constants and the pressure distribution $\pi(z)$ will be given in Section 5.2.2).

Experimentally, it has been so far impossible to quantitatively measure the pressure profile. The only experiments performed to date have yielded qualitative and partial pictures for the hydrocarbon region only. In these experiments,[11,24] changes in the lateral pressure along the bilayer normal were "sensed" using a series of di-pyrenyl phosphatidylcholine (dipyPC) fluorescence probes. DipyPCs are PC lipids carrying pyrene moieties attached to their tail ends. Ultraviolet stimulation produces both monomer and excimer fluorescence from pyrene. The excimer signal, which is entirely intramolecular at low dilutions of dipyPC, results from (excited) dimerization of adjacent pyrene groups, and depends on the frequency with which the two pyrene moieties collide to form excimers; this frequency, in turn, is proportional to the lateral pressure. The relative intensity of the excimer to monomer signal is thus a measure of the pressure; by using dipyPCs of different acyl chain lengths it is possible to qualitatively estimate the pressure variations across different depths in the bilayer.[11]

2.3.4 The dipole potential

In typical physiological conditions, the presence of ions in the water phases at the interface with both sides of a membrane, along with the orientational ordering of interfacial water dipoles and the intramembrane distribution of charged groups along the lipids (Figure 2.7),

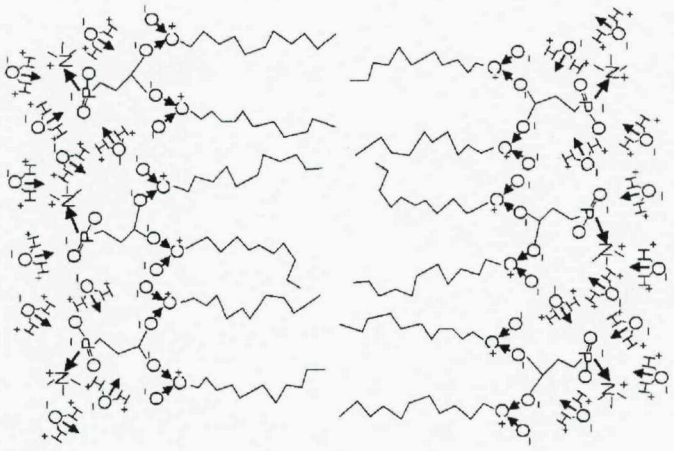


Figure 2.7: Electrostatics at the lipid/water interface. The signs and arrows represent the main charges and dipoles possessed by lipid segments and water molecules. Adapted from Shinoda et al.[99]

create a characteristic electrical potential distribution along the direction normal to the membrane plane (Figure 2.8). In particular, the membrane *dipole potential* Ψ_d originates from the alignment of dipolar residues of the lipids and water dipoles in the bilayer-water interfacial region. Ψ_d is positive inside the membrane with respect to the outer water phase; its exact value is unknown, but it is believed to be of the order of $0.2-0.5$ V. Since this potential drops across a very small distance within the headgroup region of the membrane,[101] the corresponding electric field strength is enormous, with peak values in the range $10^8 - 10^9$ V/m. The membrane dipole potential, and associated electric field, are involved in a great number of biological processes, such as membrane fusion,[15,102] permeation,[103] the regulation of membrane proteins (Na^+,K^+-ATPase,[14] gramicidin channel,[13] phospholipase A_2[12]) insertion and folding of amphiphilic peptides,[16] the kinetics of DNA-lipid complexes,[104] the kinetics of redox reactions at membrane surfaces,[17] human skin permeability,[18] general anaesthesia,[105,106] membrane partitioning of pregnanolone,[20] the binding capacity of saquinavir,[19] and the modulation of molecule-membrane interactions in lipid rafts with possible effects on cells signalling.[107,108] Despite the growing evidence for its importance, the dipole potential Ψ_d has received so far relatively little attention;[101] for instance, the overall transmembrane potential $\Delta\Psi$, which regulates numerous ion channels, is much more popular. The reason is that, while $\Delta\Psi$ can be easily measured and controlled by placing electrodes in the solution phases on each side of the membrane, the dipole potential Ψ_d cannot be directly measured, as it is impossible to insert electrodes at different depths within the membrane.[101] Therefore, the dipole potential can only be estimated by indirect measurements. One method involves studying the membrane conductivity associated with the translocation of different hydrophobic ions; due to the presence of the dipole potential, hydrophobic anions permeate much faster than hydrophobic cations, and the magnitude of this effect can be used to quantify Ψ_d.[109,110] It is also possible to consider the bilayer dipole potential to be equivalent to the potential across a monolayer, which can be directly obtained using electrodes after

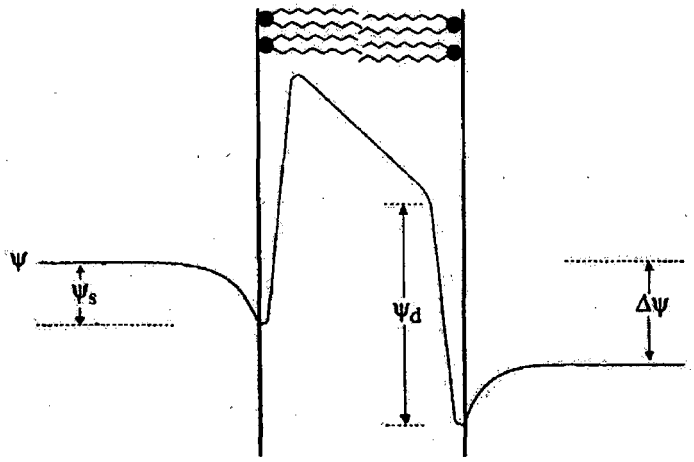


Figure 2.8: The electrical potential profile Ψ across a phospholipid membrane. The transmembrane potential $\Delta\Psi$ is due to the difference in anion and cation concentrations between the two aqueous bulk phases. The surface potential Ψ_s arises from charged residues (charged headgroups) at the membrane-solution interface. The dipole potential Ψ_d results from the alignment of dipolar residues of the lipids and associated water molecules within the membrane. From Clarke.[100]

Table 2.1: Measurements of the dipole potential Ψ_d in phosphocholine bilayers.

Method	*Lipid*	Ψ_d / V
Ion translocation[109]	DPPC	0.227
Ion translocation[110]	DPPC	0.346
Monolayer[111]	DMPC	0.449
Cryo-EM[10]	DPhPC	0.510

DPPC = dipalmitoylphosphatidylcholine, DMPC = dimyristoylphosphatidylcholine, DPhPC = diphytanoylphosphatidylcholine.

spreading a monolayer of lipids onto the surface of a Langmuir trough.[111] However, this measurement relies on the questionable assumption that such isolated monolayers are equal to each of the monolayers paired into a bilayer assembly. Recently, the dipole potential has been estimated using cryo-EM, by recording the interactions of electrons with regions of different electrostatic potentials across rapidly frozen bilayers.[10] Unfortunately, this technique also relies on a number of approximations that might affect its reliability. For instance, it is expected that the bilayer structure remains intact during the freezing, which occurs at a rate of $10^6\,\mathrm{K/s} = 1\,\mathrm{K}/\mu\mathrm{s}$; in fact, this cooling rate might be slow enough to allow artificial rearrangements of water and lipid molecules. The available experimental estimates for the dipole potential of ester-PC lipids are collected in Table 2.1.

2.3.5 Dynamics

Despite displaying structural integrity typical of solid materials, the biologically relevant state of lipid bilayers is that of a liquid crystal characterised by substantial fluidity and

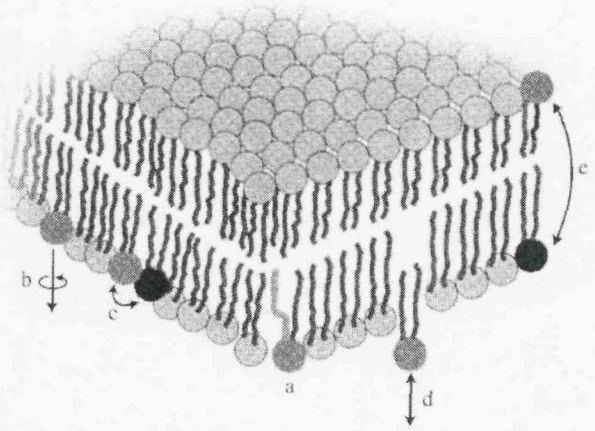


Figure 2.9: Lipid dynamics. Characteristic motions of lipid molecules inside a bilayer: a) tail conformational change, b) rotation around molecular axis, c) diffusion (swap), d) protrusion, e) flip-flop. From Mouritsen.[4]

disorder. In their biological state, lipids typically display various dynamic features: they change conformation, diffuse laterally in the membrane plane, rotate around their main molecular axis, protrude into the water phase and flip through the monolayers. These characteristic motions are represented in Figure 2.9. The lateral mobility of lipids in the plane of the membrane (diffusion) is particularly important, as it defines the liquid-like nature of membranes. The diffusion of lipids can be measured by a number of different experimental methods. The motion of a single lipid can be detected by single-particle tracking.[112] A colloidal particle of a typical diameter of 40 nm is attached to the lipid molecule and the particle's motion is followed by microscopy; the spatial resolution of this kind of experiment is $\approx$ 50 nm, and the time resolution is $\approx$ 5 ms. Diffusion coefficients have been obtained by quasi-elastic neutron scattering, which measures short-range diffusion taking place over sub-nanosecond timescales.[113] Alternatively, long-range methods such as NMR spectroscopy[8] and fluorescence recovery after photobleaching[114] can be used; these methods probe millisecond time-scales. Interestingly, the lateral diffusion coefficients measured with short-range methods turn out to be $\approx$ 2 orders of magnitude higher than those obtained by long-range observations.[115] To understand the reason of this discrepancy, it is useful to take into account the *free-volume* diffusion theory, which was proposed more than 50 years ago to describe transport in soft condensed matter.[116–118] The foundation of the theory is that diffusion is related to the average free volume per particle; molecular diffusion proceeds by jumps occurring when a large enough free volume is created (by free volume redistribution) next to the diffusing molecule. Lipid diffusion is then assumed to proceed by "hopping" of molecules into vacancies formed by lateral density fluctuations in the membrane.[119] In between jumps, a lipid molecule spends a relatively long time "rattling in a cage" formed by its neighbours. Over short ($<$ 1 ns) time-scales, the diffusion coefficient is high because it is determined by the rapid short-range lipid motion mainly due to such "rattling-about" behaviour. Over longer times, this rattling motion averages out, yielding no net displace-

ment. The true, long-range diffusion coefficient is thus determined by the lipid jumps, that give rise to effective displacement over extended ($> 10\,\mathrm{ns}$) time-scales.

2.4 Summary

Membrane lipids are driven together by water, through the hydrophobic effect, into various self-assembling structures. The most abundant lipid aggregate is the bilayer, which constitutes the backbone of the plasma membrane encapsulating cells. Owing to many peculiar molecular characteristics of lipids (such as shape and charge anisotropy) and to the fact that they interact through "weak", non-covalent forces, the resulting bilayers are multifaceted materials:

- the internal structure is highly stratified and characteristically (dis)ordered;
- there is a steeply varying depth-dependent distribution of pressures;
- the charges present give rise to a large potential difference between the hydrocarbon core and the outer water phase;
- there are various dynamic phenomena, such as lipid diffusion in the bilayer plane.

All these different bilayer characteristics, often involved in a complex interplay with proteins, are integral to many biological phenomena, such as transport, growth and signalling.

Chapter 3

The molecular dynamics method for simulation of condensed phases

Molecular dynamics (MD) is a computer simulation technique employed to compute the equilibrium and transport properties of a classical many-body system. The word *classical* is intended to mean that the motion of the constituent particles obeys the laws of classical mechanics. In this chapter, the MD method is summarised, with special focus on the model potentials, integration schemes, and computational techniques relevant for the applications presented in the following chapters. Thorough descriptions of particle-based simulation methods can be found in several excellent books,[120–124] which are also the main sources of this chapter. The most technical aspects, such as details of specific algorithms adopted and explicit derivations of forces and torques, are reported in the appendices.

3.1 Foundations

In statistical mechanics, for the canonical ensemble where the number of particles N, the volume V and the temperature T are fixed, the equilibrium average of some quantity G is expressed in terms of phase-space integrals involving the potential function $U(\mathbf{r}_1, \ldots, \mathbf{r}_N)$:

$$\langle G \rangle = \frac{\int G(\mathbf{r}_1, \ldots, \mathbf{r}_N)\, e^{-\beta U(\mathbf{r}_1, \ldots, \mathbf{r}_N)}\, \mathrm{d}\mathbf{r}_1 \cdots \mathbf{r}_N}{\int e^{-\beta U(\mathbf{r}_1, \ldots, \mathbf{r}_N)}\, \mathrm{d}\mathbf{r}_1 \cdots \mathbf{r}_N} \tag{3.1}$$

with $\mathbf{r}_i$ the coordinates, $\beta = 1/k_B T$ and k_B the Boltzmann constant. This average corresponds to a series of measurements over an ensemble of independent systems.

In MD simulation, the microscopic state of a system is defined by the positions and momenta of the particles of the system under investigation. In particular, the total energy, or Hamiltonian, H can be written as the sum of kinetic energy K and potential energy U:

$$H(\mathbf{q}, \mathbf{p}) = K(\mathbf{p}) + U(\mathbf{q}) \tag{3.2}$$

with $\mathbf{q}$ and $\mathbf{p}$ the system sets of coordinates and momenta, respectively. From the potential

THE GLOBAL MD ALGORITHM

1. Input initial conditions:

- Interaction potential U and system topology (the *force-field*)
- Positions and velocities of all sites in the system
- Simulation parameters (time-step, temperature, pressure, etc.)

2. Compute:

- Forces and torques
- Thermodynamic quantities
- Properties of interest

3. Integrate equations of motion

repeat steps **2,3** for the required number of cycles

4. Output:

- Trajectory of every particle
- Averages of properties of interest

Figure 3.1: The main steps of a typical molecular dynamics simulation.

energy it is possible to obtain the forces acting on each molecule, and from there calculating the entire time evolution of the system. It is therefore possible to calculate the average of a quantity G as:

$$\langle G \rangle = \frac{1}{M} \sum_{\mu=1}^{M} G_\mu(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{3.3}$$

over a set of M measurements taken as the (single) system evolves in time.

The *ergodic hypothesis*, which is the fundamental assumption of molecular dynamics, states that the ensemble average of Equation 3.1 is equal to the time average of Equation 3.3.

A global diagram for MD is given in Figure 3.1. The various components are described in the following sections.

3.2 Interaction potentials

The potential energy U describes the interactions between the particles of a system. U typically comprises several terms, accounting for different types of intermolecular interactions (such as van der Waals and electrostatics) and intramolecular interactions (such as covalent bonding). Considering that forces and torques are defined completely by the total potential U, the set of constituent potential functions and corresponding parameters of a given system is often called the *force-field*. The most common potential terms present in molecular simulation are described in the following sections, along with more unusual potentials employed in the work presented in the following chapters of this thesis.

3.2.1 The Lennard-Jones potential

For a pair of sites located with a distance r between them, the Lennard-Jones potential energy is:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3.4}$$

with σ the contact distance (or collision diameter) and ϵ the magnitude of the attractive energy well depth. The Lennard-Jones pair-potential is characterised by a short-range strongly repulsive core and a long-range attractive tail. It is an isotropic potential, as it defines spherically-symmetrical interacting sites.

3.2.2 The Gay-Berne potential

The Gay-Berne potential[125] can be considered a sort of generalised, anisotropic, Lennard-Jones potential. It is typically used to model molecules as prolate solid ellipsoids and oblate (disk-like) particles. Considering two particles with orientations identified by the unit vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_j$, and $\mathbf{r} \equiv r\hat{\mathbf{r}}$ being the vector joining their centres of mass, the Gay-Berne potential can be written as:

$$U(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \mathbf{r}) = 4\epsilon(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) \left[\left(\frac{\sigma_0}{r - \sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) + \sigma_0} \right)^{12} - \left(\frac{\sigma_0}{r - \sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) + \sigma_0} \right)^{6} \right] \tag{3.5}$$

with σ and ϵ defining the contact distance, and the attractive strength, respectively. These orientation-dependent terms control the range and strength of the interaction. Detailed formulae are reported in Section 5.1.2, where the Gay-Berne potential is employed in the development of the lipid model.

3.2.3 Electrostatic potentials

In molecular modelling, charge distributions are typically described by empirical partial charges and point dipoles chosen to effectively reproduce the known multipole moments. The electrostatic interaction potentials commonly employed in molecular simulation are presented

in the following paragraphs. More details, along with derived formulae for forces and torques, are reported in Appendix B.

Charge-charge potential

The interaction potential energy $U(r)$ between two point charges Q_i and Q_j located with a distance r between them is defined by Coulomb's law:

$$U(r) = \frac{Q_i Q_j}{4\pi\epsilon_0 r} \tag{3.6}$$

with ϵ_0 the dielectric constant in free space (vacuum).

Charge-dipole potential

Considering a charge Q interacting with a dipole μ, the electrostatic potential energy becomes:[126]

$$U(\mathbf{e}_\mu, \mathbf{r}) = \frac{Q\mu}{4\pi\epsilon_0 r^3} (\mathbf{e}_\mu \cdot \mathbf{r}) \tag{3.7}$$

with $\mathbf{e}_\mu$ and $\mathbf{r}$ the unit vector along the dipole, and the charge-dipole distance vector, respectively.

Dipole-dipole potential

For dipole-dipole interactions, the interaction energy is:[126]

$$U(\mathbf{e}_i, \mathbf{e}_j, \mathbf{r}) = \frac{\mu^2}{4\pi\epsilon_0 r^3} \left(\mathbf{e}_i \cdot \mathbf{e}_j - 3 \frac{\mathbf{e}_i \cdot \mathbf{r}}{r|\mathbf{e}_i|} \frac{\mathbf{e}_j \cdot \mathbf{r}}{r|\mathbf{e}_j|} \right) \tag{3.8}$$

with $\mathbf{e}_i$ and $\mathbf{e}_j$ the unit vectors along the directions of the two dipoles.

3.2.4 The Hooke (harmonic) potential

Covalent bonds can be represented by mechanical springs, which in turn are modelled via the Hooke (harmonic) potential $U(l)$:

$$U(l) = \frac{1}{2} k (l - l_0)^2 \tag{3.9}$$

where k is the rigidity constant, l is the actual bond length and l_0 is the reference bond length.*

*The "reference bond length" (or "natural bond length") is sometimes called "equilibrium bond length", but to do so can be misleading, considering that the equilibrium bond length is really the value adopted in a minimum energy structure, when all other potentials contribute. The equilibrium value may thus deviate from the reference value.[121]

3.3 Rigid bodies

The molecular models employed in the work presented in this thesis include both standard isotropic potentials (such as Lennard-Jones and Coulomb) and more complex anisotropic, orientation-dependent potentials (such as Gay-Berne and dipolar). In molecular dynamics, particles represented by isotropic potentials are simulated as simple point-masses, their motion being completely described by translational degrees of freedom (typically the mass centre coordinates). However, sites modelled by anisotropic potentials also possess orientational degrees of freedom, so that a point-mass representation becomes insufficient. Moreover, it is sometimes practical to model entire molecules as rigid, neglecting intramolecular flexibility; in this case it is convenient to treat such particles as single entities and characterise their motion also in terms of linear and orientational degrees of freedom. In such cases particles can be efficiently represented as *rigid bodies*. The linear motion of a rigid body is described by the motion of its mass centre, which can be simply treated as a point-mass equivalent to the mass of the entire body. The rotational motion is however more complex, and it requires a representation of orientational degrees of freedom. In particular, the orientation of a rigid body specifies the relation between an axis system S fixed in space and one (in general) translating and rotating attached to the body, usually the "principal" body-fixed system b in which the inertia tensor is diagonal.[120] The orientation of a rigid body can be expressed through the full rotation matrix $\mathbf{R}$. The nine components of the rotation matrix are the direction cosines of the body-fixed axis vectors in the space-fixed frame. There are two ways of interpreting the rotation described by $\mathbf{R}$:

- consider a vector $\mathbf{r}^S$ and use $\mathbf{R}$ to obtain its components in the rotated coordinate system, namely $\mathbf{r}^b = \mathbf{R}\mathbf{r}^S$;
- rotate a vector, beginning with $\mathbf{r}^b$ and applying the opposite rotations in reverse order by means of the transpose of $\mathbf{R}$, in which case the result is the rotated vector $\mathbf{r}^S = \mathbf{R}^T\mathbf{r}^b$.

Clearly, if $\mathbf{r}$ is a vector fixed in the molecular frame (for instance a bond vector of a rigid molecule) then $\mathbf{r}^b$ will not change with time; in space-fixed coordinates, though, the components of $\mathbf{r}^S$ will vary.

3.4 Forces and torques

The force $\mathbf{f}$ on the mass centre of a particle can be obtained from the gradient of the potential U:

$$\mathbf{f} = -\nabla_{\mathbf{r}} U \tag{3.10}$$

with $\mathbf{r}$ the vector defining the particle's position. For "molecular" rigid bodies comprising a number of n atomic sites $a = 1, \ldots, n$ the total force $\mathbf{f}$ acting on the molecule is:

$$\mathbf{f} = \sum_{a=1}^{n} \mathbf{f}_a \tag{3.11}$$

being $\mathbf{f}_a$ the force acting on the atomic site a. The torque $\mathbf{T}$ about the centre of mass of the particle is computed as:

$$\mathbf{T} = \sum_{a=1}^{n} (\mathbf{r}_a - \mathbf{r}_{\mathrm{COM}}) \wedge \mathbf{f}_a = \sum_{a=1}^{n} \mathbf{d}_a \wedge \mathbf{f}_a \tag{3.12}$$

with $\mathbf{r}_a$ the atom position in the system's frame of reference, $\mathbf{r}_{\mathrm{COM}}$ the molecule's centre of mass position and $\mathbf{d}_a$ the atomic position relative to the molecule's centre of mass.[120]

Single-site particles must also be treated as rigid bodies when the potential is orientation-dependent. For symmetric particles (such as Gay-Berne sites), the torque is equivalent to a force acting on a point separated by a unit distance from the mass centre and acting in a direction orthogonal to the molecular symmetry axis.[127] This equivalent force can be defined in terms of the derivative of the potential with respect to the coordinates of this point, where the mass centre is taken as the origin. These coordinates are just the components of the unit vector $\hat{\mathbf{e}}$ describing the molecular orientation, so that the torque can be calculated as:

$$\mathbf{T} = -\hat{\mathbf{e}} \wedge \nabla_{\hat{\mathbf{e}}} U. \tag{3.13}$$

The torque acting on a symmetric rigid body is thus defined as perpendicular to the symmetry axis; this relies on the assumption that the inertia moment along the symmetry axis is infinite.

3.5 Equations of motion

For simple point-mass particles, the motion is completely described by Newton's equation:

$$m\ddot{\mathbf{r}} = \mathbf{f} \tag{3.14}$$

with m the mass of the particle, $\ddot{\mathbf{r}}$ its acceleration and $\mathbf{f}$ the force acting on it. For rigid bodies, the motion also contains a rotational contribution, which can be described by Euler's equation:

$$\boldsymbol{\omega} \wedge \mathbf{I}\boldsymbol{\omega} + \mathbf{I}\dot{\boldsymbol{\omega}} = \mathbf{T} \tag{3.15}$$

with $\mathbf{I}$ the moment of inertia tensor, $\boldsymbol{\omega}$ the angular velocity, $\dot{\boldsymbol{\omega}}$ the angular acceleration and $\mathbf{T}$ the torque about the body mass centre, all these quantities being expressed in the body-fixed (principal) reference frame. In MD simulations, the equations of motion must be solved via numerical integration. The *integrator* is the beating heart of any dynamics simulation; it is the scheme which replaces a differential equation in continuous time by a difference equation defining approximate snapshots of the solution at discrete time-steps.[128] The crucial properties that a good integrator should possess, and that are possessed by the equations of motion in the first place, are:

- symplecticity, which implies preservation of phase-space volume and Hamiltonian value (energy);

- time-reversibility, that is, the system capability to trace back its trajectory by reversing the velocities.

It has indeed been proved that symplecticity and reversibility closely relate to the stability of an integrator, on extremely long-time simulations and allowing large step sizes;[129] there are now numerous examples illustrating the superior preservation of phase-space structures and qualitative dynamics by symplectic integrators.[130–135] In practice, the total energy is not preserved exactly, but the energy error remains constant over long times; this is different from non-symplectic methods, which typically display a systematic energy drift in time.[123] Rigorously, it may be shown that symplectic integrators exactly conserve a "pseudo-Hamiltonian" or "shadow-Hamiltonian" $\overline{\mathcal{H}}$ which differs from the true one by a small amount (vanishing as $\Delta t \to 0$, with Δt the integration time-step). This means that no drift in energy will occur: the system will remain on a "hypersurface" in phase space which is "close" (in the above sense) to the true constant-energy hypersurface.[136]

In this work, we employ the symplectic and time-reversible rigid body integrator developed by Dullweber, Leimkuhler and McLachlan,[132] DLM for short. The DLM method, based on a representation of the orientation of rigid bodies with rotation matrices,* comprises two parts, as described in the following paragraphs.

Part A Given the forces $\mathbf{f}(t)$ and the space-frame torques $\mathbf{T}^S(t)$ at the current time t, the momenta of all molecules are advanced from t to $t + \Delta t/2$, whereas mass centre positions $\mathbf{r}$ are moved a full time step:

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t) + \Delta t\, \mathbf{f}(t)/2m \tag{3.16}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t\, \mathbf{v}(t + \Delta t/2) \tag{3.17}$$

$$\mathbf{h}^b(t + \Delta t/2) = \mathbf{h}^b(t) + \Delta t\, \mathbf{T}^b(t)/2 \tag{3.18}$$

where $\mathbf{h}^b = \mathbf{I}\boldsymbol{\omega}^b$ is the body-frame angular momentum, with $\mathbf{I}$ the principal moments of inertia tensor and $\boldsymbol{\omega}^b$ the body-frame angular velocity, and $\mathbf{T}^b$ is the body-frame torque, which is obtained from $\mathbf{T}^b = \mathbf{Q}(t)\mathbf{T}^S(t)$, $\mathbf{Q}(t)$ being the rotation matrix. Now five consecutive body-frame rotations $\mathbf{R}_1, \ldots, \mathbf{R}_5$ are applied to all angular momenta and all orientation matrices are propagated for a full time step, from $\mathbf{Q}(t)$ to $\mathbf{Q}(t + \Delta t)$:

$$\mathbf{Q}(t + \Delta t) = \mathbf{Q}(t)\mathbf{R}_1^T\mathbf{R}_2^T\mathbf{R}_3^T\mathbf{R}_4^T\mathbf{R}_5^T$$

with the explicit computation being reported in Appendix A.

Part B After having obtained $\mathbf{r}(t + \Delta t)$ and $\mathbf{Q}(t + \Delta t)$ from the previous part, the corresponding new forces $\mathbf{f}(t + \Delta t)$ and torques $\mathbf{T}^S(t + \Delta t)$ are calculated. Subsequently, the

*The rotation matrix representation and the related rotational part of the integration scheme are really the novel features of the algorithm,[132] in that the linear integration is performed as in the popular velocity-Verlet scheme.[120]

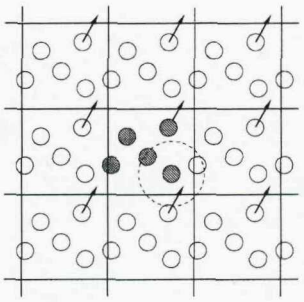


Figure 3.2: Periodic boundary conditions in two dimensions. From Allen.[136]

momenta are propagated another half time step through the following formulae:

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t+\Delta t/2) + \Delta t\,\mathbf{f}(t+\Delta t)/2m \tag{3.19}$$

$$\mathbf{h}^b(t+\Delta t) = \mathbf{h}^b(t+\Delta t/2) + \Delta t\,\mathbf{T}^b(t+\Delta t)/2 \tag{3.20}$$

where again $\mathbf{T}^b(t+\Delta t) = \mathbf{Q}(t+\Delta t)\mathbf{T}^S(t+\Delta t)$. The integration step is now complete.

The advantage of the DLM method over traditional integrators[120,137,138] is that the rotation matrix representation allows normalisation calculations to be avoided. This is necessary to preserve the time-reversibility property. Also, the DLM integrator is symplectic, whereas the traditional alternatives are not. The practical advantage of time-reversible symplectic integrators is that they prove extremely stable. In Appendix A we show examples where the DLM scheme permits integration steps to be used that are ten times larger than those possible with traditional methods.

3.6 Periodic boundary conditions

Owing to limitation of computer resources, simulations are typically performed on systems containing relatively small numbers of particles. In a typical MD simulation comprising 1000 particles, roughly half of them are in contact with the outer boundaries. Even for 10^6 atoms, the surface atoms amount to 6% of the total. Assuming, as it is often the case, that we are interested in the bulk properties of the system (and not in boundary effects), the presence of a boundary surface will introduce severe simulation artefacts. This problem can be solved by surrounding the cell with replicas of itself, thus effectively eliminating any physical boundary. This is shown in Figure 3.2. Whenever a molecule leaves the central cell passing through a particular face of the central simulation region, a "replacement" particle will enter the central cell through the opposite face. Only the coordinates in the central box need to be recorded; in the course of a simulation, when a particle leaves the central simulation box, its coordinates are updated with the values of the corresponding incoming image.

3.6.1 Minimum image convention

Periodic boundaries solve the problem of surface effects but introduce a "computational paradox": all the infinite images of any given particle should now be considered in the interaction calculations. To avoid this (impossible) task, the *minimum image convention* is normally adopted: each atom of the main simulation cell interacts only with the nearest image of any other particle.

3.7 Truncation of nonbonded interactions

Even with the implementation of the minimum image convention, the evaluation of all nonbonded interactions is still a computationally expensive job; also, such an "all-pair" approach is often unnecessary to achieve the typical degree of accuracy required. A further approximation is therefore made to increase the computational efficiency of simulation programs. Nonbonded interaction models are normally short-range: the potential energy between a pair of particles rapidly decays with increasing interparticle distance, becoming almost negligible after some *cutoff distance* r_c. To maximise computational efficiency, the potential is thus normally ignored (truncated) after r_c. For consistency with the minimum image convention, the cutoff radius r_c must be smaller than half the length of the shortest edge of the simulation region.

A problem arising from truncating the interactions is the introduction of a discontinuity in the potential and its derivative (force), affecting both the energy of the system and the motion of the particles. This problem can be tackled by changing the form of the potential function slightly, adding a constant and a linear term so that both the potential and its derivative go smoothly to zero at the cutoff distance r_c:[120,124]

$$U^{\mathrm{SF}}(r) = U(r) - U(r_c) - (r - r_c)\frac{\mathrm{d}U(r)}{\mathrm{d}r}|_{r=r_c} \tag{3.21}$$

where $U^{\mathrm{SF}}(r)$ is the "new" model, called *shifted-force* potential, and $U(r)$ is the original potential. This removes problems in energy conservation and any numerical instability in the equations of motion.[120] A possible issue with this treatment is that the potential is modified across the entire interaction range (even if only slightly); properties sensitive to the specific form of the potential might be affected. Alternative methods involve using so-called *switching* functions, applied in the proximity of r_c to remove the discontinuity without changing the overall potential form.[124] In this case the potential is "switched off" smoothly across a (small) distance between a switching distance r_s and the cutoff distance r_c (where for instance $r_s = 0.9\,r_c$).

3.8 Thermodynamic measurements

Basic thermodynamic properties can be easily calculated from an MD simulation; measurements are averaged over time, typically after an initial equilibration stage.

3.8.1 Potential energy

The total potential energy of a system is measured through the evaluation of a double-loop over all pair interactions:

$$\mathcal{U} = \sum_i \sum_{j>i} u(r_{ij}) \tag{3.22}$$

where i and j identify the interacting particles, $u(r_{ij})$ their potential energy and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ their separation distance.

3.8.2 Kinetic energy and temperature

The instantaneous kinetic energy $\mathcal{K}$ for a system of N point-mass particles is:

$$\mathcal{K} = \frac{1}{2} \sum_{i=1}^{N} \left(m_i |\mathbf{v}_i|^2 \right) \tag{3.23}$$

with m_i and $\mathbf{v}_i$ the mass and velocity of the i-th particle. The instantaneous kinetic energy $\mathcal{K}$ for a system of N symmetric rigid bodies (such as Gay-Berne sites) is:

$$\mathcal{K} = \frac{1}{2} \sum_{i=1}^{N} \left(m_i |\mathbf{v}_i|^2 + I_i |\boldsymbol{\omega}_i^b|^2 \right) \tag{3.24}$$

with I_i and $\boldsymbol{\omega}_i^b$ the principal moments of inertia tensor and the body-frame angular velocity of the i-th particle. For general (non-symmetric) rigid bodies, the kinetic energy can be measured as:

$$\mathcal{K} = \frac{1}{2} \sum_{i=1}^{N} \left(m_i |\mathbf{v}_i|^2 + I_{x_i} (\omega_{x_i}^b)^2 + I_{y_i} (\omega_{y_i}^b)^2 + I_{z_i} (\omega_{z_i}^b)^2 \right) = \frac{1}{2} \sum_{i=1}^{N} \left(m_i |\mathbf{v}_i|^2 + |\mathbf{h}_i^b|^2 / \mathbf{I} \right) \tag{3.25}$$

with I_i, ω_i^b and $\mathbf{h}^b$ being the inertia moment, the angular velocity and the angular momenta of the i-th particle in the body-frame, respectively.

The equipartition theorem states that each degree of freedom (DOF) contributes an average of $k_B T/2$ to the kinetic energy, where k_B is the Boltzmann constant and T the temperature. In general, the instantaneous temperature $\mathcal{T}$ is therefore given by:

$$\mathcal{T} = \frac{2\mathcal{K}}{k_B \cdot DOFs} \tag{3.26}$$

with $DOFs$ the total number of degrees of freedom in the system:

$$DOFs = 3N + 2N_{\mathrm{SRB}} + 3N_{\mathrm{NSRB}} - N_{\mathrm{C}} \tag{3.27}$$

N being the total number of sites, N_{SRB} the number of symmetric rigid bodies (two non-zero moments of inertia), N_{NSRB} the number of general, non-symmetric rigid bodies (three non-zero moments of inertia) and N_{C} the total number of constraints on the system (the centre of mass is typically constrained, so N_{C} is normally at least 3).

Mixtures Considering nS species $(1, 2, \ldots, i, \ldots, nS)$ each characterised by N_i sites each carrying f_i degrees of freedom, the kinetic energy is:

$$\mathcal{K} = \frac{k_B T}{2} \sum_{i=1}^{nS} (f_i \cdot N_i - N_{\mathrm{C}}) \tag{3.28}$$

hence the instantaneous temperature is:

$$\mathcal{T} = \frac{2\mathcal{K}}{k_B \sum_{i=1}^{nS} (f_i \cdot N_i - N_{\mathrm{C}})} \tag{3.29}$$

In particular then, considering a mixture of N_{PM} point-masses (3 degrees of freedom each) and N_{SRB} symmetric rigid bodies (5 degrees of freedom each), the total kinetic energy K is:

$$\mathcal{K} = \frac{k_B T}{2} (3 \cdot N_{\mathrm{PM}} + 5 \cdot N_{\mathrm{SRB}} - N_{\mathrm{C}}) \tag{3.30}$$

Hence the instantaneous temperature is:

$$\mathcal{T} = \frac{2\mathcal{K}}{k_B (3 \cdot N_{\mathrm{PM}} + 5 \cdot N_{\mathrm{SRB}} - N_{\mathrm{C}})} \tag{3.31}$$

In the general case where the system also comprises N_{NSRB} non-symmetric rigid bodies, the temperature becomes:

$$\mathcal{T} = \frac{2\mathcal{K}}{k_B (3 \cdot N_{\mathrm{PM}} + 5 \cdot N_{\mathrm{SRB}} + 6 \cdot N_{\mathrm{NSRB}} - N_{\mathrm{C}})} \tag{3.32}$$

3.8.3 Pressure

The pressure P can be defined in terms of the virial expression:

$$PV = N k_B T + \langle \mathcal{W} \rangle \tag{3.33}$$

with V the volume, N the number of sites, k_B the Boltzmann constant, T the temperature and $\mathcal{W}$ the 'internal virial'. For pair-potentials, $\mathcal{W}$ is expressed as:

$$\mathcal{W} = \frac{1}{3}\sum_i \sum_{j>i} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \tag{3.34}$$

so we finally obtain the definition of an instantaneous pressure $\mathcal{P}$:

$$\mathcal{P} = \rho k_B \mathcal{T} + \mathcal{W}/V = \rho k_B \mathcal{T} + \frac{1}{3V}\sum_i \sum_{j>i} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \tag{3.35}$$

with ρ the system's density.

Pressure tensor

The macroscopic pressure tensor of a system can be written as:

$$\mathbf{P} = \sum_i m_i \mathbf{v}_i \otimes \mathbf{v}_i / V - \mathbf{W} \tag{3.36}$$

where $\sum_i m_i \mathbf{v}_i \otimes \mathbf{v}_i/2$ is the *kinetic energy density tensor* and $\mathbf{W} = \sum_{i<j} \mathbf{f}_{ij} \otimes \mathbf{r}_{ij}/V$ is the *system's configurational stress tensor*. The off-diagonal elements of $\mathbf{P}$ vanish in equilibrium, and for an isotropic system the diagonal elements are expected to be equal. For an anisotropic system such as a lipid bilayer, the diagonal elements need not be equal, leading to a finite *surface tension*. In particular, assuming the surface to be parallel to the xy plane, and hence normal to the z axis, the surface tension γ is related to the pressure tensor by:

$$\gamma = L_z \times [P_{zz} - (P_{xx} + P_{yy})/2] \tag{3.37}$$

where L_z denotes the length of the simulation region normal to the surface, P_{zz} is the component of the pressure tensor normal to the surface and P_{xx}, P_{yy} are the tangential components.

3.9 Statistical analysis

The measurement process in MD must undergo rigorous statistical analysis to quantify the errors due to random fluctuations of the properties investigated, and hence to establish the significance of the results.[120,124] Besides, statistical parameters (such as the variance) are sometimes needed to calculate interesting properties. From a series of M measurements of a fluctuating property A in a system at equilibrium, the mean value is:

$$\langle \mathcal{A} \rangle = \frac{1}{M}\sum_{\mu=1}^{M} \mathcal{A}_\mu \tag{3.38}$$

and if each measurement $\mathcal{A}_\mu$ is independent, with variance

$$\sigma^2(\mathcal{A}) = \frac{1}{M}\sum_\mu (\mathcal{A}_\mu - \langle\mathcal{A}\rangle)^2 = \langle\mathcal{A}^2\rangle - \langle\mathcal{A}\rangle^2 \tag{3.39}$$

then the variance of the mean $\langle\mathcal{A}\rangle$ is:

$$\sigma^2(\langle\mathcal{A}\rangle) = \frac{1}{M}\sigma^2(\mathcal{A}) \tag{3.40}$$

and the estimated error in the mean is simply $\sigma(\langle\mathcal{A}\rangle)$. In MD simulation, the variance is underestimated because successive measurements are not independent, but (highly) correlated. Luckily, the simple method of *block averaging* can be used to tackle this issue. In particular, assuming the A_μ to be correlated, if averages are evaluated over blocks of successive values, then as the block length increases the block averages will be decreasingly correlated; eventually, once the block length exceeds the (unknown) longest correlation time present in the data, the block averages will be independent from a statistical point of view.

Standard deviation The standard or Root-Mean-Squared (RMSD) deviation $\sigma(\mathcal{A})$ is simply the square root of the variance:

$$\sigma(\mathcal{A}) = \sqrt{\frac{1}{M}\sum_{\mu=1}^{M}\mathcal{A}_\mu^2 - \left(\frac{1}{M}\sum_{\mu=1}^{M}\mathcal{A}_\mu\right)^2} \tag{3.41}$$

In general the standard deviation gives an indication of the spread of data: in most distributions, the bulk of the distribution lies within two standard deviations from the mean, i.e., within the interval $[\langle\mathcal{A}\rangle - 2\sigma(\mathcal{A}), \langle\mathcal{A}\rangle + 2\sigma(\mathcal{A})]$. The RMSD can also be employed to calculate physical properties, such as the specific heat or compressibility moduli.

3.10 Summary

Molecular dynamics is a powerful method to simulate matter at the molecular scale. The fundamental components of a molecular dynamics model are:

- a position-dependent potential, to describe the interaction (forces) between particles;
- an integrator, to evolve the system in time according to the forces experienced by each particle.

The main data generated by a simulation comprise the trajectories of every particle, which can be used to calculate dynamic and thermodynamic properties of the system.

Chapter 4

Coarse-grain modelling of lipid bilayers: literature background

Particle-based computer simulation is a powerful tool to study the behaviour of membranes at molecular resolution. Atomic-level models have been employed for decades now, and have given an understanding of many membrane phenomena. However, these studies are computationally very expensive, for an enormous amount of calculation has to be performed to model the interactions between all atoms in the system. This problem can be tackled by adopting simplified, *coarse-grain* descriptions, in which the number of interacting particles is significantly reduced. In this review, we summarise and discuss the most representative work reported in the literature concerning coarse-grain computer models of lipid bilayers. Every model is analysed in terms of the force-field employed, parameterisation procedure, and predictive power in relation to the corresponding experimental observables. We also highlight general advantages and drawbacks of the coarse-grain approach with respect to the traditional atomic-level methodology. The material presented in this chapter has been recently published elsewhere.[139] Other reviews of the coarse-grain field are also available in the literature.[140–145]

4.1 Introduction

Lipid bilayers are extremely complex systems, characterised by highly heterogeneous structure and dynamics; membrane-related phenomena take place on a wide range of spatial and temporal scales. Considering the hugely differing nature of the structures, interactions and phenomena present in lipid bilayers, it is clear that the development of realistic membrane models is a challenging task. The scope of this review is limited to the nanosecond to microsecond temporal scale, and to the nanometre spatial scale, as this is the realm of the models that we will consider. The nature of the modelling methodology further limits the treatment to *non-specific* membrane phenomena, in the sense that biochemical reactions will not be involved. It is important to note that these restrictions do not compromise in any way the possibility of obtaining a general understanding of bilayers: experiments have in-

deed shown that the most fundamental membrane properties depend exclusively on basic *physical* principles governing lipid structure and dynamics *at the molecular level.*[4] In particular, we will describe the two main *particle-based* approaches to the simulation of membranes by molecular dynamics. First, in Section 4.2, we summarise the basics of the traditional methodology, which involves an *atomistically-detailed* description of the system. Second, in Section 4.3, we describe the recently developed *coarse-grain* approach, which is based on a simplified representation: this is the main focus of this review, and therefore a number of representative coarse-grain models that have appeared in the literature will be described in some detail.

4.2 Atomic-level modelling

The traditional methodology employed to simulate lipid bilayers is based on an atomic-level (AL) representation: every atom of the system is explicitly represented as a point-mass. The inter-atomic interactions are described by molecular mechanics force-fields:[121] non-bonded interactions are modelled using Lennard-Jones and Coulomb potentials, whereas bonded terms are considered via harmonic, angle and dihedral potentials.

Over the past two decades, several AL models of lipid bilayers have been developed and validated on experimental structural and dynamic data.[27,146–152] More recently, AL membrane simulations have been used to study important biological phenomena, such as undulations,[153] self-diffusion,[29,31,32,34,154] electrostatic interactions,[36–38,155,156] cholesterol function,[41,157] the permeation of small molecules,[35,158] the lateral pressure distribution[39–45] and the appearance of transient ordered domains.[159] Despite their popularity, AL membrane models are affected by a limiting efficiency issue: obtaining data comparable to the experimental measurements is hugely time-consuming, due to the computational cost inherent in the simulation of every atom in the system. This drawback of the AL approach makes difficult the study of (relatively) large-scale phenomena, such as membrane fusion or lipid rafts. Furthermore, there are efficiency issues associated even with the standard (relatively small) membrane sizes. For instance, it has become clear that several tens of nanoseconds are required for some crucial parameters (such as the lipid area) to converge;[160] even with parallel computing, several weeks of computation may be needed just to equilibrate the AL system. Also, the measurement of some properties such as elastic moduli or the lateral pressures, typically require extended simulation times to obtain accurate and precise data. Another issue is related to the prohibitive cost of simulating the fundamental phenomenon of the self-assembly of lipids into a bilayer from a random solution. With the single exception of Marrink et al.,[28] all AL studies reported in the literature to date have been conducted on pre-assembled bilayers: these systems are therefore not guaranteed to be thermodynamically stable. A more general problem concerns the force-field reliability: AL lipid force-fields are still far from accurate, in terms of being able to *quantitatively* reproduce experimental data. For example, Benz et al.[161] conducted a thorough testing on the popular CHARMM and

GROMACS force-fields via constant-pressure simulation, and concluded that neither parameter set can capture within experimental error the experimentally determined structure of a DOPC bilayer in the fluid state. Furthermore, de Vries et al.[162] showed through constant-volume simulation of a standard hydrated DPPC bilayer that the GROMOS parameter set yields a total pressure of about -140 bar (the proper equilibrium value being 1 bar). A refinement of the current lipid force-fields could in principle be performed by trial-and-error adjustments of the parameters, although in practice the computational cost associated with this procedure may be too high.

4.3 Coarse-grain modelling

A possible solution to the AL efficiency problem involves the use of simpler, *coarse-grain* (CG) models. The general CG strategy involves grouping together selected clusters of atoms into single super-sites, to reduce the number of interactions calculated, and hence also the computational cost. A typical lipid molecule, which in reality comprises more than one hundred atoms, is thus typically reduced to a collection of only a dozen CG sites. The representation of water also undergoes a simplification process, that typically results in the replacement of groups of three/four water molecules by single CG units. To further improve efficiency, electrostatic interactions are highly simplified or completely absent.

In the following sections, we review the most representative coarse-grain models for biological membranes that have been reported in the literature, in chronological order. We focus on *specific* CG models, i.e., those models that retain a connection with specific chemical systems: such methods allow a direct comparison with the experimental results for the corresponding bilayer systems. However, a section is also devoted to a brief summary of "coarser", idealised models: in this case, there is no effort to reproduce any particular real system, the objective instead being to capture general membrane phenomena (see Section 4.3.5).

4.3.1 Smit and Groot models

In the early Nineties, Smit et al.[163,164] developed a seminal CG model for systems comprising oil, water and surfactant particles: simulations showed for the first time the spontaneous formation of micelles. Some years later, Groot and Warren[165] discussed the use of the dissipative particle dynamics (DPD) technique to simulate the dynamics of mesoscopic systems, and also proposed parameterisation methods. The DPD technique is a coarse-grain scheme by construction: the forces due to clusters of individual molecules are lumped together to yield effective friction and a fluctuating force between the interacting sites. In particular, beads interact pairwise via a combination of three contributions: a conservative-repulsive, a dissipative, and a random force.[122] On these bases, Venturoli and Smit[166] employed DPD to model single-chain surfactant bilayers, while Groot[167] simulated the spontaneous formation of surfactant micelles and the formation of polymer-surfactant mixtures.

In all these early studies however, the parameters had not been related to molecules of

specific chemistry. This issue was addressed by Groot and Rabone,[168] who developed a model for phosphatidylethanolamine (PE) membranes; they also included similar models of non-ionic surfactants. The parameterisation method of Groot and coworkers was subsequently used and extended by Smit and coworkers to develop a model of dimyristoylphosphatidylcholine (DMPC) lipids in water.[50]

Force-field and parameterisation

Smit et al.[164] modelled idealised surfactant molecules as chains of identical Lennard-Jones sites. Venturoli and Smit[166] employed different parameters to construct surfactants as linear molecules composed of one headgroup (hydrophilic) site and six to ten tail (hydrophobic) sites. Groot and Rabone[168] parameterised a DPD force-field to represent PE lipids. In particular, triplets of methyl groups were coarse-grained into individual tail beads, whereas the glycerol-headgroup region was described by three CG sites. Parameters were derived from compressibility and solubility data. Triplets of water molecules were collected into individual DPD solvent beads, the parameters being fitted to give the correct compressibility of water. Many subsequent DPD studies have employed this model of solvation.[168]

A model of DMPC consisting of three hydrophilic head beads and two tails, each consisting of five hydrophobic tail beads was implemented by Kranenburg et al.[50] The chain length of this model was varied by Kranenburg and Smit[169] to model the dilaureoylphosphatidylcholine (DLPC) and distearoylphosphatidylcholine (DSPC) lipid types.

A common feature of the Smit and Groot CG membrane force-fields is the complete lack of electrostatic interactions. While Groot[170] has recently extended the DPD methodology to incorporate electrostatics in simulations of a cationic polyelectrolyte solution, no biomembrane model has been developed to date.

Results

The very simple oil/water/surfactant model of Smit et al.[164] was able to (qualitatively) reproduce experimental observations such as micelle formation, density profiles and order parameters.

Venturoli and Smit[166] obtained for the first time the self-assembly of (model) surfactant bilayers via DPD simulations. Studies on the effect of changes in the chain length and stiffness of the single-tail surfactants on the properties of the model membranes were carried out. The lateral stress profile across the model bilayer was also computed: the distribution is qualitatively reasonable.

Groot and Rabone[168] made the first attempt to simulate a realistic biomembrane with DPD using a molecular-specific parameterisation: the resulting CG membrane structure matched that of corresponding AL simulations, and the essential experimental thermodynamics was reproduced. They also included non-ionic surfactants to investigate morphology changes and bilayer rupture processes.

Smit and coworkers worked on further tuning of Groot's parameter set. Kranenburg et al.[171] studied the phase behaviour of model membranes and were able to induce interdigitation; the self-assembly process was also simulated.[172] Kranenburg et al.[50,173] reproduced the experimental values of area per lipid and the hydrophobic thickness of a DMPC bilayer. The phase behaviour of bilayers comprising different lipid species was also characterised.[169]

Kranenburg et al.[174] also developed a DPD system composed of a DSPC bilayer incorporating model alcohol molecules, represented as idealised amphiphiles. The study reproduced the experimental phase diagrams, as well as the alcohol-induced interdigitated phase shown by experiments.[56] Smit's model has also been extended to study the structural changes resulting from the inclusion of a rod-like object taken as an idealised protein.[175]

4.3.2 Klein model

Klein and coworkers developed a model for simulating hydrated DMPC lipid bilayers which for the first time included an explicit, though incomplete, treatment of electrostatic interactions.[48]

Force-field and parameterisation

In the Klein CG model, the 118 atoms of a DMPC lipid are reduced to a 13-site model. The two choline and phosphate head-groups were assigned charges of $+e$ and $-e$, respectively: these are the only charges present in the force-field, and they interact via a dielectric constant set to $\epsilon_r = 78$. The potentials employed were systematically parameterised to mimic structural properties obtained from atomistic simulations (radial distribution functions) and experimental data (density, surface tension).

Klein et al. modelled water through spherically symmetric sites each representing a loose grouping of three water molecules; site-site interactions were described using a Lennard-Jones potential, the parameters being chosen to reproduce the correct density. The electrostatic properties of water were not included in the solvent model.[48]

Results

Shelley et al.[176] qualitatively reproduced the density profiles of a hydrated liquid-phase DMPC bilayer; simulations were conducted at constant-volume. Lopez et al.[52] further simulated Klein's model and studied the lipid lateral diffusion: the CG diffusion coefficient was about two orders of magnitude higher than the experimental measurement.

The Klein model has been extended to incorporate the anaesthetic halothane inside the DMPC bilayer.[57,176] A single site was used to represent each halothane molecule. The system studied was characterised by a 2:1 ratio of the phospholipid to halothane, equivalent to an atomistic simulation considered for comparison. After several adjustments of the parameters, the distribution of the halothane perpendicular to the membrane was brought into qualitative agreement with that found in the atomistic studies. Pickholz et al.[57] observed that increasing

the anaesthetic concentration resulted in an increase of the lipid area and order parameters and a decrease in the inter-lamellar spacing.

Srinivas and Klein[177] studied the interaction of a synthetic pore-promoting "hydraphile" molecule with the CG phospholipid bilayer; the system was simulated for 5 ns, during which the initially fully extended trans-membrane hydraphile adjusted its end-to-end distance to match the bilayer thickness.

Srinivas et al.[178] employed Klein's CG model to simulate diblock copolymer self-assembly. The Klein bilayer was also extended to incorporate a model nanotube.[140,179] In particular, the nanotube is modelled as a hydrophobic rod (made of sites identical to the lipid tail sites) capped at its termini with hydrophilic sites (identical to the water sites). Simulations were carried out in the NPT ensemble for several tens of nanoseconds to study the structure and dynamics of spontaneous insertion into the CG membrane.[180] First the nanotube fuses with a bilayer leaflet, then penetrates the interior while rotating to assume a transverse orientation.

Nielsen et al.[181] studied the lipid bilayer perturbations around the trans-membrane nanotube, focusing on the contact angle at the bilayer-nanotube interface and on the orientation of the lipid molecules in the vicinity of the inclusion.

Nielsen et al.[54] studied the trans-membrane peptide-induced lipid sorting (the phenomenon by which integral proteins attract the lipid type which better matches their hydrophobic surface) and the mechanism of L_α-to-inverted phase transition.

The spontaneous insertion of antimicrobial polymers has also been simulated: the average orientation of the antimicrobial molecules was found to be parallel to the bilayer plane.[182]

4.3.3 Marrink model

Marrink et al.[46] developed a CG model for lipid simulation which has become very popular due to its high efficiency, flexibility and simplicity.

Force-field and parameterisation

In the Marrink model, DPPC molecules were coarse-grained into 12 sites.[46] Water is represented by Lennard-Jones sites accounting for groups of four water particles. A trial and error procedure was used to optimise the parameters to reproduce the experimental densities of pure water and alkane systems around room temperature, the mutual solubility of oil and water, and the relative diffusion rates. Electrostatic interactions are only present between lipid headgroups, where they are treated through a Coulombic potential in a manner similar to the Klein model. A relative dielectric constant $\epsilon_r = 20$ is employed for explicit screening.[46]

Results

Marrink's model showed spontaneous bilayers formation: the final structures were consistent with corresponding AL results, in terms of density profiles and order parameters. A number of elastic parameters were computed: results for the area compressibility modulus, bending

rigidity and line tension were within an order of magnitude of the experimental data.[46] The calculation of the bending rigidity modulus proved particularly demanding: the simulation of undulatory modes (from which the bending rigidity can be extracted) required the simulation of a bilayer patch comprising 6400 lipids for 250 ns. The lipid lateral diffusion coefficient was also computed: it was found to be about four times larger than the experimental data.[46]

The phase transformation into the gel phase was simulated by Marrink et al.:[183] the observed drop in lateral mobility by two orders of magnitude is consistent with experiments.

The Marrink CG model has been used to simulate a variety of phenomena and extended-systems, such as vesicle formation[184] and fusion,[55] and mixed-lipid systems.[185]

By increasing the temperature or reducing the hydration level, Marrink et al.[53] simulated the complete transition pathway from a multi-lamellar to an inverted hexagonal phase: stalk intermediates were identified, in agreement with experimental observations.

Faller and Marrink[186] studied binary mixtures of two phosphatidylcholines of different chain lengths: the experimental phase behaviour was qualitatively reproduced. Shi and Voth[187] also employed Marrink's methodology to simulate a binary mixed system at the liquid-gel phase coexistence condition.

Dickey and Faller[188] studied the interaction of Marrink's bilayer with an alcohol. The Marrink model has also been extended to contain CG representations of polyamidomine dendrimers[189] and membrane proteins.[59,190,191]

Adhangale and Gaver[192] employed the Marrink model to study a DPPC monolayer at the air/water interface.

May et al.[193] adopted Marrink's CG approach to model phosphatidylinositol-4-phosphate (PI4P) lipid: the elastic properties of mixed DPPC/PI4P bilayers were investigated.

The Marrink model has been recently revised and applied to the study of the lateral pressure profile in a DPPC bilayer and to the simulation of a DPPC/cholesterol system.[194] However, the reported CG pressure profile[194] displays magnitudes that are markedly different from the other published AL results[39,43,44] for DPPC bilayers.

4.3.4 Voth model

Izvekov and Voth[51] developed a CG model for hydrated DMPC bilayers using a multi-scale approach in which explicit atomistic forces are propagated in scale to the coarse-grained level.

Force-field and parameterisation

The Voth model relies on a force-field which is obtained from a corresponding AL simulation via a so-called "force-matching" procedure. In particular, the force-matching yields potentials of mean force (PMFs) that are fitted using a spline interpolation of the AL forces, where atoms are grouped according to the chosen CG subdivision. This method is unique in the CG field, as it is not dependent on the matching of selected thermodynamic data, but it makes use of the calculated atomic forces from an underlying AL model.

Izvekov and Voth[51] coarse-grained each DMPC lipid to a 13-site model, in a way similar to the Klein and Marrink models. Water molecules are represented explicitly through individual CG sites. Electrostatic interactions are not present in this CG force-field.

Results

A pre-assembled membrane bilayer comprising 64 DMPC molecules solvated by 1312 water sites was simulated in the NVT ensemble. Radial distribution functions and density profiles of the underlying AL model were reproduced.[51]

The "force-matching" procedure has subsequently been applied to the study of a DMPC-cholesterol mixture simulated at constant temperature and pressure.[58] CG order parameters were consistent with the underlying AL data. CG diffusion coefficients turned out to be about four times higher than the corresponding experimental measurements.

Shi et al.[61] constructed a mixed AL-CG model of a membrane-protein system: in particular, an AL model of the gramicidin A ion channel was embedded in hydrated CG lipid bilayer. The system was simulated at constant temperature and volume for 10 ns; the radial distribution functions of the AL simulation employed in the "force-matching" parameterisation were reproduced.

4.3.5 Idealised membrane models

This section briefly presents CG models that are more phenomenological and simplistic then the ones previously summarised: bilayers are composed of amphiphiles with no specific chemical identity. These models are sometimes called "bead-and-spring" models, as the interactions represented often reduce to Lennard-Jones and harmonic potentials only. Electrostatic interactions are indeed typically not included. In some cases, even the solvent is sacrificed (see following paragraph "Implicit-solvent models").

Lipowsky model

In the late Nineties, Lipowsky and coworkers developed a simple, idealised CG bilayer model that proved capable of qualitatively capturing a number of fundamental membrane characteristics.[195] In particular, solvated aggregates of surfactant molecules were obtained employing only two types of Lennard-Jones sites: hydrophilic sites, used to describe both solvent and surfactant head-group particles, and hydrophobic sites, employed to model surfactant tail segments. Simulations of the model allowed bilayer self-aggregation, diffusion, interfacial tension and area compressibility to be studied. The trans-bilayer lateral pressure profile was also calculated: the distribution is qualitatively reasonable, apart from the unphysical negative pressure peak at the bilayer mid-plane. The Lipowsky model was further investigated in terms of undulations: in particular, from the spectral analysis of the bilayer shape fluctuations, Goetz et al.[196] were able to extract a value for the bending rigidity modulus. Imparato et al.[197] extended the model to simulate a mixture of two different types of

molecules: lateral diffusion was measured, and the activation barrier of flip-flop processes was estimated. Imparato et al.[198] investigated further the two-component membrane to study shape fluctuations and elastic properties. The Lipowsky model was employed by den Otter et al.[199] to study the bilayer thermally-induced undulations and to ultimately extract the bending rigidity using different methods: undulatory modes either arose naturally during equilibrium simulations, or were imposed through a number of non-equilibrium methods. The results of this work have proved consistent with the bending free energy prediction from the Helfrich theory.[77,97,98] Boek et al.[200] performed additional simulations on Lipowsky's model to analyse structure factors: they propose refining the intermolecular potential parameters to yield fluctuation spectra that coincide with the atomistic results.

Other bead-and-spring models

Loison et al.[201] proposed a model where amphiphiles are represented by linear tetramers, each composed of two solvophobic (tail) and two solvophilic (head) beads. A self-assembled bulk lamellar phase was investigated from a mechanical perspective. Pores, fluctuations and defects of the lamellar stack were also analysed.[202,203]

Stevens[204] developed a CG model where each lipid is composed of eleven beads, four in each aliphatic tail and three for the headgroup; the solvent is represented by particles equivalent to the headgroup sites. Stevens' model system spontaneously self-assembled into a lamellar bilayer; measurements of lipid lateral diffusion and material elastic parameters yielded qualitatively reasonable values.[204] Fusion simulations have also been performed: at the contact edge between liposomes, lipids were observed to splay their tails into the opposing leaflets, thus progressively producing a new hydrophobic core (a "stalk") that eventually opened to complete the fusion process.[205]

Lenz and Schmid[206] presented a simple CG bead-spring model with the aim of reproducing the main (gel-liquid) phase transition of biomembranes. Lipids are modelled as single-tail amphiphiles, each composed of six tail beads and one slightly larger head bead. Molecules are defined via bond and bond-angle harmonic potential; non-bonded interactions are taken into account via Lennard-Jones potentials. Solvent was represented by explicit "phantom" solvent beads: they behave exactly like head beads, except that they do not interact with each other. The model proved well suited for simulating lipid bilayers in the regime of the liquid-gel transition.

Implicit-solvent models

In this section we briefly consider a number of CG membrane models that have been designed to work without the presence of explicit solvent (for a thorough review, see Brannigan et al.[142]).

Although water is generally regarded as an essential component of bilayer systems, in specific cases it may not play a fundamental role while still representing a computational

bottle-neck. Hence, solvent-free models might be useful to gain insight into selected membrane phenomena at a very low computational cost.

Whitehead et al.[207] developed a CG model for the bilayer hydrocarbon region only, based on the Gay-Berne model of liquid crystals:[125] in the absence of water and headgroup, the lipid packing was maintained through the use of a restraint potential. The experimental phase behaviour of typical phospholipid bilayers was reproduced.

Noguchi and Takasu[208] represented lipids as rigid trimers consisting of one hydrophilic and two hydrophobic units interacting through multi-body potentials. This model has been used to simulate the self-assembly[208] and fusion[209] of vesicles, adhesion of nanoparticles to vesicles[210] and pulled vesicles.[211]

Lyubartsev[47] developed a 10-site lipid model parameterised exclusively from corresponding AL structural data processed through the inverse Monte Carlo method.[212] Several simulations of the CG model obtained were performed, both within Monte Carlo and molecular dynamics simulations, such as a periodic sample of lipid molecules ordered in a bilayer, a free sheet of such a bilayer without periodic boundary conditions, formation of vesicle from a plain membrane and self-assembly of lipids.[47]

Brannigan and Brown[213] developed a model of bilayers where entire lipids are represented by single soft spherocylinder. Through the combination of three simple pair potentials, a rich assortment of self-assembled phases was recovered, including micelles, fluid bilayers and gel-like bilayers. Brannigan et al.[214] extended the model to allow for variations in lipid length and simulations under constant surface tension conditions.

Brannigan et al.[215] developed a new model with flexible lipids that was employed to extract compressibility and bending moduli, and the lateral pressure profile. The pressure distribution is qualitatively reasonable apart from an unrealistic pronounced pressure trough located at the centre of the bilayer; interestingly, the curve is very similar to the result from the explicit-solvent model by Goetz and Lipowsky.[195]

The model by Brannigan et al.[215] was extended to incorporate multiple lipid species: in particular, the elasticity of idealised heterogeneous bilayers has been analysed.[216]

Cooke, Kremer and Deserno[217] developed a model where each lipid comprises three connected beads; bilayer assembly is triggered by effective tail attractions. The model reproduces key elastic properties,[218] and it has been recently applied to study how curvature-mediated interactions can lead to aggregation and vesiculation of idealised membrane proteins.[219]

4.4 Summary and conclusion

The traditional methodology for simulating biomembranes involves an atomic-level description of lipids and hydrating water. This approach has yielded fairly precise and accurate predictions of a number of experimental data. However, the associated large computational cost results in two major issues:

- collective, large-scale phenomena such as membrane fusion or micro-domain formation (lipid rafts) are typically intractable;
- the reliability of the calculation of some important properties, such as the lipid area or the lateral pressure profile, is often undermined by insufficient sampling.

Coarse-grain modelling techniques have shown a number of advantages over atomistic models:

- orders of magnitude more efficient, resulting in the possibility to study phenomena characterised by much larger temporal and spatial scales;
- enhanced sampling, resulting in statistically-reliable measurements;
- bilayers are usually self-assembled, and hence the simulated systems are more likely to be at thermodynamic equilibrium than atomic-level bilayer systems (which are normally preassembled).

On the other hand, CG models are also affected by issues:

- limited force-field transferability;
- experimental data are typically captured at a qualitative level only;
- the over-simplification or complete lack of electrostatics precludes the proper representation of the membrane potential, which has indeed never been computed for CG models;
- hydration is achieved through physically-questionable water models, which do not include any explicit electrostatic feature;
- no model published to date has proved able to reasonably quantify the trans-bilayer lateral pressure profile, which is crucial for the accurate modelling of many membrane processes.

There is a clear need, therefore, for CG models to be developed that incorporate electrostatic interactions more accurately. Such models will likely increase the range of membrane phenomena that may be accurately studied using the CG approach. Considering that the CG field is still in its infancy, the preliminary results obtained so far are encouraging. In the future, we foresee that coarse-grain models will play an increasingly important role in the understanding of lipid membranes.

Chapter 5

A quantitative coarse-grain model for lipid bilayers

This chapter describes the development of our new simplified particle-based computer model for hydrated phospholipid bilayers and its application to the quantitative prediction of the major physical features of fluid-phase biomembranes. Our method is based on the fundamental simplification idea present in all coarse-grain models previously developed: the representation of hydrated lipid systems is simplified by substantially reducing the number of interacting particles. However, we have introduced a number of advantageous features compared with the existing methods. First, electrostatics are explicitly incorporated. Second, water is realistically described, on an individual level. Third, hydrocarbon tails are accurately modelled using the anisotropic Gay-Berne potential. These characteristics provide for a physically realistic model capable of quantitative prediction; they also allow the multiscale mixing with standard atomic-level models to be straightforwardly implemented. Most of the material presented here has been recently published.[220]

5.1 Construction and parameterisation

5.1.1 Lipid model

We have designed a CG model for dimyristoylphosphatidylcholine (DMPC): Figure 5.1 shows both the AL structure and our simplified CG representation. Each lipid molecule, in reality comprising more than one hundred atoms, has been reduced to ten macrounits. The lipid headgroup is coarse-grained into two Lennard-Jones spherical units, accounting for the choline and phosphate moieties, respectively. Headgroup electrostatics are represented by a positive point-charge embedded in the choline group and a negative one in the phosphate group. The glycerol and hydrocarbon regions are modelled by soft uniaxial ellipsoids through the Gay-Berne potential.[125] The Gay-Berne potential can be seen as an extension of the (isotropic) Lennard-Jones potential, where extra terms are included to allow the modelling of non-spherical (anisotropic) particles. In particular, the glycerol-ester region is described

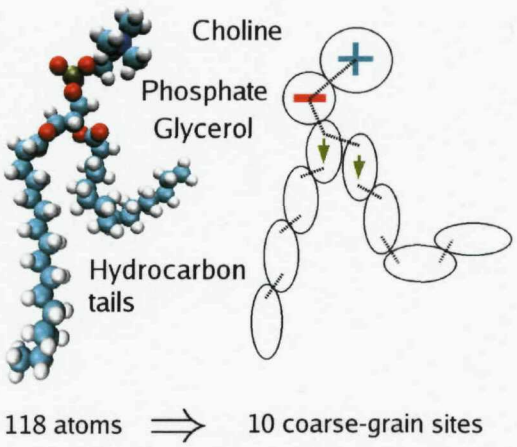


Figure 5.1: Lipid coarse-graining strategy. The left molecule is an all-atom representation of a DMPC lipid. The corresponding CG model is depicted on the right. CG electrostatics are highlighted: they comprise positive (*"+" sign*) and negative (*"-" sign*) point-charges, and point-dipoles (*arrows*). Harmonic springs, representing CG covalent bonding, are also shown (*dotted lines*).

by two Gay-Berne ellipsoidal units, each embedded with a point-dipole to account for the dipolar charge distribution in this region. Hydrocarbon tails are modelled by chains of three neutral Gay-Berne ellipsoids, each representing a segment of four consecutive methyl groups. The shape of the Gay-Berne ellipsoids can be tuned to capture accurately the underlying real elongated structure of the tail segments considered. The Gay-Berne representation of lipid tails has already been used to simulate an idealised bilayer, without headgroups and solvation: despite the simplicity of that model, order parameters and diffusion coefficients proved consistent with experiment.[207] Mixed Lennard-Jones/Gay-Berne interactions (between headgroup and tail sites) are consistently treated using the generalised Gay-Berne potential.[221] Intra-lipid bonds are modelled by the Hooke (harmonic) potential, as is standard practice. No angle or torsional potentials are present.

5.1.2 The Gay-Berne potential

The Gay-Berne potential[125] is a sort of generalised, anisotropic version of the simple Lennard-Jones potential. In fact, the Gay-Berne pair potential has the same distance dependence as the Lennard-Jones potential, but with a contact distance and a well depth which depend on the particles' mutual orientation. Considering two particles with principal axes identified by the unit vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_j$, and $\mathbf{r} \equiv r\hat{\mathbf{r}}$ the vector joining the centres of mass (Figure 5.2), the Gay-Berne potential $\mathcal{U}$ can be written as:

$$\mathcal{U}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \mathbf{r}) = 4\epsilon(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) \left[\left(\frac{\sigma_{\mathrm{GB}}}{r - \sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) + \sigma_{\mathrm{GB}}} \right)^{12} - \left(\frac{\sigma_{\mathrm{GB}}}{r - \sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) + \sigma_{\mathrm{GB}}} \right)^{6} \right] \tag{5.1}$$

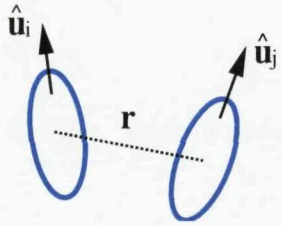


Figure 5.2: Pair of Gay-Berne particles.

Figure 5.3: Representative configurations of a pair of Gay-Berne particles. From left to right: end-to-end **e**, side-by-side **s**, cross **x** and tee **t**.

In explaining the various terms of Equation 5.1, we will sometimes refer to the representative pair configurations depicted in Figure 5.3. The range parameter $\sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}})$ is:

$$\sigma(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) = \sigma_{\mathrm{GB}} \left\{ 1 - \frac{\chi}{2} \left[\frac{(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}} + \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}})^2}{1 + \chi\, \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j} + \frac{(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}} - \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}})^2}{1 - \chi\, \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j} \right] \right\}^{-\frac{1}{2}}, \tag{5.2}$$

where σ_{GB} is the contact distance when the molecules are in either the side-by-side or the cross configuration. The shape anisotropy is controlled by the parameter χ, which is defined as:

$$\chi = \frac{(\sigma_{ee}/\sigma_{\mathrm{GB}})^2 - 1}{(\sigma_{ee}/\sigma_{\mathrm{GB}})^2 + 1} = \frac{\kappa^2 - 1}{\kappa^2 + 1}, \tag{5.3}$$

with σ_{ee} representing the contact separation distance for the end-to-end configuration, and $\kappa = \sigma_{ee}/\sigma_{\mathrm{GB}}$. The parameter κ represents the length-to-breadth ratio of the particles (values of $\kappa > 1$ are appropriate for prolate, that is, rod-like, molecules and values $\kappa < 1$ correspond to oblate, that is, disk-like, molecules). The well depth energy anisotropy function $\epsilon(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}})$, is:

$$\epsilon(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) = \epsilon_{\mathrm{GB}} \epsilon_1^{\nu}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) \epsilon_2^{\mu}(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}), \tag{5.4}$$

with ϵ_{GB} the depth of the attractive well (potential minimum) when the particles are arranged in the cross configuration. The exponents μ and ν are adjustable parameters. The remaining functions are:

$$\epsilon_1(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = [1 - \chi^2 (\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j)^2]^{-1/2}, \tag{5.5}$$

and:

$$\epsilon_2(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j, \hat{\mathbf{r}}) = 1 - \frac{\chi'}{2} \left[\frac{(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}} + \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}})^2}{1 + \chi'\, \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j} + \frac{(\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}} - \hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}})^2}{1 - \chi'\, \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j} \right]. \tag{5.6}$$

The parameter χ' controls the energy anisotropy, and can be written as:

$$\chi' = \frac{(\epsilon_{ss}/\epsilon_{ee})^{\frac{1}{\mu}} - 1}{(\epsilon_{ss}/\epsilon_{ee})^{\frac{1}{\mu}} + 1} = \frac{\kappa'^{1/\mu} - 1}{\kappa'^{1/\mu} + 1}, \tag{5.7}$$

where ϵ_{ss} is the potential minimum for a pair of particles placed side-by-side, ϵ_{ee} is the potential minimum for particles placed end-to-end, and $\kappa' = \epsilon_{ss}/\epsilon_{ee}$ is simply the ratio of the two characteristic well depths. Overall, the Gay-Berne potential is usually characterised by the four parameters κ, κ', μ and ν.

Generalised Lennard-Jones/Gay-Berne interaction

If one of the two interacting particles is spherical, that is, a Lennard-Jones site, whereas the other one is a Gay-Berne site, we need a special set of equations to model the interaction.[221] First, we define σ_{LJ} as the diameter of the spherical particle, and σ_{GB} as the breadth of the ellipsoidal particle. We then define the mixed range constant σ_{GBLJ} using the standard mixing rule as:

$$\sigma_{\mathrm{GBLJ}} = (\sigma_{\mathrm{LJ}} + \sigma_{\mathrm{GB}})/2$$

The shape anisotropy constant χ_{GBLJ} can be defined as:

$$\chi_{\mathrm{GBLJ}} = \frac{\sigma_{\mathrm{GB}}^2(\kappa^2 - 1)}{(\kappa\sigma_{\mathrm{GB}})^2 + \sigma_{\mathrm{LJ}}^2} \tag{5.8}$$

The range parameter for this mixed Lennard-Jones/Gay-Berne interaction can now be written as:[222]

$$\sigma(\hat{\mathbf{r}}_{ij}, \hat{\mathbf{u}}_j) = \sigma_{\mathrm{GBLJ}}[1 - \chi_{\mathrm{GBLJ}}(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_j)^2]^{-1/2}, \tag{5.9}$$

with $\hat{\mathbf{r}}_{ij}$ a unit vector parallel to the distance between the mass centres of the two sites and $\hat{\mathbf{u}}_j$ the unit vector defining the orientation of the Gay-Berne site.
The energy function is expressed as:

$$\epsilon(\hat{\mathbf{r}}, \hat{\mathbf{u}}_j) = \epsilon_{\mathrm{GBLJ}}[1 - \chi'_{\mathrm{GBLJ}}(\hat{\mathbf{r}} \cdot \hat{\mathbf{u}}_j)^2]^{\mu} \tag{5.10}$$

where ϵ_{GBLJ} represents the potential minimum for the side-by-side arrangement, which in this case of sphere-ellipsoid pair is also equivalent to the cross arrangement.[221] We set the value of ϵ_{GBLJ} by using the standard mixing rule:

$$\epsilon_{\mathrm{GBLJ}} = \sqrt{\epsilon_{\mathrm{GB}} \times \epsilon_{\mathrm{LJ}}}$$

being ϵ_{GB} and ϵ_{LJ} the potential minima of the two interacting sites. The energy anisotropy constant χ'_{GBLJ} is defined as:[221]

$$\chi'_{\mathrm{GBLJ}} = 1 - \left(\frac{1}{\kappa'_{\mathrm{GBLJ}}}\right)^{1/\mu} \tag{5.11}$$

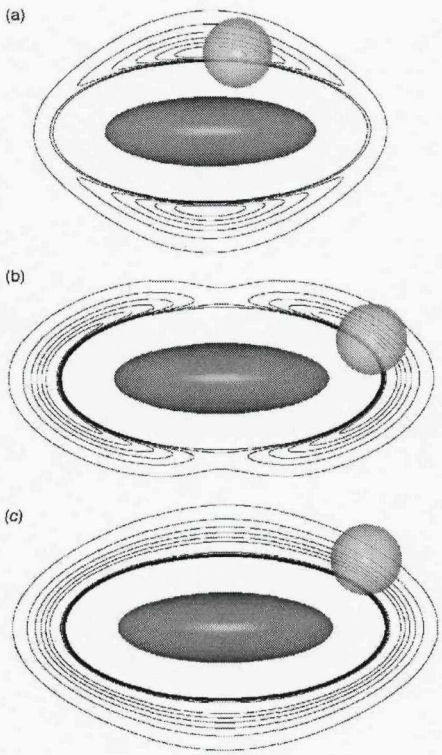


Figure 5.4: Equipotential lines of the Lennard-Jones/Gay-Berne potential with different values of the ratio $\kappa'_{\mathrm{GBLJ}} = \epsilon_{\mathrm{GBLJ}}/\epsilon_E$: (a) $\epsilon_{\mathrm{GBLJ}}/\epsilon_E = 5$, (b) $\epsilon_{\mathrm{GBLJ}}/\epsilon_E = 1/5$, (c) $\epsilon_{\mathrm{GBLJ}}/\epsilon_E = 1$. From Antypov and Cleaver.[223]

with $\kappa'_{\mathrm{GBLJ}} = \epsilon_{\mathrm{GBLJ}}/\epsilon_E$, the parameter ϵ_E representing the potential minimum of the end-to-end arrangement. The parameter κ'_{GBLJ} can be set to control the interaction anisotropy. In particular, as it is exemplified in Figure 5.4, three possibilities exist:

- for $\kappa'_{\mathrm{GBLJ}} > 1$, the side-by-side arrangement is favoured, that is, the spherical site favours the sides of the ellipsoidal site (Figure 5.4 a);
- for $\kappa'_{\mathrm{GBLJ}} < 1$, the end-to-end arrangement is favoured, that is, the spherical site favours the ends of the ellipsoidal site (Figure 5.4 b);
- for $\kappa'_{\mathrm{GBLJ}} = 1$, no preferential mutual orientation is favoured (Figure 5.4 c).

The well-depth anisotropy function can now be written as:

$$\epsilon(\hat{\mathbf{u}}_j, \hat{\mathbf{r}}_{ij}) = \epsilon_{\mathrm{GBLJ}}[1 - \chi'_{\mathrm{GBLJ}}(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_j)^2]^{\mu} \tag{5.12}$$

Explicit formulae of derived forces and torques for both the original Gay-Berne model and its generalised version are reported in the literature.[224]

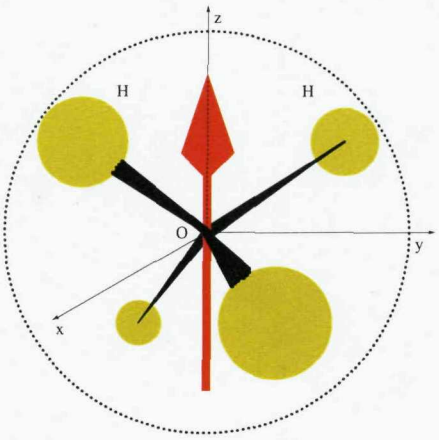


Figure 5.5: Schematic of the SSD model. The coordinate system is centred on the oxygen atom; the origin is in fact the centre of a soft LJ sphere, a point dipole (red arrow), and a tetrahedral octopolar sticky potential (dotted yellow patches).

5.1.3 Water model

Water molecules are represented by the soft sticky dipole (SSD) model.[225] The SSD water is a single-site model: the three atoms of individual water molecules are coarse-grained into a single interaction centre, which comprises a Lennard-Jones core providing excluded-volume, a point-dipole to account for electrostatics, and a tetrahedral "sticky" term to model hydrogen bonding (Figure 5.5). The total interaction energy between two SSD water molecules i and j is:

$$\mathcal{V}_{ij} = \mathcal{V}_{ij}^{\mathrm{LJ}}(r_{ij}) + \mathcal{V}_{ij}^{\mathrm{dp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) + \mathcal{V}_{ij}^{\mathrm{sp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) \tag{5.13}$$

where r_{ij} is the distance and $\mathbf{r}_{ij}$ the separation vector between the molecular centres, and Ω_i, Ω_j define the orientation of the two water molecules. The first term in Equation 5.13 is a standard Lennard-Jones term:

$$\mathcal{V}_{ij}^{\mathrm{LJ}}(r_{ij}) = 4\epsilon_w \left[\left(\frac{\sigma_w}{r_{ij}} \right)^{12} - \left(\frac{\sigma_w}{r_{ij}} \right)^{6} \right] \tag{5.14}$$

with σ_w the contact distance and ϵ_w the energy well depth. The second term is the dipole-dipole potential:

$$\mathcal{V}_{ij}^{\mathrm{dp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = \frac{1}{4\pi\varepsilon_0} \left[\frac{\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j}{r_{ij}^3} - \frac{3(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} \right] \tag{5.15}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are the dipole moment vectors of molecules i and j and ε_0 is the electric permittivity of vacuum. The third term in Equation 5.13 is the tetrahedral sticky potential:

$$\mathcal{V}_{ij}^{\mathrm{sp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = v_0 \left[s(r_{ij}) w(\mathbf{r}_{ij}, \Omega_i, \Omega_j) + s'(r_{ij}) w'(\theta_{ij}) \right] / 2 \tag{5.16}$$

where v_0 determines the strength of the sticky potential. The function w is the proper attractive (sticky) potential with tetrahedral geometry:

$$w(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = \sin\theta_{ij} \sin 2\theta_{ij} \cos 2\varphi_{ij} + \sin\theta_{ji} \sin 2\theta_{ji} \cos 2\varphi_{ji} \qquad (5.17)$$

in which $(\theta_{ij}, \varphi_{ij})$ is the set of spherical polar angles of the position of molecule j in the frame fixed on molecule i and with an orientation such that the z axis is parallel to the dipole moment of molecule i and the x axis is perpendicular to the molecular plane. The function w can also be expressed as:[226]

$$w(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = \frac{2}{r_{ij}^3}[(x_{ij}^2 - y_{ij}^2)z_{ij} + (x_{ji}^2 - y_{ji}^2)z_{ji}] \qquad (5.18)$$

with x_{ij}, y_{ij}, z_{ij} the x, y, z coordinates of the position of molecule j in the frame fixed on molecule i, and vice versa for the x_{ji}, y_{ji}, z_{ji} coordinates. The term $w'(\theta_{ij})$ is an empirical correction that counters the normal aligned and anti-aligned structures favoured by point dipoles:

$$w'(\theta_{ij}) = (\cos\theta_{ij} - 0.6)^2 (\cos\theta_{ij} + 0.8)^2 + (\cos\theta_{ji} - 0.6)^2 (\cos\theta_{ji} + 0.8)^2 - 2w^\circ \qquad (5.19)$$

Finally, $s(r_{ij})$ and $s'(r_{ij})$ are modulating cubic functions that switch off the sticky interaction beyond the first solvation shell:

$$s(r) = \begin{cases} 1 & \text{if} \quad r < r_L \\ (r_U - r)^2(r_U + 2r - 3r_L)/(r_U - r_L)^3 & \text{if} \quad r_L \leqslant r < r_U \\ 0 & \text{if} \quad r \geqslant r_U \end{cases}$$

$$s'(r) = \begin{cases} 1 & \text{if} \quad r < r_{L'} \\ (r_{U'} - r)^2(r_{U'} + 2r - 3r_{L'})/(r_{U'} - r_{L'})^3 & \text{if} \quad r_{L'} \leqslant r < r_{U'} \\ 0 & \text{if} \quad r \geqslant r_{U'} \end{cases}$$

where $r = |\mathbf{r}_{ij}| = |\mathbf{r}_{ji}|$. The SSD forces and torques derived from the interaction potential are reported in the literature.[226]

The SSD model was originally parameterised on the assumption that long-range electrostatic interactions are considered through Ewald methods.[225] However, a different parameter set has later been proposed to use with either a reaction field or with a simple cutoff.[227] Interestingly, at biologically relevant temperatures the simulations carried out with the simple cutoff parameterisation gave a density and a diffusion constant closer to the experimental values than the ones from the original SSD parameterisation.[227] The two different parameters settings are reported in Table 5.1, and Figure 5.6 offers a visual comparison.

Parameterisation	Liu and Ichiye[225]	Fennell and Gezelter[227]
μ (D)	2.350	2.420
σ/Å	3.051	3.035
ϵ/kcal mol^{-1}	0.152	0.152
v_0/kcal mol^{-1}	3.738	3.900
r_L /Å	2.750	2.400
r_U /Å	3.350	3.800
r'_L /Å	2.750	2.750
r'_U /Å	4.000	3.350
$r_c^{dipoles}$ /Å	-	9.000
$r_{switch}^{dipoles}$ /Å	-	7.650

Table 5.1: SSD parameterisations. The second and third columns report respectively the original parameterisation[225] and the adjusted model[227] to be used with a cutoff scheme for long-range electrostatics.

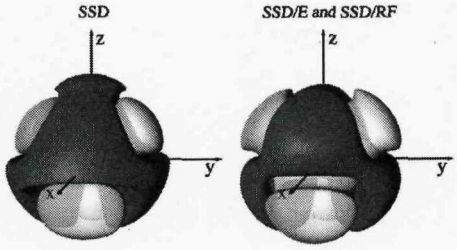


Figure 5.6: Potential isosurfaces of the sticky potential. On the left it is shown the parameterisation of Liu and Ichiye,[225] and on the right that by Fennell and Gezelter.[227] Light areas correspond to the tetrahedral attractive component, and darker areas correspond to the dipolar repulsive component. Figure from Fennell and Gezelter.[227]

The SSD model is about one order of magnitude computationally cheaper than the traditional AL multi-site water models, (such as the popular three-site models SPC,[228] SPC/E[229] and TIP3P[230]); this comes at no surprise, considering that the computational efficiency of a molecular model is approximately proportional to $1/n^2$, n being the number of sites in the molecule (e.g. $n_{SPC} = 3$, $n_{SSD} = 1$). Yet the SSD model accurately reproduces structural, thermodynamic, dielectric, dynamic, and temperature-dependent properties of liquid water.[225–227,231]

The Lennard-Jones term of the SSD potential interacts with the Gay-Berne lipid terms (tail and glycerol sites) through the generalised Gay-Berne potential.[221]

5.1.4 Treatment of electrostatic interactions

All electrostatic terms in our model interact with each other, through either charge-charge, charge-dipole or dipole-dipole potentials.[126] A relative dielectric constant $\epsilon_r = 1$ is assumed, that is, no artificial explicit screening is introduced. Long-range electrostatics are treated using site-based cutoff schemes: a charged site interacts with all and only the other charged

sites separated by a distance less than the cutoff.* In particular, charge-charge and charge-dipole interactions are implemented using the shifted-force cutoff method.[120] We employ the SSD parameters optimised to treat long-range dipole-dipole interactions with a cubic switching cutoff scheme:[227] for consistency, all dipole-dipole interactions are treated in this manner. Explicit formulae for the electrostatic potentials employed are reported in Appendix B, along with the derivation of forces and torques corresponding to the cutoff schemes adopted.

We are aware that using cutoffs to approximate long-range electrostatics might introduce simulation artifacts. In AL simulation, long-range electrostatic interactions are typically included by Ewald techniques,[232] which however are also known to introduce artifacts.[152,233–235] In fact, it has been argued that cutoff schemes can be as good as[152] or better[235] than Ewald methods. We have chosen the cutoff alternative as its simplicity and efficiency is consistent with the overall spirit of simplification of the model.

5.1.5 Simulation details

To study our model by molecular dynamics, we have specifically developed the software BRAHMS. The basic structure of BRAHMS has been designed following Rapaport.[124] Interactions are computed using a combined cell-subdivision/neighbour-list algorithm, which relies on standard periodic-boundary and minimum-image conventions. Dedicated routines have been implemented for the calculation of energies, forces and torques, the integration of rigid-body dynamics, the control of pressure and temperature, as well as for the analysis of the trajectory. The translational motion of all particles is described by Newton's second law. Lennard-Jones particles (lipid headgroups) are represented as point masses, their position being defined by the coordinates of the mass centres, as is standard practice. Gay-Berne particles (lipid tails) are represented as symmetric rigid bodies, whereas SSD molecules (water) as general, nonsymmetric rigid bodies: the rotational motion is described by Euler's equation, the orientations being represented with rotation matrices. To numerically integrate the equations of motion, we have implemented an advanced symplectic and time-reversible method,[132] previously described in Section 3.5. Implementation details of BRAHMS are reported in Appendix A.

The initial membrane configuration was constructed from a set of AL coordinates for hydrated DMPC.[236] Following our CG strategy (Figure 5.1), groups of atoms were mapped into CG interaction sites as depicted in Figure 5.7. The number of water molecules was adjusted to 26.6 water/lipid;† this ratio is consistent with that reported in a recent experi-

*This is different from group-based cutoff schemes, often employed to collect together charged sites into neutral groups. For instance, the interaction between pairs of atomistically-modelled water molecules is typically treated as group-based, meaning that all or none the atoms of the water molecule pair are taken into account for the interaction calculation. This is done in general to avoid the risk of creating artificial charges by cutting the interaction across molecular models comprising a distribution of (partial) charges. For the SSD model, since water electrostatics in each molecules are represented by single point-dipoles, there is no such risk.

†This corresponds to a hydration level of ≈ 42 wt% water content, indicating a condition of full hydration (see also Figure 2.5). The hydration level H can be readily computed as $H = 100 \times weight_{water}/weight_{total} \approx 42\%$, considering a molar weight of 18 amu/water and 648 amu/DMPC.

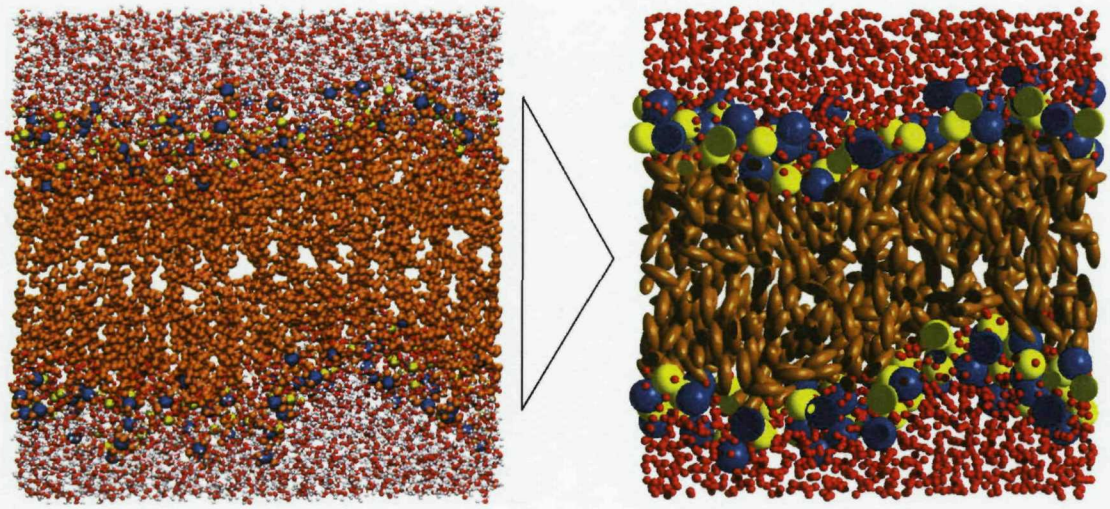


Figure 5.7: Coarse-grain mapping from atomistic representation.

mental characterisation of the structure of fully hydrated DMPC lipid bilayers.[6] We will in fact parameterise our model to fit the basic structural measurements obtained in that investigation.[6] Our final model system comprises 128 DMPC lipids and 3400 water molecules.

Coordinates are defined in a Cartesian frame with the origin in the centre of the simulation region: the x and y axes lie parallel to the bilayer interfacial plane, the z axis thus being perpendicular to it. The z axis will be also referred to as (interfacial) "normal". Molecular dynamics simulations are carried out with an integration time-step of 20 fs. This time-step is larger than those normally employed in AL simulation ($1 - 5$ fs), and hence it allows more efficient sampling. This improvement is determined by the removal of fast degrees of freedom (such as those, typical in AL models, due to stiff harmonic potentials or light particles) and the excellent stability properties of the integration algorithm implemented.[132] Pressure and temperature are maintained at 1 atm and 30°C using the weak-coupling scheme.[237] Lipid and water temperatures are coupled separately with time constant $\tau_T = 0.2$ ps; for rigid-body sites, translational and rotational degrees of freedom are coupled independently. The pressure is controlled by semi-isotropic volume scaling: the normal and tangential components of the pressure tensor are regulated separately. In particular, the pressure along the z-axis, that is, along the direction normal to the interface, is controlled by rescaling the z-dimension of the simulation region, whereas the tangential pressure is controlled by rescaling the xy area, with the constraint that the interface remains a square. The pressure-coupling time constant is $\tau_P = 0.5$ ps, and the isothermal compressibility is $\beta = 4.6 \times 10^{-5}\,\mathrm{atm}^{-1}$. The cutoff radius for both Lennard-Jones and electrostatic water-water interactions is 0.9 nm.[227] All other nonbonded cutoff radii, both for Lennard-Jones and electrostatic interactions, have been set to 1.2 nm. The net mass centre velocity of the entire system is set to zero at every step.[238] To avoid artefacts in the evaluation of lipid diffusion,[31] the net lateral translation of each of the two monolayers is removed at every step.

5.1.6 Sampling enhancement

To estimate the efficiency gain of our simplified CG methodology over traditional AL modelling, we compared the sampling speed of our code BRAHMS with the popular AL software CHARMM.[239] We selected an AL test-system,[35] and constructed a corresponding CG configuration: both membranes comprised 72 lipids and 2094 water molecules. In particular, there were 2814 interaction sites in the CG system and 15210 atoms in the AL system. Our model was simulated with a 20 fs time-step and cutoff treatment of long-range interactions, whereas the AL simulation was conducted with a 2 fs time-step and PME scheme[232] for evaluation of long-range electrostatics. By simulating on the same AMD 1400 MHz processor, we measured sampling speeds of 324 ps/CPU-hour for BRAHMS and 2.5 ps/CPU-hour for CHARMM, corresponding to a CG speed-up factor of ≈ 130. The choice of a 2 fs time-step in the AL run is typical for AL membrane simulations.[27,33,35–41,43–45] We are aware that the use of constraint and multi-step algorithms can allow time-steps of 4 fs,[29,32] and even 5 fs,[28,31] hence more than twice larger than that employed in our test. Using a 5 fs time-step in the AL run would bring the CG speed-up factor down to ≈ 50. However, there are also some AL membrane studies that report the use of a 1 fs time-step.[30,34,42] Overall, our choice to consider an "average" AL time-step of 2 fs seems reasonable. We are also aware that the reported CG speed-up factor benefits in part from the cutoff treatment of CG electrostatic interactions with respect to the more costly PME scheme used in the AL CHARMM simulation. Also, CHARMM is not the most efficient biological molecular dynamics code; however, neither has significant effort been made to optimise the performance of BRAHMS. In summary, it seems acceptable to claim that the reduced number of interactions, the use of larger integration time-steps and the simplified treatment of electrostatics make our CG technique roughly two orders of magnitude less demanding of computational resources than traditional AL methodology. All data reported in the rest of the chapter refer to the larger membrane (128 DMPC lipids and 3400 water molecules) described in the previous section: single-processor simulations of this system have been carried out on the Iridis high-performance computational cluster at the University of Southampton,[240] where BRAHMS runs at up to 22 ns/CPU-day.

5.1.7 Force-field parameterisation

We chose to parameterise the CG force-field of our lipid model to reproduce the experimental volume and area per lipid,[6] and the average segmental tail order parameter[7,75] of fluid-phase DMPC. The volume per lipid is computed as $V_L = (V_T - N_W V_W)/N_L$, with V_T the total volume of the system, N_W the number of water molecules, V_W the volume per water and N_L the number of lipids. The area per lipid is computed as $A_L = A/N_L^m$, with A the total interfacial area and N_L^m the number of lipids per monolayer. Second-rank order parameters are calculated for the CG tail sites as $(3\cos^2\theta - 1)/2$, with θ the angle between the ellipsoid axis and the bilayer normal. The six tail ellipsoids of each lipid are sorted, in pairs, into three layers at different depth along the tails: the "top-layer" comprises the tail sites connected to

the glycerol units, the "end-layer" comprises the terminal tail sites, and the "mid-layer" is between the previous two. We will compute segmental order parameters for each of the three layers so defined, and the average order parameter for the entire hydrocarbon tail region.

Setting of bond rigidity, masses and moments of inertia

The rigidity constant of the Hooke potential, modelling intra-lipid covalent bonds as harmonic springs, is set to $3\,\mathrm{kcal/(mol\,\AA^2)}$, a typical value for CG models.[46] The reference bond lengths have been set to zero for simplicity, although they could have also been empirically assigned. Both approaches seem equally reasonable, as the CG springs are extremely soft: the selected CG rigidity constant is two orders of magnitude smaller than those normally used for AL bond stretching ($300 - 800\,\mathrm{kcal/(mol\,\AA^2)}$).[121]

The masses of the CG lipid sites are set to the sum of the atomic masses of the corresponding AL groups. For the ellipsoidal rigid-body units, the principal moments of inertia are assigned assuming uniform density. In particular, we have followed Routh's rule: *For solid bodies of the ellipsoidal type, the moment of inertia about a principal axis through the centre is equal to:*

$$\frac{m\,(a^2+b^2)}{5} \tag{5.20}$$

where m is the mass of the body and a, b are the lengths of the semi-axes perpendicular to the principal axis in question.[241]

For Gay-Berne particles, taking z as the principal axis parallel to the particle's orientation vector, the principal moments of inertia I_x and I_y can therefore be calculated as:

$$I_x = I_y = m\,[(\sigma_{\mathrm{GB}}/2)^2 + (\kappa\sigma_{\mathrm{GB}}/2)^2]/5 = m\,[\sigma_{\mathrm{GB}}^2 + (\kappa\sigma_{\mathrm{GB}})^2]/20 \tag{5.21}$$

The mass and principal moments of inertia of water sites are increased to optimise the stability of molecular dynamics integration.[242] In particular, water mass is set to 50 amu (the real value being 18 amu). The chosen principal moments of inertia correspond to a redistribution of water masses as follows: 15 amu for each hydrogen, and 20 amu for the oxygen. Thermodynamic properties are not affected by such an alteration of the inertial features of water. However, dynamics are intuitively predicted to be somewhat slower.

To quantify this effect, we ran a simulation of a pure water system comprising 500 molecules and computed the diffusion coefficient. Our result of $1.8 \times 10^{-5}\,\mathrm{cm^2/s}$ is only slightly lower than the experimental value[243] of $2.3 \times 10^{-5}\,\mathrm{cm^2/s}$ and that obtained from simulation of the SSD model with normal masses[227] of $2.5 \times 10^{-5}\,\mathrm{cm^2/s}$, indicating that the dynamic behaviour of our "heavy" water remains realistic.

Nonbonded terms: initial setting and refinement

Preliminary Lennard-Jones parameters for the headgroup sites were set by fitting the potential energy of the CG sites to that of the corresponding AL clusters of particles. Gay-Berne

Table 5.2: Hydrogen bonding capabilities of phospholipid-water systems.

Choline group	no hydrogen bonding
Phosphate group	hydrogen bond acceptor
Glycerol-ester group	hydrogen bond acceptor
Methyl tail group	no hydrogen bonding
Water	hydrogen bond acceptor and donor

parameters were initially taken from our previous model of the hydrocarbon tails.[207] Lennard-Jones and Gay-Berne cross-terms are calculated by standard rules,[120,221] with the following exceptions, where scaling factors are introduced: (i) To promote the self-assembly process, the mixed Lennard-Jones/Gay-Berne energy parameters between hydrophilic (water and headgroup) and hydrophobic (tail) sites are decreased. (ii) To mimic the hydrogen bonding capabilities of phosphate-water and glycerol-water pairs, the corresponding Lennard-Jones/Gay-Berne energy cross-terms are increased. General hydrogen bonding capabilities of phospholipid-water systems are collected in Table 5.2. The magnitudes of the headgroup charges and glycerol dipoles were chosen to reproduce the corresponding net charges and dipoles computed from the underlying AL distribution of partial charges. Initial simulations were performed on a pre-assembled bilayer. Parameters were optimised by trial-and-error molecular dynamics to reproduce the targeted experimental data. In each trial-and-error simulation, the area per lipid (which is the slowest-converging quantity) typically reached equilibrium in $\approx 20\,\mathrm{ns}$, corresponding to one CPU-day of simulation with BRAHMS. By running several tests in parallel, the parameter space could be explored quite efficiently. The experimental volume and area per lipid could be matched by tuning the Lennard-Jones parameters of the headgroup particles, whereas tail order parameters were reproduced by adjusting the Gay-Berne potential. Self-assembly runs were then prepared. To generate an initial random solution of lipids and water, the pre-assembled bilayer was brought to high temperature (up to 1000°C) while switching off electrostatics; runs were continued until visual inspection and order parameters confirmed a completely random configuration. By varying the temperature and the run length, several different starting configurations were obtained (typically, less than 1 CPU-hour of simulation is sufficient for each "disassembling" run). We then restored the electrostatics and started the self-assembly simulations at a temperature of 30°C; self-aggregation was typically completed over a time-scale of $\approx 100\,\mathrm{ns}$. Further trial-and-error runs were carried out to refine the force-field until we consistently obtained stable, defect-free bilayers matching the targeted experimental structural properties. The final complete parameter set is reported in Table 5.3.

Table 5.3: Force-field parameters

Parameter	Value
σ_{CC}	5.0 Å
σ_{PP}	4.9 Å
σ_{GG}	3.8 Å
σ_{TT}	3.8 Å
σ_{WW}	3.035 Å
ϵ_{CC}	1.9 kcal/mol
ϵ_{PP}	2.0 kcal/mol
ϵ_{GG}	1.3 kcal/mol
ϵ_{TT}	1.3 kcal/mol
ϵ_{WW}	0.152 kcal/mol
ϵ_{TW}	$\sqrt{\epsilon_{TT}\,\epsilon_{WW}}/3$
ϵ_{TC}	$\sqrt{\epsilon_{TT}\,\epsilon_{CC}}/2$
ϵ_{TP}	$\sqrt{\epsilon_{TP}\,\epsilon_{PP}}/2$
ϵ_{WP}	$1.5\sqrt{\epsilon_{WW}\,\epsilon_{PP}}$
ϵ_{WG}	$1.5\sqrt{\epsilon_{WW}\,\epsilon_{GG}}$
μ	2
ν	1
κ	1.77
κ'	20
κ'_{GBLJ}	1
Q_C	+0.7 e
Q_P	−0.7 e
μ_G	3 D
μ_W	2.42 D
k	3 kcal/(mol Å^2)

Subscripts C, P, G, T and W stand for the site types *choline*, *phosphate*, *glycerol*, *tail* and *water*, respectively. Lennard-Jones cross-terms are calculated by a standard rule[120] except for $\epsilon_{TW}, \epsilon_{TC}, \epsilon_{TP}, \epsilon_{WP}, \epsilon_{WG}$, which have been set as reported. The constants μ, ν, κ and κ' refer to Gay-Berne parameters.[125] As for the mixed Gay-Berne/Lennard-Jones potential,[221] the shape anisotropy cross terms χ_{GBLJ} are calculated by Equation 5.8, whereas the energy constants χ'_{GBLJ} (see Equation 5.11) are set to zero. Charges and dipoles are identified by Q and μ; cross terms are obtained via a standard rule.[126] The rigidity of the Hooke spring potential is identified by k; reference lengths are zero. Springs are anchored at the mass centre for C and P sites, and at a distance $\kappa_{\mathrm{GB}}\,\sigma_{TT}/4$ from the mass centre along the symmetry axes for G and T sites.

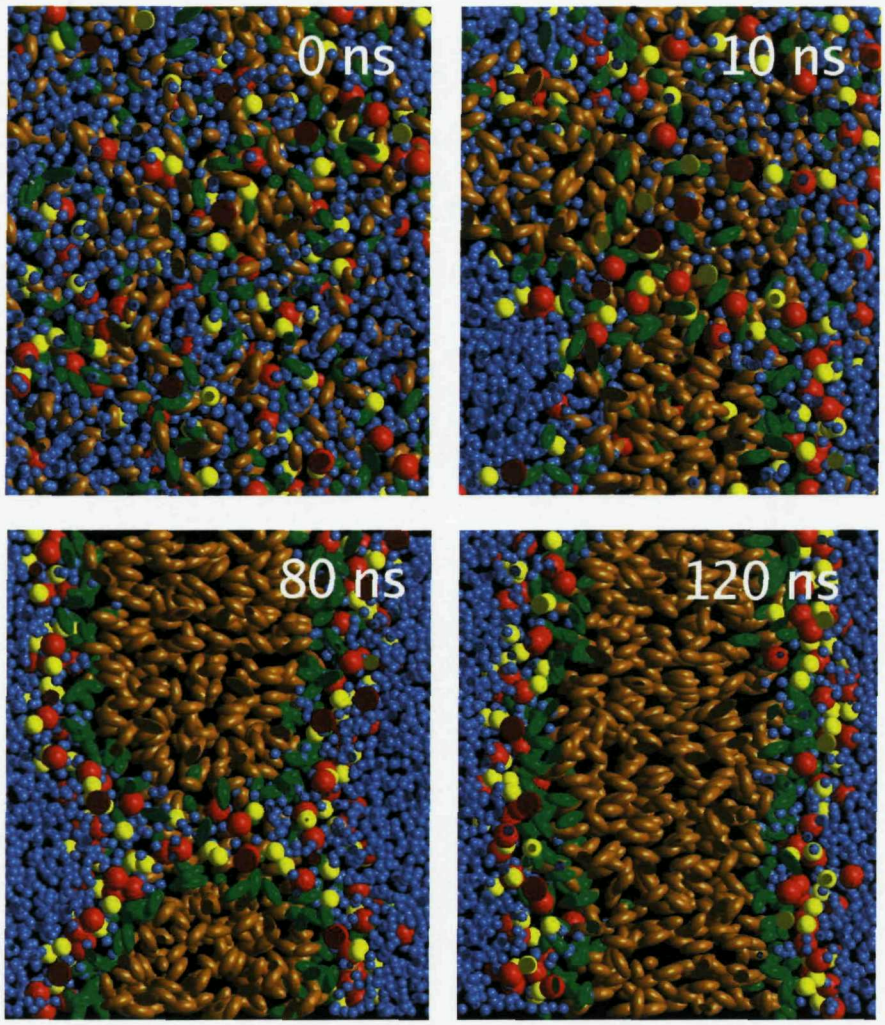


Figure 5.8: Self-assembly simulation snapshots. The choline, phosphate, glycerol and tail sites of lipid molecules are represented in red, yellow, green and orange, respectively. Water molecules are coloured in blue. A quick phase separation (0 to 10 ns) is followed by the formation of a transient water pore (80 ns) which eventually disappears leading to the stabilisation of a defect-free bilayer (120 ns).

Fitted properties

Snapshots from a representative self-assembly simulation are displayed in Figure 5.8. A fast phase separation between lipids and water is followed by a slow rearrangement of the bilayer to expose the headgroups while burying the hydrocarbon tails in the interior. A transient water pore is also observed. The time-scale and the overall aggregation mechanism are consistent with the only AL self-assembly simulation reported to date.[28] We then analysed the self-assembled bilayers to check the correct reproduction of the targeted experimental data. The time evolution of the lipid volume V_L and area A_L over a 100-ns time window (after self-assembly) is shown in Figure 5.9, together with the experimental estimates. Our measurements $V_L = 1.104 \pm 0.002\,\mathrm{nm}^3$ and $A_L = 0.594 \pm 0.003\,\mathrm{nm}^2$ are consistent with the experimental data[6] $V_L = 1.101\,\mathrm{nm}^3$ and $A_L = 0.606\,\mathrm{nm}^2$. The fluctuations are similar to

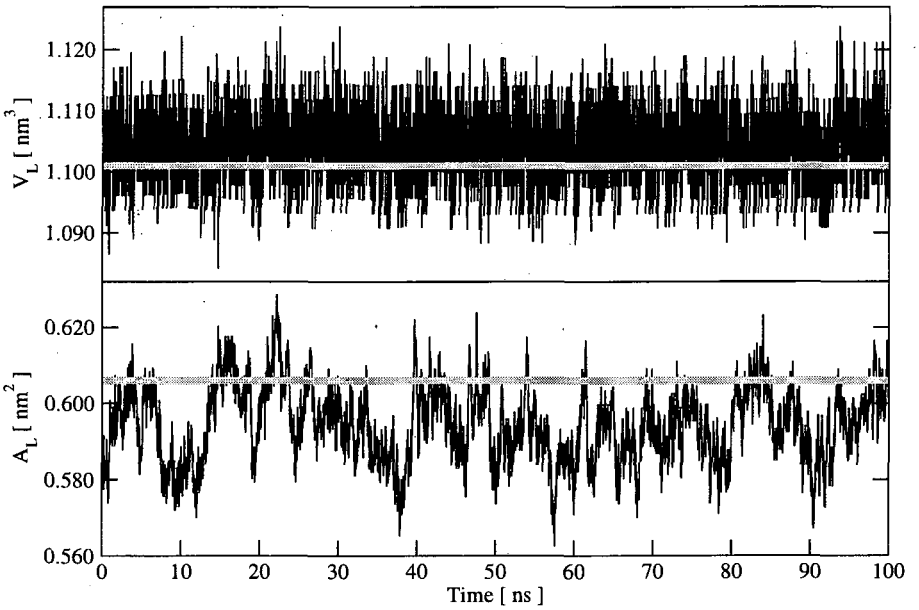


Figure 5.9: Time evolution of the volume V_L and the area A_L per lipid: the curves connect 10-ps running averages. The experimental values[6] are plotted as straight grey lines.

those observed with AL models.[29,31] Segmental order parameters for the three layers previously defined are plotted in Figure 5.10, along with the experimental data. The agreement is rather good, apart from the mid-layer order parameter which is somewhat lower than the experimental values. We calculate an average segmental order parameter for the entire tail region $S_{\mathrm{mol}} = 0.36 \pm 0.01$, consistent with the value of 0.38 deduced from the experimental data.[7,75] Standard errors have been estimated using the block averaging method[124] using ten consecutive 10-ns blocks.

5.2 Results

In this section, we thoroughly investigate the physical properties of our self-assembled CG bilayer model. Simulation results will be primarily compared to experimental data, although in some cases AL and CG models, and alternative theoretical approaches, will also be considered. If not otherwise stated, all data reported for comparison refer to systems and conditions consistent with our simulations, that is, fully-hydrated liquid-phase DMPC bilayers at 30°C. We simulated a self-assembled bilayer for several hundreds of nanoseconds, over which time the system remained stable and all the properties observed fluctuated around their equilibrium values. Most of the measurements have been taken over a 200-ns time window; for each parameter measured, the reported average value and standard error have been computed from two subaverages taken over the two 100-ns consecutive blocks of the trajectory, unless otherwise stated. However, to compute the water permeability coefficient, the detection of water permeation events has been carried out for 900 ns: this long simulation time is necessary to collect enough statistics on such relatively rare events. The analysis of a number of

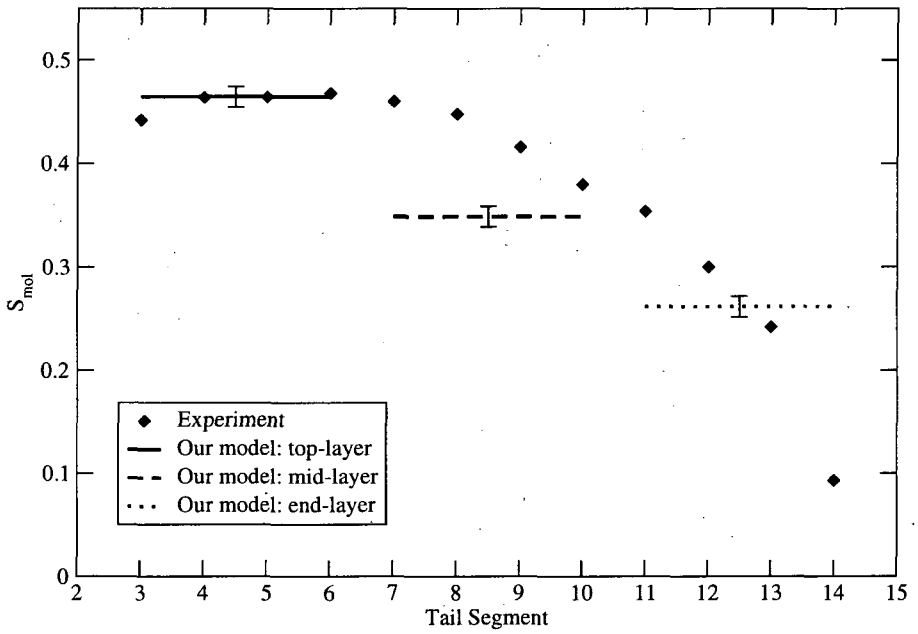


Figure 5.10: Tail order parameters. The estimated segmental order parameters[75] from experiment[7] are superimposed on our simulation results obtained by averaging over layers of tail sites as defined in the text. Standard errors are reported.

properties is carried out following a general process that involves "slicing" the system along planes perpendicular to the z axis (interface normal). In particular, we have defined 600 slices of thickness $\Delta z \approx 0.1\,\text{Å}$, the actual value of Δz being evaluated at every step to account for the fluctuations of the $z-$dimension of the simulation region. Several bilayer properties are homogeneous inside a particular slice, due to the intrinsic axial symmetry of the system. Therefore single curves, *profiles* evaluated as a function of z, provide full characterisation. The slicing procedure has been employed to calculate the following profiles: electron density, lateral pressure, electric field, water polarisation, and electrostatic potential. We report the average profiles obtained from calculation at every molecular dynamics integration step over the 200-ns measurement time. No extra processing has been done, i.e., no filters have been applied to smooth the curves, and we have not averaged over the two monolayers. It will be seen that the profiles are nonetheless extremely smooth (almost noise-free) and symmetrical. This is indicative of a well-equilibrated system and adequate sampling. We again emphasise that all properties considered in the following did not directly enter the parameterisation process.

5.2.1 Structure

Electron density profile

In general, to estimate the electron distribution from particle-based models, electrons have to be arbitrarily assigned to each site. Care should be taken to match as closely as possible

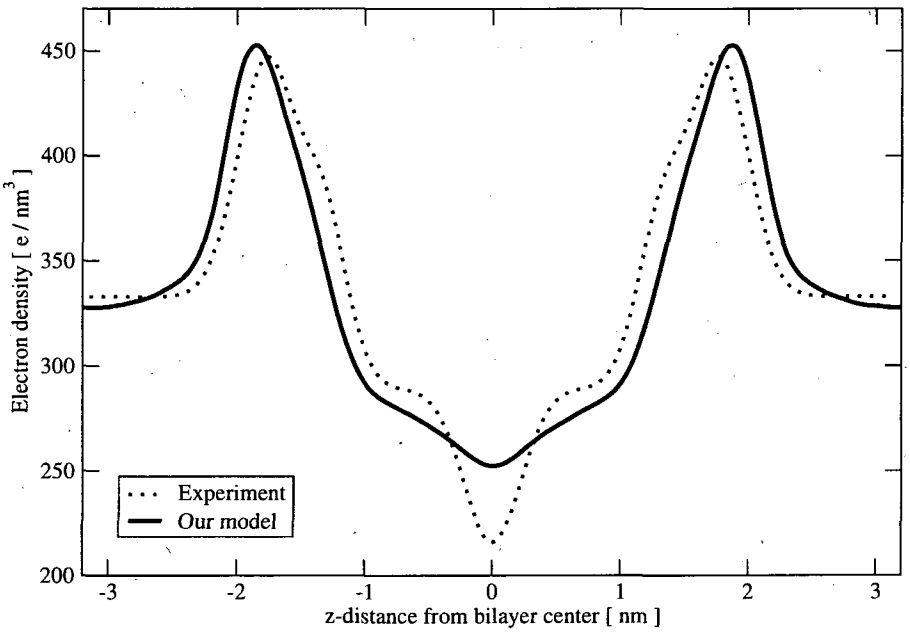


Figure 5.11: Total electron density profile. The distribution calculated with our model is superimposed on the profile obtained from experiment.[6]

the underlying real electron locations, with the obvious requirement for the total number of electrons per lipid in the model to be equal to the real value (374 for DMPC). For our model, we assume all electrons belonging to the choline and phosphate clusters to be located at the mass centre of the choline and phosphate CG units, respectively. As for the ellipsoidal particles that constitute the rest of the lipid in our representation, we have resolved four positions evenly spaced along the principal axis of each ellipsoid; electrons are assigned to these positions to match as closely as possible the underlying real distribution. The total electron density profile of our model is reported in Figure 5.11 along with the experimental profile:[6] the agreement is rather good. In fact, our profile matches the experimental result with an accuracy which is comparable to the results obtained with AL models.[30–32] The only significant difference involves the central region, corresponding to the terminal methyl groups, where the density of our model is slightly higher than that obtained from experiment. The pronounced central trough in the electron density is possibly determined by the "curling up" of the terminal methyl segment: experimental order parameters[7] indeed indicate that the terminal methyl segment is significantly tilted with respect to the neighbouring segment. Our model (by construction) cannot capture this feature, as the four consecutive methyl groups at the end of each tail, represented by a single rigid ellipsoid, are assumed to be aligned. From the distance between the maxima of the profile we compute a membrane thickness $d_{\mathrm{HH}} = 3.71 \pm 0.02\,\mathrm{nm}$, close to the experimental value[6] of 3.53 nm. Single-site profiles are shown in Figure 5.12. We observe broad peaks, significant headgroup hydration, and water penetration down to the glycerol region. These findings are consistent with AL simulation data.[30–32]

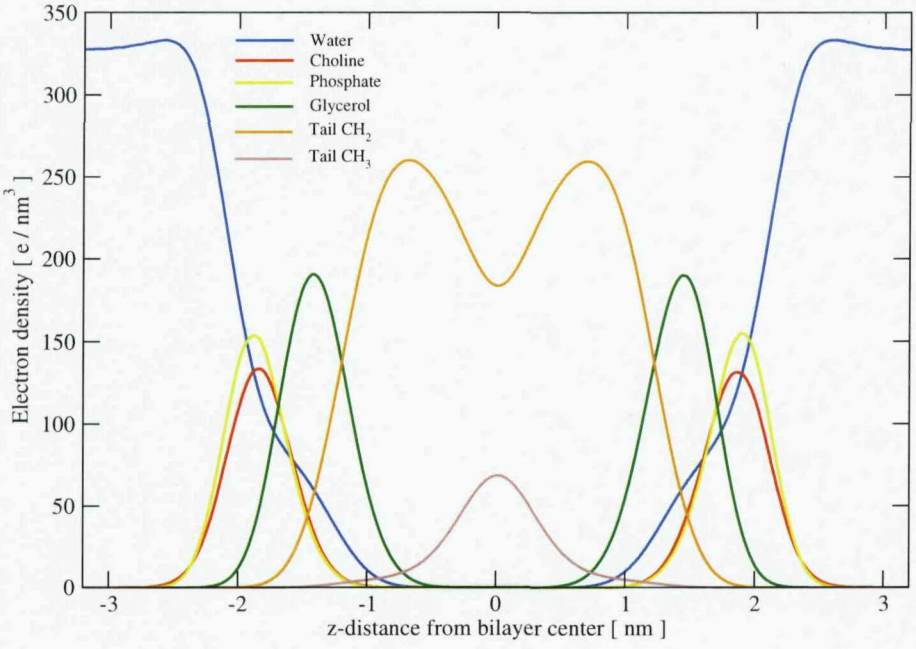


Figure 5.12: Single-site electron density profiles. Distributions are displayed for each site of our model. For the tail sites, electrons belonging to the CH_2 groups have been differentiated from the ones belonging to the terminal CH_3 group.

Headgroup dipole

We define the headgroup dipole moment as $\vec{\mu}_{\mathrm{HG}} = q\vec{d}$, with $\vec{d}$ the instantaneous vector connecting the phosphate to the choline mass centre and q the (equal) absolute magnitude of each headgroup charge. Owing to the lack of experimental data for the headgroup dipole of DMPC, we compare here our model with experiments on dipalmitoylphosphatidylcholine (DPPC) bilayers. Apart from slightly longer tails (two more carbons), DPPC is identical to DMPC; also, the experiments considered[74,244] were carried out on fully-hydrated bilayers in the fluid-phase, under the same conditions as our simulation. Therefore it is reasonable to compare our results for DMPC with these experimental data for DPPC. We calculate a dipole magnitude $\mu_{\mathrm{HG}} = 16.1 \pm 0.1\,\mathrm{D}$, broadly consistent with the experimental estimate[244] for DPPC of 18.7 D. As for the orientation, the headgroup dipole of our model is inclined, on average, by 92.3° from the membrane normal, marginally pointing toward the bilayer interior. Experiment on fluid-phase DPPC estimated the preferred conformation at $\approx 72°$ from the membrane normal, hence slightly pointing toward the water phase.[74] Despite a small discrepancy regarding the average angle, our result agrees with experiment by describing the headgroup dipole as being roughly parallel to the membrane plane.[74] Recent AL simulations of DMPC also reported the headgroup dipole to lie essentially flat on the bilayer plane, the average angle being estimated at[31] 72.3° and at[245] 79.8° from the interface normal.

5.2.2 Mechanical properties

Compressibility moduli

The area compressibility K_A can be computed from simulation as $K_A = k_B T A/\sigma^2(A)$, A and $\sigma^2(A)$ being respectively the mean and mean squared fluctuation of the interfacial area. We have calculated for our model $K_A = 297 \pm 22\,\mathrm{dyn/cm}$, a value consistent with the experimental measurement[9] of 234 dyn/cm. The volume compressibility modulus K_V can be computed as $K_V = k_B T V/\sigma^2(V)$, V and $\sigma^2(V)$ being respectively the mean and mean squared fluctuation of the volume of the simulation region. We have calculated for our model $K_V = 13.7 \pm 0.2\,\mathrm{kbar}$, consistent with the experimental range $10-30\,\mathrm{kbar}$ reported as typical for fluid-phase phospholipid bilayers.[2]

Bending rigidity modulus

The bilayer bending modulus κ^b can be related to the area compressibility modulus K_A via[9] $\kappa^b = K_A\, d_t^2/24$, the "effective" bilayer thickness d_t being $d_t = d_{HH} - 1\,\mathrm{nm}$, with d_{HH} the peak-to-peak distance in the electron density profile. Using K_A and d_{HH} from our previous results, we obtain $\kappa^b = 21.9 \pm 1.6\,k_B T$. Experiments have measured $16.7\,k_B T$ from x-ray data,[246] $13.5\,k_B T$ from pipette aspiration,[9] $31.4\,k_B T$ from thermally excited shape fluctuation[247] and $32.1\,k_B T$ from all-optical measurement.[248] Our estimate lies inside the range of the experimental values.

Lateral pressure profile

To calculate the lateral pressure profile along the membrane normal, we have to spatially resolve the pressure tensor along that direction. The macroscopic pressure tensor in a system of N particles can be written as:*

$$\mathbf{P} = \frac{1}{V_{\mathrm{total}}} \left(\sum_{i=1}^{N} m_i \mathbf{v}_i \otimes \mathbf{v}_i + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathbf{F}_{ij} \otimes \mathbf{r}_{ij} \right) \tag{5.22}$$

the first term inside the brackets being the *kinetic energy density tensor* and the second being the *system's configurational stress tensor*, with V_{total} the total volume of the system, m_i and $\mathbf{v}_i$ the mass and velocity of the i-th particle, and $\mathbf{F}_{ij}$ and $\mathbf{r}_{ij}$ the force and distance between particles i and j. To study spatial variations in $\mathbf{P}$ along the interface normal, which in our case corresponds to the z-axis, it is possible to define a *local* pressure by splitting the total tensor in equal slabs, of volume V_{slab}, along the z-axis:

$$\mathbf{p}_{\mathrm{local}}(z) = \frac{1}{V_{\mathrm{slab}}} \left(\sum_{i \in \mathrm{slab}} m_i \mathbf{v}_i \otimes \mathbf{v}_i + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathbf{F}_{ij} \otimes \mathbf{r}_{ij} f(z, z_i, z_j) \right) \tag{5.23}$$

*This definition assumes the convention $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, which is the one implemented in BRAHMS. Other codes (for example GROMACS[249]) use the opposite convention $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$; in this case the "+" signs in Equation 5.22 becomes a "-" sign.

The weighting function $f(z, z_i, z_j)$ must be chosen in order to properly distribute the virial over the slabs involved. Two (similar) definitions of $f(z, z_i, z_j)$ have been reported in the literature, by Lindahl and Edholm[39] and by Gullingsrud and Schulten.[42] In particular, both approaches involve setting:

$$f(z, z_i, z_j) = \begin{cases} 1.0 & \text{if both sites are in the slab} \\ \Delta z/|z_i - z_j| & \text{if neither site is in the slab} \end{cases}$$

with Δz the slab thickness. We also conform to this choice: this function places the entire virial in a slab when both particles interact therein, otherwise spreads the virial equally over the (possible) slabs between the sites. Complications arise when the virial must be distributed in a slab occupied only by one of the two interacting particles. This particular case can be treated by simply setting:[39]

$$f(z, z_i, z_j) = 0.5\,\Delta z/|z_i - z_j| \tag{5.24}$$

Such an implementation is not ideal, as when the two sites belong to adjacent slabs and are both close to the edge, their distance $|z_i - z_j|$ can easily become very small, leading to a divergent value of the term $1/|z_i - z_j|$ in Equation 5.24. A better solution involves using:[42]

$$f(z, z_i, z_j) = dz/|z_i - z_j| \tag{5.25}$$

with dz the distance between the particle and the edge of the slab. In particular, this choice avoids divergences, as if the two sites are very close along the $z-$axis, both dz and $|z_i - z_j|$ will be small so the division will remain finite.

Our approach is equivalent to that reported by Lindahl and Edholm[39] and by Gullingsrud and Schulten[42] for the cases where they agree, but it treats in a slightly different way the "pathological" case where divergences may occur. In particular, we split the virial as follows:

$$f(z, z_i, z_j) = \begin{cases} 1.0 & \text{if both sites are in the slab} \\ \Delta z/|z_i - z_j| & \text{if neither site is in the slab} \\ 0.5\,\Delta z/|z_i - z_j| & \text{if one is in the slab and the other is at least one slab apart} \\ 0.5 & \text{if the sites are located in adjacent slabs} \end{cases}$$

Hence our approach is robust, as it does not contain divergence sources. Slight inaccuracies occur in treating the case of particles in adjacent slabs, but the error should be negligible, and it decreases with increasing number of slabs (higher resolution). The pressure tensor $\mathbf{P}(z)$ for a system at equilibrium can be expressed as:

$$\mathbf{P}(z) = \begin{pmatrix} P_{xx}(z) & 0 & 0 \\ 0 & P_{yy}(z) & 0 \\ 0 & 0 & P_{zz}(z) \end{pmatrix}$$

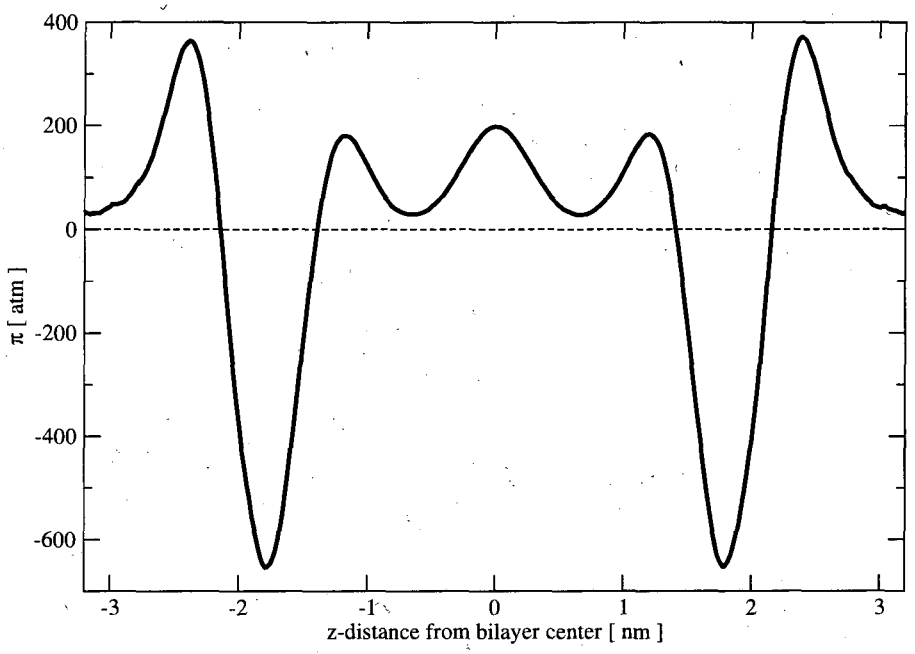


Figure 5.13: Lateral pressure profile.

Considering our frame of reference, we define $P_N(z) = P_{zz}(z)$ as the "normal" pressure (acting perpendicular to the bilayer) and $P_L(z) = [P_{xx}(z)+P_{yy}(z)]/2$ as the "lateral" pressure (acting tangential to the bilayer plane). The transmembrane lateral pressure profile $\pi(z)$ is finally calculated as the difference between the lateral and the normal pressures: $\pi(z) = P_L(z) - P_N(z)$.

The lateral pressure profile of our model is reported in Figure 5.13. Distinctive features can be highlighted and related to their molecular origin. Corresponding to the interfacial water regions, at the left and right extremes of the profile, the net lateral pressure is almost zero, as it should be for bulk water at mechanical equilibrium. In fact, it is possible to observe a slightly positive value, which indicates repulsive interactions: these arise from the partial alignment of water dipoles as a consequence of the strong membrane electric field (as analysed in the following section on water polarisation) and the disturbance of interfacial water due to the nearby lipid headgroups (with occasional headgroup protrusions into the water phase). On entering the membrane, at the headgroup region, we observe large positive lateral pressures, of peak magnitude ≈ 370 atm, which reflect a strong desire of the bilayer interfacial area to expand. This results from repulsive interactions between the headgroups due to steric, electrostatic, and hydration forces.[2] At the polar/apolar interface, roughly corresponding to the lipid glycerol groups, the profile is characterised by deep troughs of lateral pressure, characterised by minimum values of ≈ -650 atm (Figure 5.13). Here the interfacial tension tries to contract the bilayer, to minimise the exposure of the hydrocarbon core to the polar environment. In the bilayer hydrocarbon interior, the pressure profile displays positive values. The tight lipid packing conditions result in stretched tails with respect to isolated "free" tails: the corresponding entropy loss causes significant inter-tail repulsion.[250]

In particular, the hydrocarbon region is characterised by three local maxima and two local minima. Two outer maxima, of peak magnitude ≈ 180 atm, are located corresponding to the top segments of the lipid tails: high ordering here (see Figure 5.10) is indicative of high entropy loss which results in large repulsive forces. The two pressure minima are observed corresponding to the mid-tail segments: in this region, the packing-related entropy loss is relatively small, since the connectivity to neighbouring segments already limits the mid-tail conformational entropy in free tails. A broad lateral pressure maximum, of peak magnitude ≈ 200 atm, is located at the very centre of the bilayer (Figure 5.13). Here the ordering is minimal (see again Figure 5.10), and hence the presence of such large repulsive forces might seem unjustified. However, the *relative* entropy loss of the tail ends in the hydrocarbon core, with respect to free tails, is predicted to be maximal,[251] thus leading to sharp pressure increases toward the bilayer centre.

Experimental investigation of the lateral pressure distribution is extremely difficult: the few attempts made to date provide qualitative and partial pictures for the hydrocarbon region only.[11,24] Unfortunately, the lateral pressure inside DMPC bilayers has never been experimentally investigated. However, the pressure distribution has been probed for dioleylphosphatidylcholine (DOPC) bilayers.[11] DOPC tails are slightly longer than DMPC (18 vs. 14 carbons), and contain one double bond (or unsaturation), whereas DMPC is fully-saturated. The headgroup and glycerol regions are identical for the two species. The experiment[11] was carried out with fully hydrated fluid-phase bilayers, as in our simulation. Overall, it seems reasonable to compare our results for DMPC with the data for DOPC. The experimental data indicate that the hydrocarbon tails generate positive pressures nonuniformly distributed: in particular, local minima roughly at mid-depth along the tails of each leaflet were observed.[11] The qualitative experimental picture is therefore consistent with our calculation. Our profile can also be compared to the distributions obtained from AL bilayer models of a variety of phospholipid species.[39–45] We note a general agreement on the number and locations of the pressure peaks and troughs. In particular, the mid-tail minima are also reproduced by these models: they are observed for fully-saturated[39–44] as well as for unsaturated[42–45] lipids. In terms of magnitude, AL models are broadly consistent with ours in the peak values recorded. In particular, peak pressures recorded for fully-saturated lipids[39–44] are of the order 200 to 600 atm for the headgroup maxima, -500 to -1000 atm for the polar/apolar interfacial troughs and 200 to 300 atm for the central peak. Our results lie inside these intervals. The only AL pressure profile of DMPC available[40] displays the qualitative features present in our profile, although the peak magnitudes are slightly different. The AL profile[40] is rather noisy, and it was computed from a relatively short run (14.3 ns after only 5 ps of equilibration). Hence it is probably undersampled and not fully converged. Also, the membrane area was held fixed, whereas in our simulation the xy interfacial plane fluctuates according to the lateral components of the pressure tensor. It is therefore not appropriate to compare in detail our profile to that AL curve.[40]

Regarding the other specific CG models developed to date, the lateral pressure profile has

been calculated so far only once, for a DPPC bilayer simulated by Marrink et al.;[194] there are issues, however, with the reported data, because that CG pressure profile[194] does not feature the mid-tail minima and displays magnitudes that are markedly different from the other published AL results[39,43,44] for DPPC bilayers, especially corresponding to the headgroup region and the polar/apolar interface. Those AL works[39,43,44] report results quite consistent with each other, and the most recent data[43,44] were obtained from a fairly long simulation (80 ns after 20 ns of equilibration). Corresponding to the headgroup region, the CG profile by Marrink et al.[194] displays pressure maxima of ≈ 100 atm, where the AL profiles feature instead much higher values of[39] ≈ 400 atm and[44] ≈ 350 atm. Moreover, the pressure troughs at the polar/apolar interfaces have minima of ≈ -270 atm for that CG model against[39] ≈ -500 atm and[44] -850 atm for the AL models. Since electrostatic and hydration forces play a significant role in the generation of lateral pressures at the headgroups and polar/apolar interfaces,[2] these discrepancies could be due to the oversimplified description of electrostatics and water in the Marrink model.[46,194] The only agreement in terms of magnitude is present at the bilayer centre, where the CG profile by Marrink et al.[194] displays a peak value of ≈ 150 atm and the AL models yield values of[39] ≈ 150 atm and[43,44] ≈ 200 atm. Pressure profiles have also been computed from particle-based simulations of *generic* coarse-grain models.[166,195,213,215,252–255] It has been argued[41] that the level of simplification of these models (which do not incorporate electrostatics and contain highly simplified water representations) makes their pressure profile predictions questionable. In fact, some qualitative features are captured, although none of these models reproduce a central pressure maximum as observed in our profile and in the AL studies: central minima of nearly zero[166,213,252,253] or negative[195,215,254,255] magnitude are instead reported. Predictions of the pressure profile have also been obtained by analytical theory.[251,256–258] These methods typically consider only the tail region, and they assume uniform density. Two of these models[251,258] reproduce the mid-tail minima. With one exception,[251] the pressure profiles obtained by analytical theory do not feature the central maximum corresponding to the tail ends.

Membrane curvature elasticity

In the popular formalism developed by Helfrich,[77,97,98] the surface curvature elastic energy per unit area g is expressed as: $g = \kappa\,(c_1 + c_2 - 2c_0)^2\,/2 + \kappa_G\,c_1\,c_2$, with κ the bending rigidity, c_1 and c_2 the (local) principal curvatures, c_0 the spontaneous (or intrinsic) curvature, and κ_G the Gaussian curvature modulus. In the following sections, we will calculate the constants c_0 and κ_G by evaluating the first and second integral moments of the pressure profile over each monolayer.[250] The final values and standard errors are estimated from four separate averages: two (one for each monolayer) for each of the two consecutive 100-ns measurement blocks. Similar calculations have been previously performed by analytical theory,[256,259] with a generic CG model,[260,261] and with AL models,[42,44] although not for DMPC. Experimental data for DMPC are also lacking. However, we will be able to compare our results to experimental data for similar lipids and general theoretical predictions.

Torque tension and spontaneous curvature

We have calculated the first integral moment P_1 of the lateral pressure profile π as

$$P_1 = \int_0^h z\,\pi(z)\,\mathrm{d}z$$

where $z = 0$ at the centre of the bilayer and $z = h$ in the water phase.[250] In particular, we have integrated the pressure profile over each of the two monolayers, with $z = 0$ and $z = \pm h$, h being half the z–dimension of the simulation region. In practice, considering the slab subdivision previously employed to calculate the lateral pressure profile, P_1 is estimated for each monolayer by summing up the contribution from each slab i:

$$P_1 = \Delta z \cdot \sum_{i \in \text{monolayer}} (z_i \cdot \pi_i) \tag{5.26}$$

with Δz the (uniform) slab thickness.

The first moment of the lateral pressure is also called the torque tension τ^{m}, with the superscript "m" indicating "monolayer". We have calculated a monolayer torque tension $\tau^{\mathrm{m}} = P_1 = -0.020 \pm 0.003\,\mathrm{k_B T/\AA}$. We can then write[256] $\tau^{\mathrm{m}} = \kappa^{\mathrm{m}} c_0^{\mathrm{m}}$, κ^{m} and c_0^{m} being the monolayer bending rigidity and spontaneous curvature, respectively. Considering that[22] $\kappa^{\mathrm{m}} = \kappa^{\mathrm{b}}/2$, we obtain a monolayer spontaneous curvature $c_0^{\mathrm{m}} = -0.018 \pm 0.003\,\mathrm{nm}^{-1}$. The spontaneous curvature has been measured for a number of *type-II* lipids, i.e., lipids forming inverted nonlamellar structures: typical values in the range $-0.05\,\mathrm{nm}^{-1}$ to $-1\,\mathrm{nm}^{-1}$ have been reported.[22] Lower absolute values are expected for bilayer-forming, *type-0* lipids, such as DMPC. Our estimate is therefore reasonable.

Gaussian curvature modulus

The monolayer Gaussian curvature modulus can be determined as:

$$\kappa_{\mathrm{G}}^{\mathrm{m}} = -\int_0^h (z - \xi)^2\,\pi(z)\,\mathrm{d}z$$

ξ being the distance to the *pivotal surface*, defined as the surface at which there is no change in the molecular cross-sectional area upon bending.[22] Defining the second integral moment P_2 of the lateral pressure profile as:

$$P_2 = \int_0^h z^2\,\pi(z)\,\mathrm{d}z$$

we can rewrite the expression for the Gaussian curvature modulus as:

$$\kappa_{\mathrm{G}}^{\mathrm{m}} = 2\xi P_1 - P_2$$

P_1 being the first moment. In practice, P_2 is obtained for each monolayer by summing up the contribution from each slab i:

$$P_2 = \Delta z \cdot \sum_{i \in \text{monolayer}} (z_i^2 \cdot \pi_i) \tag{5.27}$$

with Δz the (uniform) slab thickness. The pivotal surface ξ has been experimentally identified close to the polar/apolar interface.[262] Considering the lateral pressure profile (Figure 5.13), we assume the polar/apolar interfaces of the two monolayers to be located at the two main troughs of the curve. In fact, the global minima of the pressure profile identify the regions of largest surface tension, situated at the hydrophobic/hydrophilic interfacial regions. Hence, by computing the half-distance between the global minima of the pressure profile, we set $\xi = 1.79\,\text{nm}$. The second integral moment of the pressure profile from simulation is $P_2 = 5.3 \pm 0.1\,\text{k}_B\text{T}$. With P_1 and ξ as defined previously, we finally obtain $\kappa_G^m = -5.4 \pm 0.1\text{k}_B\text{T}$. It has been shown[262] that the monolayer Gaussian curvature modulus is generally related to the bending rigidity according to $-1 < \kappa_G^m/\kappa^m < 0$: this prediction has been confirmed experimentally.[22] The more stringent relation $\kappa_G^m/\kappa^m \approx -0.8$ has also been proposed.[98] For our model, assuming $\kappa^m = \kappa^b/2$,[22] we calculate $\kappa_G^m/\kappa^m \approx -0.5$, a value consistent with the reported theoretical predictions. We have also estimated the *bilayer* Gaussian curvature modulus as[77] $\kappa_G^b = 2(\kappa_G^m - 2\kappa^m c_0^m \xi) = -10.9 \pm 2.3\,\text{k}_B\text{T}$.

5.2.3 Electrostatic properties

Electric field

We have computed the transbilayer electric field by integrating the charge density of the system along the bilayer normal.[263] For point-charges, considering a volume charge density ρ_q, the corresponding electric field is given by:

$$E_z(z) = \frac{1}{\epsilon_0} \int_0^z \mathrm{d}z' \rho_q(z') \tag{5.28}$$

with ϵ_0 the permittivity of free space. The electric field can be computed from the charge density by summing up over the different charge species:[263]

$$E_z(z) = \frac{\Delta z}{\epsilon_0} \sum_i q_i \rho_i(z) \tag{5.29}$$

with Δz the slab thickness, and q_i and ρ_i the partial charge and number density of species i, respectively. For point-dipoles, considering a dipolar moment density $\rho_d(z)$, the z projection of the electric field E_z is simply:[263]

$$E_z(z) = -\frac{1}{\epsilon_0} \rho_d(z) \tag{5.30}$$

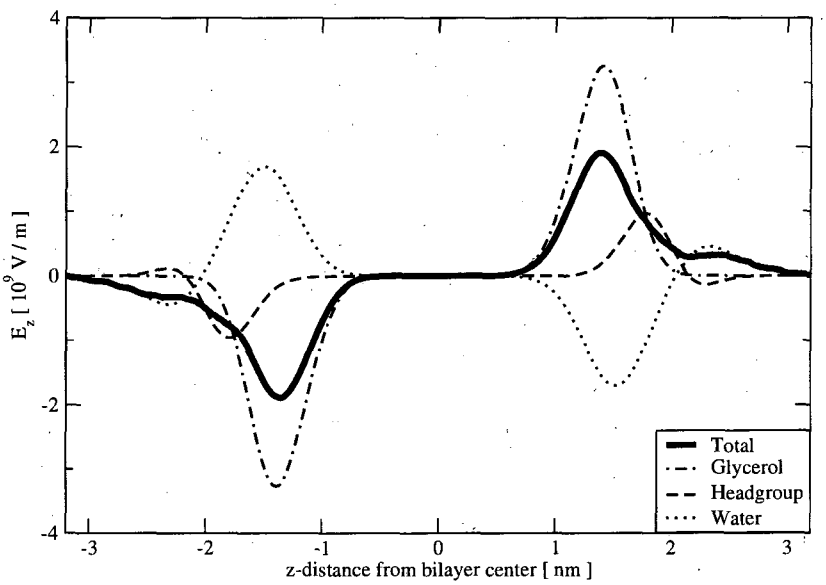


Figure 5.14: Transbilayer electric field. The $z-$projection of the total transmembrane electric field is plotted together with single-site profiles.

In practice, we compute E_z with:

$$E_z(z) = -\frac{1}{\epsilon_0} \sum_i \mu_{i_z}(z)\rho_i(z) \tag{5.31}$$

with μ_{i_z} and ρ_i the dipole z-projection and ρ_i the number density of species i, respectively. The $z-$projection of the electric field as a function of the position along the bilayer normal is reported in Figure 5.14: the total profile is displayed together with single-site profiles of the various charged groups. The calculated magnitude of the total electric field within the membrane is enormous, with local values of the order of $\approx 10^9$ V/m. Although no direct experimental measure is available, such a large figure is expected[101] considering the magnitude of the dipole potential (see next section). The total electric field arises predominantly from the glycerol-ester dipoles, through their alignment along the interface normal. The headgroup contribution is relatively small; in fact, the headgroup dipole lies almost parallel to the membrane plane, and hence, despite its large magnitude, its projection along the normal, and consequently the $z-$component of the electric field, will be small on average. A significant contribution to the net total field is due to interfacial water: water dipoles generate a strong field which counteracts the total field to lower the overall magnitude. We have further investigated interfacial water by quantifying water ordering. In particular, we have calculated the transbilayer water polarisation profile as the first-rank $P_1(z)$ order parameter[120] of the water dipoles with respect to the interface normal. The order parameter P_1 in each of the system's slabs is calculated as:

$$P_1 = \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_z \tag{5.32}$$

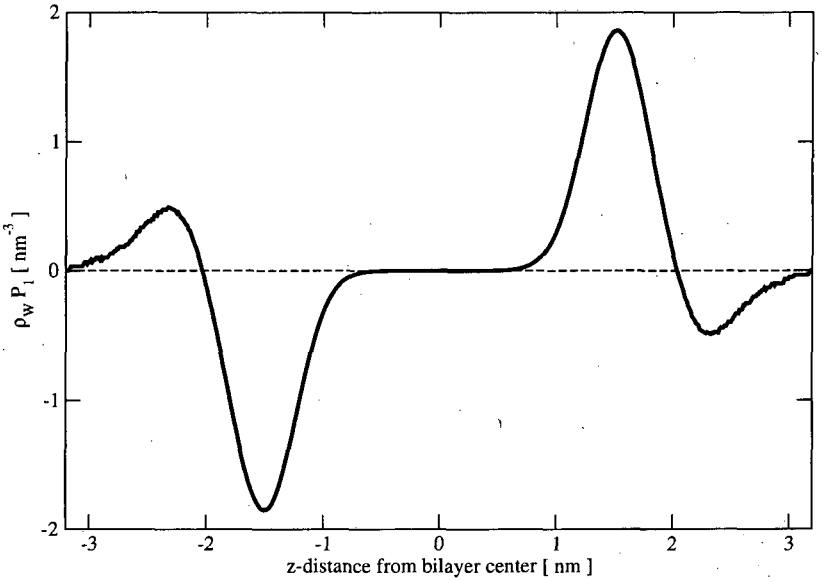


Figure 5.15: Transbilayer water polarisation profile. The first-rank order parameter P_1 of the water dipoles with respect to the z axis unit vector is weighted by the water number density ρ_W. Considering the frame of reference employed, the interface normal originates at the centre of the system and is oriented along the z axis. Hence, positive values in the left half of the water polarisation profile correspond to water dipoles oriented toward the bilayer interior, whereas negative values correspond to water dipoles oriented toward the water phase. For the right half of the curve, the convention is opposite.

with N the number of dipoles in the slab, $\hat{\mathbf{e}}_i$ the unit vector along the i-th water dipole and $\hat{\mathbf{e}}_z$ the (constant) unit vector along the cartesian z-axis (normal to the interface). P_1 in turn is the mean z-projection of the (unit) dipole vector: it measures the asymmetry of the molecular polar alignment near the interface.[263] Local non-zero values of the water dipolar rotational order parameter reveal the presence of a macroscopic electric field, which in turns indicates a change in the electrostatic potential across the interface. To better represent the physical significance of interfacial water ordering, we have also computed the density-weighted profile by multiplying $P_1(z)$ with the water number density profile ρ_W (deduced from the electron density of Figure 5.12). The density-weighted water polarisation profile is shown in Figure 5.15: the correlation with the water electric field (Figure 5.14) is evident.

Dipole potential

We have calculated the electrical potential profile $\Psi(z)$ by integrating the electric field along the interface normal:[263]

$$\Psi(z) = -\int_0^z \mathrm{d}z' E_z(z') \tag{5.33}$$

where E_z is the net sum of the electric fields generated by the various charged groups present in the system. Figure 5.16 reports the total transmembrane electric potential together with single-site profiles. The molecular origin of the membrane dipole potential can be clearly

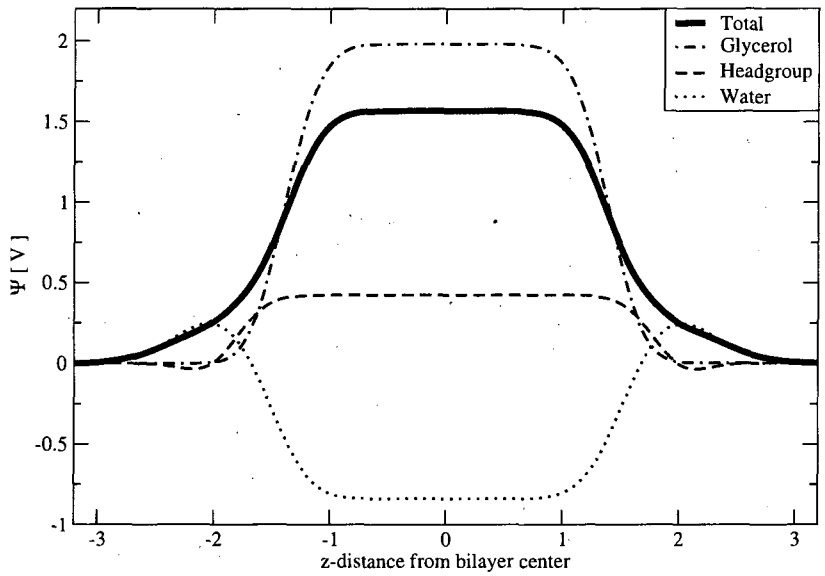


Figure 5.16: Electrical potential profile. The total transmembrane potential is plotted together with single-site profiles.

identified: according to our results, the overall potential largely originates from the glycerol groups. Water ordering generates a negative potential that lowers the overall value, whereas the headgroup contribution is comparatively small. The total membrane dipole Ψ_d for our model, measured in the hydrocarbon core with respect to the water phase, is $+1.57 \pm 0.03$ V. Experimentally, a value of $+0.45$ V for DMPC was obtained with the monolayer method.[111] Cryo-EM experiments have recently estimated a magnitude of $+0.51$ V for a diphytanoylphosphatidylcholine (DPhPC) bilayer:[10] this result can be reasonably considered as representative for ester-PC lipid bilayers in general (thus including DMPC), due to the presence of the same charged groups (glycerol-ester and headgroup).

Our model correctly predicts the sign of the dipole potential, although the magnitude is larger than those estimated experimentally. Part of the reason for the discrepancy observed is specific to our model. In our simulation, the headgroup dipole points on average slightly toward the bilayer interior, thus generating a small positive contribution to the dipole potential (Figure 5.16). As already noted, experiments suggest instead an orientation slightly pointing toward the water phase,[74] which would give rise to a negative contribution to the dipole potential. A possible solution, which we are indeed considering, involves introducing an angular bonded term, of the form typically present in both AL and CG force-fields, to restrain the headgroup dipole to a configuration more similar to experiment: this would yield a more realistic headgroup configuration and in turn a lower total potential. It is worth pointing out that large values for the membrane dipole have also been obtained with a number of AL models:[36–38,99] the overestimation of internal electrical potentials might be partly due to a general weakness of nonpolarizable force-fields.

A discrepancy between our results and corresponding AL calculation regards the single-site contributions to the membrane potential. In fact, AL simulations[37,99,156,245,264,265] typi-

Lipid (reference)	Ψ_d^{TOTAL}/V	Ψ_d^{water}/V	Ψ_d^{lipid}/V
DMPC (this work)	1.6	-0.7	2.3
DMPC ([245])	0.6	5.3	-4.7
DPPC ([156])	0.7	5.4	-4.7
DPPC ([99])	1	4	-3
DPPC ([265])	0.55	5.5	-4.95
DPPC ([264])	0.6	4.6	-4
DPhPC ([37])	1	2.4	-1.4

Table 5.4: Membrane dipole potential. The total potential and individual contributions as calculated by molecular dynamics simulations. The data represent potential differences in the bilayer centre with respect to a reference value of 0 V in the water phase.

cally identify water as the leading contributor in determining the overall potential; in particular, the AL water potential is positive in the hydrocarbon core with respect to the water phases, while the lipid (headgroup + glycerol) potential is negative and partially cancels the water term. The separate contributions to the membrane dipole potential of our model are compared in Table 5.4 to corresponding data obtained in the AL studies.

A number of modifications can be introduced to correct the membrane dipole potential. As already mentioned, we can adopt an angular force-field term to restrain the headgroup dipole to the experimental orientation. This would reverse the direction of the headgroup dipole with respect to the membrane normal, thus reversing the sign of its contribution to the overall potential. Then we can lower the magnitude of the glycerol-ester dipoles; this can be justified considering that the current dipolar magnitude of 3 D, obtained from averaging over an AL distribution, is larger than the experimental measurement[266,267] for ester groups of 1.8 D. We should hence obtain an overall negative potential due to the lipid (glycerol and headgroup), in agreement with the AL data; hopefully, the water potential will also then behave as in the AL models and overcompensate the lipid contribution, to yield a final overall positive value for the membrane dipole potential.

5.2.4 Dynamics

Lipid lateral diffusion

We have calculated the lipid lateral diffusion coefficient using the Einstein expression,[124] as is standard practice in membrane simulations.[29,31,32] Figure 5.17 shows the lipid lateral diffusion coefficient D_{lat} computed for two different measurement times. In particular, the diffusion coefficient measured on a sub-nanosecond scale is reported in the top panel of Figure 5.17: at 0.5 ns, a representative value of $\approx 60\,\text{nm}^2/\mu\text{s}$ can be observed. This result is consistent with those obtained from experimental methods such as quasi-elastic neutron scattering,[113] which measure lipid displacements over time periods of less than a nanosecond, and yield *short-range* diffusion constants in the range $10 - 100\,\text{nm}^2/\mu\text{s}$. In the bottom panel

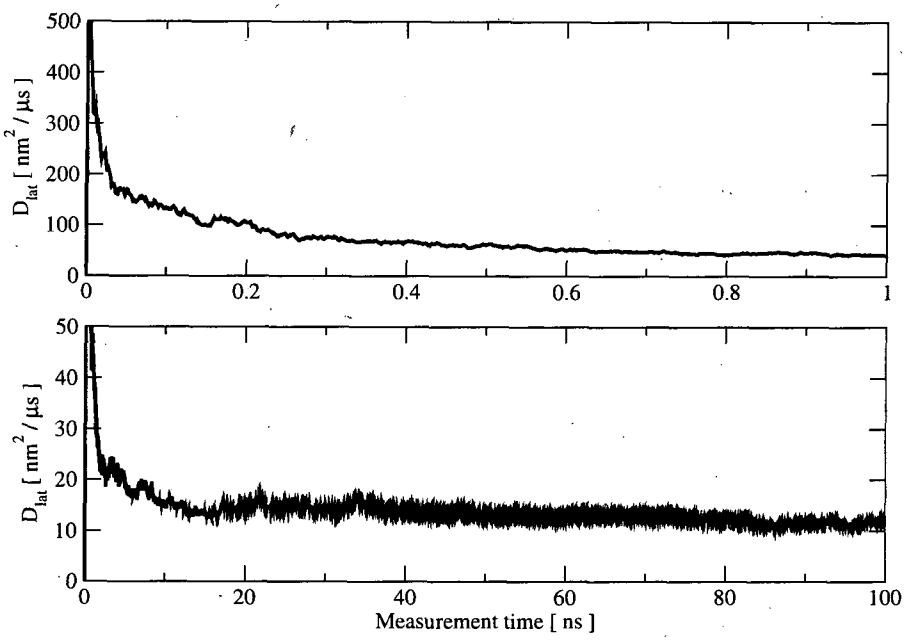


Figure 5.17: Lateral diffusion coefficients of lipid mass-centres calculated using the Einstein relation as a function of the measurement time. The top panel (1-ns temporal scale) refers to short-range diffusion, whereas the bottom panel (100-ns temporal scale) to long-range diffusion.

of Figure 5.17 we report the diffusion measurement over a 100-ns scale: the converged value is $12 \pm 3\,\mathrm{nm}^2/\mu\mathrm{s}$. This result is consistent with the value of $9\,\mathrm{nm}^2/\mu\mathrm{s}$ measured by pulsed field gradient NMR spectroscopy.[8] Our diffusion results, both for short and long time-scales, are also consistent with recent AL simulation data.[29]

The reason why lipid diffusion rates are different over short and long ranges can be understood by considering the *free-volume* theory.[115] According to the free-volume model, lateral diffusion occurs by discrete jumps of lipid molecules into nearby vacancies formed by lateral density fluctuations; in between jumps, a lipid molecule spends a relatively long time rattling in a cage formed by its neighbours. Over short ($< 1\,\mathrm{ns}$) time-scales, the diffusion coefficient is high because it is determined by the fast short-range lipid motion mainly due to a "rattling-about" behaviour. However, considering longer times, this rattling motion averages out yielding no net displacement. The true, long-range diffusion coefficient is determined by the lipid jumps, that give rise to effective displacement over longer ($> 10\,\mathrm{ns}$) time-scales. From our simulation, we can directly visualise the lipid diffusion mechanism by recording single-lipid trajectories and projecting them onto the bilayer plane. Figure 5.18 reports the mass centre lateral motion of selected lipids over a 100-ns time window: the "rattling" pattern, accompanied by occasional jumps, is clearly observable. To our knowledge, there are no experimental data directly showing the free-volume diffusive mechanism at this level of resolution. No explicit single-lipid trace has also yet been reported for other CG models. Results similar to ours have been obtained using AL models,[32–34] although from shorter simulations ($10 - 50\,\mathrm{ns}$).

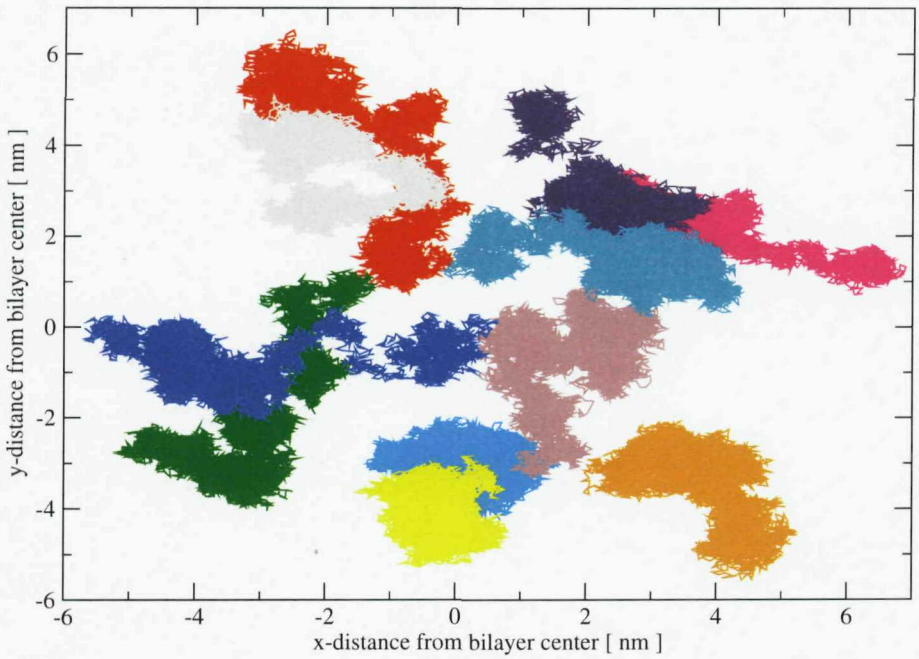


Figure 5.18: Mass-centre traces of selected lipids over a 100-ns time window. Each trace, identified by a different colour, represents the mass-centre motion of a single lipid projected onto the xy plane.

Water permeation

By simulating over relatively long times, we observe statistically-significant numbers of water molecules penetrating through the bilayer from one water phase to the other. In particular, over 900 ns we have detected 208 crossing events: 107 along the z axis positive direction, and 101 along the opposite (negative) direction. We have then computed the water permeability coefficient using Fick's law. The water concentration gradient is calculated as $\Delta C_{\mathrm{W}} = C_{\mathrm{W}}^{\text{water phase}} - C_{\mathrm{W}}^{\text{hydrocarbon core}} = 33\,\mathrm{nm}^{-3}$, assuming $C_{\mathrm{W}}^{\text{hydrocarbon core}} = 0\,\mathrm{nm}^{-3}$. Considering the interfacial area A and the unidirectional flux of water J_{W}, we compute a permeability coefficient $P_{\mathrm{W}} = J_{\mathrm{W}}/\Delta C_{\mathrm{W}} A = 92 \pm 4\,\mu\mathrm{m/s}$, in good agreement with the experimental estimate[3] of 70 μm/s. Such a calculation has been previously reported once for another CG model.[46] In that model however, groups of four water molecules are represented by generic solvent particles, without electrostatics. Also, the final permeation rate was reported to agree with the experimental value only after an empirical rescaling of the simulation time.[46] It is therefore difficult to consider realistic such a description of the spontaneous water permeation phenomenon. Passive water transport has never been quantified with AL models, due to the very demanding computational efforts required to simulate long trajectories. For example, in a recent AL study,[43] four phosphatidylcholine bilayers, each comprising 128 lipids, have been simulated for 50 ns. In the four simulations, 2, 4, 6 and 7 crossing events were respectively observed. It is evidently not possible to attempt an estimation of the permeability coefficient on the basis of such poor statistics.

5.3 Discussion

We have presented a new specific CG model for biological membrane systems. To some extent, our approach is similar to a number of CG membrane models that have recently been developed.[46–51] However, three main characteristics set our model apart from the other CG methodologies developed to date: explicit incorporation of the main electrostatics, realistic description of water, and use of anisotropic potentials to accurately capture lipid shape.

We employed our methodology to simulate a fluid-phase DMPC bilayer. The model was parameterised to reproduce basic experimental data (volume and area per lipid, order parameters) and the self-assembly process. The time-scale and overall self-aggregation mechanism are consistent with the available atomic-level data.[28] Simulating the self-aggregation phenomenon is computationally almost prohibitive for AL models (indeed only a single study has been reported so far[28]). Even among the specific CG methods, for three models such a result has not yet been reported,[49–51] and for another it proved very demanding, requiring "months of simulation".[48] With our method, the self-assembly process can be reproduced in a few days of simulation. All major physical features of our self-assembled bilayer system were then quantitatively evaluated: we stress the fact that these properties did not directly enter the parameterisation process. Structural results compared well with corresponding experimental data. In particular, the total electron density profile matched the curve obtained via experiment (Figure 5.11), and single-site profiles showed characteristic broad distribution peaks (Figure 5.12), consistent with the notion of fluid-phase lipid bilayers being highly disordered, yet distinctly stratified systems.[4] The headgroup dipole moment was also analysed: the experimental magnitude was reproduced, while the average orientation was found to be slightly different from that observed experimentally.

A thorough mechanical analysis was performed. First, compressibility moduli were evaluated. The area compressibility modulus describes the response of the bilayer surface area to an isotropic tension, whereas the volume (or bulk) compressibility modulus describes the response of the bilayer volume to uniform hydrostatic pressure.[2] Both parameters represent fundamental mechanical properties. The compressibilities computed for our model proved consistent with experimental measurement. The transbilayer lateral pressure profile was calculated (Figure 5.13). The profile was thoroughly analysed: every feature was described and related to its molecular origin. Although no quantitative experimental data is available to confirm our result, the profile of our model seems very reasonable: it shows the qualitative features obtained from experiments performed on similar systems,[11] and both shape and magnitude are consistent with AL simulation data.[39–45] Among the specific CG models, only ours has so far yielded a lateral pressure profile in agreement with experimental and AL simulation data. This crucial membrane property has been in fact recently reported for one other specific CG model: however, as mentioned in the Results section, that profile[194] shows qualitative and quantitative differences with respect to corresponding AL results.[39,43,44] Since electrostatic and hydration forces are major contributors to the transmembrane pressure distribution,[2] these discrepancies might be related to the simplified treatment of electrostatics

and water in that CG model.[194] We then analysed our simulation according to the elasticity theory developed by Helfrich.[77,97,98] The Helfrich formula, valid both for a monolayer or a bilayer, characterises the spontaneous shape of interfaces, and it predicts the stored energies that accumulate as a result of deviations from the spontaneous shape. In particular, from the first integral moment of the pressure distribution we determined the monolayer torque tension, which characterises the curvature elastic stress,[85,268] and the spontaneous curvature, which quantifies the tendency to curl and to form different phases.[77] Both these parameters are believed to specifically control protein function, membrane stability, phase behaviour and fusion.[85,268] Our model proved able to predict correctly for DMPC a monolayer curvature typical of *type-0*, bilayer-forming lipids. Through the second integral moment of the pressure profile, we computed the Gaussian curvature modulus, which in general describes the energy required to change the Gaussian curvature of a surface. From a biophysical perspective, the Gaussian curvature modulus sensitively affects the energy of intermediates in phase transitions and fusion of phospholipid bilayers,[80] and it is also predicted to strongly influence the membrane-mediated interactions between embedded proteins.[261] Our calculation proved in agreement with theoretical predictions.[22,98] It is important to highlight that the consistent results obtained for the curvature elastic parameters provide further confidence that the pressure profile obtained with our model is quantitatively credible.

Electrical properties were also investigated. In particular, we studied the membrane dipole potential and associated transbilayer electric field, for the first time using a CG model. In fact ours is the only CG method developed to date that can capture such properties, as it fully represents the electrostatics present in hydrated membrane systems: headgroup charges, and glycerol and water dipoles, are explicitly incorporated, and interactions are realistically modelled without artificial dielectric constants. Our model correctly predicted the sign of the membrane potential, the magnitude being larger than the experimental estimate. Notably, we addressed an important issue which cannot be directly studied in real systems, and which still represents an open question:[101] the molecular origin of the dipole potential. In particular, by separating individual contributions, we could identify the glycerol-ester dipoles as the major contributors to the electric field and dipole potential (Figures 5.14, 5.16), and we could quantify the strong influence of interfacial water ordering (Figure 5.15). Both these observations confirm hypotheses based on experimental data.[109,266]

Realistic lipid diffusion was also simulated. For the first time a CG model has proved able to reproduce the experimentally-measured diffusion coefficients. As mentioned previously, other CG models yield lipid diffusion coefficients which are four[46,58] to one hundred[52] times higher than experimental data. This has lead to an overall correction of the simulation time according to these lateral diffusion "speed-up" factors:[46,52] such rescaling is problematic,[60] as it neglects the high temporal heterogeneity characterising membrane processes. Our results, both for short-range and long-range diffusion coefficients, proved consistent with experiments (Figure 5.17). Furthermore, single-lipid diffusion trajectories were extracted (Figure 5.18). We observed a "rattling and jumping" behaviour of the kind predicted by free-volume the-

Table 5.5: Physical parameters of fluid-phase DMPC phospholipid bilayer

Parameter	Our model	Experiment[reference]
V_L [nm^3]	1.104 ± 0.002	1.101[6]
A_L [nm^2]	0.594 ± 0.003	0.606[6]
d_{HH} [nm]	3.71 ± 0.02	3.53[6]
S_{mol}	0.36 ± 0.01	0.38[7,75]
μ_{HG} [D]	16.1 ± 0.1	18.7[244*]
θ_{HG} [deg]	92.3 ± 0.1	≈ 72[74*]
K_A [dyn/cm]	297 ± 22	234[9]
K_V [kbar]	13.7 ± 0.2	10 to 30[2]
κ^b [k_BT]	21.9 ± 1.6	13.5,[9] 16.7,[246] 31.4,[247] 32.1[248]
τ^m [k_BT/Å]	-0.020 ± 0.003	*na*
c_0^m [nm^{-1}]	-0.018 ± 0.003	≈ 0
κ_G^m [k_BT]	-5.4 ± 0.1	-10.9 to 0[22]
κ_G^b [k_BT]	-10.9 ± 2.3	*na*
Ψ_d [V]	1.57 ± 0.03	0.45,[111] 0.51[10†]
D_{lat} [nm$^2/\mu$s]	12 ± 3	9[8]
P_W [μm/s]	92 ± 4	83,[269] 70[3]

V_L = volume per lipid, A_L = area per lipid, d_{HH} = bilayer thickness, S_{mol} = segmental order parameter, μ_{HG} = magnitude of the headgroup dipole, θ_{HG} = angle between the headgroup dipole and the bilayer normal, K_A = area compressibility, K_V = volume compressibility, κ^b = bilayer bending rigidity, τ^m = monolayer torque tension, c_0^m = monolayer spontaneous curvature, κ_G^m = monolayer Gaussian curvature, κ_G^b = bilayer Gaussian curvature, Ψ_d = dipole potential, D_{lat} = lipid lateral diffusion, P_W = water permeability, *na* = not available. *Fluid-phase DPPC. †Fluid-phase DPhPC.

ory[115] and simulated in AL studies.[32–34] Finally, long-time simulation (almost 1 μs) allowed the spontaneous permeation of water molecules to be observed and quantified: we measured a permeability coefficient consistent with experiment. To our knowledge, such a result has never been reported using AL models. The spontaneous water permeation rate has been previously computed once using a specific CG model,[46] although in that model water is represented by simple apolar particles, each accounting for four real water molecules. Also, the permeation rate was obtained after rescaling the simulation time according to the lipid diffusion "speed-up" factor.[46] Our simulation of the water passive transport phenomenon is more realistic, as water is represented on an individual level, electrostatics are explicitly included and the permeation rate is computed without any empirical correction.

The entire set of quantitative results obtained from our simulation is collected in Table 5.5, along with the corresponding experimental data. In general, experimental investigation on phospholipid bilayers suggests that the properties of membranes are governed by basic, purely-physical principles.[4] Our model is a demonstration of the validity of this view: we have consistently reproduced experiments with a simple model that includes only the fundamental physics.

A unique and very promising aspect of our CG strategy is the straightforward compatibility with AL models: the potentials present in our force-field can be readily mixed with AL force-fields. In fact, electrostatic and van der Waals interactions are represented consistently by our CG model and standard AL models using Coulombic and Lennard-Jones potentials. Our model also comprises the anisotropic Gay-Berne potential, which can be seen as a generalisation of the Lennard-Jones potential. Gay-Berne and Lennard-Jones models can be mixed through a generalised formalism,[221] as we have already done to describe headgroup-tail and water-tail interactions. Recent work in our group has been performed to validate and calibrate the multiscale interactions. In that study,[270] water-octane partition coefficients for a range of solutes have been calculated by Monte Carlo simulations. Octane molecules have been modelled as pairs of Gay-Berne ellipsoids connected by a harmonic spring, using the parameters developed here for the lipid hydrocarbon region, water has been represented with the SSD potential, and the solutes with an all-atom force-field. This mixed AL-CG study has yielded results in good agreement with experimental data, using a fraction of the computational time needed by corresponding AL models.[270] Multiscale simulations have also been conducted to study the permeability of a number of solutes (small organic molecules, large drugs and hormones) described at the all-atom level using standard force-field parameters, inserted across our CG bilayer. These simulations are presented in the following Chapters 6 and 7. Overall, it is clear how our approach can allow the efficient multiscale mixing of CG and AL models, the latter employed to represent molecules where fine chemical detail is important, such as drugs or membrane proteins.

Chapter 6

Multiscale simulation of small molecule permeability

The transmembrane permeation of small (molecular weight < 100) organic molecules across a phospholipid bilayer is investigated by multiscale molecular dynamics simulation. The bilayer and hydrating water are represented by the simplified, efficient coarse-grain models presented in Chapter 5, whereas the permeating molecules are described by standard atomic-level models. By constraining each permeant at selected depths inside the bilayer, we have sampled free energy differences and diffusion coefficients across the membrane. These data have been combined, according to the inhomogeneous solubility-diffusion model, to yield the (experimentally measurable) permeability coefficients. Results are within one order of magnitude of previous atomic-level calculations and available experimental data. In particular, the solute relative permeability coefficients are well reproduced. Computationally, our multiscale hybrid methodology proves two orders of magnitude faster than traditional atomic-level methods.

6.1 Introduction and background

Transport phenomena across biomembranes are fundamental processes in cellular biology. They are also becoming increasingly important in many medical, pharmaceutical and environmental technologies.[271] For example, drug permeation is crucial to bioavailability, and is at the basis of the technology of liposomal transport systems.[158] Although important permeation mechanisms, such as those responsible for the translocation of sugars and amino acids, are actively controlled by proteins, *passive* permeation is the most common way by which solutes cross cell membranes. Most small molecules (such as water and oxygen) and drugs are passively transported. While experiment can measure permeability coefficients, the exact mechanism of unassisted transmembrane transport is still not fully understood, as local membrane-solute interactions are difficult to probe. In fact, the current understanding of membrane permeability is still influenced by the theories developed over a century ago by Overton, who proposed that the membrane permeability coefficient of a solute is correlated

to its oil/water partition coefficient.[272] This observation lead to the crude representation of the membrane as a homogeneous oil slab: on this basis, the simple bulk solubility-diffusion model of membrane permeability was proposed.[273] In more recent years, experiments have clearly established that lipid membranes are highly heterogeneous systems, very different from uniform oil phases: for instance, density distributions, order parameters and diffusion in lipid membranes show characteristic properties that are not present in bulk oil systems. The heterogeneity present inside membranes is included in the *inhomogeneous* solubility-diffusion model[274–276] (detailed in Section 6.3.1), which relates the permeability coefficient to the variations of local properties across the membrane: these properties, such as partitioning and diffusion of a solute as a function of its position inside the membrane, are very difficult to study experimentally.

Particle-based simulations can provide insights into the understanding of transport phenomena across bilayers with the necessary resolution. In recent years, atomic-level (AL) molecular dynamics simulations have indeed been successfully employed to predict permeability coefficients and to investigate the general mechanism of passive transport across membranes. Seminal simulations were performed to study the diffusion of benzene in a lipid bilayer,[277] although the first calculation of the transbilayer permeability coefficient by molecular dynamics was conducted for water.[276] Permeability simulations have then been carried out for small molecules such as ammonia and oxygen.[278–280] The increase in computer power and the refinement of force-fields for different molecules have also allowed the calculation of permeability coefficients for several small organic molecules[35,281] and drugs.[282–285] A detailed review on AL simulation studies of membrane permeation has recently been published.[158] These investigations have been extremely useful in understanding many aspects of bilayer permeation with atomic resolution. However, the huge computational cost associated with the simulation of AL models results in a series of limitations and issues. For example, obtaining well-converged data is often problematic, as series of long simulations are required for every solute. Also, bilayer sizes must be rather small to be computationally amenable; this can induce artefacts, especially when large drugs are inserted into the membrane. Furthermore, the number of different solutes that can be investigated in a reasonable amount of time is extremely limited; this seriously hinders applications in the context of drug design, where screenings of large sets of candidate compounds are normally required. It is therefore highly desirable to develop techniques that can improve the efficiency of simulation.

A possible way to alleviate the AL computational cost involves simplifying the representation of the system via coarse-grain (CG) techniques. CG approaches generally involve grouping together selected clusters of atoms into single macro-sites, to significantly reduce the number of interactions calculated, and hence also the computational cost. CG methods can increase simulation speed by several orders of magnitude with respect to corresponding AL methods, while still retaining the most important physical features of the systems and phenomena represented.[139,142,144] However, membrane permeability is known to be extremely sensitive to the chemical identity of the permeating species: subtle variations in

the solute atomic structures often lead to differences of many orders of magnitude in the permeability coefficients.[286] Standard CG models, where groups of several atoms are lumped together in single interaction sites, are unlikely to provide a sufficiently accurate description of membrane-solute interactions. A natural compromise would involve combining the accuracy of AL force-fields to the efficiency of CG models in a multiscale fashion: the "chemically sensitive" parts of the system (the solutes in membrane permeation studies) can be modelled atomistically, and the surrounding environment can be simplified with CG representations. It is particularly desirable to employ standard AL force-fields, which have been developed, tested and successfully used for decades now. Several multiscale methods have been reported in the literature;[61,287–294] a review has also been recently published.[295] While current methods are very promising, quite a few issues can be noted. All the AL-CG multiscale approaches developed so far are characterised by rather complex algorithms developed to model specifically the interfacial region between the AL and CG parts of the system. None of these techniques has been used to simulate transmembrane permeation processes. Applications to biological systems have been so far limited to the modelling of basic structural and dynamical properties of an enzyme,[289,290] and to the study of the structure of preassembled membrane-protein systems.[61] In these approaches, problems arise whenever molecules diffuse across regions of different representations, as the models for the AL and CG particles are not compatible. Adaptive techniques, which address this issue by allowing changes of resolution across the interface, have been proposed only recently, and thus far they have been proved to work only for rather simple models, such as idealised tetrahedral solvent particles,[292] liquid methane,[287,291] water[294] and a generic solvated bead-spring polymer.[293] A general serious problem involves the preservation of realistic dynamics, as well as the consistency of relative dynamics across the different representation levels.[295]

In this study, we present multiscale simulations based on a simple, *direct* AL-CG coupling, where the different representations interact through compatible potentials. Such a straightforward interfacing is possible thanks to the characteristics of our recently-developed CG membrane model (presented in Chapter 5). The membrane permeation process will be simulated for a range of small solutes. In the next section, we summarise the theory underlying our calculation and we describe in details the multiscale methodology employed. Results obtained by multiscale AL-CG simulation are then presented, and compared to standard AL simulation results and available experimental data.

6.2 Experimental studies

The fundamental principle of passive permeation is described by Fick's first law of diffusion: a substance diffuses in the direction that eliminates its concentration gradient, at a rate proportional to the magnitude of this gradient. The permeability coefficient P, representing

this proportionality constant, can be calculated as:

$$P = \frac{J}{A\,\Delta C}$$

with J the solute's flux, ΔC its concentration gradient across the interface and A the interface area.

There is a large scatter of experimental values for the membrane permeability coefficients. While the relative permeabilities are typically well reproduced, absolute data measured in different laboratories by different techniques can vary across orders of magnitude. This can be due to difficulties in calibrating the measurements and to perturbations caused by the specific method.[296]

Water Water permeability across lipid vesicles can be quantified by measuring the net volume flow of water induced by osmotic water transport in the presence of an impermeable solute. Another method, where no net flux occurs, involves measuring the diffusional exchange of water by tracer methods using deuterated or tritiated water. These two methods do not yield consistent results: the "osmotic" permeability coefficient is typically larger, by up to 2 orders of magnitude, than the "diffusive" permeability coefficient.[297] An attempt to explain this discrepancy considers the *single-file transient pore* model.[297,298] According to this model, water permeates through pores forming from density fluctuations in the bilayer. The section of such pores is only large enough to allow a single file of water molecules to pass through. When water molecules cooperatively move in one direction across the bilayer, as in the osmotic case, the transient pores offer a major permeation route. In the diffusive permeation experiment, this process does not take place; water molecules do not have a preferential direction and cannot pass each other in the pores, so there is no pore flux. It is predicted that one pore is present every 2000 lipid molecules.[297] The transient pore mechanism is not widely accepted, in fact many researchers consider the solubility-diffusion model (presented in Section 6.3.1) more realistic.[299,300]

Small solutes Permeability coefficients for small organic molecules can be measured across planar lipid bilayers formed on a $\approx 1\,\mathrm{mm}^2$ hole in a polyethylene or Teflon partition separating two magnetically-stirred water-jacketed chambers;[299,301–303] Figure 6.1 depicts a typical system. Permeability coefficients P are calculated from the rate of change of the receiver concentration with the following equation:

$$P = s \times \frac{V_{\text{chamber}}}{AC_{\text{donor}}} \tag{6.1}$$

with s the slope of the *receiver concentration versus time interval* plot, V_{chamber} the volume of the aqueous solution in each chamber, A the bilayer area and C_{donor} the concentration of the solute in the donor chamber.[301] Permeability measurements can also be performed across lipid bilayers from large unilamellar vesicles by means of radioactive tracers, NMR[296]

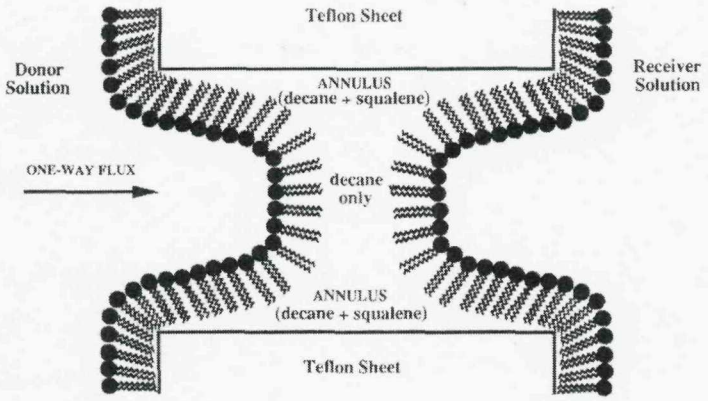


Figure 6.1: Membrane permeability experimental system. The figure highlights the increase in thickness due to the residual hydrocarbon solvent (decane + squalene) which inevitably remains in the bilayer after the formation process; in fact this solvent layer is exaggerated in the figure, as in reality the thickness increase is negligible.[301]

or fluorescence techniques,[300] by withdrawing samples of the receiving solution or collecting the vesicles at intervals of time.[304]

6.3 Methodology

6.3.1 The inhomogeneous solubility-diffusion model

The inhomogeneous solubility-diffusion model[274–276] relates the (experimentally measurable) permeability coefficient of a solute to an integral of depth-dependent parameters across the membrane. In particular, the overall membrane permeability coefficient P is expressed as:

$$P = 1/\int_{z_1}^{z_2} R(z)\,\mathrm{d}z = 1/\int_{z_1}^{z_2} \frac{\exp(\Delta G(z)/k_\mathrm{B}T)}{D_z(z)}\,\mathrm{d}z \tag{6.2}$$

with $R(z)$, $\Delta G(z)$ and $D_z(z)$ the solute resistance, excess free energy and diffusion coefficients, respectively, at position z along the direction normal to the membrane interfacial plane. The integration extremes z_1 and z_2 are taken in the water phases at the two sides of the membrane, so that the integration is performed over the entire bilayer.

6.3.2 The z-constraint method

The quantities featuring in the inhomogeneous solubility-diffusion model (Equation 6.2) can be obtained from simulation by applying the *z-constraint* method.[276] In particular, both $\Delta G(z)$ and $D_z(z)$ can be calculated from the constraining force $f_z^\mathrm{c}(z)$ required to keep the solute mass centre at selected z-positions along the interface normal. The z-constraint algorithm, applied at every MD step, consists of the following operations:

- Consider $z-$position z_t of solute molecule at the current time t

- Compute forces and integrate as usual → new position $z_{t+\Delta t}$ and velocity $\dot{z}_{t+\Delta t}$
- Compute displacement: $\Delta z := z_{t+\Delta t} - z_t$
- Constrain position $z_{t+\Delta t} := z_{t+\Delta t} - \Delta z$ and velocity $\dot{z}_{t+\Delta t} := 0$
- Compute constraining force: $f_z^c := -m\Delta z/(\Delta t)^2$

The constraining force at each step is recorded for the post-processing required to obtain free energies and diffusion coefficients. The free energy of transfer $\Delta G(z)$ from water into the membrane is computed as:

$$\Delta G(z) = \int_{water}^{z} \langle f_z^c(z') \rangle \, \mathrm{d}z' \tag{6.3}$$

with $\langle f_z^c(z') \rangle$ the average constraint force at position z' over the total simulation time. The local diffusion coefficient $D_z(z)$ along the z-dimension is calculated as:[305]

$$D_z(z) = \frac{(k_B T)^2}{\int_0^\infty \langle \Delta f_z^c(z,t) \, \Delta f_z^c(z,0) \rangle \, \mathrm{d}t} \tag{6.4}$$

with k_B the Boltzmann constant, T the temperature and $\Delta f_z^c(z,t)$ the "random" force, defined as the deviation of the instantaneous force from the average force acting on the solute: $\Delta f_z^c(z,t) = f_z^c(z,t) - \langle f_z^c(z) \rangle$.

6.3.3 Atomic-level models of small molecules

We have selected the following solute candidates, representing the most common chemical functional groups: acetamide, acetic acid, benzene, ethane, methanol, methylacetate, methylamine and water. We have chosen to test our AL-CG approach on this particular set of molecules because we will be able to make comparisons with the results obtained in a corresponding AL investigation.[35] Water is modelled by the SSD potential,[225,227] with increased masses as described in Section 5.1.7. The other solutes are described by all-atom models; in particular, these molecules have been parameterised with the GAFF force-field[306] and AM1/BCC atomic partial charges.[307] The solute models have no net charge; this is consistent with the assumption that only the neutral species diffuse across the bilayer.[301] No intramolecular degrees of freedom are taken into account, that is, molecules are rigid bodies. All solutes are small, relatively compact molecules: it seems therefore reasonable to model them as rigid bodies. To optimise the stability of molecular dynamics numerical integration, the solute principal moments of inertia were increased.[242] In particular, atomic masses in each solute were redistributed: hydrogen masses were increased from 1 to 4 amu, and the masses of heavier atoms (carbon, oxygen and nitrogen) were decreased accordingly to maintain the real value of the overall mass of each molecule. Thermodynamic properties are not affected by such an alteration of the principal moments of inertia.

In general, a rigid body is associated with an inertia tensor whose elements are typically non-zero. The inertia tensor $\mathbf{I}$ can be considered a linear operator that transforms the angular velocity $\boldsymbol{\omega}$ into the angular momentum $\mathbf{L}$:

$$\mathbf{L} = \mathbf{I}\boldsymbol{\omega} \tag{6.5}$$

with $\mathbf{L}$ and $\boldsymbol{\omega}$ column matrices, and $\mathbf{I}$ a second-rank tensor.[308] Explicitly:

$$\mathbf{I} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \tag{6.6}$$

The inertia coefficients depend upon both the location of the origin of the body set of axes and upon the orientation of these axes with respect to the body. It can be shown that there always exists a set of coordinates in which the tensor is diagonal with three principal values: the principal moments of inertia I_x, I_y, I_z. The principal moments of inertia are simply the eigenvalues of the inertia tensor, while the principal axes are the eigenvectors of the inertia tensor.

Setting up a rigid molecule for molecular dynamics Assuming to have an initial set of coordinates (obtained for instance from a pdb file) in any given reference, and knowing the mass of each atom, it is straightforward to compute the molecule's mass centre and rescale the coordinates so that the origin of the frame of reference coincides with the molecule's mass centre. The inertia tensor $\mathbf{I}$ is then computed. In particular, the diagonal elements, also known as *moment of inertia coefficients*, have the following form:

$$I_{xx} = \sum_i m_i(r_i^2 - x_i^2) \tag{6.7}$$

while the off-diagonal elements, also known as *products of inertia*, have the typical form:

$$I_{xy} = -\sum_i m_i x_i y_i \tag{6.8}$$

being m_i the mass of the i-th atom, $r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$ its distance from the origin and x_i, y_i, z_i its coordinates.[308] The eigenvalues (principal moments of inertia) and the eigenvectors (principal axes) of the inertia tensor can be computed (using for instance the program Maple[309]). The principal axes can then be used to form the rotation matrix $\mathbf{R}$, which can eventually be employed to transform the molecule coordinates from the initial (*lab*) frame onto the principal (*body*) frame of reference. The molecule can now be assigned the principal moments of inertia, and its motion can be simulated by molecular dynamics.

6.3.4 Mixed interactions

In general, all the potentials employed to represent the various interacting sites in our multiscale membrane-water-solute model are directly compatible with each other. Therefore, mixed interactions can be treated straightforwardly by available formulae.

Lennard-Jones cross terms between lipid headgroups, water and solute atoms are calculated using standard mixing rules.[120] Mixed Lennard-Jones/Gay-Berne interactions are consistently treated using the generalised Gay-Berne potential.[221] In particular, the generalised Gay-Berne potential accounts for the intralipid headgroup-tail interactions, the interaction between the Lennard-Jones term of the SSD potential and the Gay-Berne lipid term, and the interaction between the Lennard-Jones terms of the AL solutes and the Gay-Berne lipid term.

In our model we retain explicitly all relevant electrostatics of the lipid, water and solute molecules, as described previously. Therefore, monopole-monopole, monopole-dipole and dipole-dipole interactions can be simply described by standard electrostatic formulae.[126]

To optimise the mixed parameters between CG and AL sites, we recently carried out calculations of water-octane partition coefficients for a range of solutes by Monte Carlo simulations.[270] Octane molecules were modelled as pairs of Gay-Berne ellipsoids connected by a harmonic spring, using the parameters developed for the hydrocarbon tails of the CG lipid model; water was represented with the SSD potential,[225,227] and the solutes with the GAFF all-atom force-field. The solutes were 15 analogues of neutral amino acid side chain. Results could be brought in excellent agreement with experimental data by introducing two scaling factors into the standard mixing rules: i) The electrostatic potential energy term between water and solutes was increased by a factor of 0.1. ii) The generalised Gay-Berne potential energy term between lipid tail sites and solutes was decreased by a factor of 0.2. Such modifications have been adopted in this study.

6.3.5 Simulation details

Molecular dynamics is carried out with our software BRAHMS (see Section 5.1.5 in Chapter 5 for a general description and Appendix A for algorithmic details). The integration time-step is 20 fs. Pressure and temperature are maintained at 1 atm and 30°C using the weak-coupling scheme.[237] Lipid, water and solute temperatures are coupled separately with time constant $\tau_T = 0.2$ ps; for rigid-body sites, translational and rotational degrees of freedom are coupled independently. The pressure is controlled by semi-isotropic volume scaling with time constant $\tau_P = 0.5$ ps and isothermal compressibility $\beta = 4.6 \times 10^{-5}$ atm^{-1}. The cutoff radius for both Lennard-Jones and electrostatic water-water interactions is 0.9 nm, as prescribed for the SSD parameterisation adopted.[227] All other nonbonded cutoff radii, both for Lennard-Jones and electrostatic interactions, have been set to 1.2 nm. Long-range electrostatics are treated using cutoff schemes. In particular, charge-charge and charge-dipole interactions are implemented using the shifted-force cutoff method.[120] We employ the SSD

parameters optimised to treat long-range dipole-dipole interactions with a cubic switching cutoff scheme:[227] for consistency, all dipole-dipole interactions are treated in this manner. Nonbonded interactions involving solute molecules are treated as group-based with a cutoff distance of 1.2 nm: the interactions between all solute atoms and the interacting site are evaluated if, and only if, the distance between the solute mass centre and the interacting site is less than the cutoff. The net mass centre velocity of the entire system is set to zero at every step.[238] The bilayer model employed is the self-assembled CG membrane comprising 128 DMPC lipids and 3400 hydrating water molecules described in Chapter 5. For each of the AL solutes, the z-constraint method has been applied to sample 16 equally-spaced z-positions across one monolayer; results are considered valid also for the other monolayer for symmetry. In particular, we have sampled distances from 3.1 nm to 0.1 nm by 0.2 nm increments, thus covering the various intramembrane environments from the interfacial water phase to the centre of the hydrocarbon core. In each simulation, a single solute molecule was present in the bilayer. To prepare the systems, molecules were initially inserted at 1% of their real size, and with interaction parameters set to 1% of their real values. The solutes were then incrementally grown back over 1000 molecular dynamics steps, corresponding to 20 ps of simulation time. The systems were subsequently equilibrated for 1 ns. Production runs were then conducted for 80 ns, divided in two consecutive 40-ns batches. Simulations were run on the Southampton University Iridis cluster[240] in a "coarsely" parallel fashion, meaning that each simulation ran on a different single processor. The entire set of simulations could be completed in about one week.

6.4 Results

In this section, we report the results obtained from the multiscale AL-CG permeability simulations: Figure 6.2 displays a typical simulation snapshot. For the free energy, diffusion and resistance profiles, and for the overall permeability coefficients, we will report average values and standard errors computed from the averages over the two 40-ns consecutive blocks of each of the 80-ns runs. The data obtained will be compared to previous AL simulation studies and available experimental data. The CG membrane is a model for DMPC bilayers, hence ideally comparisons should be made with results obtained with this lipid species. In fact, some experimental and simulation data for DMPC will be reported. However, we will consider many results obtained for DPPC lipid bilayers, as this was the lipid species employed in the AL investigation which constitutes our primary comparison source.[35] Apart from slightly longer tails (two more carbons), DPPC is identical to DMPC; moreover, all data reported refer to fully-hydrated bilayers in the liquid-crystalline phase. Therefore it is reasonable to compare results for DMPC to corresponding data for DPPC. We will also consider experimental permeability coefficients derived from studies on lecithin bilayers, which are mixtures of different phosphatidylcholine lipids.

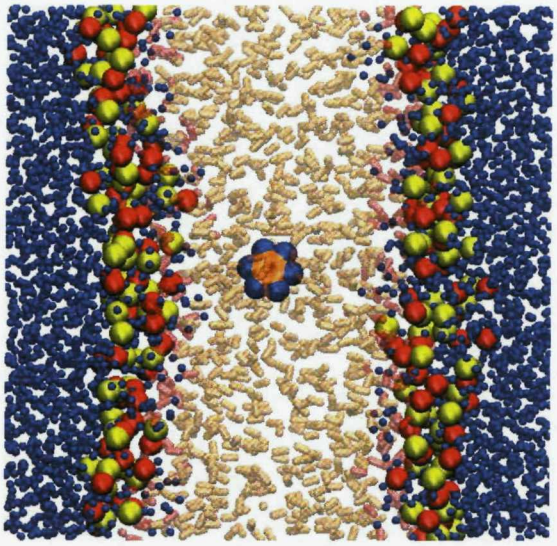


Figure 6.2: Simulation snapshot of an AL solute embedded in the CG membrane. The CG choline, phosphate, tail and water sites are represented in red, yellow, transparent orange and blue, respectively. The atoms of the AL solute, benzene in this image, are represented in orange (carbon) and blue (hydrogen). Image prepared with VMD.[310]

6.4.1 Free energies of transfer

For each of the solutes considered, the free energies of transfer from the water phase to the selected $z-$positions inside the membrane are reported in Figure 6.3. By following the free energy profiles of the solutes from the water phase to the bilayer centre, it is possible to identify clearly two different types of behaviours, which correspond to the hydrophilic or hydrophobic nature of the molecules. The hydrophilic solutes (acetamide, acetic acid, methanol, methylacetate, methylamine and water) are characterised by a net free energy increase; in particular, a shallow free energy dip corresponding to the polar/apolar interface region, observed for all hydrophilic solutes except water, is followed by a steep increase corresponding to the hydrocarbon region, culminating in a large positive free energy barrier at the bilayer centre. The hydrophobic molecules (benzene and ethane) show instead a net decrease in the free energy, reaching a negative minimum at the bilayer centre. Ethane, in particular, is characterised by an almost zero free energy difference up to the lipid glycerol region; then, in the hydrocarbon region, the free energy decreases until the negative minimum in the bilayer centre. For benzene, however, the free energy decreases almost constantly from the water/headgroup interface to the hydrocarbon region, where it eventually settles down to a negative value in the vicinity of the bilayer centre. Such a clear discrimination between hydrophilic and hydrophobic compounds is intuitively expected, considering the hydrophobic nature of the membrane hydrocarbon core. A close comparison between the AL-CG free energy profile and the corresponding curves calculated in the AL study[35] is shown in Figure 6.4. Qualitative differences can be noticed corresponding to the lipid headgroup-glycerol region.

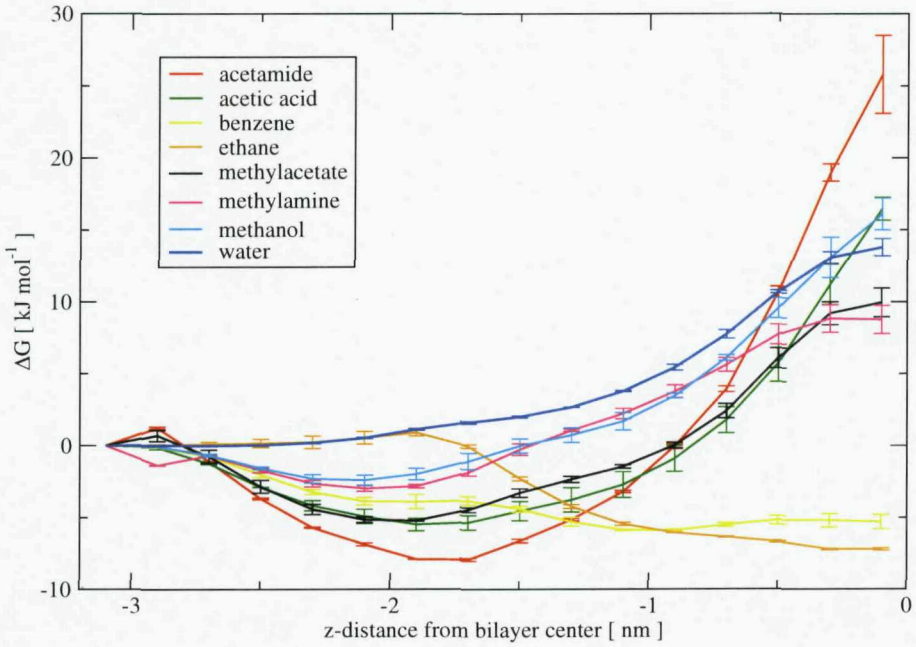


Figure 6.3: Free energies of transfer from water to selected $z-$positions along the bilayer normal.

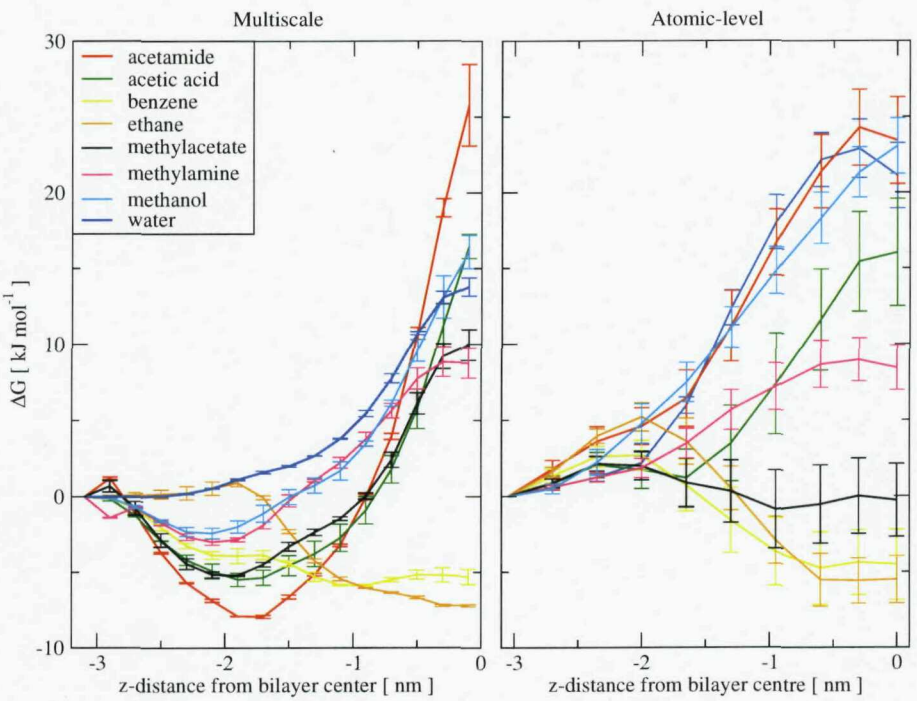


Figure 6.4: Free energies of transfer from water to selected $z-$positions along the bilayer normal. Comparison between the results from multiscale (this work, left diagram) and atomic-level[35] (right diagram) simulations.

Table 6.1: Free energy of transfer from the water phase to the bilayer centre [kJ/mol]

Solute	AL-CG (this work)	AL
acetamide	25.8 ± 2.7	24.3 (35)†
acetic acid	16.5 ± 0.8	16.0 (35)†
benzene	-5.3 ± 0.5	−4.8 (35)†
ethane	-7.2 ± 0.1	−5.6 (35)†
methanol	16.1 ± 1.1	23.1 (35)†
methylacetate	10.0 ± 1.0	0.3 (35)†
methylamine	8.8 ± 1.0	9.0 (35)†
water	13.8 ± 0.6	22.9 (35)†
		26.0 (276)†
		22.5 (280)*
		26.4 (279)†
		56.5 (311)*

References are included in round brackets. *Fluid-phase DMPC. †Fluid-phase DPPC.

In particular, all the AL profiles display a free energy increase on entering the headgroup region;[35] corresponding to the same location, the profiles reported in this work are instead characterised by a shallow dip, with the exceptions of water and ethane. To our knowledge, the only available experimental data on the free energy difference between the headgroup region and the water phase for the solutes considered has been obtained for acetic acid: from the partition coefficient in DMPC vesicles, a value $\Delta G \approx -0.45\,\mathrm{kJ/mol}$ can be calculated.[304] Such value is negative as in our AL-CG results; however, there is a discrepancy of a factor of 10 in terms of the absolute values. A quantitative comparison can be made for the free energy differences at the bilayer centre: Table 6.1 reports the results obtained in this AL-CG investigation along with corresponding AL values. It can be noticed an overall good agreement, with the single exception of methylacetate, which displays a large positive free energy value in the AL-CG simulations whereas it is almost zero in the AL study.[35] In fact, methylacetate is a hydrophilic, polar compound, as its experimental ΔG(water→hexadecane) is reported[312] to be positive (2.7 kJ/mol); our result appears therefore more realistic than the AL value. A further discrepancy between the AL-CG data and AL results concerns the free energy barrier for water. In particular, our barrier of 13.8 kJ/mol is somewhat lower than the range 22.5 – 26.4 obtained from AL simulations (Table 6.1). Part of the reason for this underestimation is accounted for by the fact that our AL-CG method is optimised for AL molecules parameterised with the GAFF force-field, whereas the water model is represented by the SSD force-field.

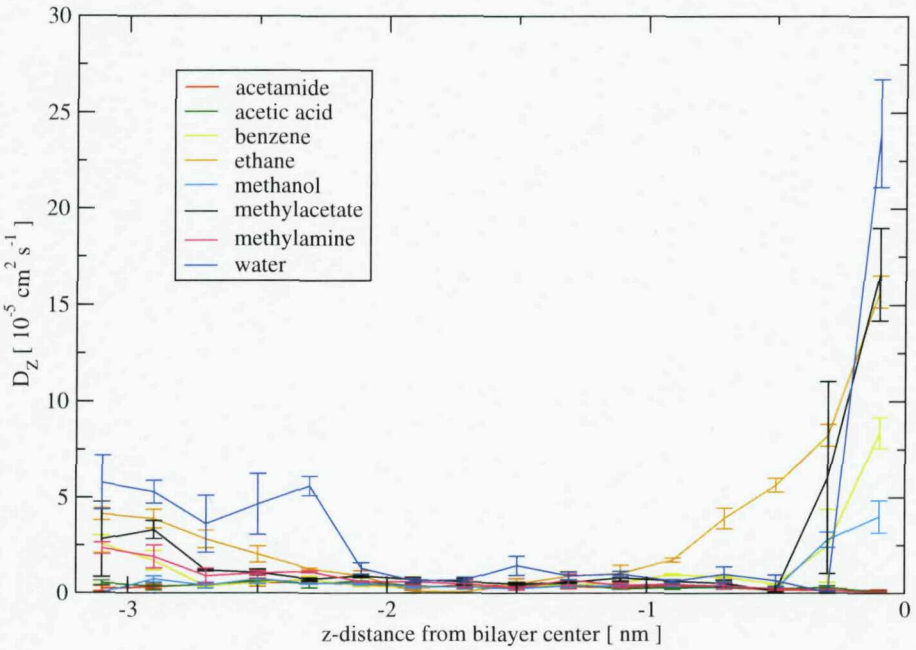


Figure 6.5: Diffusion coefficients along the $z-$dimension (normal to the membrane plane) for selected $z-$positions inside the bilayer.

6.4.2 Diffusion coefficients

Diffusion coefficients $D_z(z)$ along the z-dimension have been calculated using Equation 6.4: diffusion profiles are displayed in Figure 6.5. The values obtained with the AL-CG model are within one order of magnitude of the AL results reported in the simulation study of the same set of solutes.[35] However, qualitative discrepancies from that study[35] can be noticed for a number of solutes. In particular, Figure 6.5 shows that the diffusion coefficients of water, methylacetate, ethane, benzene and methanol are higher in the centre of the bilayer than in the outer water phase: these data do not agree with those reported in the AL simulation study[35] where, for all solutes, diffusion coefficients in the bilayer centre were found to be significantly lower than in the water phase. However, two other previous AL investigations reported results similar to those obtained here. In the water permeability simulation by Marrink and Berendsen,[276] the diffusion rate in the bilayer centre was reported to be more than twice as large as in the water phase. As noted elsewhere,[158] this discrepancy may stem from Marrink and Berendsen's modelling of the -CH_2- and -CH_3 groups in the lipid tails without explicit hydrogens, that is, using a *united-atom* representation. However, Shinoda et al.[279], using an explicit-hydrogen model, also reported the diffusion coefficient for water in a DPPC bilayer to be twice as high in the hydrocarbon centre than in the water phase. Moreover, in another molecular dynamics investigation of benzene in a DMPC bilayer with all hydrogen explicitly modelled, it was found that benzene diffused three times faster in the bilayer centre than in the interfacial region;[277] in this case though, the total diffusion coefficient was calculated, as opposed to the z-component only (which is considered in all other referenced works). In all these studies,[276,277,279] the enhanced diffusion was ascribed to

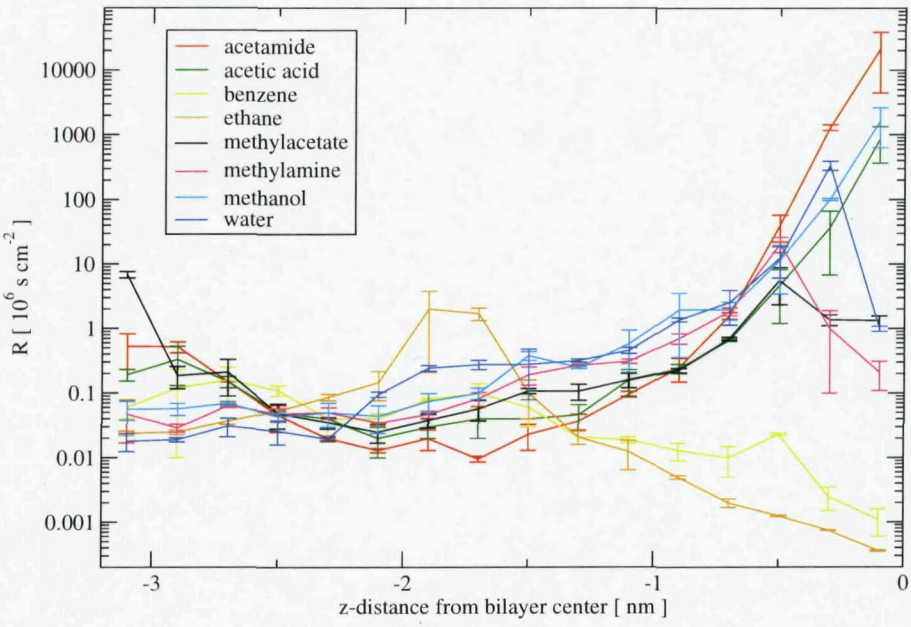


Figure 6.6: Resistance profiles.

the small size of water and benzene: they diffuse faster by jumping between the free volume pockets available in the middle of the bilayer, which is in fact the lowest-density region of the system.

6.4.3 Resistances

Resistance profiles $R(z)$ as a function of the bilayer normal, calculated from $\Delta G(z)$ and $D_z(z)$ with Equation 6.2, are displayed in Figure 6.6. It can be seen that the hydrophilic solutes (acetamide, acetic acid, methanol, methylacetate, methylamine and water) are characterised by a net resistance increase from the water phase to the hydrocarbon core, whereas the hydrophobic molecules (benzene and ethane) display a net resistance decrease. In general, the resistance profiles of Figure 6.6 show a clear correlation to the free energy profiles of Figure 6.3. This is expected, as in the solubility-diffusion model employed (Equation 6.2) the resistance $R(z)$ depends largely on the free energy difference $\Delta G(z)$, which features as the argument of an exponential, while the contribution of the diffusion coefficient $D_z(z)$ is comparatively small. However, in some resistance profiles it is possible to identify slight deviations from the shape of the corresponding free energy profiles, and these deviations can be ascribed to the diffusion coefficients (Figure 6.5). The local resistance is inversely proportional to the diffusion coefficient (Equation 6.2): the observed enhanced diffusion of some solutes in the bilayer core should therefore affect resistances in that region. Such effect can be seen for water and methylacetate, which display a resistance drop in the bilayer centre despite the free energy increase observed in Figure 6.3. The resistance decrease is clearly determined by increased diffusion. As for benzene and ethane, which also diffuse faster in the bilayer centre, it can be observed that the decrease in resistance, already dictated

by the free energy profiles, is particularly steep. The results for methanol underline the predominant effect of the free energy over the diffusion term in determining the overall resistance. Methanol displays enhanced diffusion in the bilayer centre (Figure 6.5), which favours a resistance decrease in that region. However, this effect is overcompensated by the steep free energy increase (Figure 6.3), which ultimately dictates the observed resistance increase (Figure 6.6).

6.4.4 Permeability coefficients

The free energies of transfer $\Delta G(z)$ and diffusion data $D_z(z)$ have been combined according to the inhomogeneous solubility-diffusion model[274–276] (Equation 6.2) to calculate the permeability coefficients. The data for $\Delta G(z)$ and $D_z(z)$, calculated for only one of the two monolayers, have been considered valid also for the other monolayer for symmetry. Table 6.2 reports the permeability coefficients computed for the eight solutes considered. Results obtained in this study, using a CG DMPC model mixed with AL solutes, are shown in the first column. The second column reports corresponding results obtained by standard AL simulation, whereas the last column shows available experimental data. The results obtained by the AL-CG simulations are generally within one order of magnitude of the AL and experimental data reported. This level of agreement is rather good, considering that the permeability coefficients for this set of solutes, comprising both polar and apolar molecules, span five orders of magnitude. Most importantly, the solute relative permeability coefficients, observed from both the available experiments and simulation, are well reproduced (Table 6.3). The only significant exception is methylacetate. We have already mentioned that methylacetate is a hydrophilic molecule, hence it should have a lower permeability than ethane and benzene (hydrophobic). From Tables 6.2 and 6.3 we see that this ranking is in fact observed in the AL-CG investigation carried out here, whereas there might be issues with the AL results,[35] which rank methylacetate between benzene and ethane.

6.5 Discussion

We have presented a new multiscale AL-CG model where the two levels of representation interact without the need for an interface. This is possible thanks to the unique nature of our CG model, which is constituted by potentials directly compatible with AL force-field. In particular, both lipid and water CG models contain all relevant electrostatics: this feature, which is not present in any other CG model, allows the AL-CG electrostatic interactions to be treated straightforwardly. Mixed AL-CG interactions have been calibrated by fitting to experimental oil/water partition coefficients of a number of small solutes (analogues of neutral amino acid side chain); to reproduce the experimental data, the force-field has been optimised by introducing two scaling factors into the AL-CG mixing rules.[270] Such modifications have been transferred in this study for the permeability simulations of AL solutes through the CG membrane.

Table 6.2: Permeability coefficients [cm/s]

Solute	CG DMPC (this work)	AL PC	Exp PC
acetamide	$1.5 \pm 1.1 \times 10^{-3}$	$6.6 \pm 1.9 \times 10^{-3}$ (35)†	4.5×10^{-4} (313)†
			1.7×10^{-4} (314)‡
			1.7×10^{-4} (302)‡
			2.9×10^{-4} (301)‡
			1.65×10^{-4} (273)‡
acetic acid	$3.3 \pm 1.9 \times 10^{-2}$	$1.3 \pm 0.5 \times 10^{-1}$ (35)†	6.9×10^{-3} (314)‡
			6.6×10^{-3} (303)‡
			5.0×10^{-3} (286)‡
benzene	29 ± 3	10 ± 1 (35)†	
ethane	6.2 ± 2.1	6.7 ± 0.8 (35)†	
methanol	$1.8 \pm 1.4 \times 10^{-2}$	$1.9 \pm 0.4 \times 10^{-2}$ (35)†	
methylacetate	1.9 ± 0.4	9.5 ± 1.1 (35)†	
methylamine	1.0 ± 0.2	1.2 ± 0.2 (35)†	9×10^{-1} (315)‡
			8×10^{-2} (314)‡
water	$7.0 \pm 0.1 \times 10^{-2}$	6.8×10^{-2} (280)*	8.3×10^{-3} (269)*
		4.0×10^{-2} (280)†	7.0×10^{-3} (3)*
		$1.3 \pm 0.3 \times 10^{-2}$ (35)†	2.4×10^{-2} (300)*
		1.6×10^{-2} (279)†	4.0×10^{-4} (297)*
		$7.0 \pm 3.0 \times 10^{-2}$ (276)†	1.0×10^{-3} (297)*
			6.0×10^{-4} (298)*
			2.4×10^{-3} (2)†
			1.9×10^{-2} (300)†
			3.4×10^{-3} (314)‡
			2.2×10^{-3} (273)‡
			1.9×10^{-3} (286)‡

CG = coarse-grain, AL = atomic-level, Exp = experimental data, PC = phosphatidylcholine, DMPC = dimyristoylphosphatidylcholine, DPPC = dipalmitoylphosphatidylcholine. References are included in round brackets. *Fluid-phase DMPC. †Fluid-phase DPPC. ‡Fluid-phase lecithin.

Table 6.3: Relative permeability coefficients

Solute	CG DMPC (this work)	AL DPPC (Bemporad et al.[35])
acetamide	1	1
acetic acid	22	20
benzene	19333	1515
ethane	4133	1015
methanol	12	29
methylacetate	1267	1439
methylamine	667	182
water	47	2

Overall, efficiency is improved by coarse-graining the lipid and hydrating water components (that make up for almost all of the simulation cost), whereas accuracy is retained by describing the solute molecules atomistically by a standard AL force-field. By adopting a CG description of the membrane "solvent", we could increase the simulation efficiency to explore larger systems for longer times than has been done with standard AL simulation. We studied the interaction of solutes with a bilayer system comprising 128 lipids, larger than the typical $36-72$ lipid bilayers that have been simulated by standard AL methods in permeability calculations.[35,276,278–280,282,283,311] Large membranes minimise artefacts from periodicity (due to the periodic boundary condition normally assumed in molecular dynamics) and are less likely to suffer from structural distortions caused by the inclusions (this problem being particularly serious when large solutes, such as drugs, are considered). For each solute, we have sampled each of the selected z-positions for 80 ns, a much longer time than that (< 10 ns) typically reached in corresponding AL z-constrained simulations.[35,276,278–280,283] We performed each long runs, often necessary to reach convergence, in about five days of CPU-time on single processors. By comparing the CG-AL simulation cost to corresponding AL calculations,[35] a speed-up factor of two orders of magnitude can been estimated. We could run all simulations almost simultaneously on a cluster;[240] overall, a total of $8 \times 16 \times 80$ ns $\approx 10\,\mu$s of simulation data were produced in about one week. Our AL-CG model thus provided permeation data for large sets of molecules in a fraction of the computational time required by traditional AL methods.

We evaluated the membrane permeability coefficients of the selected solutes via the inhomogeneous solubility-diffusion model.[274,276] The results obtained by the AL-CG method were within one order of magnitude of those obtained in the AL study[35] and available experimental data (Table 6.2); most importantly, the relative permeabilities, and hence the ranking order, are in good general agreement (Table 6.3). Results indicate that the overall permeability is mainly determined by the free energy component, the diffusion contribution being rather marginal; such findings agree with previous analysis.[35,285]

We reported, however, some qualitative discrepancies between the free energy and dif-

fusion profiles obtained in this work and those reported in the AL study.[35] In particular, differences can be noticed in the free energy profiles of the polar solutes. The AL-CG results display a free energy minimum at the polar/apolar interface, whereas the AL curves are characterised by free energy increases (Figure 6.4). Therefore, according to the AL-CG model, hydrophilic solutes preferentially partition at the polar/apolar interface, whereas the AL model[35] indicates that these solutes would preferentially remain in the water phase outside the bilayer. Interestingly, another previous AL simulation study[316] showed results in qualitative agreement with the AL-CG data presented here. In that investigation,[316] free energies were calculated for five small solutes (CH_4, CH_3F, CH_2F_2, CHF_3, CF_4) across a hydrated glycerol 1-monooleate (GMO) bilayer. The free energy of nonpolar molecules was found to decrease monotonically or nearly monotonically, leading to a minimum in the bilayer centre (as in our results for benzene and ethane), whereas polar molecules exhibited an interfacial minimum (as in our results for the polar solutes, except water). Such findings were explained by analysing the nonelectrostatic and electrostatic contributions: the main part of the nonelectrostatic term is the reversible work of cavity formation (larger in water than in the membrane interior), whereas the electrostatic contribution is the lowest (most negative) in water and monotonically increases toward the bilayer centre.[316] The transmembrane free energy displays an interfacial minimum when the rates of change (the average forces) of electrostatic and nonelectrostatic contributions have equal absolute values near the interface (as for polar solutes).[316] For nonpolar solutes, the electrostatic term dominates. It was also predicted[316] that for water, which is a small highly polar molecule, the electrostatic contribution would change more rapidly than the nonelectrostatic term, leading to a monotonic free energy increase, as observed in the AL-CG results presented here (Figure 6.3), and in previous AL studies.[35,276,279,280] Regarding the diffusion profiles, we have already pointed out that in the AL-CG profiles (Figure 6.5) most solutes display high diffusion coefficients in the centre of the bilayer, whereas in the AL simulations[35] diffusion coefficients in the bilayer centre for all solutes were found to be significantly lower than in the water phase. We also noticed that there are literature data supporting the AL-CG results, at least for water[276] and benzene.[277] Also, experimentally, solutes with low molecular weight (< 50) have been found to permeate significantly faster than predicted by Overton's theory.[314] This discrepancy has been explained with the free-volume model, according to which small-sized solutes diffuse abnormally fast by jumping between mobile free-volume pockets which are too small to accommodate larger solutes.[317] Free-volume voids form dynamically by lateral density fluctuations in the lipid hydrocarbon region of membranes.[115]

In general, several possible reasons can account for the discrepancies observed between the AL-CG results presented here and the corresponding AL data.[35] First, there are significant differences in the models. The AL model[35] explicitly represents every atom, whereas, in the AL-CG system, the bilayer model is significantly simplified. The solutes are modelled atomistically in both studies, but with different force-fields: CHARMM in the AL model[35] and GAFF in the AL-CG model. Second, simulation conditions are also somewhat differ-

ent. The AL-CG simulations have been carried out at a temperature of 30°C, whereas it was 50°C for the AL study.[35] These temperatures were chosen to ensure that the lipids simulated, DMPC and DPPC respectively, were in the (biologically-relevant) liquid-crystalline phase. Moreover, in the AL simulations[35] the bilayer interfacial area was artificially kept fixed to avoid known artefacts of the AL force-field employed,[318] whereas in the AL-CG study we have employed the (more realistic) semi-isotropic pressure coupling, which allows the xy interfacial plane to fluctuate according to the lateral components of the pressure tensor. Finally, there are differences in the way diffusion coefficients were computed. The integral in Equation 6.4 is typically characterised by slowly-converging tails, which can undermine the reliability of the estimation of the diffusion coefficient. To avoid such problems, the integral is sometimes fitted to an analytical function;[35,276,283,285] however, when sufficiently long simulations are performed, no fitting is required.[279,280] In the AL study,[35] the limited simulation time (10 ns per solute) prevented the diffusion coefficients from reaching convergence, so that the force fluctuation autocorrelation function had to be approximated by fitting to analytical functions.[35] The AL-CG runs were instead sufficiently long to yield converged values without resorting to approximations.

In both the AL[35] and the AL-CG models, it is difficult to clearly identify how the limitations and artefacts, inherent in each of the models, affect the results. The general tendency would be to consider the AL data more accurate. However, we noticed that in the case of methylacetate the AL calculations[35] failed to identify this molecule as hydrophilic, whereas the simplified AL-CG method yielded a correct prediction. Also, in the case of acetic acid the simplified AL-CG model yielded a permeability coefficient in better agreement with experiment than the AL model[35] (Table 6.2).

Overall, considering the general fair agreement between the AL-CG results reported in this chapter, and previous AL and experimental investigations, we have reasons to be confident of the physical accuracy of the multiscale AL-CG representation. Further permeability simulations of larger molecules, reported in the next chapter, represent an additional test to confirm the validity of our multiscale methodology.

Chapter 7

Multiscale permeability simulations of large drugs and steroid hormones

After the validation of the multiscale permeability method on a set of small molecules, reported in the last chapter, we present here permeation studies of three large drugs and two steroid hormones. The calculated permeability coefficients are in good agreement with previous atomic-level simulations and in qualitative agreement with available experimental data. In particular, the solutes' ranking order is generally reproduced. Also, the predicted location of the steroid hormones inside the membrane is consistent with experimental measurements. Possible applications of the method in the context of rational drug design are discussed, along with limitations and issues.

7.1 Introduction and background

Passive transmembrane permeability is one of the major mechanisms for drug absorption (Figure 7.1). Lipophilic compounds can rapidly partition into the cell membrane thanks to their affinity for the hydrocarbon core of lipid bilayers, and hence they typically exploit the transcellular pathway. In between cells, small water-filled pores (also called "tight junctions") are normally present; despite the much smaller surface area offered by these pores compared with the overall epithelium cell surface, it is reasonable to expect that (small)

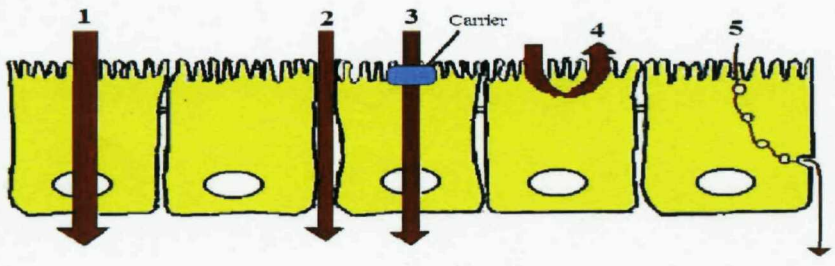


Figure 7.1: Drug transport mechanisms across the intestinal epithelium. (1) Passive transcellular route, (2) passive paracellular route, (3) carrier-mediated transport, (4) carrier-mediated efflux and (5) vesicular transport. From Shah et al.[319]

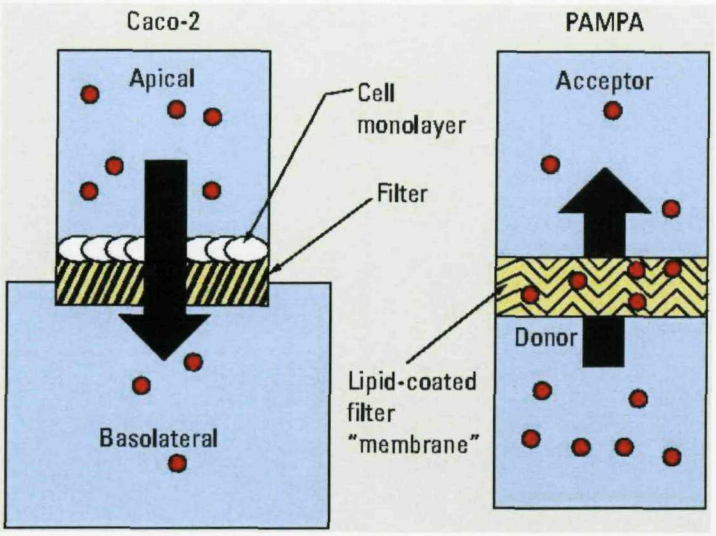


Figure 7.2: Permeability assays. Caco-2 (left) uses compartments of different volumes separated by a monolayer of cells grown on a filter. PAMPA (right) uses chambers of the same size separated by a filter coated with lipid in organic solvent. From Ruell.[320]

hydrophilic drugs can also diffuse via such paracellular route. Alternative drug transport processes, including protein-assisted and vesicle-mediated transport, are less frequently observed compared to the unassisted mechanisms. It is therefore evident that understanding passive permeation is crucial for rational drug design. The following paragraphs summarise the most important experimental and computational methods employed to predict drug permeability.

7.1.1 Experimental assays

Permeability coefficients of drug molecules are not usually measured from isolated lipid bilayers, but instead using more complex model systems that more closely correlate with the observed physiological drug absorption data.

The two most common *in vitro* permeability assays are the Caco-2 cell monolayers and the parallel artificial membrane permeability assay (PAMPA). These assays, schematically represented in Figure 7.2, are briefly described in the following paragraphs.

Caco-2 cell monolayers

Caco-2 cells are human colorectal carcinoma cells, characterised by morphological and functional similarities to the small intestinal epithelium cells. During the past few years, Caco-2 monolayers have been widely accepted by pharmaceutical companies and by regulatory authorities as a standard permeability-screening assay for prediction of drug intestinal permeability.[319] Caco-2 experiments allow the study of all major absorption routes: passive transcellular and paracellular transport, and carrier-mediated mechanisms. In typical experiments, a monolayer of cells is grown on a filter separating two stacked microwell plates

(Figure 7.2, left). The compound under investigation is then introduced on one side of the filter. Permeability coefficients can be measured by ultraviolet spectroscopy or a combination of liquid chromatography and mass spectroscopy; alternatively, radio-labelled compounds can be used in the first place. Despite being the standard for estimating drug permeability, the Caco-2 method is not without drawbacks. First, experiments are time-consuming, requiring up to 30 days. Also, since Caco-2 cells contain endogenous transporter and efflux systems, reproducibility and data interpretation can be difficult.[320]

PAMPA

Parallel artificial membrane permeability assay (PAMPA), introduced by Kansy et al.[321] in 1998, has since been gaining acceptance in pharmaceutical research as a less expensive alternative to Caco-2.[320,322–325] A PAMPA "sandwich" (Figure 7.2, right) is prepared from two plates that are similar to those used for traditional Caco-2 experiments. One plate contains a porous filter disk at the bottom of each well, whereas the other one is a reservoir plate that is molded to sit precisely under the filter plate. The filter is coated with a solution of lipid material in inert organic solvent to prepare the artificial membrane. Filters are typically $\approx 100\,\mu$m thick.[326] The phase formed by lipids in PAMPA membranes is unknown. The wells of one plate are then filled with donor solution (drug), and the other with acceptor solution (buffer); the plates are then stacked to create the sandwich and incubated. Incubation times can vary between 15 min for highly permeable molecules and 15 h for poorly permeable molecules. The sandwich is eventually separated and both the donor and acceptor compartments are assayed for the amount of material present. As for the Caco-2 method, measurements are performed by ultraviolet spectroscopy or liquid chromatography combined with mass spectroscopy. Clearly, PAMPA assays only measure passive permeation, thus eliminating the possible active contribution which can affect Caco-2 results. Another advantage of PAMPA is that experiments are much quicker to carry out than with Caco-2.

7.1.2 Computational approaches

Computational methods to predict transmembrane permeability coefficients of drugs before synthesis are increasingly desirable to minimise the investment in pharmaceutical design and development.[158] However, the accurate prediction of drug permeability represents a great challenge for *in silico* models, due to the complexity of the underlying physiological mechanism.[327]

QSAR methods

A quantitative structure-activity relation (QSAR) relates numerical properties of the drug molecular structure to its activity via a mathematical model.[121] An early example was the discovery by Meyer and Overton of a correlation between anaesthetic potency of a compound and its oil/water partition coefficient. Partition coefficients are also traditionally correlated

to permeation: within chemical series, this is often verified. However, computational models based on molecular properties typically fail when large sets of diverse compounds are analysed.[320]

Simulation methods

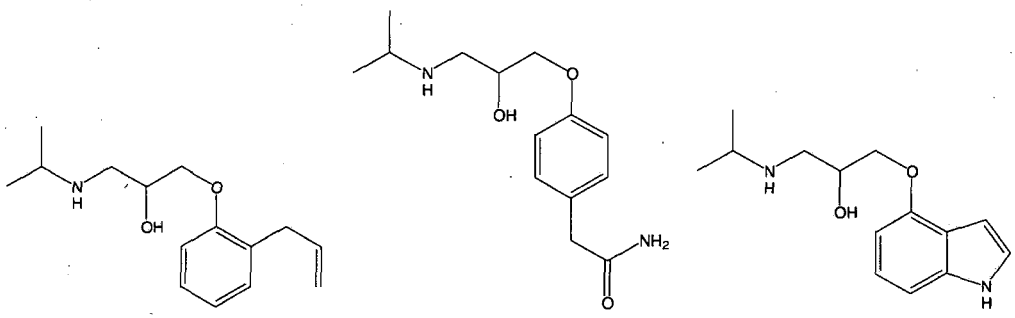
Studying the permeation process by particle-based computer models is particularly attractive: simulations have the potential to unveil this mechanism with atomic detail, while also providing accurate estimates of the permeability coefficient. An early attempt to simulate the permeation process by molecular dynamics was performed by Alper and Stouch,[317] who studied a nifedipine analogue in a lipid bilayer. Due to limited computer resources, the system was simulated for only 4 ns. While such time is not sufficient to allow a direct estimation of the permeability coefficient, it was possible to study the drug orientation and diffusion, and to analyse the role of hydrogen bonding.[317] Transbilayer permeation has been recently simulated, using the *z*-constraint method, for the amphiphilic drug valproic acid,[283] β-blockers[284,328] and psoralen derivatives.[285] Free energy profiles, diffusion coefficients and eventually permeability coefficients were calculated.[283–285,328] These investigations have been extremely useful in understanding many aspects of bilayer permeation with atomic resolution. However, the huge computational cost associated with the simulation of these models makes it difficult to obtain converged data. Also, applications in the context of drug design are seriously hindered, as screenings of large sets of candidates would be unfeasible in a reasonable amount of time.

7.2 Methodology

The permeability simulations presented in the following sections have been conducted following the general z-constraint methodology already employed for the small molecule simulations as reported in Section 6.3 of Chapter 6. The specific molecules investigated in this study, and minor modifications to the simulation protocol, are described in the following sections.

7.2.1 Atomic-level solute models

The solutes considered in this study comprise three β-blockers (alprenolol, atenolol and pindolol) and two steroid hormones (progesterone and testosterone). The structures of the atomic-level solutes were obtained from atomic-level molecular dynamics permeability simulations;[284,328,329] for each solute, a single representative structure was employed. The models have then been parameterised using the same procedure employed for the modelling of the small molecules as reported in Section 6.3.3. In particular, like in that study, no intramolecular degrees of freedom are taken into account, that is, solute molecules are represented as rigid bodies. Furthermore, solutes are again assigned no net charge, consistent with the assumption that only the neutral species diffuse across the bilayer.[301]



(a) Alprenolol (b) Atenolol (c) Pindolol

Figure 7.3: β-blocker structures.

7.2.2 Simulation details

Simulations are conducted following a protocol similar to that for the small molecule permeation study, reported in Section 6.3.5. However, in the simulations presented here, the group-based cutoff radius for the solute molecule, originating in the solute mass centre, has been increased to 1.6 nm. Since the maximum cross-section of the solute considered is $\approx$ 1.2 nm, the cutoff chosen ensures that the shortest cutoff distance at which an atom of the solute begins to interact with a lipid or water site is not less than 1 nm. Since the drugs and hormones are rather large and "heavy" molecules, simulations are stable even without artificially redistributing the masses to increase the principal moments of inertia, as was done for the small molecule simulations. Equilibration runs for each system have been carried out for 5 ns. Production runs have then been conducted for 60 ns, divided in two consecutive 30-ns batches. Simulations were run on the Southampton University Iridis cluster[240] in a "coarsely" parallel fashion, meaning that each simulation ran on a different single processor. The entire set of simulations could be completed in about one week.

7.3 Results

7.3.1 β-blockers: alprenolol, atenolol and pindolol

Alprenolol, atenolol and pindolol are β-blockers, an important cardiovascular drug class, recommended as first-line treatment of numerous diseases such as heart failure, hypertension, and angina, as well as treatment after myocardial infarction.[330] The chemical structures of these molecules are shown in Figure 7.3. The free energies of transfer from the water phase into the bilayer are reported in Figure 7.4. The three molecules are characterised by free energy profiles of similar shape. The free energy difference decreases monotonically on entering the membrane up to a minimum reached in the hydrocarbon region. The free energy than increases monotonically up to a maximum corresponding to the bilayer centre. For atenolol, the free energy minimum can be identified at a distance of 1.9 nm from the bilayer centre. For pindolol, the free energy minimum is located at 1.7 nm from the bilayer centre. Alprenolol

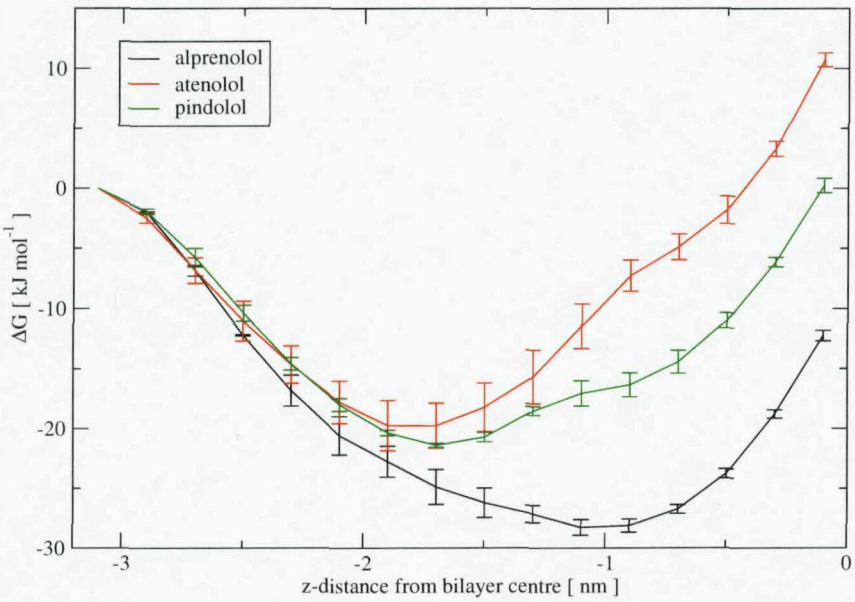


Figure 7.4: Free energies of transfer from the water phase into the bilayer.

displays a free energy minimum at 1.1 nm from the bilayer centre. According to these results, atenolol and pindolol preferentially partition in the interface region, close to the lipid glycerol backbone. Alprenolol instead prefers to partition deeper in the lipid hydrocarbon region. From the free energy barrier values at the bilayer centre, we can qualitatively predict that atenolol, characterised by the highest value, will be the slowest permeant, followed by pindolol and alprenolol (fastest). Compared to previous AL simulations,[328] we observe a similar qualitative behaviour; in particular, the free energy barrier ranking is consistent. In terms of free energy minima, there is consistency in locating the preferential partitioning location for atenolol and pindolol roughly at the interface between the tail and headgroup region, however, for pindolol the AL result indicate a minimum exactly corresponding the the bilayer centre.[328] Diffusion coefficients are reported in Figure 7.5. Diffusion values are typically in the range $0.1 - 0.3 \times 10^{-5}\,\mathrm{cm^2 s^{-1}}$; as expected due to the large size of the drugs, such values are considerably lower than those calculated for the small organic molecules (Chapter 6). It is possible to observe a slight decrease in the diffusion coefficients as the solutes are located closer and closer to the bilayer centre (Figure 7.5). Compared with the AL data,[328] similar values inside the membrane can be observed; however, the AL diffusion coefficients in the water phase are somewhat larger than those calculated here. Resistance profiles are displayed in Figure 7.6. Resistances tend to follow the free energy curves. In particular, atenolol is characterised by the highest permeation barrier, followed by pindolol and alprenolol.

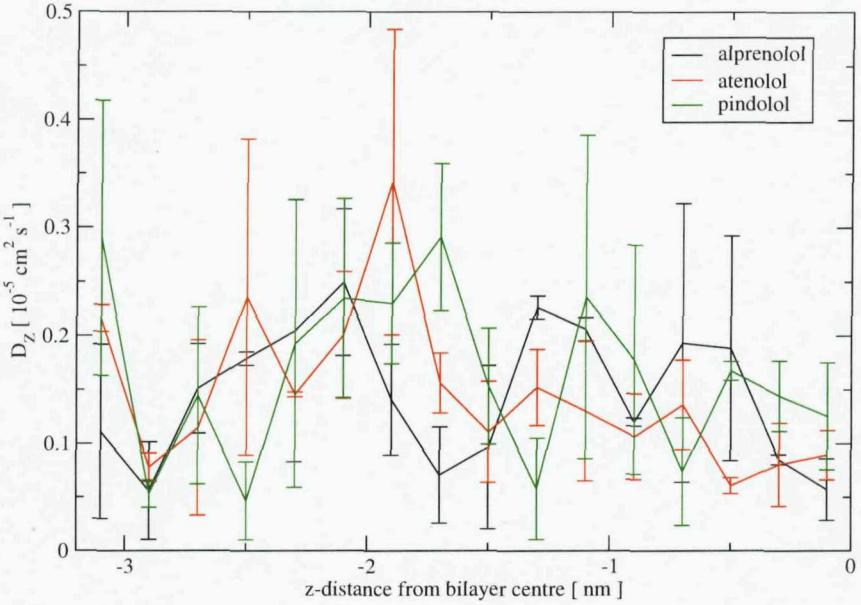


Figure 7.5: Diffusion profiles.

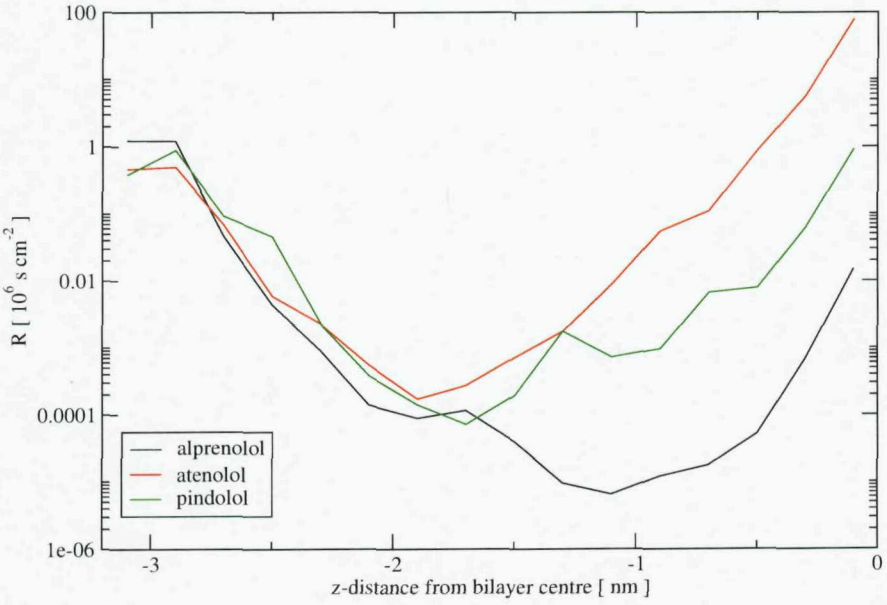


Figure 7.6: Resistance profiles.

Table 7.1: β-blocker permeability coefficients [cm/s]

Method (reference)	alprenolol	atenolol	pindolol
AL-CG simulation, DMPC bilayer	14 ± 4	$2.8 \pm 0.2 \times 10^{-1}$	11 ± 1
AL simulation, DPPC bilayer ([328])	10 ± 10	$3.3 \pm 4.2 \times 10^{-1}$	3.0 ± 3.5
Experiment, Caco-2 ([331])	4.8×10^{-3}	3.7×10^{-5}	3.0×10^{-4}
Experiment, Caco-2 ([332])	2.4×10^{-4}	1.0×10^{-6}	5.1×10^{-5}
Experiment, Caco-2 ([326])	1.8×10^{-2}	8.8×10^{-5}	2.6×10^{-3}
Experiment, PAMPA ([326])	1.0	8.6×10^{-6}	1.8×10^{-2}
Experiment, PAMPA ([333])	1.1×10^{-5}	-	4.9×10^{-6}
Experiment, PAMPA ([322])	2.5×10^{-3}	-	1.4×10^{-3}

AL = atomic-level, CG = coarse-grain, DMPC = dimyristoylphosphatidylcholine, DPPC = dipalmitoylphosphatidylcholine.

The final permeability coefficients, obtained by integration of the resistances, are collected in Table 7.1, along with corresponding previous AL simulation results and experimental measurements. The results obtained in this study by the multiscale approach are in good agreement with the previous AL simulation data,[328] with the exception of the coefficient estimated for pindolol, which is larger than the corresponding AL value. Thanks to the efficiency of the multiscale approach, simulations at each z-location could be run for 30 ns, a much longer time than that achieved by the AL approach[328] of 3 ns. This improvement in sampling is reflected in the error estimates of the final permeability coefficients, which are lower for the AL-CG than for the AL results (Table 7.1). Regarding the experimental data, it can be noticed how they typically differ from the simulation coefficients by several orders of magnitude. This is to be expected considering the differences between the simulation and the experimental materials and conditions. Simulations are conducted on simple, "minimal" pure lipid bilayers. Experiments are instead carried out on layers of entire cells (Caco-2) or on thick solutions of lipids of unknown phase (PAMPA). In both experimental systems, solutes must cross a much thicker barrier compared to that represented by the single lipid bilayer in our simulations. It is therefore reasonable to observe much larger permeability coefficients in our systems compared to experiments. In general, despite being far from ideal, PAMPA data are the most appropriate to be used in comparison with our simulation results, as they represent purely passive permeability measurements (as opposed to the Caco-2 data which may contain contributions from active transport mechanisms). It is also worth noting how the experimental measurements show great variability even amongst each other. Again this is not surprising considering how sensitive permeability coefficients are to slightly different conditions and setup details. However, it is most important to look at *relative* permeabilities; this also has a high practical value for drug design, as it is the ranking order among a set of compounds which is crucial rather than the knowledge of the absolute individual magnitudes. Relative permeability coefficients, calculated for each complete set of data with respect to the permeability coefficient of atenolol, are collected in Table 7.2. The ranking order obtained by experimental and AL methods ($P_{alprenolol} > P_{pindolol} > P_{atenolol}$) is

Table 7.2: Relative permeability coefficients

Method (reference)	alprenolol	atenolol	pindolol
AL-CG simulation, DMPC bilayer	50	1	39
AL simulation, DPPC bilayer ([328])	30	1	9
Experiment, Caco-2 ([331])	130	1	8
Experiment, Caco-2 ([332])	240	1	51
Experiment, Caco-2 ([326])	204	1	30
Experiment, PAMPA ([326])	120790	1	2091

AL = atomic-level, CG = coarse-grain, DMPC = dimyristoylphosphatidylcholine, DPPC = dipalmitoylphosphatidylcholine.

also reproduce by the multiscale AL-CG approach. There is an overall good consistency between the simulation data and the Caco-2 experimental measurements, whereas the PAMPA experiment considered reports larger differences between the relative values.

Additional observation

The orientation of the β-blocker molecules has been qualitatively analysed by observing the trajectories during the simulations at the different $z-$locations probed. When the solutes are constrained in the water and in the headgroup/glycerol layers, no preferential orientation could be determined. Also in the hydrocarbon region the solutes were observed to rotate continuously, however, in this region it was possible to identify recurring orientations. Examples of such orientations for each of the three β-blockers are displayed in Figures 7.7, 7.8 and 7.9. It can be seen that in these "typical" orientations the β-blockers tend to expose their central oxygen atoms outwards, toward the headgroups. Previous AL simulations[284] of the same drugs could not identify any conclusive equilibrium distribution of orientations inside the membrane due to the short simulation times achieved with those atomistic models (3 ns). However, it was noticed that the orientations of these β-blockers were mainly tilted with respect to the bilayer normal;[284] such findings agree with our observations from the multiscale simulations.

Figure 7.8, representing a drug located almost in the centre of the bilayer, also shows some headgroups and water molecules which penetrated in the hydrocarbon core to interact with the drug. Such a situation is reproduced in all cases when drugs are constrained in the hydrocarbon region. In fact, Figure 7.9, which represents a drug constrained in the hydrocarbon core close to the glycerol region, also shows perturbed headgroups attracted by the drug. Large solutes hence seem to have a disturbing effect on the bilayer structure, causing local headgroup and water intrusions into the hydrocarbon region.

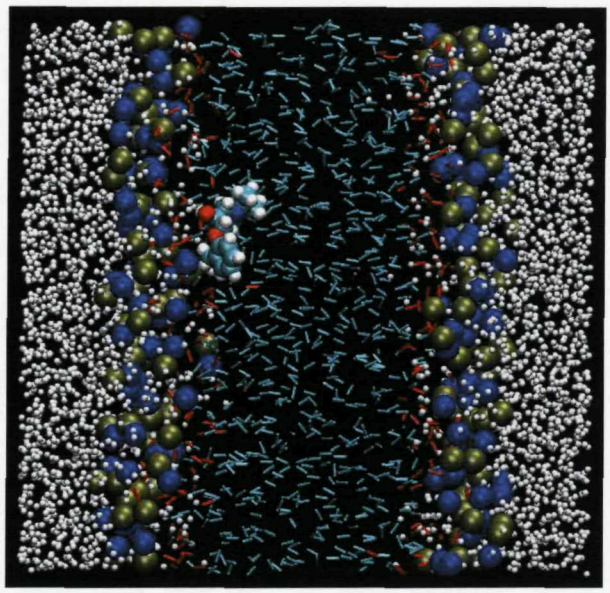


Figure 7.7: Alprenolol orientation. Typical orientation of alprenolol at a distance of 0.9 nm from the bilayer centre. CG membrane colour code: water molecules are white, choline headgroup sites are blue, phosphate headgroup site are yellow, glycerol-ester sites are red, tail sites are green. AL solute colour code: carbon atoms are green, hydrogen atoms are white, oxygen atoms are red, nitrogen atoms are blue.

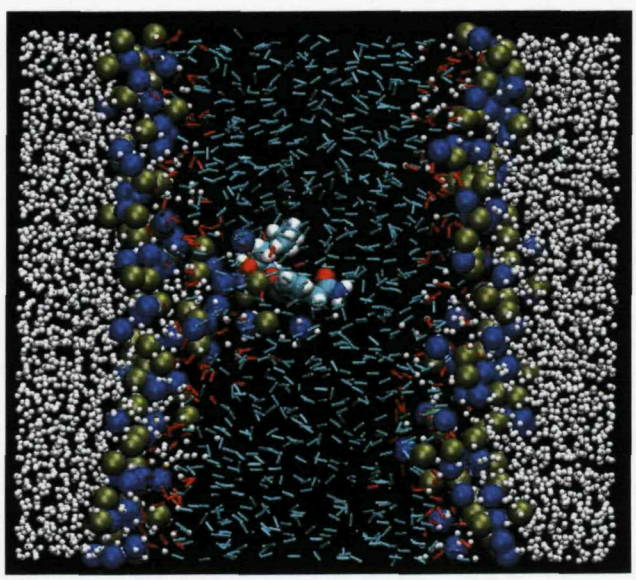


Figure 7.8: Atenolol orientation. Typical orientation of atenolol at a distance of 0.3 nm from the bilayer centre. Colour codes are reported in the caption of Figure 7.7.

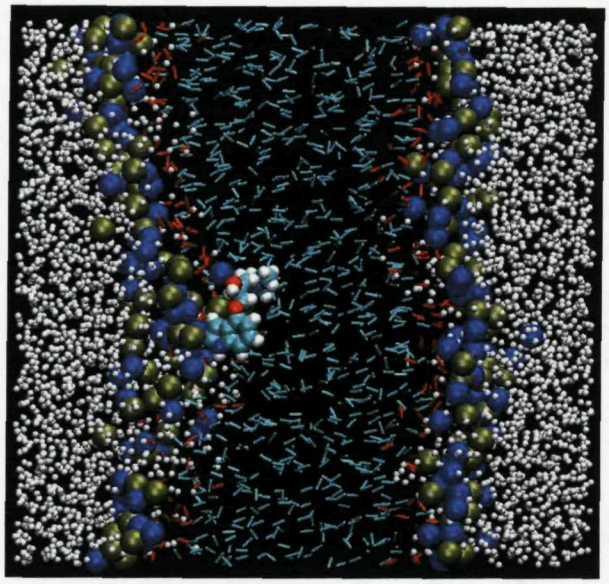


Figure 7.9: Pindolol orientation. Typical orientation of pindolol at a distance of 0.7 nm from the bilayer centre. Colour codes are reported in the caption of Figure 7.7.

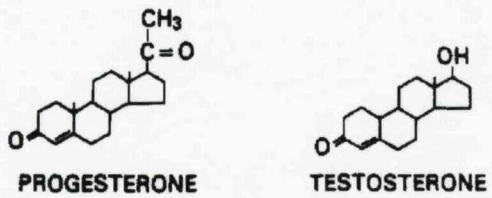


Figure 7.10: Steroid hormone structures.

7.3.2 Steroid hormones: progesterone and testosterone

Steroid hormones are substances that regulate a great variety of physiological functions, including growth, sexual development and carbohydrate metabolism.[62] We have studied the permeability of progesterone and testosterone, two steroid hormones which also have anaesthetic properties.[334] The chemical structures of these molecules are shown in Figure 7.10. Free energy profiles are reported in Figure 7.11. For both solutes, the free energy curves display deep minima corresponding to the lipid tail region. In particular, the free energy minimum of progesterone, and hence its preferential location, is at ≈ 0.9 nm from the bilayer center, whereas for testosterone this distance is ≈ 1.1 nm. Both hormones thus preferentially partition at a depth corresponding to the bilayer hydrocarbon core, close to the lipid glycerol region. This result is consistent with Raman spectroscopic measurements, which broadly localised the hormones within the tail region of the bilayer.[334] Diffusion profiles are reported in Figure 7.12. For both solutes, it is possible to observe a decrease in the diffusion coefficients on entering the headgroup region. Both profiles then remain almost constant across the headgroup and glycerol regions.

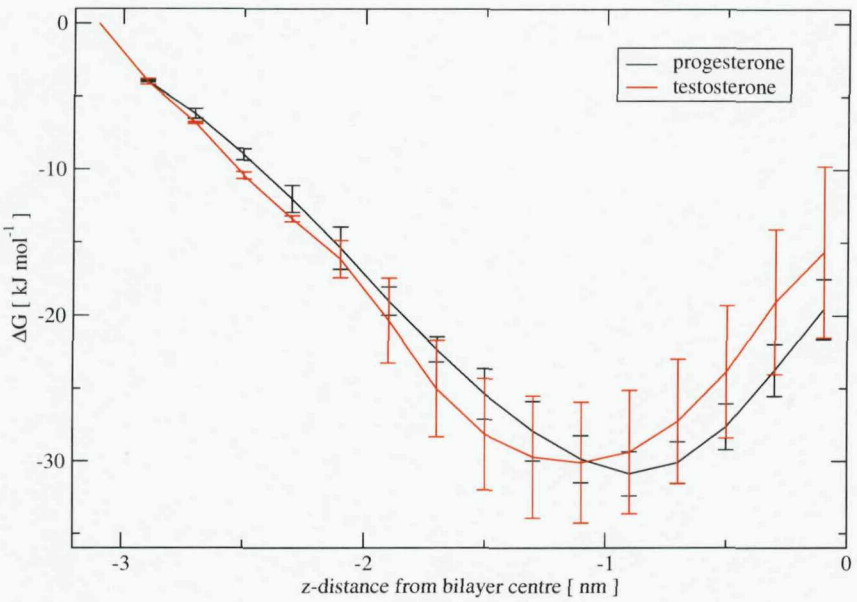


Figure 7.11: Free energies of transfer from the water phase into the bilayer.

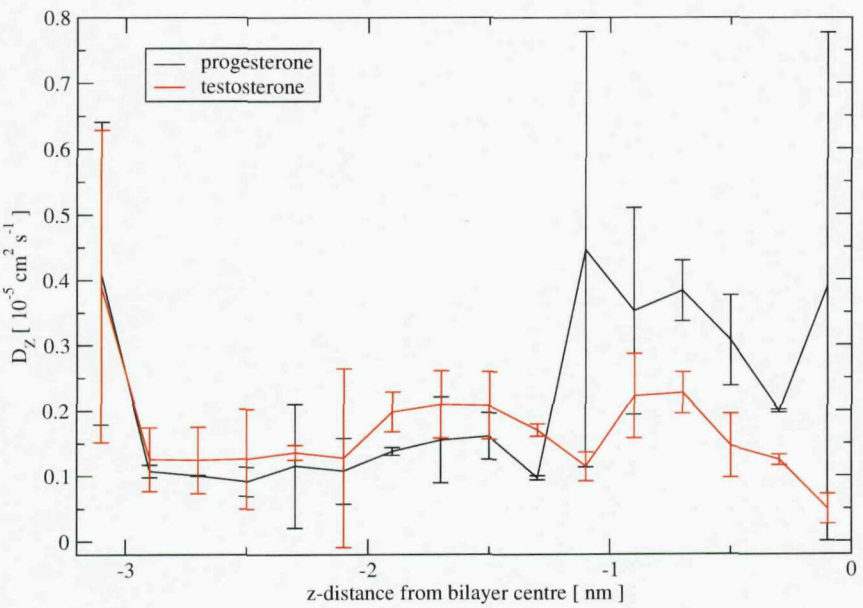


Figure 7.12: Diffusion profiles.

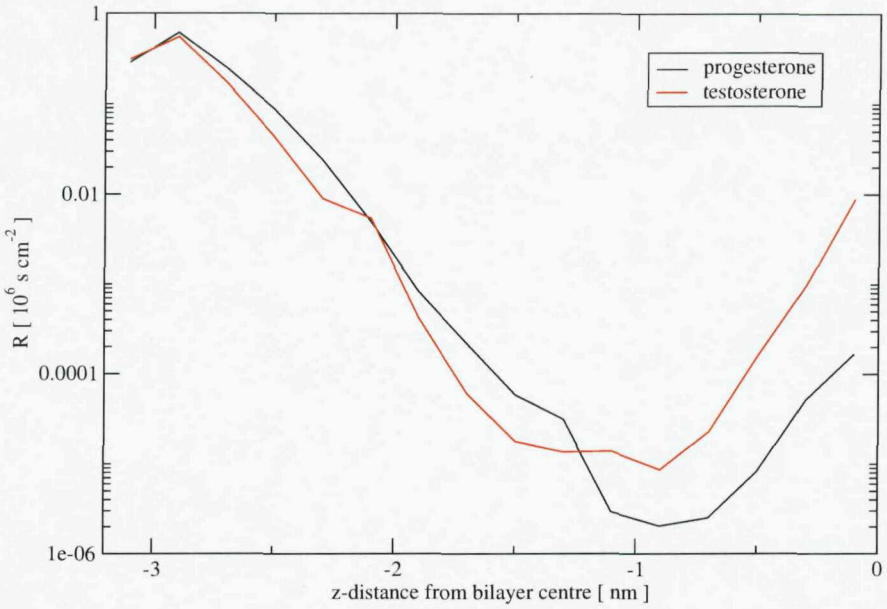


Figure 7.13: Resistance profiles.

Table 7.3: Permeability coefficients [cm/s]

Method (reference)	progesterone	testosterone
AL-CG simulation (this work), DMPC bilayer	22.3 ± 1.9	26.3 ± 0.7
Experiment, PAMPA ([335])	2.7×10^{-5}	7.8×10^{-5}

In the hydrocarbon core, progesterone displays increased diffusion, whereas testosterone is characterised by moderate variations and eventually by a slight decrease towards the bilayer centre. Resistance profiles are reported in Figure 7.13. As we already observed for the other solutes studied, it is clear how the resistance profiles follow the shape of the corresponding free energy curves. The overall permeability coefficients are reported in Table 7.3, along with the available experimental data. Considering the high structural similarities of progesterone and testosterone, we expect the permeability coefficients to be similar. In fact, our results predict very close values, with testosterone permeating slightly faster than progesterone; the experimental data[335] also predict close permeabilities (well within one order of magnitude), and the ranking order is consistent with our findings. There is a clear discrepancy of six orders of magnitude between our results and those obtained from experiment. As already noted for the similar discrepancy regarding the β-blocker data, our simple bilayer membrane is very different from the much thicker experimental PAMPA "membrane"; it is therefore not surprising to observe this difference. It is however encouraging that our results reproduce the ranking order obtained experimentally.

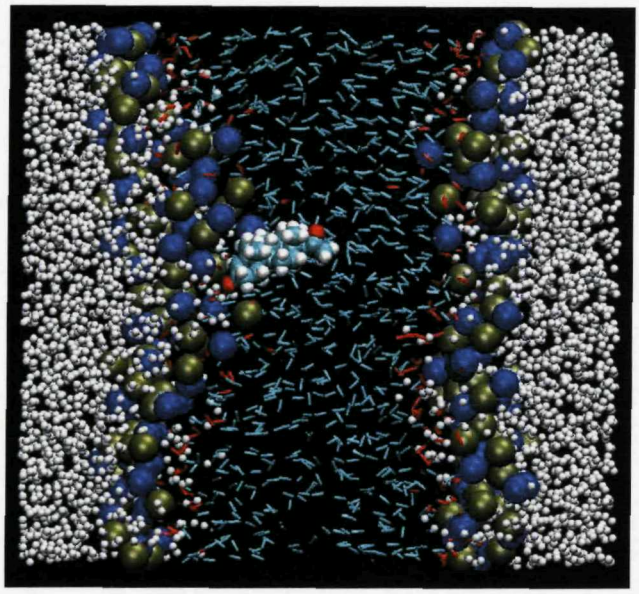


Figure 7.14: Progesterone orientation. Typical progesterone orientation at a distance of 0.5 nm from the bilayer center.

Additional observations

By visualising the progesterone trajectories at different depth across the bilayer, we observe no preferential orientation when the molecule is constrained in the water and interface region. However, in the hydrocarbon core progesterone is clearly preferentially oriented with its main axis parallel to the membrane normal direction. Also, we observe that this steroid preferentially points its methyl group towards the bilayer center, as depicted in Figure 7.14. For testosterone, similar observations can be reported. There is no preferential orientation of this solute as long as it is constrained in the water and interfacial regions. In the hydrocarbon region however, testosterone is preferentially aligned along the bilayer normal; in particular, the alcohol group preferentially points towards the bilayer centre (Figure 7.15). Both hormones therefore display an analogous preferential orientation; in general, they tend to align along the bilayer normal pointing their alpha-beta unsaturated ketone groups towards the lipid headgroups.

7.4 Discussion

To reach their biological target, drugs must typically cross cell membranes. A major route is passive permeation; the understanding of this phenomena, and the prediction of the permeability coefficients, are therefore crucial factors in rational drug design.[284] In this study we have applied a novel multiscale molecular dynamics method to investigate the permeation process of three large drugs (the β-blockers alprenolol, atenolol and pindolol) and two steroid hormones (progesterone and testosterone). Using the z-constraint algorithm,[276] we have calculated free energy profiles, diffusion constants and resistance profiles across the

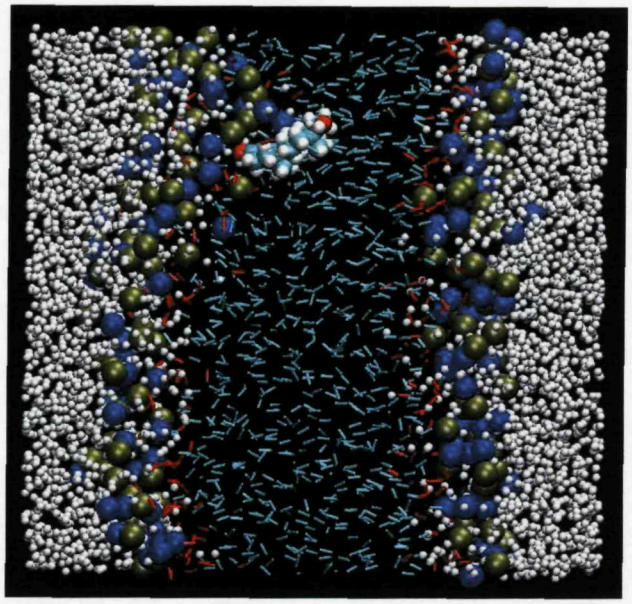


Figure 7.15: Testosterone orientation. Typical testosterone orientation at a distance of 0.3 nm from the bilayer center.

bilayer for each solute. By integration of the resistance, we calculated the absolute permeability coefficients as well as relative values. These final permeability coefficients reproduce the ranking order obtained by AL simulation[284] and experiment,[322,326,331–333,335] although the actual magnitudes can differ by several orders of magnitude. In particular, the permeabilities calculated with our method are typically $3-6$ orders of magnitude faster than the corresponding experimental data. As already noted, this is not surprising, considering the very different systems involved; we compute permeability coefficients through a single bilayer membrane of ≈ 5 nm thickness, whereas experimental data are obtained from assays comprising layers of entire cells (Caco-2) or lipid phases of $\approx 100\,\mu$m thickness (PAMPA). It is therefore reasonable to observe such large differences, with the simulated rates being much faster than the experimental values. The main advantage of the multiscale approach is its efficiency, which makes it a promising tool in a drug design context. We could complete the permeability simulations in less than one week. Given the current (growing) availability of clusters for supercomputing, it would be possible to apply our methodology to conduct screening studies on large sets of compounds. This would allow the cheap and fast estimation of the permeability ranking for a set of candidates, thus offering a simple way to identify possible lead compounds.

7.4.1 Limitations and issues

The multiscale methodology proposed is rather promising, but a number of drawbacks exist and must be highlighted.

The assumption of intramolecular rigidity is quite severe, and although it does not seem to compromise the reliability of the permeability prediction, it is rather unrealistic for such large molecules. In fact, AL simulations highlighted depth-dependent conformational changes: the β-blockers tend to stretch and become elongated as they move from the water phase toward the bilayer centre.[284] Evidently, this phenomenon cannot be captured by the current multiscale method. A related issue concerns the drug structures employed. They were selected as representative of "average" conformations, however, it is clear from the AL results that there should be different "average" structures at different locations across the membrane. Modelling intramolecular flexibility for AL solutes would involve additional force-field terms to account for bonded interactions (stretching, bending, dihedral motions). Such extensions could introduce efficiency issues, because rapid intramolecular motions, especially due to the vibration of hydrogen atoms, would severely limit the molecular dynamics integration timestep. However this problem can be tackled by implementing a multiple timestep method, allowing different degrees of freedom to be integrated at different frequencies. Typically, fast atoms are integrated with a small timestep, whereas the position of coarse-grain sites is updated with a larger timestep, that is, the force calculation is performed at a lower frequency.

Another issue with the multiscale model presented is the lack of an explicit hydrogen bonding representation, due to the absence of hydrogens in the coarse-grain description of lipids and water. Although our explicit treatment of electrostatics probably allows the fundamental features of hydrogen bonding to be captured, the exact effect of the approximation made remains to be established. The hydrogen bonding ability of drug molecules is recognised as an important factor for drug permeation and absorption.[336] Previous AL data indicated a correlation between the number of drug-water hydrogen bonds and the ranking of the permeability coefficients: the more hydrogen bonds formed, the less the permeability coefficient.[284]

Finally, the approach must be validated on a larger set of molecules, as the correct ranking prediction from a data set of only five compounds may be fortuitous.

Chapter 8

Conclusions and perspectives

This thesis has presented work focused on both methodology and application. We have reported the development of a new CG model for biomembranes and its application to the study of the most important physical properties of a representative fluid-phase phospholipid bilayer (Chapter 5). The model has proved capable of consistently and quantitatively reproducing a large number of experimental observables, in terms of structure, elasticity, electrostatics and dynamics. In this work, for the first time a coarse-grain membrane model has yielded quantitative data for the headgroup dipole, Gaussian curvature modulus, electric field and electrostatic potential profiles, interfacial water ordering, short-range and long-range lateral diffusion coefficients, single-lipid diffusive paths and spontaneous permeation of explicit water. Molecular-level insights into crucial membrane properties and mechanisms have been obtained and rationalised. This study generally demonstrates that the fundamental behaviour of lipid bilayers does not depend on fine chemical detail, but on basic physical interactions. The direct compatibility of our model with standard atomic-level representations allows the study of problems where a multiscale approach is desirable. In Chapter 6 we have presented a new multiscale hybrid AL-CG methodology based on the mixing between the coarse-grain membrane and atomic-level solutes. The method has been applied to simulate the transmembrane permeation process of eight small AL solutes embedded in a hydrated CG lipid bilayer. Our results indicate that nonpolar, hydrophobic solutes preferentially partition in the lipid tail region, whereas polar compounds preferentially accumulate at the water/membrane interface. The method has also been applied to study the permeability of three large drugs and two steroid hormones (Chapter 7). For both the small molecules and the drugs, permeability coefficients predicted by applying the inhomogeneous solubility-diffusion model are in good agreement with previous computational investigation. The calculated magnitudes of the permeability coefficients generally differ from available experimental data. However, the relative permeabilities, and hence the permeability ranking order, is consistently reproduced. Also, the predicted location of the steroid hormones from simulation is in agreement with the available experimental data.[334] Overall, the multiscale AL-CG methodology presented allows simulations to benefit from a speed-up of two orders of magnitude over atomic-level methods, while still retaining a good degree of accuracy and

generality.

There are potentially many future extensions to the model presented. In fact, preliminary investigations into the modelling of different lipid species (such as DOPC) have been carried out. Future work will focus on substituting the choline CG headgroup with a smaller-sized site, to allow type-II phosphatidylethanolamine (PE) lipids to be modelled. We could then study bilayers composed of lipid mixtures comprising such "high-curvature" lipids, as is common in real systems. When studied individually, more than half of the lipids naturally present in biomembranes are type-II lipids, which do not form lamellar phases but rather cubic or inverted-hexagonal structures. However, functional biomembranes rarely deviate from lamellar phases. The high abundance of non-lamellar lipids thus generates curvature stress fields in bilayer membranes, through which lipids seem to control protein function.[4] There is a growing number of experimental studies addressing such phenomena, and it would be very interesting to reproduce these experiments by molecular dynamics simulations.

Cholesterol is also a primary candidate for future inclusion in our CG model. The cholesterol molecule comprises a small (OH) head-group and a rather bulky hydrophobic steroid ring; it therefore displays a propensity for promoting non-lamellar structures. There is still no consensus, despite over 70 years of study, on the role played by cholesterol in biological membranes, although several credible hypothesis have been put forward.[337] For instance, cholesterol seems to affect membrane permeability by increasing the membrane dipole potential,[21,338] and it is known to decrease lipid diffusion.[339,340] The Gay-Berne potential can allow cholesterol to be coarse-grained with disk-like sites. Alternatively, AL models of cholesterol can be mixed with the CG lipids in a multiscale approach.

In general, simulating mixtures of different lipids could help understanding one of the simple questions about lipids[341] that still needs to be answered:[342] why do cell membranes contain thousands of different molecular lipid species? Is it a left-over from evolution, or a structural and functional necessity? Other related fascinating questions concern how cells maintain the unique lipid and protein composition of their organelles and how the synthesis of the right quantities of individual proteins and lipids is controlled.[341]

From a more technical point of view, the software implementing the model will be recoded to include parallel calculations. Parallelisation allows the computing time to be reduced by distributing the calculation over a number of processors. This will allow the investigation of longer timescales and larger systems, to address problems related to complex phenomena such as undulations, fusion, raft formation.

Appendix A

The BRAHMS molecular dynamics code: a Biomembrane Reduced-ApproacH Molecular Simulator

A.1 Introduction

For the project presented, the molecular dynamics software BRAHMS* has been specifically designed and developed de novo. BRAHMS is written in the C language;[343,344] the software comprises $\approx$ 10000 lines, organised in 24 modules. The main loop that allows the efficient interaction computation using a combined cell-subdivision/neighbour-list algorithm has been adapted from the excellent text by Rapaport.[124] Dedicated routines have been implemented for the calculation of energies, forces and torques, the integration of rigid-body dynamics, the control of pressure and temperature, as well as for the analysis of the trajectory. The dipolar energy, forces and torques have been implemented from the formulae given by Allen and Tildesley.[120] The expressions for the Gay-Berne energy, forces and torques have been adapted from the Fortran code CHAMPAGNE.[345] The SSD model has been implemented following Chandra and Ichiye.[226] Details on the integration algorithms, as well as on the schemes adopted to control pressure and temperature, are reported in the following sections. BRAHMS incorporates the following interaction models:

- Lennard-Jones potential, employed to model lipid headgroups, SSD water[225,227] and any atom site as point-masses;
- Gay-Berne potential,[125] used to model tail and glycerol sites as symmetric rigid bodies;
- Lennard-Jones/Gay-Berne potential,[221] employed to model mixed interactions;
- Coulomb potential, used to describe the interaction between point-charges;
- dipolar potential, employed to model water and glycerol electrostatic dipoles;

*The name BRAHMS is meant to be both a sensible acronym and a tribute to a great composer.

- charge/dipole mixed potential;
- sticky-tetrahedral (octopolar) potential, employed in the SSD water model;[225,227]
- harmonic potential, used to model inter-site covalent bonding.

A characteristic feature of BRAHMS is the ability to efficiently carry out the integration of rigid-body rotational dynamics. The program performs three main operations: start-up from a given configuration of molecules, generation of new configurations in the desired ensemble and calculation of observable properties by averaging over a finite number of configurations. In the following sections, the most important algorithms implemented in BRAHMS are described in detail.

A.2 Integration of the equations of motion: the DLM scheme

To integrate the equations of motion, we have implemented in BRAHMS the integration algorithm developed by Dullweber, Leimkuhler and McLachlan,[132] that we shall call DLM. In this method, the orientational dynamics is propagated by a sequence of planar rotations, in terms of matrix evaluations that update the rotation matrix associated with each rigid body. The integration step is split in two parts, as detailed in the following paragraphs.

Part A Given the forces $\mathbf{f}(t)$ and the space-frame torques $\mathbf{T}^S(t)$ at the current time t, the momenta of all molecules are advanced from t to $t + \Delta t/2$, whereas mass centre positions $\mathbf{r}$ are moved a full time step:

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t) + \Delta t\, \mathbf{f}(t)/2m \tag{A.1}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t\, \mathbf{v}(t + \Delta t/2) \tag{A.2}$$

$$\mathbf{h}^b(t + \Delta t/2) = \mathbf{h}^b(t) + \Delta t\, \mathbf{T}^b(t)/2 \tag{A.3}$$

where $\mathbf{h}^b = \mathbf{I}\boldsymbol{\omega}^b$ is the body-frame angular momentum, with $\mathbf{I}$ the principal moments of inertia tensor and $\boldsymbol{\omega}^b$ the body-frame angular velocity, and $\mathbf{T}^b$ is the body-frame torque, which is obtained from $\mathbf{T}^b = \mathbf{Q}(t)\mathbf{T}^S(t)$, $\mathbf{Q}(t)$ being the rotation matrix. Now five consecutive body-frame rotations $\mathbf{R}_1, \ldots, \mathbf{R}_5$ are applied to all angular momenta and all orientation matrices are propagated for a full time step, from $\mathbf{Q}(t)$ to $\mathbf{Q}(t + \Delta t)$:

$$\mathbf{Q}(t + \Delta t) = \mathbf{Q}(t)\mathbf{R}_1^T\mathbf{R}_2^T\mathbf{R}_3^T\mathbf{R}_4^T\mathbf{R}_5^T$$

with the explicit computation being:

$$\mathbf{R}_1 := \mathbf{R}_x\left(\frac{1}{2}\Delta t\frac{h_1}{I_1}\right), \quad \mathbf{h}^b = \mathbf{R}_1\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_1^T;$$
$$\mathbf{R}_2 := \mathbf{R}_y\left(\frac{1}{2}\Delta t\frac{h_2}{I_2}\right), \quad \mathbf{h}^b = \mathbf{R}_2\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_2^T;$$
$$\mathbf{R}_3 := \mathbf{R}_z\left(\Delta t\frac{h_3}{I_3}\right), \quad \mathbf{h}^b = \mathbf{R}_3\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_3^T;$$
$$\mathbf{R}_4 := \mathbf{R}_y\left(\frac{1}{2}\Delta t\frac{h_2}{I_2}\right), \quad \mathbf{h}^b = \mathbf{R}_4\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_4^T;$$
$$\mathbf{R}_5 := \mathbf{R}_x\left(\frac{1}{2}\Delta t\frac{h_1}{I_1}\right), \quad \mathbf{h}^b = \mathbf{R}_5\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_5^T;$$

where I_1, I_2, I_3 are elements of the diagonal inertia tensor of a molecule and h_1, h_2, h_3 are the corresponding components of $\mathbf{h}^b$ in the (body-fixed frame of reference) principal axes system. $\mathbf{R}_x(\phi)$ denotes a rotation* around the (body-frame) x-axis by an angle ϕ, and $\mathbf{R}_i^T$ is the transpose of $\mathbf{R}_i$.

Part B After having obtained $\mathbf{r}(t+\Delta t)$ and $\mathbf{Q}(t+\Delta t)$ from the previous part, the corresponding new forces $\mathbf{f}(t+\Delta t)$ and torques $\mathbf{T}^S(t+\Delta t)$ are calculated. Subsequently, the momenta are propagated another half time step through the following formula:

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t+\Delta t/2) + \Delta t\,\mathbf{f}(t+\Delta t)/2m \tag{A.4}$$
$$\mathbf{h}^b(t+\Delta t) = \mathbf{h}^b(t+\Delta t/2) + \Delta t\,\mathbf{T}^b(t+\Delta t)/2 \tag{A.5}$$

where again $\mathbf{T}^b(t+\Delta t) = \mathbf{Q}(t+\Delta t)\mathbf{T}^S(t+\Delta t)$. This is the end of one integration step; now we can start again from Part A. Note that the constraint $\mathbf{Q}^T\mathbf{Q} = \mathbf{1}$, that is, the orthogonality condition,[308] never needs to be artificially enforced, for it is intrinsic in the algorithm; also note that if we consider only the translational dynamics the DLM algorithm reduces to the popular velocity-Verlet scheme.

*A computationally efficient representation of $\mathbf{R}(\phi)$ is achievable by setting $\cos\phi \approx (1-\phi^2/4)/(1+\phi^2/4)$ and $\sin\phi \approx \phi/(1+\phi^2/4)$, and use:

$$\mathbf{R}_x(\phi) \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix}$$
$$\mathbf{R}_y(\phi) \approx \begin{pmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{pmatrix}$$
$$\mathbf{R}_z(\phi) \approx \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The above rational orthogonal approximation formula are reliable only when dealing with small angles (which is the case in MD simulation).

Symmetric rigid-bodies The DLM method can be modified to specifically treat the symmetric rigid body, with (say) $I_1 = I_2$ in principal coordinates. In this case only three rotations are needed:

$$\mathbf{R}_1 := \mathbf{R}_x\left(\frac{1}{2}\Delta t\frac{h_1}{I_1}\right), \quad \mathbf{h}^b = \mathbf{R}_1\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_1^T;$$
$$\mathbf{R}_2 := \mathbf{R}_y\left(\Delta t\frac{h_2}{I_2}\right), \quad \mathbf{h}^b = \mathbf{R}_2\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_2^T;$$
$$\mathbf{R}_3 := \mathbf{R}_x\left(\frac{1}{2}\Delta t\frac{h_1}{I_1}\right), \quad \mathbf{h}^b = \mathbf{R}_3\mathbf{h}^b, \quad \mathbf{Q}^T = \mathbf{Q}^T\mathbf{R}_3^T;$$

These three consecutive body-frame rotations are applied to update all orientation matrices a full time step, from $\mathbf{Q}(t)$ to $\mathbf{Q}(t+\Delta t)$:

$$\mathbf{Q}(t+\Delta t) = \mathbf{Q}(t)\mathbf{R}_1^T\mathbf{R}_2^T\mathbf{R}_3^T$$

These formulae allow the treatment of, e.g., point-dipoles and Gay-Berne solid ellipsoids.

A.3 Integration of rotational motion: symplectic vs non-symplectic schemes

In this section, we compare different integration schemes to simulate rigid-body dynamics. In particular, we show that the DLM matrix-based symplectic* scheme permits integration steps to be used that are ten times larger than those possible with traditional quaternion-based† methods. Such traditional methods are neither symplectic nor time-reversible. In our software BRAHMS we initially implemented one of those methods: the standard quaternion-based schemes by Fincham,[137,138] also reported in Allen and Tildesley's canonical text on molecular simulation.[120] However, poor energy conservation resulted in stability problems, that in turn severely limited the integration step size; similar problems have been observed elsewhere for non-symplectic methods.[132,227,346–348] Over the last ten years, more sophisticated

*The term *symplectic* comes from the Greek for "intertwined", particularly appropriate for Hamilton's equations where $\dot{q}$ is matched with a derivative with respect to p and $\dot{p}$ similarly with the negative of a q derivative.[308]

†The *quaternion* is a quartet of numbers whose components can be expressed in terms of Euler angles:

$$\begin{aligned} q_1 &= \sin(\theta/2)\cdot\cos((\phi-\psi)/2) \\ q_2 &= \sin(\theta/2)\cdot\sin((\phi-\psi)/2) \\ q_3 &= \cos(\theta/2)\cdot\sin((\phi+\psi)/2) \\ q_4 &= \cos(\theta/2)\cdot\cos((\phi+\psi)/2) \end{aligned}$$

The rotation matrix $\mathbf{R}$ can in turn be expressed in terms of the quaternion components:

$$\mathbf{R} = 2\begin{pmatrix} q_1^2+q_4^2-\frac{1}{2} & q_1q_2+q_3q_4 & q_1q_3-q_2q_4 \\ q_1q_2-q_3q_4 & q_2^2+q_4^2-\frac{1}{2} & q_2q_3+q_1q_4 \\ q_1q_3+q_2q_4 & q_2q_3-q_1q_4 & q_3^2+q_4^2-\frac{1}{2} \end{pmatrix}$$

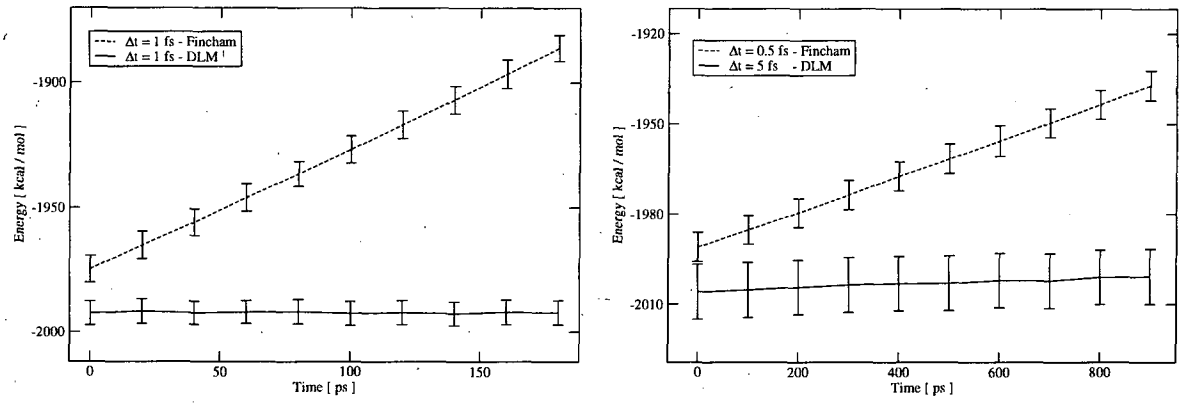
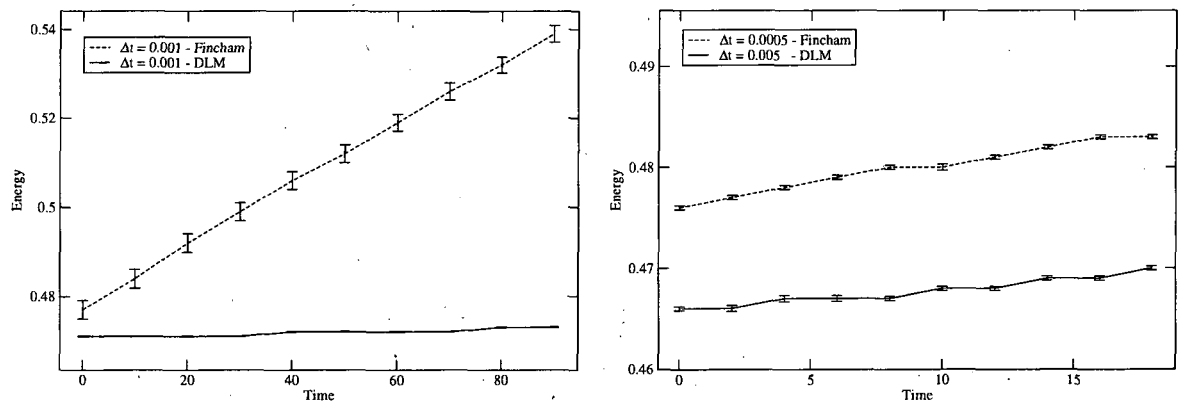


Figure A.1: Energy conservation, for a system of SSD non-symmetric rigid-bodies. Panel (a) shows a comparison between Fincham's and DLM schemes for an equal integration step size Δt. Panel (b) refers to a stability test where DLM is used with a step one order of magnitude larger than Fincham's.

but far superior rotational algorithms have been developed, the key property being symplecticity.[132,133,348] Excellent energy conservation has been reported, thus allowing a substantially larger integration step. Here we apply both the standard Fincham method[137,138] and the recent DLM algorithm[132] to the simulation of simple test systems (SSD water molecules and Gay-Berne particles).

A system of 256 SSD rigid water molecules was equilibrated at a temperature of $T =$ 298 K. Subsequently, NVE molecular dynamics simulations were carried out, and the averages and standard deviations of the energy collected. Figure A.1(a) shows a comparison between Fincham's[137] and the DLM[132] integrators using the same step size: the DLM scheme accurately preserves the total energy value, whereas the employment of Fincham's method results in a significant energy drift. Figure A.1(b) further underlines the stability improvement: the symplectic DLM algorithm allows a time-step an order of magnitude larger than Fincham's to be used, while still being more stable.

A system of 500 Gay-Berne symmetric rigid-bodies was equilibrated at a temperature $T = 1$, and density $\rho = 0.2$ (reduced units). NVE simulations were carried out: integration was again performed with both Fincham's[138] and the DLM scheme.[132] Figure A.2(a) reports the energy behaviour using the two integrators with the same step size: again, DLM conserves energy much more accurately than Fincham's algorithm. Also in this case, as it can be seen in figure A.2(b), the symplectic matrix-based DLM method allows for a step size about ten times larger than the largest possible with the traditional quaternion-based Fincham's scheme.



(a) (b)

Figure A.2: Energy conservation, for a system of Gay-Berne symmetric rigid-bodies, using both Fincham's and DLM schemes. Panel (a) shows a comparison for an equal integration step size Δt. Panel (b) refers to a stability test where DLM is used with a step ten times larger than Fincham's. Standard reduced units are used.

A.4 Temperature control

Conventionally, MD simulations are carried out in the constant energy ensemble; on the other hand, it would be desirable to perform simulations in conditions closer to the real world, i.e., constant-temperature and constant-pressure. In the following sections some methods to control the temperature in MD calculations are presented and discussed.

A.4.1 Velocity scaling

A trivial method to control the temperature is via velocity scaling. At periodic intervals linear and angular velocities are multiplied by a factor:

$$\lambda = \sqrt{\frac{g k_B T}{2\langle \mathcal{K} \rangle}} \tag{A.6}$$

where T is the desired temperature, $\mathcal{K}$ is the kinetic energy and g is the number of degrees of freedom (e.g., 3 for Lennard-Jones sites, 5 for Gay-Berne sites and 6 for general, non-symmetric molecules). In case of mixtures, the kinetic energy needs to be split into different contributions. For example, in a system with N_{LJ} Lennard-Jones point-mass sites and N_{GB} Gay-Berne axially symmetric rigid bodies, the velocities of the Lennard-Jones particles must be rescaled with:

$$\lambda_{LJ} = \sqrt{\frac{3 k_B T}{2\langle \mathcal{K}_{LJ} \rangle}} \tag{A.7}$$

whereas both linear and angular velocities of the Gay-Berne particles with:

$$\lambda_{GB} = \sqrt{\frac{5k_BT}{2\langle \mathcal{K}_{GB} \rangle}} \tag{A.8}$$

where $\mathcal{K}_{LJ}$ and $\mathcal{K}_{GB}$ are the kinetic energy of the Lennard-Jones and Gay-Berne particles, respectively. It must be noticed that the an MD with scaling does not generate a valid statistical ensemble, therefore this control should be switched off before any calculation of thermodynamic averages is performed.

A.4.2 Weak-coupling method - Berendsen thermostat

Berendsen et al.[237] proposed to control the temperature by rescaling the velocities at each step by a factor χ:

$$\chi = \sqrt{1 + \frac{\Delta t}{\tau_T}\left(\frac{T}{\mathcal{T}} - 1\right)} \tag{A.9}$$

with Δt the integration timestep, τ_T a time constant, T the desired temperature and $\mathcal{T}$ the current temperature. This algorithm forces the system towards the desired temperature T at a rate determined by the time constant τ_T, while only slightly perturbing the forces on each molecule. This method does not generate states in the canonical ensemble.[120] Instead, the weak-coupling scheme can be shown[349] to produce an ensemble with properties intermediate between the canonical (NVT) and the microcanonical (NVE). The velocity-Verlet implementation of this algorithm is straightforward: at the end of the second part, velocities are rescaled according to Equation A.9.

A.5 Pressure (and temperature) control

The system's pressure can be controlled (along with the temperature) by a variety of methods. The region shape can be cubic, orthorhombic or general triclinic, as long as it is space-filling. Some constant-pressure methods allow for size and shape changes of the simulation box; this possibility is particularly helpful in the study of solids, since it allows for phase changes in the simulation which may involve changes in the unit cell dimensions and angles.[120] Here we will describe the weak-coupling method,[237] as this is the scheme implemented in BRAHMS.

A.5.1 Weak-coupling method - Berendsen barostat

Berendsen et al.[237] proposed a simple technique to control the pressure by coupling to a "pressure bath". An extra term is added to the equation of motion to produce a pressure change. The system is made to obey the equation:

$$\mathrm{d}P(t)/\mathrm{d}t = [P_{\text{ext}} - P(t)]/\tau_P \tag{A.10}$$

Isotropic

At each step, the coordinates and box edges are scaled by a factor μ:

$$\mu = 1 - \frac{\beta \Delta t}{3\tau_P} [P_{\text{ext}} - P(t)] \tag{A.11}$$

with β the isothermal compressibility of water: $\beta \sim 4.6 \times 10^{-5}$ Atm.

Anisotropic

For the general case of anisotropic triclinic systems, Equation A.11 becomes a tensorial equation:

$$\boldsymbol{\mu} = \mathbf{1} - \frac{\beta \Delta t}{3\tau_P} [\mathbf{P}_{\text{ext}} - \mathbf{P}(t)] \tag{A.12}$$

Mass centres are scaled as:

$$\mathbf{r}' = \boldsymbol{\mu} \mathbf{r} \tag{A.13}$$

And the simulation region is scaled as:

$$\mathbf{H}' = \boldsymbol{\mu} \mathbf{H} \tag{A.14}$$

where in particular $\mathbf{H} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$ is a transformation matrix whose columns are the three vectors $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ representing the edges of the simulation box. Berendsen's algorithm only slightly alters the trajectories and is easy to program, but does not generate states in the NPT ensemble.[120]

A.6 Improving the interaction computations

The site-site interactions can be simply computed through the examination of all possible (different) pairs of sites: for a system of N particles, $N(N-1)/2$ pair distances are evaluated, and eventually forces and torques are computed for those particles separated by a distance shorter than the cutoff radius r_c. This method is however extremely inefficient when the interaction range r_c is small compared with the linear size of the simulation region: the fact that the amount of computation grows as $O(N^2)$ rules out this method for all but the smallest values of N. Two techniques for reducing this growth rate to $O(N)$ are presented in the following subsections.

A.6.1 Cell subdivision

The simulation region is divided into a lattice of small cells, and the cell edges all exceed r_c in length. Then if atoms are assigned to cells on the basis of their current positions it is clear that interactions are only possible between atoms that are either in the same cell or in immediately adjacent cells (Figure A.3). Obviously the region size must be at least $4\,r_c$ for

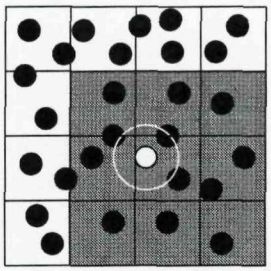


Figure A.3: Cell subdivision. The cutoff range for the particle in white is represented by the white circle; in searching for neighbours of that particle, it is only necessary to examine the particle's own cell and its adjacent cells (shaded). Figure from Allen.[136]

the method to be useful. The cell subdivision method involves a general organisation of data known as a *linked list*:[350] rather than accessing data sequentially, the linked list associates a pointer p_n with each data item x_n, the purpose of which is to provide a non-sequential path through the data. Each linked list requires a separate pointer f to access the first data item, and the item terminating the list must have a special pointer value, such as -1, that cannot be mistaken for anything else. Thus $f = a$ points to x_a as the first item in the list, $p_a = b$ points to x_b as the second item, and so on until a pointer value $p_z = -1$ terminates the list.[124] In the cell subdivision algorithm, linked lists are used to associate atoms with the cells in which they reside at any given instant; a separate list is required for each cell. All data are eventually sorted in a one-dimensional array of integer number. The cell-subdivision method has also been successfully used in the simulation of plasmas, galaxies and ionic crystals.[120]

A.6.2 Neighbour List

Verlet[351] suggested a technique for improving the speed of a molecular dynamics program by maintaining a list of the neighbours of a particular molecule, which is updated at intervals: between updates the program does not check through all the possible pairs, but only through neighbours (Figure A.4). The cell subdivision method can be used to speed up the list construction; in the end all neighbour pairs are consecutively sorted in a (rather long) $2 \times k_{NL} \times N_{sites}$ one-dimensional array of integer numbers, where k_{NL} is a parameter controlling the (predicted) maximum number of neighbours per particle.

A.7 Dimensionless reduced units

In MD simulation all physical quantities are typically expressed in terms of dimensionless *reduced* units. Careful choices of units may lead to increased efficiency; incidentally, reduced units also permits working with numerical values that are not too distant from unity, thus removing any risk of encountering values lying outside the range that is representable by the computer hardware.

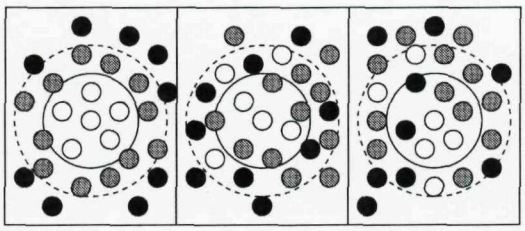


Figure A.4: Verlet neighbour list. The left panel shows the initial list construction: the "neighbours" of the central particle, enclosed by the dashed circle, are depicted as white and light-grey particles. The white particles are inside the cutoff radius (solid circle), and hence they represent the only particles interacting with the central particle at this initial stage. The central panel shows a possible later configuration of the system: now some of the grey particles have entered the cutoff. Since they were recorded on the neighbour list, they are properly taken into account for the interaction computation with the central particle. The right panel shows a potentially problematic situation: some of the black particles, not listed on the neighbour list of the central particle and hence not considered in the interaction calculation, have penetrated into the cutoff zone. The list must be reconstructed before the system reaches such a configuration. Figure from Allen.[136]

Choosing the LJ parameters σ and ϵ and the particle mass m to be the fundamental units of length, energy and mass respectively, we have:[120]

$$
\begin{aligned}
\text{length:} \quad & L^* = L/\sigma \\
\text{energy:} \quad & E^* = E/\epsilon \\
\text{mass:} \quad & M^* = M/m \\
\text{number density:} \quad & \rho^* = \rho\sigma^3 \\
\text{time:} \quad & t^* = t\sqrt{\epsilon/m\sigma^2} \\
\text{temperature:} \quad & T^* = Tk_B/\epsilon \\
\text{pressure:} \quad & P^* = P\sigma^3/\epsilon \\
\text{charge:} \quad & z^* = z/\sqrt{4\pi\varepsilon_0\sigma\epsilon} \\
\text{dipole moment:} \quad & \mu^* = \mu/\sqrt{4\pi\varepsilon_0\sigma^3\epsilon} \\
\text{rigidity:} \quad & k^* = k
\end{aligned}
$$

Definition of reduced units in BRAHMS Considering that in (explicitly) solvated biomembrane systems the majority of the computing time is employed for the evaluation of the interactions between water molecules, BRAHMS takes $\sigma_{unit} = \sigma_{SSD} = 3.035\,\text{Å}$, $\epsilon_{unit} = \epsilon_{SSD} = 0.152\,\text{kcal/mol}$ and $m_{unit} = m_{SSD} = 18.01\,\text{amu} = 2.99\times10^{-26}\,\text{kg}$ as the fundamental units. Derived units can be easily calculated according to the expressions reported in the previous paragraph; for instance, the number density unit is $\rho_{unit} = 0.03577\,\text{molecules/Å}^3$ and the time unit is $t_{unit} = 1.606\,\text{ps}$.

Appendix B

Electrostatic potentials, forces and torques

In the following, the derivation of forces and torques from the electrostatic potentials employed in our model is presented. General treatments on the derivation of forces and torques from anisotropic potentials can be found elsewhere.[126,352,353]

B.1 Electrostatic interactions

B.1.1 Charge-charge (Coulombic) interactions

Charge-charge potential

The interaction potential energy u^{QQ} between two point charges Q_i and Q_j located with a distance r between them is defined by Coulomb's law:

$$u^{QQ}(r) = \frac{Q_i Q_j}{4\pi\epsilon_0 r} \tag{B.1}$$

with ϵ_0 the dielectric constant in free space (vacuum).

Charge-charge forces

The electrostatic force corresponding to the potential of Equation B.1 is:

$$\mathbf{f}_{ij}(r) = \frac{Q_i Q_j}{4\pi\epsilon_0 r^3}\mathbf{r} \tag{B.2}$$

with $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j$.

Shifted-force cutoff scheme The discontinuity at the cutoff distance r_c affects both the apparent energy conservation and the actual atomic motion. This discontinuity can be smeared out by changing the form of the potential function slightly, adding a small linear

term so that its derivative is zero at the cutoff distance:[120,124]

$$u^{\mathrm{SF}}(r) = u(r) - u_{\mathrm{c}} - (r - r_{\mathrm{c}})\frac{\mathrm{d}u(r)}{\mathrm{d}r}|_{r=r_{\mathrm{c}}} \tag{B.3}$$

For the Coulomb potential, defining a multiplicative shifting function:

$$S^M(r) = \left(1 - \frac{r}{r_c}\right)^2 \tag{B.4}$$

the Coulomb energy is modified as:

$$u(r) = \frac{q_i q_j}{r} S^M(r) \tag{B.5}$$

and the force becomes:

$$\mathbf{f} = \frac{q_i q_j}{r^2}\left(1 - \frac{r}{r_c}\right)\left(\frac{1}{r} + \frac{1}{r_c}\right)\mathbf{r} = \frac{q_i q_j}{r^3}\left(1 - \frac{r}{r_c}\right)\left(1 + \frac{r}{r_c}\right)\mathbf{r} = \frac{q_i q_j}{r}\left(\frac{1}{r^2} - \frac{1}{r_c^2}\right)\mathbf{r} \tag{B.6}$$

B.1.2 Charge-dipole interaction

Interaction between a dipole μ with associated orientation vector $\mathbf{e}_\mu$ and a charge Q. In the following we will assume site j to be a point-dipole and site i a point-charge. The angle θ between the dipole orientation vector $\mathbf{e}_\mu$ and the interparticle distance vector $\mathbf{r}$ can be obtained from the relation $\cos\theta = \mathbf{e}_\mu \cdot \mathbf{r}/r$. Also, we define $C = Q\mu/4\pi\epsilon_0$.

Charge-dipole potential

Considering site j a dipole μ, with corresponding orientation vector $\mathbf{e}_\mu$, and site i a charge Q, the electrostatic interaction energy is:

$$u^{Q\mu} = \frac{C}{r^3}(\mathbf{e}_\mu \cdot \mathbf{r}) = \frac{C}{r^2}\cos\theta \tag{B.7}$$

with $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j = \mathbf{r}_Q - \mathbf{r}_\mu$.

Charge-dipole force

Pair force:

$$\mathbf{f}_{ij} = -\mathbf{f}_{ji} = \frac{C}{r^3}\left(\frac{3\cos\theta}{r}\mathbf{r} - \hat{\mathbf{e}}\right) = \frac{C}{r^5}\left[3(\hat{\mathbf{e}}_\mu\mathbf{r})\mathbf{r} - r^2\hat{\mathbf{e}}_\mu\right] \tag{B.8}$$

with $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j = \mathbf{r}_Q - \mathbf{r}_\mu$.

Charge-dipole torque

Pair torque on site j:

$$\mathbf{T}_{ji} = \frac{C}{r^3}(\mathbf{r} \times \mathbf{e}_j) = \mathbf{T}_{\mu Q} = \frac{C}{r^3}(\mathbf{r} \times \mathbf{e}_\mu) \tag{B.9}$$

with $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j = \mathbf{r}_Q - \mathbf{r}_\mu$.

B.1.3 Charge-dipole interaction: shifted-force form

Considering a charge Q and a dipole μ, being $r = r_Q - r_\mu$, the interaction energy is:

$$u^{\mathrm{SF}}(r) = \frac{C}{r^2}\left[1 - \left(\frac{r}{r_\mathrm{c}}\right)^2 + 2\frac{r - r_\mathrm{c}}{r_\mathrm{c}}\left(\frac{r}{r_\mathrm{c}}\right)^2\right]\cos\theta \tag{B.10}$$

Using the chain rule:

$$\mathbf{f} = -\frac{\partial u}{\partial r}\nabla_\mathbf{r} r - \frac{\partial u}{\partial \cos\theta}\nabla_\mathbf{r}\cos\theta \tag{B.11}$$

$$\frac{\partial u}{\partial r} = -\frac{2C\cos\theta}{r^3}\left[1 - \left(\frac{r}{r_\mathrm{c}}\right)^3\right] \tag{B.12}$$

Finally we obtain:

$$\mathbf{f}_{\mu Q} = \frac{C}{r^3}\left\{3\left[1 - \left(\frac{r}{r_\mathrm{c}}\right)^2\right]\frac{\mathbf{r}}{r}\cos\theta - \left[1 - 3\left(\frac{r}{r_\mathrm{c}}\right)^2 + 2\left(\frac{r}{r_\mathrm{c}}\right)^3\right]\mathbf{e}\right\} = \frac{2C}{r^3}\left[1 - \left(\frac{r}{r_\mathrm{c}}\right)^3\right]\mathbf{e} \tag{B.13}$$

$$\mathbf{T}_{\mu Q} = \frac{C}{r^3}\left[1 - 3\left(\frac{r}{r_\mathrm{c}}\right)^2 + 2\left(\frac{r}{r_\mathrm{c}}\right)^3\right](\mathbf{r}\times\mathbf{e}) \tag{B.14}$$

Consistency check: for $r_\mathrm{c} \to \infty$ we recover the unshifted case.

B.1.4 Charge-dipole interaction: linear-switch form

The switched potential energy is:

$$u^{\mathrm{S}}(r) = u(r)\cdot s(r) \tag{B.15}$$

The modulating linear function s is:

$$s(r) = \begin{cases} 1 & \text{if} \quad r < r_s \\ (r_c - r)/(r_c - r_s) & \text{if} \quad r_s \leqslant r \leqslant r_c \\ 0 & \text{if} \quad r > r_c \end{cases}$$

The switched force (for $r_s \leqslant r \leqslant r_c$) is:

$$\mathbf{f} = \frac{C}{r^3(r_\mathrm{c} - r_\mathrm{s})}\left[\left(\frac{3r_\mathrm{c}}{r} - 2\right)\cos\theta\,\mathbf{r} + (r - r_\mathrm{c})\,\mathbf{e}\right] \tag{B.16}$$

The switched torque (for $r_s \leqslant r \leqslant r_c$) is:

$$\mathbf{T} = \frac{C(r_c - r)}{r^3(r_c - r_s)}\mathbf{r}\times\mathbf{e} \tag{B.17}$$

B.1.5 Dipole-dipole interactions

Here we consider a pair of (symmetric molecules and, in particular) dipoles i and j. The orientations are defined by the unit vectors $\mathbf{e}_i$ and $\mathbf{e}_j$; the angles between these orientation vectors and the interparticle separation vector $\mathbf{r}_{ij}$ are defined respectively as θ_i and θ_j. Also, we define γ_{ij} as the angle between $\mathbf{e}_i$ and $\mathbf{e}_j$, so that $\cos\gamma_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$.

Dipolar potential

The electrostatic interaction potential energy is:

$$u_{ij}^{\mu\mu} = \frac{\mu^2}{r^3}(\cos\gamma_{ij} - 3\cos\theta_i \cos\theta_j) \tag{B.18}$$

where $r = |\mathbf{r}_{ij}|$. Cosines are computed through:

$$\cos\theta_i = \frac{\mathbf{e}_i \cdot \mathbf{r}_{ij}}{r|\mathbf{e}_i|} \qquad \cos\theta_j = \frac{\mathbf{e}_j \cdot \mathbf{r}_{ij}}{r|\mathbf{e}_j|} \tag{B.19}$$

Dipolar forces

Pair force:

$$\mathbf{f}_{ij} = -\mathbf{f}_{ji} = \frac{3\mu^2}{r^4}\left[(\cos\gamma_{ij} - 5\cos\theta_i\cos\theta_j)(\mathbf{r}_{ij}/r + \cos\theta_j\mathbf{e}_i + \cos\theta_i\mathbf{e}_j)\right] \tag{B.20}$$

Dipolar torques

Pair torques:

$$\mathbf{T}_{ij} = -\frac{\mu^2}{r^3}\left[\mathbf{e}_i \times \mathbf{e}_j - 3\cos\theta_j(\mathbf{e}_i \times \mathbf{r}_{ij})/r\right] \tag{B.21}$$

$$\mathbf{T}_{ji} = -\frac{\mu^2}{r^3}\left[\mathbf{e}_j \times \mathbf{e}_i - 3\cos\theta_i(\mathbf{e}_j \times \mathbf{r}_{ij})/r\right] \tag{B.22}$$

A complete treatment of the dipolar potential, along with the explicit derivation to obtain forces and torques, can be found elsewhere.[120]

References

(1) Berg, J.; Tymoczko, J.; Stryer, L. *Biochemistry - 5th ed.;* W. H. Freeman and Company, 2002.

(2) Cevc, G.; Marsh, D. *Phospholipid Bilayers - Physical Principles and Models;* John Wiley & Sons, first ed., 1987.

(3) Bloom, M.; Evans, E.; Mouritsen, O. G. *Q Rev Biophys* **1991**, *24*, 293–397.

(4) Mouritsen, O. G. *Life - As a Matter of Fat. The Emerging Science of Lipidomics;* Springer, first ed., 2005.

(5) `http://kentsimmons.uwinnipeg.ca/cm1504/Image127.gif`. Accessed February 2008.

(6) Kučerka, N.; Liu, Y. F.; Chu, N. J.; Petrache, H. I.; Tristram-Nagle, S. T.; Nagle, J. F. *Biophys J* **2005**, *88*, 2626–2637.

(7) Douliez, J.-P.; Leonard, A.; Dufourc, E. J. *Biophys J* **1995**, *68*, 1727–1739.

(8) Filippov, A.; Orädd, G.; Lindblom, G. *Langmuir* **2003**, *19*, 6397–6400.

(9) Rawicz, W.; Olbrich, K. C.; McIntosh, T.; Needham, D.; Evans, E. *Biophys J* **2000**, *79*, 328–339.

(10) Wang, L.; Bose, P. S.; Sigworth, F. J. *Proc Natl Acad Sci USA* **2006**, *103*, 18528.

(11) Templer, R. H.; Castle, S. J.; Curran, A. R.; Rumbles, G.; Klug, D. R. *Faraday Discuss* **1998**, *111*, 41–53.

(12) Maggio, B. *J Lipid Res* **1999**, *40*, 930–939.

(13) Rokitskaya, T. I.; Kotova, E. A.; Antonenko, Y. N. *Biophys J* **2002**, *82*, 865–873.

(14) Starke-Peterkovic, T.; Turner, N.; Else, P. L.; Clarke, R. J. *Am J Physiol Regul Integr Comp Physiol* **2005**, *288*, R663–R670.

(15) Cladera, J.; Martin, I.; O'Shea, P. *Embo J* **2001**, *20*, 19–26.

(16) Cladera, J.; O'Shea, P. *Biophys J* **1998**, *74*, 2434–2442.

(17) Alakoskela, J.-M. I.; Kinnunen, P. K. J. *Biophys J* **2001**, *80*, 294–304.

(18) Cladera, J.; O'Shea, P.; Hadgraft, J.; Valenta, C. *J Pharm Sci* **2003**, *92*, 1018–1027.

(19) Asawakarn, T.; Cladera, J.; O'Shea, P. *J Biol Chem* **2001**, *276*, 38457–38463.

(20) Alakoskela, J.-M. I.; Soderlund, T.; Holopainen, J. M.; Kinnunen, P. K. J. *Mol Pharmacol* **2004**, *66*, 161–168.

(21) Starke-Peterkovic, T.; Turner, N.; Vitha, M. F.; Waller, M. P.; Hibbs, D. E.; Clarke, R. J. *Biophys J* **2006**, *90*, 4060–4070.

(22) Shearman, G. C.; Ces, O.; Templer, R. H.; Seddon, J. M. *J Phys: Condens Matter* **2006**, *18*, S1105–S1124.

(23) van den Brink-van der Laan, E.; Killian, J. A.; de Kruijff, B. *Biochim Biophys Acta* **2004**, *1666*, 275–288.

(24) Kamo, T.; Nakano, M.; Kuroda, Y.; Handa, T. *J Phys Chem B* **2006**, *110*, 24987–24992.

(25) Curnow, P.; Lorch, M.; Charalambous, K.; Booth, P. J. *J Mol Biol* **2004**, *343*, 213–222.

(26) Mohr, J. T.; Gribble, G. W.; Lin, S. S.; Eckenhoff, R. G.; Cantor, R. S. *J Med Chem* **2005**, *48*, 4172–4176.

(27) Essex, J. W.; Hann, M. M.; Richards, W. G. *Phil Trans R Soc Lond B* **1994**, *344*, 239–260.

(28) Marrink, S.-J.; Lindahl, E.; Edholm, O.; Mark, A. E. *J Am Chem Soc* **2001**, *123*, 8638–8639.

(29) Wohlert, J.; Edholm, O. *J Chem Phys* **2006**, *125*, 204703.

(30) Klauda, J. B.; Kučerka, N.; Brooks, B. R.; Pastor, R. W.; Nagle, J. F. *Biophys J* **2006**, *90*, 2796–2807.

(31) Högberg, C.-J.; Lyubartsev, A. P. *J Phys Chem B* **2006**, *110*, 14326–14336.

(32) Moore, P. B.; Lopez, C. F.; Klein, M. L. *Biophys J* **2001**, *81*, 2484–2494.

(33) Shinoda, W.; Mikami, M.; Baba, T.; Hato, M. *Chem Phys Lett* **2004**, *390*, 35–40.

(34) Klauda, J. B.; Brooks, B. R.; Pastor, R. W. *J Chem Phys* **2006**, *125*, 144710.

(35) Bemporad, D.; Essex, J. W.; Luttmann, C. *J Phys Chem B* **2004**, *108*, 4875–4884.

(36) Sachs, J. N.; Crozier, P. S.; Woolf, T. B. *J Chem Phys* **2004**, *121*, 10847–10851.

(37) Shinoda, K.; Shinoda, W.; Baba, T.; Mikami, M. *J Chem Phys* **2004**, *121*, 9648–9654.

(38) Song, Y. H.; Guallar, V.; Baker, N. A. *Biochemistry* **2005**, *44*, 13425–13438.

(39) Lindahl, E.; Edholm, O. *J Chem Phys* **2000**, *113*, 3882–3893.

(40) Gullingsrud, J.; Babakhani, A.; McCammon, J. A. *Mol Simulat* **2006**, *32*, 831–838.

(41) Patra, M. *Eur Biophys J* **2005**, *35*, 79–88.

(42) Gullingsrud, J.; Schulten, K. *Biophys J* **2004**, *86*, 3496–3509.

(43) Ollila, S.; Hyvönen, M. T.; Vattulainen, I. *J Phys Chem B* **2007**, *111*, 3139 –3150.

(44) Ollila, O. H. S.; Róg, T.; Karttunen, M.; Vattulainen, I. *J Struct Biol* **2007**, *159*, 311–323.

(45) Carrillo-Tripp, M.; Feller, S. E. *Biochemistry* **2005**, *44*, 10164–10169.

(46) Marrink, S.-J.; de Vries, A. H.; Mark, A. E. *J Phys Chem B* **2004**, *108*, 750–760.

(47) Lyubartsev, A. P. *Eur Biophys J* **2005**, *35*, 53–61.

(48) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J Phys Chem B* **2001**, *105*, 4464–4470.

(49) Elezgaray, J.; Laguerre, M. *Comput Phys Commun* **2006**, *175*, 264–268.

(50) Kranenburg, M.; Nicolas, J. P.; Smit, B. *Phys Chem Chem Phys* **2004**, *6*, 4142–4151.

(51) Izvekov, S.; Voth, G. A. *J Phys Chem B* **2005**, *109*, 2469–2473.

(52) Lopez, C. F.; Moore, P. B.; Shelley, J. C.; Shelley, M. Y.; Klein, M. L. *Comput Phys Commun* **2002**, *147*, 1–6.

(53) Marrink, S.-J.; Mark, A. E. *Biophys J* **2004**, *87*, 3894–3900.

(54) Nielsen, S. O.; Lopez, C. F.; Ivanov, I.; Moore, P. B.; Shelley, J. C.; Klein, M. L. *Biophys J* **2004**, *87*, 2107–2115.

(55) Marrink, S.-J.; Mark, A. E. *J Am Chem Soc* **2003**, *125*, 11144–11145.

(56) Kranenburg, M.; Vlaar, M.; Smit, B. *Biophys J* **2004**, *87*, 1596–1605.

(57) Pickholz, M.; Saiz, L.; Klein, M. L. *Biophys J* **2005**, *88*, 1524–1534.

(58) Izvekov, S.; Voth, G. A. *J Chem Theory Comput* **2006**, *2*, 637–648.

(59) Bond, P. J.; Sansom, M. S. P. *J Am Chem Soc* **2006**, *128*, 2697–2704.

(60) Chang, R.; Ayton, G. S.; Voth, G. A. *J Chem Phys* **2005**, *122*, 244716.

(61) Shi, Q.; Izvekov, S.; Voth, G. A. *J Phys Chem B* **2006**, *110*, 15045–15048.

(62) Voet, D.; Voet, J. G. *Biochemistry;* John Wiley & sons, inc., third ed., 2004.

(63) Dowhan, W.; Bogdanov, M. *Functional roles of lipid in membranes;* Elsevier, 2004.

(64) Simons, K.; Toomre, D. *Nat Rev Mol Cell Bio* **2000**, *1*, 31–39.

(65) `http://www.ualr.edu`. Accessed October 2006.

(66) `http://www.scq.ubc.ca`. Accessed January 2008.

(67) `http://www.steve.gb.com/images/science`. Accessed January 2008.

(68) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J Phys Chem B* **2002**, *106*, 521–533.

(69) Chandler, D. *Nature* **2005**, *437*, 640–647.

(70) Hamley, I. W. *Introduction to Soft Matter;* Wiley, 2007.

(71) Nagle, J. F.; Tristram-Nagle, S. *Biochim Biophys Acta* **2000**, *1469*, 159–195.

(72) Koenig, B. W.; Strey, H. H.; Gawrisch, K. *Biophys J* **1997**, *73*, 1954–1966.

(73) `http://persweb.wabash.edu/facstaff/fellers/coordinates/jcp-a68.pdb`. Accessed March 2008.

(74) Akutsu, H.; Nagamori, T. *Biochemistry* **1991**, *30*, 4510–4516.

(75) Seelig, J.; Niederberger, W. *J Am Chem Soc* **1974**, *96*, 2069–2072.

(76) Janiak, M. J.; Small, D. M.; Shipley, G. G. *J Biol Chem* **1979**, *254*, 6068–6078.

(77) Seddon, J. M.; Templer, R. H. In *Structure and Dynamics of Membranes*; Lipowsky, R., Sackmann, E., Eds.; Elsevier, 1995; pages 97–160.

(78) Yang, L.; Ding, L.; Huang, H. W. *Biochemistry* **2003**, *42*, 6631–6635.

(79) Kozlovsky, Y.; Efrat, A.; Siegel, D. P.; Kozlov, M. M. *Biophys J* **2004**, *87*, 3619–3619.

(80) Siegel, D. P.; Kozlov, M. M. *Biophys J* **2004**, *87*, 366–374.

(81) Curran, A. R.; Templer, R. H.; Booth, P. J. *Biochemistry* **1999**, *38*, 9328–9336.

(82) Meijberg, W.; Booth, P. J. *J Mol Biol* **2002**, *319*, 839–853.

(83) Hong, H. D.; Tamm, L. K. *Proc Natl Acad Sci USA* **2004**, *101*, 4065–4070.

(84) Bowie, J. U. *Nature* **2005**, *438*, 581–589.

(85) Attard, G. S.; Templer, R. H.; Smith, W. S.; Hunt, A. N.; Jackowski, S. *Proc Natl Acad Sci USA* **2000**, *97*, 9032–9036.

(86) Davies, S. M. A.; Epand, R. M.; Kraayenhof, R.; Cornell, R. B. *Biochemistry* **2001**, *40*, 10522–10531.

(87) Fanani, M. L.; Topham, M. K.; Walsh, J. P.; Epand, R. M. *Biochemistry* **2004**, *43*, 14767–14777.

(88) Sen, A.; Isac, T. V.; Hui, S. W. *Biochemistry* **1991**, *30*, 4516–4521.

(89) Ruiz-arguello, M. B.; Goni, F. M.; Alonso, A. *Biochemistry* **1998**, *37*, 11621–11628.

(90) Botelho, A. V.; Gibson, N. J.; Thurmond, R. L.; Wang, Y.; Brown, M. F. *Biochemistry* **2002**, *41*, 6354–6368.

(91) Wang, Y.; Botelho, A. V.; Martinez, G. V.; Brown, M. F. *J Am Chem Soc* **2002**, *124*, 7690–7701.

(92) Botelho, A. V.; Huber, T.; Sakmar, T. P.; Brown, M. F. *Biophys J* **2006**, *91*, 4464–4477.

(93) Keller, S. L.; Bezrukov, S. M.; Gruner, S. M.; Tate, M. W.; Vodyanoy, I.; Parsegian, V. A. *Biophys J* **1993**, *65*, 23–27.

(94) Perozo, E.; Kloda, A.; Cortes, D. M.; Martinac, B. *Nat Struct Biol* **2002**, *9*, 696–703.

(95) Jensen, M. O.; Mouritsen, O. G.; Peters, G. H. *Biophys J* **2004**, *86*(6), 3556–3575.

(96) Rostovtseva, T. K.; Kazemi, N.; Weinrich, M.; Bezrukov, S. M. *J Biol Chem* **2006**, *281*, 37496–37506.

(97) Helfrich, W. *Z Naturforsch* **1973**, *28c*, 693–703.

(98) Marsh, D. *Chem Phys Lipids* **2006**, *144*, 146–159.

(99) Shinoda, W.; Shimizu, M.; Okazaki, S. *J Phys Chem B* **1998**, *102*, 6647–6654.

(100) Clarke, R. J. *Biochim Biophys Acta* **1997**, *1327*, 269–278.

(101) Clarke, R. J. *Adv Colloid Interface Sci* **2001**, *89*, 263–281.

(102) Cladera, J.; Martin, I.; Ruysschaert, J. M.; O'Shea, P. *J Biol Chem* **1999**, *274*, 29951–29959.

(103) Franklin, J. C.; Cafiso, D. S. *Biophys J* **1993**, *65*, 289–299.

(104) Gelbart, W. M.; Bruinsma, R. F.; Pincus, P. A.; Parsegian, V. A. *Physics Today* **2000**, *53*, 38–44.

(105) Qin, Z. H.; Szabo, G.; Cafiso, D. S. *Biochemistry* **1995**, *34*, 5536–5543.

(106) Cafiso, D. S. *Toxicol Lett* **1998**, *101*, 431–439.

(107) Luker, G. D.; Flagg, T. P.; Sha, Q.; Luker, K. E.; Pica, C. M.; Nichols, C. G.; Piwnica-worms, D. *J Biol Chem* **2001**, *276*, 49053–49060.

(108) O'Shea, P. *Biochem Soc Trans* **2003**, *31*, 990–996.

(109) Gawrisch, K.; Ruston, D.; Zimmerberg, J.; Parsegian, V. A.; Rand, R. P.; Fuller, N. *Biophys J* **1992**, *61*, 1213–1223.

(110) Schamberger, J.; Clarke, R. J. *Biophys J* **2002**, *82*, 3081–3088.

(111) Lairion, F.; Disalvo, E. A. *Langmuir* **2004**, *20*, 9151–9155.

(112) Sonnleitner, A.; Schutz, G. J.; Schmidt, T. *Biophys J* **1999**, *77*, 2638–2642.

(113) Tabony, J.; Perly, B. *Biochim Biophys Acta* **1991**, *1063*, 67–72.

(114) Almeida, P. F. F.; Vaz, W. L. C.; Thompson, T. E. *Biochemistry* **1992**, *31*, 6739–6747.

(115) Vaz, W. L. C.; Almeida, P. F. *Biophys J* **1991**, *60*, 1553–1554.

(116) Cohen, M. H.; Turnbull, D. *J Chem Phys* **1959**, *31*, 1164.

(117) Turnbull, D.; Cohen, M. H. *J Chem Phys* **1961**, *34*, 120.

(118) Turnbull, D.; Cohen, M. H. *J Chem Phys* **1970**, *52*, 3038.

(119) O'Leary, T. J. *Proc Natl Acad Sci USA* **1987**, *84*, 429–433.

(120) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids;* Oxford Science Publications, first ed., 1987.

(121) Leach, A. R. *Molecular Modelling - Principles and Applications;* Prentice Hall, second ed., 2001.

(122) Frenkel, D.; Smit, B. *Understanding molecular simulation;* Academic Press, 2002.

(123) Schlick, T. *Molecular Modeling and Simulation - An Interdisciplinary Guide;* Springer, 2002.

(124) Rapaport, D. C. *The Art of Molecular Dynamics Simulation;* Cambridge University Press, second ed., 2004.

(125) Gay, J. G.; Berne, B. J. *J Chem Phys* **1981**, *74*, 3316–3319.

(126) Price, S. L.; Stone, A. J.; Alderton, M. *Mol Phys* **1984**, *52*, 987–1001.

(127) Luckhurst, G. R.; Stephens, R. A.; Phippen, R. W. *Liq Cryst* **1990**, *8*, 451–464.

(128) Leimkuhler, B.; Reich, S. *Simulating Hamiltonian Dynamics;* Cambridge University Press, 2004.

(129) Tuckerman, M. E.; Martyna, G. J. *J Phys Chem B* **2000**, *104*, 159–178.

(130) McLachlan, R. I. *Phys Rev Lett* **1993**, *71*, 3043–3046.

(131) Gray, S. K.; Noid, D. W.; Sumpter, B. G. *J Chem Phys* **1994**, *101*, 4062–4072.

(132) Dullweber, A.; Leimkuhler, B.; McLachlan, R. *J Chem Phys* **1997**, *107*, 5840–5851.

(133) Miller III, T. F.; Eleftheriou, M.; Pattnaik, P.; Ndirango, A.; Newns, D.; Martyna, G. J. *J Chem Phys* **2002**, *116*, 8649–8659.

(134) Omelyan, I. P.; Mryglod, I. M.; Folk, R. *Cond Matt Phys* **2002**, *5*, 369–390.

(135) Praprotnik, M.; Janezic, D. *J Chem Phys* **2005**, *122*.

(136) Allen, M. P. *Computational Soft Matter* **2004**, *23*, 1–28.

(137) Fincham, D. *CCP5 Quarterly* **1981**, *2*, 6–10.

(138) Fincham, D. *CCP5 Quarterly* **1984**, *12*, 47.

(139) Orsi, M.; Sanderson, W.; Essex, J. W. In *Molecular Interactions - Bringing Chemistry to Life*; Hicks, M. G., Kettner, C., Eds.; Beilstein-Institut, Frankfurt, 2007; pages 185–205.

(140) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J Phys: Condens Matter* **2004**, *16*, R481–R512.

(141) Tozzini, V. *Curr Opin Struc Biol* **2005**, *15*, 144–150.

(142) Brannigan, G.; Lin, L. C. L.; Brown, F. L. H. *Eur Biophys J* **2006**, *35*, 104–124.

(143) Shillcock, J. C.; Lipowsky, R. *J Phys: Condens Matter* **2006**, *18*, S1191–S1219.

(144) Muller, M.; Katsov, K.; Schick, M. *Phys Rep* **2006**, *434*, 113–176.

(145) Venturoli, M.; Sperotto, M. M.; Kranenburg, M.; Smit, B. *Phys Rep* **2006**, *437*, 1–54.

(146) van der Ploeg, P.; Berendsen, H. J. C. *J Chem Phys* **1982**, *76*, 3271–3276.

(147) van der Ploeg, P.; Berendsen, H. J. C. *Mol Phys* **1983**, *49*, 233–248.

(148) Egberts, E.; Marrink, S.-J.; Berendsen, H. J. C. *Eur Biophys J* **1994**, *22*, 423–436.

(149) Tieleman, D. P.; Marrink, S.-J.; Berendsen, H. J. C. *Biochim Biophys Acta* **1997**, *1331*, 235–270.

(150) Merz, K. M. *Curr Opin Struc Biol* **1997**, *7*, 511–517.

(151) Feller, S. E.; Venable, R. M.; Pastor, R. W. *Langmuir* **1997**, *13*, 6555–6561.

(152) Scott, H. L. *Curr Opin Struc Biol* **2002**, *12*, 495–502.

(153) Marrink, S.-J.; Mark, A. E. *J Phys Chem B* **2001**, *105*, 6122–6127.

(154) Lindahl, E.; Edholm, O. *J Chem Phys* **2001**, *10*, 4938–4950.

(155) Saiz, L.; Klein, M. L. *J Chem Phys* **2002**, *116*, 3052–3057.

(156) Wohlert, J.; Edholm, O. *Biophys J* **2004**, *87*, 2433–2445.

(157) Falck, E.; Patra, M.; Karttunen, M.; Hyvonen, M. T.; Vattulainen, I. *Biophys J* **2004**, *87*, 1076–1091.

(158) Xiang, T.-X.; Anderson, B. D. *Adv Drug Deliver Rev* **2006**, *58*, 1357–1378.

(159) Murtola, T.; Rog, T.; Falck, E.; Karttunen, M.; Vattulainen, I. *Phys Rev Lett* **2006**, *97*.

(160) Anézo, C.; Vries, A. H.; Holtje, H. D.; Tieleman, D. P.; Marrink, S.-J. *J Phys Chem B* **2003**, *107*, 9424–9433.

(161) Benz, R. W.; Castro-Roman, F.; Tobias, D. J.; White, S. H. *Biophys J* **2005**, *88*, 805–817.

(162) de Vries, A. H.; Chandrasekhar, I.; van Gunsteren, W. F.; Hunenberger, P. H. *J Phys Chem B* **2005**, *109*, 11643–11652.

(163) Smit, B.; Hilbers, P. A. J.; Esselink, K.; Rupert, L. A. M.; van Os, N. M.; Schlijper, A. G. *Nature* **1990**, *348*, 624–625.

(164) Smit, B.; Hilbers, P. A. J.; Esselink, K.; Rupert, L. A. M.; van Os, N. M.; Schlijper, A. G. *J Phys Chem* **1991**, *95*, 6361–6368.

(165) Groot, R. D.; Warren, P. B. *J Chem Phys* **1997**, *107*, 4423–4435.

(166) Venturoli, M.; Smit, B. *Phys Chem Comm* **1999**, *10*, 9/064721.

(167) Groot, R. D. *Langmuir* **2000**, *16*, 7493–7502.

(168) Groot, R. D.; Rabone, K. L. *Biophys J* **2001**, *81*, 725–736.

(169) Kranenburg, M.; Smit, B. *J Phys Chem B* **2005**, *109*, 6553–6563.

(170) Groot, R. D. *J Chem Phys* **2003**, *118*, 11265–11277.

(171) Kranenburg, M.; Venturoli, M.; Smit, B. *J Phys Chem B* **2003**, *107*, 11491–11501.

(172) Kranenburg, M.; Venturoli, M.; Smit, B. *Phys Rev E* **2003**, *67*, 060901.

(173) Kranenburg, M.; Laforge, C.; Smit, B. *Phys Chem Chem Phys* **2004**, *6*, 4531–4534.

(174) Kranenburg, M.; Smit, B. *FEBS Lett* **2004**, *568*, 15–18.

(175) Venturoli, M.; Smit, B.; Sperotto, M. M. *Biophys J* **2005**, *88*, 1778–1798.

(176) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Moore, P. B.; Klein, M. L. *J Phys Chem B* **2001**, *105*, 9785–9792.

(177) Srinivas, G.; Klein, M. L. *Nanotechnology* **2004**, *15*, 1289–1295.

(178) Srinivas, G.; Shelley, J. C.; Nielsen, S. O.; Discher, D. E.; Klein, M. L. *J Phys Chem B* **2004**, *108*, 8153–8160.

(179) Lopez, C. F.; Nielsen, S. O.; Moore, P. B.; Klein, M. L. *Proc Natl Acad Sci* **2004**, *101*, 4431–4434.

(180) Lopez, C. F.; Nielsen, S. O.; Ensing, B.; Moore, P. B.; Klein, M. L. *Biophys J* **2005**, *88*, 3083–3094.

(181) Nielsen, S. O.; Ensing, B.; Ortiz, V.; Moore, P. B.; Klein, M. L. *Biophys J* **2005**, *88*, 3822–3828.

(182) Lopez, C. F.; Nielsen, S. O.; Srinivas, G.; Degrado, W. F.; Klein, M. L. *J Chem Theory Comput* **2006**, *2*, 649–655.

(183) Marrink, S.-J.; Risselada, J.; Mark, A. E. *Chem Phys Lipids* **2005**, *135*, 223–244.

(184) Marrink, S.-J.; Mark, A. E. *J Am Chem Soc* **2003**, *125*, 15233–15242.

(185) de Joannis, J.; Jiang, F. Y.; Kindt, J. T. *Langmuir* **2006**, *22*, 998–1005.

(186) Faller, R.; Marrink, S.-J. *Langmuir* **2004**, *20*, 7686–7693.

(187) Shi, Q.; Voth, G. A. *Biophys J* **2005**, *89*, 2385–2394.

(188) Dickey, A. N.; Faller, R. *J Polym Sci* **2005**, *43*, 1025–1032.

(189) Lee, H.; Larson, R. G. *J Phys Chem B* **2006**, *110*, 18204–18211.

(190) Shih, A.; Arkhipov, A.; Freddolino, P.; Schulten, K. *Journal of Physical Chemistry B* **2006**, *110*(8), 3674–3684.

(191) Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J Am Chem Soc* **2007**, *129*, 10126–10132.

(192) Adhangale, P. S.; Gaver, D. P. *Mol Phys* **2006**, *104*, 3011–3019.

(193) May, E. R.; Narang, A.; Kopelevich, D. I. *Mol Simulat* **2007**, *33*, 787–797.

(194) Marrink, S.-J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J Phys Chem B* **2007**, *111*, 7812–7824.

(195) Goetz, R.; Lipowsky, R. *J Chem Phys* **1998**, *108*, 7397–7409.

(196) Goetz, R.; Gompper, G.; Lipowsky, R. *Phys Rev Lett* **1999**, *82*, 221–224.

(197) Imparato, A.; Shillcock, J. C.; Lipowsky, R. *Eur Phys J E* **2003**, *11*, 21–28.

(198) Imparato, A.; Shillcock, J. C.; Lipowsky, R. *Europhys Lett* **2005**, *69*, 650–656.

(199) den Otter, W. K.; Briels, W. J. *J Chem Phys* **2003**, *118*, 4712–4720.

(200) Boek, E. S.; den Otter, W. K.; Briels, W. J.; Iakovlev, D. *Phil Trans R Soc Lond A* **2004**, *362*, 1625–1638.

(201) Loison, C.; Mareschal, M.; Kremer, K.; Schmid, F. *J Chem Phys* **2003**, *119*, 13138–13148.

(202) Loison, C.; Mareschal, M.; Schmid, F. *J Chem Phys* **2004**, *121*, 1890–1900.

(203) Loison, C.; Mareschal, M.; Schmid, F. *Comput Phys Commun* **2005**, *169*, 99–103.

(204) Stevens, M. J. *J Chem Phys* **2004**, *121*, 11942–11948.

(205) Stevens, M. J.; Hoh, J. H.; Woolf, T. B. *Phys Rev Lett* **2003**, *91*, 188102.

(206) Lenz, O.; Schmid, F. *J Mol Liq* **2005**, *117*, 147–152.

(207) Whitehead, L.; Edge, C. M.; Essex, J. W. *J Comput Chem* **2001**, *22*, 1622–1633.

(208) Noguchi, H.; Takasu, M. *Phys Rev E* **2001**, *64*, 041913.

(209) Noguchi, H.; Takasu, M. *J Chem Phys* **2001**, *115*, 9547–9551.

(210) Noguchi, H.; Takasu, M. *Biophys J* **2002**, *83*, 299–308.

(211) Noguchi, H.; Takasu, M. *Phys Rev E* **2002**, *65*, 051907.

(212) Lyubartsev, A. P.; Laaksonen, A. *Lect Notes Phys* **2004**, *640*, 214–244.

(213) Brannigan, G.; Brown, F. L. H. *J Chem Phys* **2004**, *120*, 1059–1071.

(214) Brannigan, G.; Tamboli, A. C.; Brown, F. L. H. *J Chem Phys* **2004**, *121*, 3259–3271.

(215) Brannigan, G.; Philips, P. F.; Brown, F. L. H. *Phys Rev E* **2005**, *72*, 011915.

(216) Brannigan, G.; Brown, F. L. H. *J Chem Phys* **2005**, *122*, 074905.

(217) Cooke, I. R.; Kremer, K.; Deserno, M. *Phys Rev E* **2005**, *72*, 011506.

(218) Harmandaris, V. A.; Deserno, M. *J Chem Phys* **2006**, *125*, 204905.

(219) Reynwar, B. J.; Illya, G.; Harmandaris, V. A.; Muller, M. M.; Kremer, K.; Deserno, M. *Nature* **2007**, *447*, 461–464.

(220) Orsi, M.; Haubertin, D. Y.; Sanderson, W.; Essex, J. W. *J Phys Chem B* **2008**, *112*, 802.

(221) Cleaver, D. J.; Care, C. M.; Allen, M. P.; Neal, M. P. *Phys Rev E* **1996**, *54*, 559–567.

(222) Berne, B. J.; Pechukas, P. *J Chem Phys* **1972**, *56*, 4213–4216.

(223) Antypov, D.; Cleaver, D. J. *J Chem Phys* **2004**, *120*, 10307–10316.

(224) Golubkov, P. A.; Ren, P. *J Chem Phys* **2006**, *125*, 064103.

(225) Liu, Y.; Ichiye, T. *J Phys Chem* **1996**, *100*, 2723–2730.

(226) Chandra, A.; Ichiye, T. *J Chem Phys* **1999**, *111*, 2701–2709.

(227) Fennell, C. J.; Gezelter, J. D. *J Chem Phys* **2004**, *120*, 9175–9184.

(228) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces;* Reidel, Dordrecht, 1981.

(229) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J Phys Chem* **1987**, *91*, 6269–6271.

(230) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* **1983**, *79*, 926–935.

(231) Tan, M.-L.; Brooks, B. R.; Ichiye, T. *Chem Phys Lett* **2006**, *421*, 166–170.

(232) Darden, T.; York, D.; Pedersen, L. *J Chem Phys* **1993**, *98*, 10089–10092.

(233) Tobias, D. J. *Curr Opin Struc Biol* **2001**, *11*, 253–261.

(234) Fennell, C. J.; Gezelter, J. D. *J Chem Phys* **2006**, *124*, 234104.

(235) Beck, D. A. C.; Armen, R. S.; Daggett, V. *Biochemistry* **2005**, *44*, 609–616.

(236) `http://moose.bio.ucalgary.ca/files/dmpc_npat.pdb`. Biocomputing group, University of Calgary. Accessed April 2007.

(237) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J Chem Phys* **1984**, *81*, 3684–3690.

(238) Harvey, S. C.; Tan, R. K. Z.; Cheatham III, T. E. *J Comput Chem* **1998**, *19*, 726–740.

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

(240) Iridis is a Beowulf cluster based on AMD Opteron processors, running RedHat Enterprise Linux. `http://www.soton.ac.uk/iss/computing/hpc/iridis/`.

(241) Synge, J. L.; Griffith, B. A. *Principles of Mechanics;* McGraw-Hill, 1959.

(242) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J Comput Chem* **1999**, *20*, 786–798.

(243) Mills, R. *J Phys Chem* **1973**, *77*, 685–688.

(244) Shepherd, J. C. W.; Büldt, G. *Biochim Biophys Acta* **1978**, *514*, 83–94.

(245) Högberg, C.-J.; Lyubartsev, A. P. *Biophys J* **2008**, *94*, 525–531.

(246) Chu, N.; Kučerka, N.; Liu, Y. F.; Tristram-Nagle, S.; Nagle, J. F. *Phys Rev E* **2005**, *71*, 041904.

(247) Méléard, P.; Gerbeaud, C.; Pott, T.; Fernandez-Puente, L.; Bivas, I.; Mitov, M. D.; Dufourcq, J.; Bothorel, P. *Biophys J* **1997**, *72*, 2616–2629.

(248) Lee, C.-H.; Lin, W.-C.; J., W. *Phys Rev E* **2001**, *6402*, 020901.

(249) Gromacs User Manual version 3.1. van der Spoel, D.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Hess, B.; Feenstra, K. A.; Lindahl, E.; van Drunen, R.; Berendsen, H. J. C.; Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: www.gromacs.org, **2001**.

(250) Ben-Shaul, A. In *Structure and Dynamics of Membranes*; Lipowsky, R., Sackmann, E., Eds.; Elsevier, 1995; pages 359–401.

(251) Mukhin, S. I.; Baoukina, S. *Phys Rev E* **2005**, *71*, 061918.

(252) Xiang, T.-X.; Anderson, B. D. *Biophys J* **1994**, *66*, 561–572.

(253) Frink, L. J.; Frischknecht, A. L. *Biophys J* **2006**, *91*, 4081.

(254) Shillcock, J. C.; Lipowsky, R. *J Chem Phys* **2002**, *117*, 5048–5061.

(255) Frischknecht, A. L.; Frink, L. J. D. *Phys Rev E* **2005**, *72*, 041924.

(256) Szleifer, I.; Kramer, D.; Ben-Shaul, A.; Gelbart, W. M.; Safran, S. A. *J Chem Phys* **1990**, *92*, 6800–6817.

(257) Harries, D.; Ben-Shaul, A. *J Chem Phys* **1997**, *106*, 1609–1619.

(258) Cantor, R. S. *Biophys J* **1999**, *76*, 2625–2639.

(259) Cantor, R. S. *Chem Phys Lipids* **1999**, *101*, 45–56.

(260) Brannigan, G.; Brown, F. L. H. *Biophys J* **2006**, *90*, 1501–1520.

(261) Brannigan, G.; Brown, F. L. H. *Biophys J* **2007**, *92*, 864–876.

(262) Templer, R. H.; Khoo, B. J.; Seddon, J. M. *Langmuir* **1998**, *14*, 7427–7434.

(263) Sokhan, V. P.; Tildesley, D. J. *Mol Phys* **1997**, *92*, 625–640.

(264) Smondyrev, A. M.; Berkowitz, M. L. *J Comput Chem* **1999**, *20*, 531–545.

(265) Cordomi, A.; Edholm, O.; Perez, J. J. *J Phys Chem B* **2008**.

(266) Flewelling, R. F.; Hubbell, W. L. *Biophys J* **1986**, *49*, 541–552.

(267) Smyth, C. P. *Dielectric behavior and structure;* McGraw-Hill, 1955.

(268) Bezrukov, S. M. *Curr Opin Colloid Interface Sci* **2000**, *5*, 237–243.

(269) Nagle, J. F.; Mathai, J. C.; Zeidel, M. L.; Tristram-Nagle, S. *J Gen Physiol* **2008**, *131*, 77–85.

(270) Michel, J.; Orsi, M.; Essex, J. W. *J Phys Chem B* **2008**, *112*, 657.

(271) Pohorille, A.; Wilson, M. A. *Cell Mol Biol Lett* **2001**, *6*, 369–374.

(272) Overton, E. *Vierteljahrsschr Naturforsch Ges Zurich* **1896**, *41*, 383.

(273) Finkelstein, A. *J Gen Physiol* **1976**, *68*, 127–135.

(274) Diamond, J. M.; Katz, Y. *J Membrane Biol* **1974**, *17*, 121–154.

(275) Diamond, J. M.; Szabo, G.; Katz, Y. *J Membrane Biol* **1974**, *17*, 148–152.

(276) Marrink, S.-J.; Berendsen, H. J. C. *J Phys Chem* **1994**, *98*, 4155–4168.

(277) Bassolino-Klimas, D.; Alper, H. E.; Stouch, T. R. *Biochemistry* **1993**, *32*, 12624–12637.

(278) Marrink, S.-J.; Berendsen, H. J. C. *J Phys Chem* **1996**, *100*, 16729–16738.

(279) Shinoda, W.; Mikami, M.; Baba, T.; Hato, M. *J Phys Chem B* **2004**, *108*, 9346–9356.

(280) Sugii, T.; Takagi, S.; Matsumoto, Y. *J Chem Phys* **2005**, *123*, 184714.

(281) Bemporad, D.; Luttmann, C.; Essex, J. W. *Biophys J* **2004**, *87*, 1–13.

(282) Grossfield, A.; Woolf, T. B. *Langmuir* **2002**, *18*, 198–210.

(283) Ulander, J.; Haymet, A. D. J. *Biophys J* **2003**, *85*, 3475–3484.

(284) Bemporad, D.; Luttmann, C.; Essex, J. W. *Biochim Biophys Acta* **2005**, *1718*, 1–21.

(285) dos Santos, D. J. V. A.; Eriksson, L. A. *Biophys J* **2006**, *91*, 2464–2474.

(286) Xiang, T.-X.; Anderson, B. D. *J Membrane Biol* **1994**, *140*, 111–122.

(287) Abrams, C. F. *J Chem Phys* **2005**, *123*.

(288) Werder, T.; Walther, J. H.; Koumoutsakos, P. *J Comput Phys* **2005**, *205*, 373–390.

(289) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. *Phys Rev Lett* **2005**, *95*, 218102.

(290) Neri, M.; Anselmi, C.; Carnevale, V.; Vargiu, A. V.; Carloni, P. *J. Phys:: Condens. Matter* **2006**, *18*, S347–S355.

(291) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. *J Chem Theory Comput* **2007**, *3*, 1100–1105.

(292) Praprotnik, M.; Delle Site, L.; Kremer, K. *Phys Rev E* **2006**, *73*, 066701.

(293) Praprotnik, M.; Delle Site, L.; Kremer, K. *J Chem Phys* **2007**, *126*.

(294) Praprotnik, M.; Matysiak, S.; Delle Site, L.; Kremer, K.; Clementi, C. *J Phys: Condens Matter* **2007**, *19*, 292201.

(295) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr Opin Struc Biol* **2007**, *17*, 192–198.

(296) Huster, D.; Jin, A. J.; Arnold, K.; Gawrisch, K. *Biophys J* **1997**, *73*, 855–864.

(297) Jansen, M.; Blume, A. *Biophys J* **1995**, *68*, 997–1008.

(298) Carruthers, A.; Melchior, D. L. *Biochemistry* **1983**, *22*, 5797–5807.

(299) Finkelstein, A. *J Gen Physiol* **1976**, *68*, 127–135.

(300) Paula, S.; Volkov, A. G.; Vanhoek, A. N.; Haines, T. H.; Deamer, D. W. *Biophys J* **1996**, *70*, 339–348.

(301) Xiang, T.-X.; Chen, X. L.; Anderson, B. D. *Biophys J* **1992**, *63*, 78–88.

(302) Orbach, E.; Finkelstein, A. *J Gen Physiol* **1980**, *75*, 427–436.

(303) Walter, A.; Gutknecht, J. *J Membrane Biol* **1984**, *77*, 255–264.

(304) Xiang, T.-X.; Anderson, B. D. *J Membrane Biol* **1995**, *148*, 157–167.

(305) Roux, B.; Karplus, M. *J Phys Chem* **1991**, *95*, 4856–4868.

(306) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* **2004**, *25*, 1157–1174.

(307) Jakalian, A.; Bush, B. L.; Jack, D. B. *J Comput Chem* **2000**, *21*, 132–146.

(308) Goldstein, H.; Poole, C.; Safko, J. *Classical Mechanics - third edition;* Pearson, 2002.

(309) Maple 8.00, Waterloo Maple Inc.

(310) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(311) Jedlovszky, P.; Mezei, M. *J Am Chem Soc* **2000**, *122*, 5125–5131.

(312) Zhu, T. H.; Li, J. B.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Chem Phys* **1998**, *109*, 9117–9133.

(313) Hill, W. G.; Rivers, R. L.; Zeidel, M. L. *J Gen Physiol* **1999**, *114*, 405–414.

(314) Walter, A.; Gutknecht, J. *J Membrane Biol* **1986**, *90*, 207–217.

(315) Bar-On, Z.; Degani, H. *Biochim Biophys Acta* **1985**, *813*, 207–212.

(316) Pohorille, A.; Wilson, M. A. *J Chem Phys* **1996**, *104*, 3760–3773.

(317) Alper, H. E.; Stouch, T. R. *J Phys Chem* **1995**, *99*, 5724–5731.

(318) Sonne, J.; Jensen, M. Ø.; Hansen, F. Y.; Hemmingsen, L.; Peters, G. H. *Biophys J* **2007**, *92*, 4157–4167.

(319) Shah, P.; Jogani, V.; Bagchi, T.; Misra, A. *Biotechnol Progr* **2006**, *22*, 186–198.

(320) Ruell, J. *Mod Drug Discov* **2003**, *Jan*, 28–30.

(321) Kansy, M.; Senner, F.; Gubernator, K. *J Med Chem* **1998**, *41*, 1007–1010.

(322) Avdeef, A.; Nielsen, P. E.; Tsinman, O. *Eur J Pharm Sci* **2004**, *22*, 365–374.

(323) Carrara, S.; Reali, V.; Misiano, P.; Dondio, G.; Bigogno, C. *Int. J. Pharmaceut.* **2007**.

(324) Verma, R. P.; Hansch, C.; Selassie, C. D. *J Comput Aid Mol Des* **2007**, *21*, 3–22.

(325) Galinis-Luciani, D.; Nguyen, L.; Yazdanian, M. *J Pharm Sci* **2007**, *96*, 2886–2892.

(326) Avdeef, A.; Artursson, P.; Neuhoff, S.; Lazorova, L.; Grasjo, J.; Tavelin, S. *Eur J Pharm Sci* **2005**, *24*, 333–349.

(327) Penzotti, J. E.; Landrum, G. A.; Putta, S. *Curr Opin Drug Discovery Dev* **2004**, *7*, 49–61.

(328) Bemporad, D. *Computer Simulation of Biological Membranes and Small Molecule Permeation.* PhD thesis, University of Southampton, **2003**.

(329) Unpublished data. Cheney, B. **2007**.

(330) Shin, J.; Johnson, J. A. *Pharmacotherapy* **2007**, *27*, 874–887.

(331) Adson, A.; Burton, P. S.; Raub, T. J.; Barsuhn, C. L.; Audus, K. L.; Ho, N. F. H. *J Pharm Sci* **1995**, *84*, 1197–1204.

(332) Palm, K.; Luthman, K.; Ungell, A. L.; Strandlund, G.; Artursson, P. *J Pharm Sci* **1996**, *85*, 32–39.

(333) Fujikawa, M.; Ano, R.; Nakao, K.; Shimizu, R.; Akamatsu, M. *Bioorgan Med Chem* **2005**, *13*, 4721–4732.

(334) O'Leary, T. J.; Ross, P. D.; Levin, I. W. *Biochemistry* **1984**, *23*, 4636–4641.

(335) Ottaviani, G.; Martel, S.; Carrupt, P. A. *J Med Chem* **2007**, *50*, 742–748.

(336) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv Drug Deliver Rev* **1997**, *23*, 3–25.

(337) Mouritsen, O. G.; Zuckermann, M. J. *Lipids* **2004**, *39*, 1101–1113.

(338) Szabo, G. *Nature* **1974**, *252*, 47–49.

(339) Orädd, G.; Lindblom, G.; Westerman, P. W. *Biophys J* **2002**, *83*, 2702-2704.

(340) Filippov, A.; Orädd, G.; Lindblom, G. *Biophys J* **2003**, *84*, 3079–3086.

(341) Sackmann, E. In *Structure and Dynamics of Membranes*; Lipowsky, R., Sackmann, E., Eds.; Elsevier, 1995; pages 1–62.

(342) Bagatolli, L. A. *Biochim Biophys Acta* **2006**, *1758*, 1541–1556.

(343) Oualline, S. *Practical C Programming, Third Edition;* O'Reilly, 1997.

(344) Harbison III, S. P.; Steele Jr., G. L. *C - A Reference Manual;* Pearson Education, 2002.

(345) Haubertin, D. Y. *The Simulation of Biomembranes and Drug Transport Therein using a Gay-Berne model* PhD thesis, University of Southampton, **2003**.

(346) Kol, A.; Laird, B. B.; Leimkuhler, B. J. *J Chem Phys* **1997**, *107*, 2580–2588.

(347) Ratanapisit, J.; Isbister, D. J.; Ely, J. F. *Fluid Phase Equil* **2001**, *183*, 351–361.

(348) Kamberaj, H.; Low, R. J.; Neal, M. P. *J Chem Phys* **2005**, *122*, 224114-2.

(349) Morishita, T. *J Chem Phys* **2000**, *113*, 2976–2982.

(350) Knuth, D. E. *The Art Of Computer Programming;* Addison-Wesley, 1968.

(351) Verlet, L. *Phys Rev* **1967**, *159*, 98–103.

(352) Cheung, P. S. Y. *Chem Phys Lett* **1976**, *40*, 19–22.

(353) Popelier, P. L. A.; Stone, A. J. *Mol Phys* **1994**, *82*, 411–425.