UNIVERSITY OF SOUTHAMPTON

The Simulation of Biomembranes and Drug Transport Therein Using a Gay-Berne Model

David Yan Haubertin

A thesis submitted for the qualification of
Doctor of Philosophy at the University of Southampton
Department of Chemistry
September 2003

University of Southampton

ABSTRACT

FACULTY OF SCIENCE

CHEMISTRY

Doctor of Philosophy

THE SIMULATION OF BIOMEMBRANES AND DRUG TRANSPORT THEREIN USING A GAY-BERNE MODEL

by David Yan Haubertin

In recent years the number of simulations of hydrated, fluid phase, lipid bilayers has grown rapidly. All the simulations generally employ molecular dynamics and/or Monte Carlo methods to study a system usually constituted of 100 lipids and a few thousand waters. The study of biological membranes has been extended to include small molecules diffusion within the membrane. However the simulations are time-consuming and require powerful computational resources due to the size of the system. The development of membrane models that are less computationally demanding and still allow to access membrane behaviour, are therefore highly desirable. In this thesis the concepts used to model the liquid crystalline mesophases are applied to the modelling of a lipid membrane. The most popular model for the simulation of the liquid crystalline mesophase is that developed by Gay and Berne,[1] in which molecules are modelled by ellipsoids of various length and breadth ratios. In this model the individual atoms of the molecule are subsumed into a single ellipsoid that interacts with its neighbours through an anisotropic intermolecular potential. This model has already been applied to model the hydrocarbon region of a lipid bilayer[2] and has proved to give a good fit with experiments. The work in this thesis consists of extending this model to build an entire membrane (tail, glycerol, head group and solvent), and then studying the diffusion of small drugs within this environment. Chapter 4 describes the construction of the model. A new parameterization technique for the Gay-Berne potential is introduced. Chapter 5 presents simulations performed with this model as well as a self aggregation study. Analysis of preassembled bilayer simulated with this model clearly demonstrates the existence of a phase transition, and shows the presence of lipid long-range lateral diffusion. The results obtained have been favourably compared to the available experimental data in most cases. The self aggregation study failed to identify the formation of a bilayer of correct structure. Rather, the formation of a two phases system was observed, water and a low-hydration bilayer. The small molecule permeation study, presented in Chapter 6, shows promising results. However, systematic deficiencies of the membrane model in the interfacial region have been identified.

Acknowledgements

I would like to thank my boss, Jonathan Essex for his help and support. I would also like to thanks my industrial supervisor, Trevor Howe as well as Novartis Pharma UK for funding this work.

Invariably, when people read a thesis, the only part that is always read is the acknowledgements. So one should try to make it as enjoyable to read as the remaining of the thesis is painful. I tried to do so, but I could not avoid long list of names, so please accept my apologies if you fall asleep halfway through. Be reassured though, I succeeded in making the remainder of the reading painful.

I would like to thank all the members of the group for their support and many intersting social outings. Namely: Adrian, Caterina, Chris, Daniele, Donna (some people should not be allowed to cycle), Francesca, Ian, Julen, Julien, Lewis, Loredana, Luca, Martin, Phil, Richard H., Richard M., Richard T. (some people should not be allowed to wear pink lycra), Rob, Sebastien (some people should not be allowed to cycle either), Steve, Stuart (some people should not be allowed anything). Special thanks goes to Daniele.

I would have liked to finish this PhD on time, unfortunately I discovered the full meaning of the quote *"a plan is just a list of things that don't happen"*, and had to stay a bit longer. This gave me the opportunity to meet people I would not have met otherwise. Thank you, especially to Deborah.

I would also like to thank my friends David, Emilie, Hervé and Samuel, and especially Isabelle who helped me to identify yet with another Brel song, *"Ma mère arrête tes prières; Ton Jacques retourne en enfer; Mathilde m'est revenue"*, Bonne route.

Finally I would like to thank my family; namely Lionel, Maïté and Carole, for having helped me to get this far; let's not mention all the smelly cheeses and saucissons send by post all the way from France.

> *"I was twenty then. What the hell, I used to say, take your time, Bandini. You got ten years to write a book, so take it easy, get out and learn about life, walk the streets. That's your trouble: your ignorance of life."*
>
> John Fante in *Ask the dust*

Contents

Chapter 1

Introduction and Overview

1.1 Introduction

The ease with which small molecule pharmaceutical agents can pass through biological membranes is a critical factor influencing their *in vivo* potency. Although techniques exist for measuring these effects (either directly with artificial membranes or indirectly by consideration of partition coefficients measurements) there is still a need to investigate these diffusional processes at a molecular level. Computer simulations offer a powerful way of studying diffusional processes. These methods, based on united atom and all-atom representations, have been extensively applied to the study of biological membranes,[3–6] the diffusion of small molecules,[7–13] the simulation of membrane bound proteins,[5, 14–16] lipid-DNA complexes[17] and even spontaneous aggregation of phospholipids into micelles or bilayers.[18, 19] Recent advances in computer power have allowed the simulation of mixed micelles modelling human bile, with simulation times extending to 150 ns.[20] Atomistic simulations are able to reproduce and predict many fundamental properties of phospholipid membranes.[21–23] However, these simulations are time-consuming and require powerful computational resources due to the size of the system. The development of membrane models that are less computationally demanding but that still allow access to membrane behaviour is therefore highly desirable. In this project the concepts used to model liquid crystalline mesophases by computer simulation will be applied to the modelling of a lipid membrane.

Liquid crystal materials exhibit not only solid, liquid and gaseous states but also a liquid crystalline phase or mesophase.[24] This phase has properties consistent with a liquid (flow) but also with a solid (optical birefringence for example). Arguably the most popular model for the simulation of the liquid crystalline mesophase is that developed by Gay and Berne,[1] in which molecules are modelled by ellipsoids of various length and breadth ratios. In this model the individual atoms of the molecule are subsumed into a single ellipsoid that interacts with its neighbours through an anisotropic intermolecular potential. This potential is basically a shifted version of the Lennard-Jones potential where σ and ϵ depend on the orientation of the ellipsoid.[25] As the temperature increases, a liquid crystal material will pass through a solid crystalline phase, a smectic or nematic mesophase and an isotropic liquid. Biological lipids are also able to adopt a range of phases depending on temperature and level of hydration, the most important phase being the L_α phase, that is characterized by a high degree of disorder in the long alkyl chains of the hydrophobic core. The application of the Gay-Berne potential to model biomembranes is therefore very attractive ; the potential is known to yield the sort of phase behaviour shown by a biological membrane, but because the number of interactions is reduced, longer simulations are possible compared with a conventional all-atom representation.

The membrane we have focused on consists of 72 molecules of Dimyristoylphosphatidylcholine (DMPC). The Gay-Berne model of the hydrocarbon region of DMPC has previously been developed by Lewis Whitehead,[2] where each Gay-Berne particle is used to model four methylene groups. Three beads are therefore needed to model each hydrocarbon chain. These beads were linked using a harmonic spring potential and the two chains were linked at the top again using a harmonic spring. The 'head' Gay-Berne particles are restrained in a plane parallel to the membrane surface using a harmonic potential, thus allowing the integrity of the bilayer to be maintained in the absence of the glycerol part, head group and solvation. Simulations were previously performed using this model on a system containing 72 lipids. His work was based on a modified version of an academic program written by Professor Palke.[26] Building on this work, our goal has been to extend this model to include head group representation as well as solvation, and to study the permeation of small molecules across this bilayer.

1.2 Overview

The overall aims of this project are the following:

1. Construction of a computer model of a biological membrane using a Gay-Berne simplified model, including the head group and solvent representation.
2. Simulation of a set of small molecules within this membrane to determine depths of partitioning, rates of diffusion and the energy profile for diffusion.
3. Correlation of collected data with experimental permeability.

Chapter 2

Computer Simulation Methodology

2.1 Introduction

This chapter describes the basis of the theoretical techniques used in this thesis. It does not correspond to an exhaustive discussion, as this field as already been comprehensively covered in many books, for example 'Computer Simulation of Liquids' by Allen and Tildesley[27] and 'Molecular Modelling' by Leach.[28] A reader intending further, more in-depth study, should therefore refer to these books.

2.2 Statistical Thermodynamics

Thermodynamics is the study of macroscopic materials (made of a large number of particles). Its goals are twofold:

- to describe equilibrium states of systems (temperature, pressure, volume, energy...),
- to predict the evolution of systems with time.

Classical thermodynamics is able to predict various global properties of matter without considering the microscopic structure. It is complemented by the study of *sta-*

tistical thermodynamics, which is based on the microscopic description of matter (particles constantly in motion) and of statistical mechanics. Equilibrium states defined by *classical* thermodynamic are therefore statistical equilibrium states; only the average value of physical quantities stay constant, as the particles are continuously subject to thermal agitation.

2.2.1 Microscopic and Macroscopic States of an Assembly of Identical Particles

Consider a system consisting of N identical particles, with a total energy U, where each particle can only access certain energy levels u_1, u_2, ..., u_n. The microscopic state of the system is described by the instantaneous value of the energy of each particle taken individually. The macroscopic state of the system is defined by the number of particles n_i at energy level u_i, without paying attention to which particle it may be. This description is therefore less detailed than the former. Two conditions needs to be respected :

- conservation of the number of particles : $N = \sum_i n_i$
- conservation of the total energy : $U = \sum_i n_i u_i$

Particles constantly move from one energetic state to another due to the thermal agitation, and the possible microscopic states are therefore modified at every instant. There are $N!$ possible energetic states as there are $N!$ possible permutations of the N particles of the system; among these $N!$ microscopic states the $n_1!$ states obtained by permuting the n_1 particles at the energetic level u_1 are identical, the same thing can be said about the $n_2!$ states obtained permuting the n_2 particles. The total number of microscopic states is therefore

$$W = \frac{N!}{n_1!n_2!...n_N!} \tag{2.1}$$

$$\ln(W) = \ln(N!) - \sum_i \ln(n_i!) \tag{2.2}$$

The thermal agitation can be characterised by this number W which is also called weight of the configuration. The introduction of $\ln W$ enables the simplification of the factorial using Stirling's approximation : $\ln x! = x\ln x - x$.

2.2.2 Equilibrium State of an Isolated System. Partition Function

The dominating configuration of a system is such that the number of microscopic states W is largest. This state of thermal equilibrium can be determined by looking at the values of n_i that lead to a maximum value of W i.e. $\mathrm{d}W = 0$. Practically it is equivalent and simpler to find the criterion for which $\mathrm{d}lnW = 0$. In the most probable configuration the following relation must therefore be met by the n_n populations:

$$\mathrm{d}(\ln W) = \sum_i \left(\frac{\partial \ln W}{\partial n_i}\right) \mathrm{d}n_i = 0. \tag{2.3}$$

However, when the n_i change, they do so subject to the constraints :

$$\sum_i u_i \mathrm{d}n_i = 0 \quad \text{and} \quad \sum_i \mathrm{d}n_i = 0, \tag{2.4}$$

i.e. the total energy and the number of particles of the system are constant. The constraints described in Equation 2.4 are introduced through the use of Lagrange multipliers α and β. Multiplying the first constraint by $-\beta$, the second by α and adding them to Equation 2.3, we obtain:

$$\begin{aligned} \mathrm{d}(\ln W) &= \sum_i \left(\frac{\partial \ln W}{\partial n_i}\right) \mathrm{d}n_i + \alpha \sum_i \mathrm{d}n_i - \beta \sum_i u_i \mathrm{d}n_i \\ &= \sum_i \left[\left(\frac{\partial \ln W}{\partial n_i}\right) + \alpha - \beta u_i\right] \mathrm{d}n_i \end{aligned} \tag{2.5}$$

This relation must be true for any $\mathrm{d}n_i$ variation, therefore:

$$\left(\frac{\partial \ln W}{\partial n_i}\right) + \alpha - \beta u_i = 0. \tag{2.6}$$

Using Stirling's approximation we can write $\ln W = N\ln N - \sum_i n_i \ln n_i$, leading to

$$\frac{\partial \ln W}{\partial n_i} = -\ln n_i - 1 \tag{2.7}$$

where -1 can be neglected as $\ln n_i$ is large. Equation 2.6 therefore becomes

$$-\ln n_i + \alpha - \beta u_i = 0. \tag{2.8}$$

The most probable population of the state of energy u_i is therefore :

$$n_i = e^{-\alpha} e^{-\beta u_i}, \tag{2.9}$$

hence

$$N = \sum_i n_i = e^{-\alpha} \sum_i e^{-\beta u_i}. \tag{2.10}$$

The equilibrium distribution law can be obtained by dividing equations 2.9 and 2.10:

$$\frac{n_i}{N} = \frac{e^{-\beta u_i}}{\sum_j e^{-\beta u_j}}. \tag{2.11}$$

This corresponds to the probability of finding a particle in the energetic level u_i, and is referred to as the Boltzmann distribution. The denominator corresponds to the molecular partition function q. The probability p_i of finding a particle at the energetic level u_i is therefore

$$p_i = \frac{e^{-\beta u_i}}{q}. \tag{2.12}$$

It can be demonstrated that $\beta = 1/kT$ where T is the thermodynamic temperature. The partition function contains all the information needed to calculate the thermodynamic properties of a system of independent molecules at thermal equilibrium such as the internal energy and entropy, and is therefore of critical importance.

2.2.3 Internal Energy and Entropy

The internal energy of the system can be derived from q. The total energy of the system is:

$$U = \sum_i n_i u_i \tag{2.13}$$

As the most probable configuration is so dominant, the Boltzmann expression for the population can be introduced:

$$U = \frac{N}{q} \sum_i u_i e^{-\beta u_i}. \tag{2.14}$$

This equation can be manipulated to a form that involves only q:

$$U = -\frac{N}{q}\frac{\mathrm{d}q}{\mathrm{d}\beta} \tag{2.15}$$

U is the internal energy relative to its value at $T = 0$, therefore

$$U - U(0) = -\frac{N}{q}\left(\frac{\partial q}{\partial \beta}\right)_V \tag{2.16}$$

Since the molecular partition function may depend on variables other than temperature (volume), the derivative with respect to β is a partial derivative with other variables held constant.

The partition function contains all thermodynamic informations and can therefore be used to calculate the entropy. The entropy is a measure of the disorder and is hence a growing function of the weight W : $S = f(W)$. If we consider two separate independent subsystems A and B, the weight of the ensemble will be $W = W_A W_B$, whereas the entropy will be $S = S_A + S_B$ or $f(W) = f(W_A) + f(W_B)$, these equations allow

$$S = k \ln W \tag{2.17}$$

as a solution, k being the Boltzmann constant. Using Equation 2.2 we can relate this relation, known as the Boltzmann formula for the entropy, to the partition function:

$$\begin{aligned} S &= k \sum_i (n_i \ln N - n_i \ln n_i) \\ &= -Nk \sum_i \frac{n_i}{N} \ln \frac{n_i}{N} \\ &= -Nk \sum_i p_i \ln p_i \end{aligned} \tag{2.18}$$

According to Equation 2.12, $\ln p_i = -\beta u_i - \ln q$, therefore:

$$S = -Nk \left(\sum_i \beta p_i u_i - \sum_i p_i \ln q \right) . \tag{2.19}$$

As the sum over p_i is unity, the sum over $N p_i u_i$ is equal to $U - U(0)$, and $\beta = 1/kT$, the final expression for the entropy is therefore:

$$S = \frac{(U - U(0))}{T} + Nk \ln q \tag{2.20}$$

From these derivations, of U and S, all remaining thermodynamic quantities may be derived. The partition function is important as it enables to all thermodynamic quantities to be determined. Nevertheless, its analytical determination is impossible for the sort of system we are studying, as it implies the knowledge of all possible energy levels. One has therefore to resort to sampling techniques, namely computer simulations, to determine these thermodynamic quantities.

2.3 Statistical Mechanics

Through computer simulations, one can generate data at the microscopic level (atomic positions, velocities ...), and these data can be converted into macroscopic information using statistical mechanics. Computer simulations are microscopic systems by their very nature; simulation systems are typically of the order of 10^3 to 10^5 particles in size and statistical mechanics is able to bridge the gap between the micro- and the macroscopic environments. This section will describe the essential part that statistical mechanics plays in computer simulation methodologies.

The validity of any computational model results from the successful reproduction and prediction of experimental observables, typically thermodynamic and structural properties. Statistical mechanics provides a route by which thermodynamic observables can be directly obtained, by calculating the averages over molecular properties. If one considers the instantaneous mechanical state of a chemical system consisting of N interacting particles to be characterised by their positions in space and their momenta (assuming a classical approximation), which is the set of $6N$ coordinates and momenta $(\boldsymbol{r}^N, \boldsymbol{p}^N)$, these instantaneous states are linked by a trajectory that can be followed using Newton's second law. These positions and momenta that define the different snapshots of the chemical system are described as locations in phase space.

This description requires the introduction of the concept of *ensemble*. An ensemble is a collection of replicas of the system. Even though each system is allowed to evolve independently from the others, they share constraints. The nature of these constraints determine the type of the ensemble. For example in the *canonical* ensemble or NVT, all the identical closed systems are regarded as being in thermal contact with each other, they can exchange energy with each other, they therefore have the same temperature T. Each replica is constrained to a constant volume V, number of particles N and temperature T. Each ensemble has its own partition function designed by Q.

To follow the progress of a particularly interesting system property, A, then the value of the property A at a point in phase space can be written as the function $A(\boldsymbol{p}^N, \boldsymbol{r}^N)$. As this system will evolve in real time, $A(\boldsymbol{p}^N, \boldsymbol{r}^N)$ will change due to the interactions between the particles. The value of A we measure experimentally can be considered as an average of $A(\boldsymbol{p}^N, \boldsymbol{r}^N)$ over time, and therefore referred to as a time average:

$$A_{obs} = \langle A \rangle_{time} = \lim_{t\to\infty} \frac{1}{t} \int_{t=0}^{t_{obs}} A(\boldsymbol{p}^N, \boldsymbol{r}^N) dt \qquad (2.21)$$

To calculate the average values of the properties of this system the simulation of the dynamic behaviour of the system is required. This can be achieved by describing an energy function that calculates the forces inherent within the system due to inter-atomic interactions, and from the force generated by this energy function the acceleration can be determined. Integration of Newton's laws of motion yields a trajectory that describes the positions, velocities and accelerations of the particles in the system over time, and this in turn will enable the calculation of system properties to be determined directly using equation 2.21. However, using a computer to calculate these macroscopic properties is fraught with difficulties, one major problem being that to simulate a macroscopic system is currently impossible, due to intrinsic computational constraints, and another is that the time integral cannot be calculated to infinity. Boltzmann and Gibbs recognised these problems and developed statistical mechanics to overcome them.

These problems are solved by considering a microscopic system in the order of 10^2-10^5 particles, where this single system evolving in time is replaced by a large number of replications that are considered simultaneously. The time average calculation in equation 2.21 can then be replaced by an ensemble average:

$$\langle A \rangle_{ensemble} = \int\int d\boldsymbol{p}^N d\boldsymbol{r}^N A(\boldsymbol{p}^N, \boldsymbol{r}^N) \rho(\boldsymbol{p}^N, \boldsymbol{r}^N) \qquad (2.22)$$

where $\rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$ is the probability density of the ensemble. This ergodic hypothesis is a fundamental axiom of statistical mechanics that assumes that the ensemble average is equal to the time average, allowing the transformation of equation 2.21 to 2.22. The angle brackets indicate that it is the average value of the property A over all the components of the ensemble, that is being calculated. This describes an ensemble average. Integrating over all possible configurations of the system determines the ensemble average of the property A. $\rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$ is the probability density of the ensemble, indicating the probability of finding a system in a specified state with momenta $\boldsymbol{p}^N$ and positions $\boldsymbol{r}^N$. In an equilibrium ensemble, $\rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$ is not time-dependent, that is at each time instant the particles of the system are distributed among the different energy states according to the probability density: when one leaves a particular state $\Gamma(\boldsymbol{p}_i, \boldsymbol{r}_i)$, another arrives to replace it. Such a system is called ergodic and this is the reason why the ensemble average can replace the time average. Such distribution of states is called a *Boltzmann distribution*. The probability density of each state, on which the Boltzmann distribution is based, is calculated differently depending on the constraints applied to the ensemble.

The probability density of a state i is the ratio between the number of particles in that state, n_i, and the total number of particles N in the system. In the canonical ensemble (NVT), that is an ensemble where the number of particles, the volume and the temperature of the system remain constant, this ratio is equal to:

$$\rho_{NVT}\left(\boldsymbol{p}^N, \boldsymbol{r}^N\right) = \frac{n_i}{N} = \frac{\exp(-E_i/k_B T)}{Q_{NVT}} \tag{2.23}$$

where E_i is the energy of state i, k_B the Boltzmann constant, T the temperature, and Q_{NVT} the so called canonical partition function. The classical partition function in the canonical ensemble is defined as:

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int\int d\boldsymbol{p}^N d\boldsymbol{r}^N \exp(E(\boldsymbol{p}^N, \boldsymbol{r}^N)/k_B T) \tag{2.24}$$

where $E(\boldsymbol{p}^N, \boldsymbol{r}^N)$ is E_i. The factor $N!$ arises from the indistinguishability of the particles and $1/h^{3N}$ is required to ensure that the partition function is equal to the quantum mechanical result for a particle in a box. The total energy of the system E is the sum of the potential energy, E^p and of the kinetic energy, E^k. The partition function can be expressed as a product of kinetic and potential parts : $Q = Q^p \cdot Q^k$. The kinetic contribution is known as the *ideal gas* part and can be solved analytically for the systems we are studying. The potential partition function is know as the *excess part* and cannot be solved analytically. The probability density $\rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$ can therefore not be accessed analytically. Through the choice of an adequate energy function that calculates the forces within the system due to inter atomic interactions, computer simulations can sample the phase space according to $\rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$. If the simulation time is sufficient the time average (Equation 2.21) and the ensemble average (Equation 2.22) are equal.

2.4 The Potential Energy Function for Atomic Models

Simulations performed at the atomistic scale are generally termed molecular mechanics simulations. From the positions and momenta of the atoms, a molecular mechanics force field is used to calculate the approximate energy and forces of a system. These forces are then used to advance the positions of the particles using Newton's second law (cf. 2.6.3). This force field only typically considers the nuclei positions; electron motion is ignored, and usually bond lengths, valence angles, torsions, and non-bonded interactions are included. Several force fields exist (AMBER,[29] OPLS,[30] and CHARMM[31] for example), using the same components and differing through parameterization or definition of these components. Force field can be united atom (UA) or all atom (AA). In united atom force fields, the hydrogen atoms bonded to carbon atoms are not modelled explicitly, rather the mass of the hydrogen atom is added to the carbon atom to form a "united atom". The non-bonded parameters associated with the carbon atom are also modified to take this change into account. The functional form of the total potential energy, U_{total}, in the AMBER force field is as follows:

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{non-bonded}. \tag{2.25}$$

The bond and angle contributions are described by harmonic potentials :

$$U_{bond} = \sum_{bonds} K_b(r - r_{eq})^2 \tag{2.26}$$

$$U_{angle} = \sum_{angles} K_\theta(\theta - \theta_{eq})^2 \tag{2.27}$$

where r corresponds to the bond length, θ to the valence angle, and r_{eq} and θ_{eq} to the associated equilibrium values. K_b and K_θ are force constants. The torsional term $U_{dihedral}$ is computed as:

$$U_{dihedral} = \sum_{dihedrals} A_n(1 + \cos(n\phi - \delta)), \tag{2.28}$$

where ϕ is the dihedral angle, n is the multiplicity (which gives the number of minimum points in the function as the torsion angle changes from 0 to 2π), δ is the phase angle and A_n is the force constant. Finally, the non-bonded energy is composed of an electrostatic and a Lennard-Jones term:

$$U_{non-bonded} = \sum_i \sum_{j>i} \left\{ \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \right\}, \tag{2.29}$$

where the sum is over all atom pairs i, j. The q_i are the partial atomic charges, ϵ_{ij} and σ_{ij} are the Lennard-Jones well-depth energy and collision-diameter parameters, and r_{ij} is the inter-atomic distance. The AMBER force field described here is just an illustration of a typical force field that can be used in a molecular simulation, our model use a different force field that will be described later in this thesis.

2.5 Reduced Units

For systems consisting of identical particles, it is sensible to use the mass of the molecule as a fundamental unit, i.e. set $m_i = 1$. As a consequence, the particle momenta and velocities become numerically identical, as do the forces and accelerations. For molecules interacting via a relatively simple pair potential this approach can be extended further. For example, as the Lennard-Jones potential is completely defined by the parameters ϵ and σ, then further fundamental units of energy, length, etc. may be defined. Units of other quantities follow directly.[27] Since the Gay-Berne potential is an anisotropic form of the Lennard-Jones potential, the reduced units used in this work are all defined in the same way.

$$\text{Density} \quad \rho^* = \rho\sigma^3 \tag{2.30}$$

$$\text{Temperature} \quad T^* = k_B T/\epsilon \tag{2.31}$$

$$\text{Energy} \quad E^* = E/\epsilon \tag{2.32}$$

$$\text{Force} \quad \mathrm{f}^* = \mathrm{f}\sigma/\epsilon \tag{2.33}$$

$$\text{Pressure} \quad P^* = P\sigma^3/\epsilon \tag{2.34}$$

$$\text{Time} \quad t^* = t\sqrt{(\epsilon/m\sigma^2)} \tag{2.35}$$

The use of reduced units avoids duplicate simulations being run by accident. There are also technical advantages in that when parameters such as ϵ and σ are given a value of unity, they will not appear in the program at all, saving time in the calculation of energies, forces, etc.

2.6 Molecular Dynamics or Monte Carlo Simulations

2.6.1 Stochastic Versus Deterministic

The way in which molecular positions are computed enables two simulation methods to be discriminated. In Molecular Dynamics the positions are obtained by numerically solving differential equations of motion, and positions are connected in time. Molecular Dynamics is a deterministic method. In *Monte-Carlo* positions are not temporally related; they are stochastically generated such that a molecular configuration only depends on the previous configuration. When the outcome of a random event depends only on the outcome of the immediately previous event, the sequence is called a *Markov chain*.

2.6.2 The Monte Carlo Method

The Monte Carlo method aims to generate a trajectory in phase space which samples from a chosen statistical ensemble depending only on the positions of the particles (and eventually their orientations). The Monte Carlo method thus samples from a $3N$-dimensional space for the positions of the particles. There is no momentum contribution, in contrast with Molecular Dynamics methods. A thermodynamic property

A can be evaluated through the integral:

$$< A >= \iiint_V \rho(\mathbf{r}^N) A(\mathbf{r}^N) d\tau \tag{2.36}$$

where $d\tau = d\mathbf{r}_1 \cdots d\mathbf{r}_N$ is the unit volume. The integral over momenta is not present in this equation as they are not evaluated in Monte Carlo methods and in any case the momenta contribution may be calculated analytically. In Monte Carlo simulations, ensemble averages such as Equation 2.36 are evaluated by accumulating the integrand at randomly generated values of the positions $\mathbf{r}^N$. Some configurations make large contributions to the integral due to the Boltzmann factor. The Metropolis method enables the sampling to be biased in favour of the configurations most likely to occur and hence to contribute to the integrand. This method involves the following steps. First positions are attributed to the N particles and the total potential energy $\mathcal{U}$ is computed. A new configuration is then generated by arbitrarily choosing a particle and moving its position by a randomly chosen distance and direction from $\mathbf{r}$ to $\mathbf{r}'$. The new total potential energy $\mathcal{U}'$ is then computed. If the new configuration's potential energy is lower than the old one, the move is accepted. If $\mathcal{U}' > \mathcal{U}$, the move is accepted with a probability proportional to the factor $exp[-\beta\Delta\mathcal{U}]$, where $\Delta\mathcal{U} = \mathcal{U}' - \mathcal{U}$. The procedure is repeated with either the old configuration, if the move has been rejected or the new configuration, if the move has been accepted. Under these conditions the value obtained for the thermodynamic property, A_{MC} obtained in the simulation is equal the the ensemble average of the property, $< A >_{ensemble}$, i.e. $A_{MC} =< A >_{ensemble}$ providing a sufficient number of moves has been performed.

2.6.3 The Molecular Dynamics Method

In molecular dynamics a simulation trajectory is generated by solving Newton's equation of motion, namely the second law,

$$\mathbf{f}_i = m_i \mathbf{a}_i, \tag{2.37}$$

where m_i and $\mathbf{a}_i$ are respectively the mass and acceleration of atom i and $\mathbf{f}_i$, the force acting on atom i, is defined as

$$\mathbf{f}_i = -\mathbf{grad} V_i, \tag{2.38}$$

where V_i, is the potential of atom i. These forces will allow the system to change by collective motions of particles, and the integration of the equations of motion is therefore necessary to follow this trajectory. Analytical integration of this trajectory is impossible due to the size of the system. Differential equations can be solved using the finite difference approach. Given the atomic positions, velocities and other dynamic information at time t, positions, velocities, etc. can be obtained at a later time $t + \delta t$ where δt is a small but finite time-step. The integration schemes commonly assume that the atomic positions, velocities, accelerations can be approximated as Taylor series expansions:

$$\begin{aligned}
\mathbf{r}(t+\delta t) &= \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2!}\delta t^2\mathbf{a}(t) + \frac{1}{3!}\delta t^3\mathbf{b}(t) + \dots \\
\mathbf{v}(t+\delta t) &= \mathbf{v}(t) + \delta t\mathbf{a}(t) + \frac{1}{2!}\delta t^2\mathbf{b}(t) + \dots \\
\mathbf{a}(t+\delta t) &= \mathbf{a}(t) + \delta t\mathbf{b}(t) + \dots
\end{aligned} \tag{2.39}$$

Where t is the time, $\mathbf{r}$, $\mathbf{v}$ and $\mathbf{a}$ are the position, velocity and acceleration of a particle. $\mathbf{b}$ is the third time derivative of $\mathbf{r}$ and is often neglected. From these equations, the Verlet leap-frog equations may be derived. It uses the following relationships to derive the dynamic trajectory of the system

$$\begin{aligned}
\mathbf{r}(t+\delta t) &= \mathbf{r}(t) + \delta t\mathbf{v}(t+\frac{\delta t}{2}) \\
\mathbf{v}(t+\frac{\delta t}{2}) &= \mathbf{v}(t-\frac{\delta t}{2}) + \delta t\mathbf{a}(t).
\end{aligned} \tag{2.40}$$

The velocities $\mathbf{v}(t+\frac{\delta t}{2})$ are initially calculated from the velocities at time $t-\frac{\delta t}{2}$, and using these velocities the new positions at time $t+\delta t$ can then be calculated. The velocity at time t may be calculated as the average of $\mathbf{v}(t-\frac{\delta t}{2})$ and $\mathbf{v}(t+\frac{\delta t}{2})$. The stored quantities are the current position $\mathbf{r}(t)$ and acceleration $\mathbf{a}(t)$ as well as the mid-step velocities $\mathbf{v}(t-\frac{\delta t}{2})$. The time step δt is chosen to be significantly smaller than the fastest frequency motion, e.g. that of the water O-H stretch is 3700 cm^{-1} or $11.09 \times 10^{15}\text{s}^{-1}$, so a time-step of 1 fs would be chosen. A too large time-step will result in the algorithm becoming unstable, causing eventual numerical overflow. A condition for molecular dynamics stability is that the force must be approximately constant in between time-steps.

2.7 The Application of Constraints in the Simulation

The molecular dynamics method described so far corresponds to the microcanonical ensemble or NVE ensemble, where the number of particles N, the volume V, and the total energy E are constant, i.e. obey Newton's law of conservation of energy and momenta. Although this ensemble is useful particularly in determining the integrity of the force field and of the integration algorithms, some other ensembles might also be used. In experimental conditions, the temperature and the pressure are constant, and therefore studying a system in the NPT ensemble or isothermal-isobaric ensemble could be of interest. To access such ensembles, algorithms exist that constraint the temperature or pressure to the desired values; they are called constraint algorithms.

2.7.1 Constant Temperature

The instantaneous value of the temperature is related to the kinetic energy via the particle's momenta as follows:

$$\mathcal{K} = \frac{1}{2}\sum_{i=1}^{N}\frac{|\mathbf{p}_i|^2}{m_i} = \frac{k_B T}{2}(3N - N_c), \qquad (2.41)$$

where N_c is the number of constraints and so $3N - N_c$ is the total number of degrees of freedom. This equation applies only in the case where the particles considered are spherical, and hence in the absence of rotational motion. The simplest way to control the temperature is to scale the velocities.[32] Alternative and more elaborate ways to control the temperature exist. Berendsen *et al.*[33] coupled the system to an external heat bath fixed at the correct temperature. The bath acts as a source of thermal energy, supplying or removing heat from the system as appropriate. Other methods such as stochastic collisions[34] and extended system[35,36] methods also exist.

2.7.2 Constant Pressure

Just as one might wish to maintain a constant temperature in a molecular dynamics simulation, a constant pressure may also be desired. Simulations performed in the isothermal-isobaric ensemble are more directly linked to experimental data as most experiments are carried out at constant temperature and pressure. A macroscopic

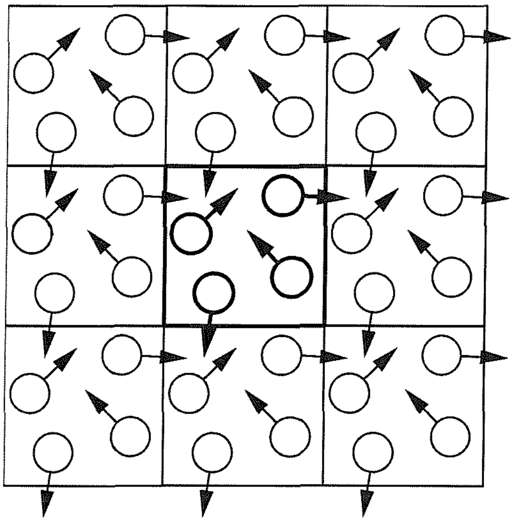


Figure 2.1: A two-dimensional periodic system.

system maintains constant pressure by adjusting its volume. Hence, the simplest way to run a constant pressure simulation is to scale the volume. Alternatively the system can be coupled to an external pressure bath as described by Berendsen *et al.*[33] Other methods, including pressure-coupling system methods[34] or extended system methods[35] can also be used.

2.8 Techniques Used in Molecular Dynamics

2.8.1 Periodic Boundary Conditions

Periodic boundary conditions allow surface effects to be avoided where particles on the boundary of the box experience different forces to those in the bulk, owing to the presence of vacuum. Under periodic boundary conditions the box is replicated throughout space to form an infinite lattice.[27] In the course of the simulation, as a molecule moves in the original box, its periodic image in each of the neighbouring boxes move in exactly the same way. Hence as a molecule leaves the box on one side, its image molecule will enter the box on the opposite side, as pictured in Figure 2.1. There are no boundaries and therefore no surface effects.

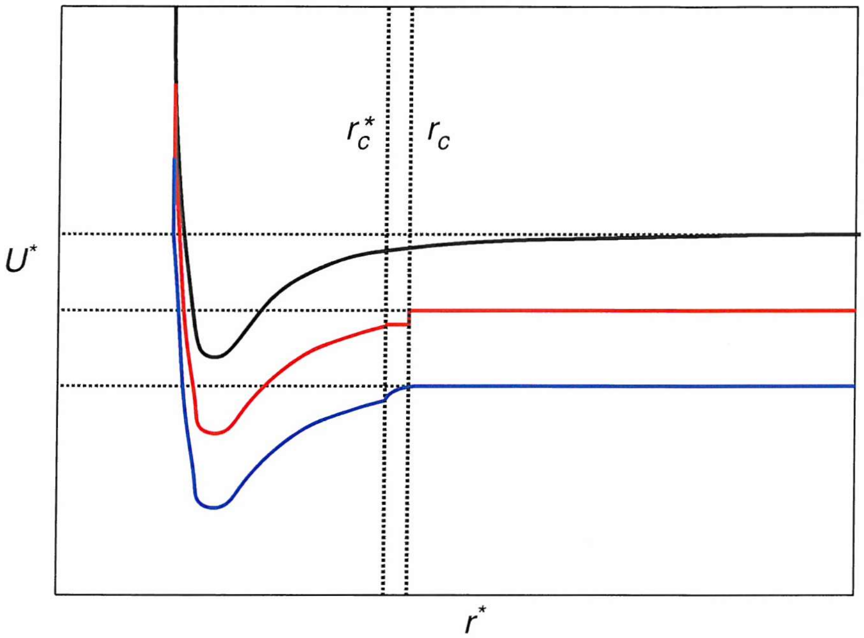


Figure 2.2: The influence of cutoff on the interaction pair potential. The untrucated potential is represented in black, the truncated potential in red and truncated smoothed potential in blue. For clarity the different potentials are shifted.

2.8.2 Truncated Potentials

In a system of N particles, $\frac{1}{2}N(N-1)$ pair interactions have to be calculated. Most interaction potentials become negligible after a certain distance, and thus considerable savings in computer time can be achieved by neglecting pair interactions beyond some distance r_C. Truncated and untrucated pair potentials are illustrated in Figure 2.2, respectively in red and black. The use of truncation can lead to a force discontinuity around r_C, leading to an instability in the integrator. To avoid such problem, the potential is often "smoothed" in the truncation region. Instead of reaching the value of zero at the cutoff value r_C, the potential is scaled in between r_C^* and r_C in order to take the value of zero at r_C. Before r_C^*, which is proportional to r_C (e.g. $r_C^* = 0.95r_C$), the potential is not modified. In between r_C^* and r_C the potential is multiplied by $(r - r_C)/(r_C^* - r_C)$ in the case of a linear smoothing, where r is the intermolecular distance. After r_C the potential is zero. Polynomial smoothing can also be used.[37]

2.8.3 The Ewald Summation

The development of our model required the inclusion of charges for the head group representation. These charges interact with each other through an electrostatic potential. While the Lennard-Jones potential falls off very rapidly with distance, due to the r^{-6} distance dependence of the dispersive interaction, the electrostatic potential has a r^{-1} distance dependence and therefore falls off slowly. In the presence of a non bonded cut-off, the long range nature of electrostatic forces can result in energy and force discontinuities at the boundary. The Ewald sum has therefore been adopted to treat long range coulombic interactions.

The Ewald sum is a technique to calculate the interaction between an ion and all its periodic images; it was originally devised for the study of ionic crystals.[38] The advantage of the Ewald sum is to treat long range interactions as two short ranged summations. In this method each charge interacts with all the charges within the simulation box and with all of their images in an infinite array of periodic cells. The position of each box (we assume here that each simulation cell is defined by three vectors **a**,**b** and **c** and contains N charges) can be related to the central box by the vector $(n_a\mathbf{a}, n_b\mathbf{b}, n_c\mathbf{c})$ where n_a, n_b and n_c are integers. The potential energy due to the interaction of all charges in the central box and all their images is given by equation 2.42.

$$\mathcal{V} = \frac{1}{2}\sum_{nbox=1}^{\infty}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\epsilon_0|\mathbf{r_{ij}} + \mathbf{r_{box}}|} \tag{2.42}$$

Where $\mathbf{r_{ij}}$ is the minimal distance between charges i and j, and $\mathbf{r_{box}}$ is the vector describing the image box. If this box is positioned at $\mathbf{n}=(n_a\mathbf{a}, n_b\mathbf{b}, n_c\mathbf{c})$ then:

$$\mathcal{V} = \frac{1}{2}\sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\epsilon_0|\mathbf{r_{ij}} + \mathbf{n}|} \tag{2.43}$$

Where $|\mathbf{n}|$ takes the values $1,\sqrt{2},\ldots$ The expression can then be rewritten as presented in equation (2.44).

$$\mathcal{V} = \frac{1}{2}\sum_{|\mathbf{n}|=0}^{\infty}{}' \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{4\pi\epsilon_0\,|\mathbf{r}_{ij} + \mathbf{n}|} \tag{2.44}$$

The prime on the first summation indicates that for $\mathbf{n}= 0$, the series does not include the pairs where i = j. The slowly decaying long range part of the Coulomb potential renders a straightforward summation of equation (2.44) impracticable. The equation

(2.44) has an extremely slow convergence, and moreover it is conditionally convergent, i.e. its convergence depends on the order in which its terms are considered. The solution is to split the problem in two series, each of which converges more rapidly, using the following relationship:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \tag{2.45}$$

The idea is to distribute the two parts of the coulomb energy using an adequate $f(r)$. There is a large choices of functions for f, however the traditional choice is to use the complementary error function $erfc$.

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} exp(-t^2)dt \tag{2.46}$$

The summation over $f(r)/r$ is performed in the real space whereas the summation over $(1 - f(r))/r$ is performed in the reciprocal space. The physical interpretation of this decomposition is that each ion is effectively neutralised (at long range) by the superposition of a spherical Gaussian charge density of opposite charge and centred on the ion. The corresponding Gaussian charge distribution $\rho_i(\mathbf{r})$ is presented in equation (2.47).

$$\rho_i(\mathbf{r}) = \frac{q_i \alpha^3}{\pi^{\frac{3}{2}}} exp\left(-\alpha^2 r^2\right) \tag{2.47}$$

Where α is a positive parameter that determines the width of the distribution. This charge distribution screens the interaction between neighbouring point-charges. The ensemble of point ions and Gaussian charges becomes the *Real Space* part of the Ewald sum, which is now short ranged. A second set of Gaussian charges with the same charges as the original ions is superimposed and centred on the ion (in order to cancel the effect of the first set of Gaussians). The potential due to these Gaussians is obtained from Poisson's equation and is solved as a Fourier series in *Reciprocal Space*. The complete Ewald sum requires additional correction terms. The total energy contains four contributions:

$$\mathcal{V} = \mathcal{V}^{(r)} + \mathcal{V}^{(k)} + \mathcal{V}^{(s)} + \mathcal{V}^{(d)} \tag{2.48}$$

- $\mathcal{V}^{(r)}$ is the contribution from the real space. The sum over point charges is now converted into a sum of the interactions between the charges and the neutralising distributions.

$$E^{(r)} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} {\sum_{|\mathbf{n}|=0}^{\infty}}' \frac{q_i q_j}{4\pi\epsilon_0} \frac{erfc(\alpha|\mathbf{r_{ij}} + \mathbf{n}|)}{|\mathbf{r_{ij}} + \mathbf{n}|} \tag{2.49}$$

This new summation converges much more rapidly with a rate of convergence dependent upon the width of the Gaussian distribution α. The wider the Gaussian, the faster the convergence.

- $\mathcal{V}^{(k)}$ is the contribution from reciprocal space and is added to cancel the neutralising distribution. This summation is performed in reciprocal space, meaning that the Fourier transforms of the cancelling distribution are added (one for each charge) and the total transformed back into real space. It corresponds to the interactions between the neutralising Gaussian density functions.

$$\mathcal{V}^{(k)} = \frac{1}{2} \frac{1}{\pi V_0} \frac{1}{4\pi\epsilon_0} \sum_{k \neq 0} \frac{4\pi^2}{k^2} exp\left(-\frac{k^2}{4\alpha^2}\right) \sum_{i=1}^{N} |q_i exp(-i\mathbf{k} \cdot \mathbf{r_i})|^2 \tag{2.50}$$

V_0 is the simulation cell volume, k (modulus of $\mathbf{k}$) theoretically range from $-\infty$ to $+\infty$ but practically a limit is fixed. $\mathbf{k}$ is a reciprocal lattice vector defined by :

$$\mathbf{k} = l\mathbf{u} + m\mathbf{v} + n\mathbf{w} \tag{2.51}$$

l,m,n are integers and $\mathbf{u}$,$\mathbf{v}$,$\mathbf{w}$ are the reciprocal space basis vectors. Both V_0 and $\mathbf{u}$,$\mathbf{v}$,$\mathbf{w}$ are derived from the vectors ($\mathbf{a}$,$\mathbf{b}$,$\mathbf{c}$) defining the simulation cell. Thus

$$V_0 = |\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}| \tag{2.52}$$

and

$$\begin{aligned} \mathbf{u} &= 2\pi \tfrac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} \\ \mathbf{v} &= 2\pi \tfrac{\mathbf{a} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} \\ \mathbf{w} &= 2\pi \tfrac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} \end{aligned} \tag{2.53}$$

With these definitions, the Ewald formula is applicable to general periodic systems. The $\mathcal{V}^{(k)}$ expression described in equation 2.50 is the Fourier series development of the potential. The Fourier series coefficients of the potential energy are derived from the charge density using the Poisson equation (2.54) as they can be derived directly. The charge density coefficients can be calculated using equation 2.47.

$$\nabla^2 V = -\frac{\rho}{\epsilon_0 \epsilon} \tag{2.54}$$

The reciprocal space sum converges much more rapidly than the original (2.44), but the number of terms that must be included (number of k vectors) increases with the width of the Gaussian. The real space sum converges more rapidly for a large value of α, whereas the reciprocal space sum requires a small value. A balance must therefore be found.

- $\mathcal{V}^{(s)}$ is the self-energy term corresponding to the subtraction of the interaction of each Gaussian with itself which is included in $\mathcal{V}^{(r)}$.

$$\mathcal{V}^{(s)} = -\frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\epsilon_0} \tag{2.55}$$

- $\mathcal{V}^{(d)}$ is the dipole correction term. It must be added if the surrounding medium is a vacuum (relative permittivity of 1). If the medium has an infinite relative permittivity then no correction is required.

$$\begin{aligned} &\mathcal{V}(\epsilon_s = 1) = \mathcal{V}(\epsilon_s = \infty) + \tfrac{2\pi}{3L_xL_yL_z} \left|\textstyle\sum_i z_i \mathbf{r_i}\right|^2 \\ &\mathcal{V}^{(d)} = \tfrac{2\pi}{3V_0} \left|\textstyle\sum_i z_i \mathbf{r_i}\right|^2 \end{aligned} \tag{2.56}$$

Then the final expression is :

$$
\begin{aligned}
\mathcal{V} &= \\
&\underbrace{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N\,\prime}\sum_{|\mathbf{n}|=\mathbf{0}}^{\infty}\frac{q_iq_j}{4\pi\epsilon_0}\frac{erfc(\alpha|\mathbf{r_{ij}}+\mathbf{n}|)}{|\mathbf{r_{ij}}+\mathbf{n}|}}_{\text{Real-space term }:\mathcal{V}^{(r)}} \\
&+\underbrace{\frac{1}{2}\frac{1}{\pi V_0}\frac{1}{4\pi\epsilon_0}\sum_{k\neq 0}\frac{4\pi^2}{k^2}exp\left(-\frac{k^2}{4\alpha^2}\right)\sum_{i=1}^{N}|q_iexp(-i\mathbf{k}\cdot\mathbf{r_i})|^2}_{\text{Reciprocal-space term }:\mathcal{V}^{(k)}} \\
&-\underbrace{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\alpha}{\sqrt{\pi}}\sum_{k=1}^{N}\frac{q_k^2}{4\pi\epsilon_0}}_{\text{Self term }:\mathcal{V}^{(s)}} \\
&+\underbrace{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{2\pi}{3V_0}\left|\sum_{k=1}^{N}\frac{q_k}{4\pi\epsilon_0}r_k\right|^2}_{\text{Dipole term }:\mathcal{V}^{(d)}}
\end{aligned}
\tag{2.57}
$$

The force on atom i is obtained by differentiation and is given by : $F_i = -\nabla_{r_i}\mathcal{V}$. There are three contributions to the overall force $F_i = F_i^{(r)} + F_i^{(k)} + F_i^{(d)}$.

$$
\begin{aligned}
F_i &= \\
&\underbrace{\frac{q_i}{4\pi\epsilon_0}\sum_{\mathbf{n}}{}'\sum_{\substack{j=1\\ j\neq i}}^{N} q_j\left\{\frac{erfc(\alpha\,|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|}+\frac{2\alpha}{\sqrt{\pi}}e^{-\alpha^2|\mathbf{r}_{ij}+\mathbf{n}|^2}\right\}\frac{\mathbf{r}\boldsymbol{ij}+\mathbf{n}}{|\mathbf{r}_{ij}+\mathbf{n}|^2}}_{\text{Real-space term }:F_i^{(r)}} \\
&+\underbrace{-\frac{2q_i}{\epsilon_0V_0}\sum_{\mathbf{k}\neq 0}i\mathbf{k}exp(\mathbf{k}\cdot\mathbf{r}_i)exp\left(-\frac{k^2}{4\alpha^2}\right)\sum_{j=1}^{N}q_jexp(-i\mathbf{k}\cdot\mathbf{r}_j)}_{\text{Reciprocal-space term }:F_i^{(k)}} \\
&-\underbrace{\frac{4\pi q_i}{3V_0}\sum_{j=1}^{N}\frac{q_j}{4\pi\epsilon_0}\mathbf{r}_j}_{\text{Surface Dipole term }:F_i^{(d)}}
\end{aligned}
\tag{2.58}
$$

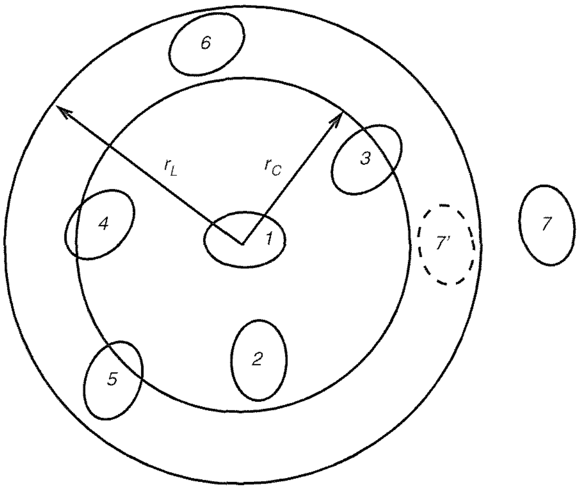


Figure 2.3: The cutoff sphere and its skin, around molecule 1. Molecules 2,3,4,5 and 6 are on the list of molecule 1; molecule 7 is not. Only molecules 2,3 and 4 are in the range of the potential at the time the list is constructed.

2.8.4 Neighbour Lists

The use of a simple non-bonded cut-off is not without problems. At each time step, the distances between particle i and all other particles in the simulation box has to be computed. Even if the potential and forces are not calculated for particles separated by a distance bigger than the potential cut-off, the distance calculation is still computationally expensive with the calculation time proportional to N^2. Verlet[39] suggested a way to avoid this problem by maintaining a list of the neighbours of a particular molecule, which is updated at intervals. Between updates only the distances between particle i and the ones in the list are checked. The number of pair separations explicitly considered is therefore reduced. In the original Verlet method, the potential cut-off sphere, of radius r_C, around a particular molecule is surrounded by a 'skin', to give a larger sphere of radius r_L as shown in Figure 2.3. At the first step in a simulation, a list is built of all the neighbours of each molecule, for which the pair separation is within r_L. These neighbours are stored in an array called LIST. At the same time an indexing array of size N called POINT is constructed. POINT(i) points to the position in the array LIST where the first neighbour of molecule i can be found. Neighbours of particle i therefore goes from POINT(i) to POINT(i+1)-1. Over the next time steps the LIST array is used to identify the neighbours of particle i. From time to time the LIST array is reconstructed and the cycle is repeated. The layer around r_C is chosen to be thick enough so that between list updates a molecule such as 7 cannot penetrate through the skin into the r_C sphere. Molecules 3 and 4 can move in or out of the sphere, since they are on the list of molecule 1, they are always considered regardless of their positions, until the list is next updated. The list can be updated at fixed intervals (say every 10-20 steps) or updated automatically by checking the particle displacements.

2.9 Problems Associated with Rigid Molecules

The Gay-Berne potential was, in our model, used for the lipid representation. As opposed to the Lennard-Jones potential, the Gay-Berne potential requires not only translational motion but rotational motion as well, due to the rigid nature of the particles involved. In the permeation study, solutes were also held rigid, therefore requiring algorithms for non-linear rigid body motion as well. Section 2.9.1 describes the use of the quaternion leap-frog algorithm which was used for the solute motion. Section 2.9.2 describes the linear-particles' algorithms that were used for the Gay-Berne particles' motion.

While for spherical particles motion definition is straightforward, the addition of rotational motion for rigid molecules makes the whole simulation process much more complicated. Equations of motion need to be written for the rotational contribution and these equations are more complex than their translational counterparts. Moreover some thermodynamic properties such as temperature require an additional rotational term.

In classical mechanics it is natural to divide molecular motion into translation of the centre of mass and rotation about the centre of mass.[40] The rotational motion is governed by the torque $\boldsymbol{\tau}_i$, about the centre of mass. The torque plays the same role as the force in the translation equations. If we consider a site i with forces f_{ia} acting on it, the torque is defined by:

$$\boldsymbol{\tau}_i = \sum_a (\mathbf{r}_{ia} - \mathbf{r}_i) \times \mathbf{f}_{ia} = \sum_a \mathbf{d}_{ia} \times \mathbf{f}_{ia}, \tag{2.59}$$

where $\mathbf{r}_i$ is the position of the centre of mass of the molecule, $\mathbf{r}_{ia}$ is an atom coordinate, or the coordinate at which the force is acting, and $\mathbf{d}_{ia}$ defines the positions of the atoms with respect to the centre of mass of the molecule.

2.9.1 Non-linear Molecules

In the permeation study described in Chapter 6 the solutes were treated as rigid entities. While the equation of motion of linear molecules can be simplified due to the use of a single moment of inertia (cylindrical symmetry), non-linear molecule motion is a less straightforward business. The orientation of a rigid body specifies the relation between an axis system fixed in space and one fixed with respect to the body, usually the "principal" body-fixed system in which the inertia tensor is diagonal. The unit vector $\mathbf{e}$ is denoted $\mathbf{e}^{\mathrm{b}}$ in the body-fixed frame and $\mathbf{e}^{\mathrm{s}}$ in the space-fixed frame. They are related by the rotation matrix $\mathbf{A}$

$$\mathbf{e}^{\mathrm{b}} = \mathbf{A} \cdot \mathbf{e}^{\mathrm{s}}. \tag{2.60}$$

The components of the rotation matrix are the direction cosines of the body-fixed axis vectors in the space-fixed frame, and they completely define the molecular orientation. Only three independent quantities taken to be the *Euler angles* ϕ, θ and ψ are used to define the matrix

$$\mathbf{A} = \begin{bmatrix} \cos\phi\cos\psi - \sin\phi\cos\theta\sin\psi & \sin\phi\cos\psi + \cos\phi\cos\theta\sin\psi & \sin\theta\sin\psi \\ -\cos\phi\sin\psi - \sin\phi\cos\theta\cos\psi & -\sin\phi\sin\psi + \cos\phi\cos\theta\cos\psi & \sin\theta\cos\psi \\ \sin\phi\sin\theta & -\cos\phi\sin\theta & \cos\theta \end{bmatrix} \tag{2.61}$$

in the suitable convention. The *Euler angles* are often used to describe the orientation of a molecule. ϕ is a rotation about the Cartesian z axis (moving x and y axes). θ is a rotation about the new x axis and ψ is a rotation about the new z axis. The matrix $\mathbf{A}$ can be used to derive equations of motion of the Euler angles. Unfortunately they suffer from a serious drawback, as the presence of the $\sin\theta$ terms means that a divergence occurs whenever θ approaches 0 or π. The angles ϕ and ψ become degenerate. An elegant solution to this problem was suggested by Evans,[41] who proposed the use of four quaternion parameters as generalised coordinates. A quaternion $\mathbf{Q}$ is a set of four scalar quantities

$$\mathbf{Q} = (q_0, q_1, q_2, q_3). \tag{2.62}$$

The quaternions satisfy the constraint

$$q_0{}^2 + q_1{}^2 + q_2{}^2 + q_3{}^2 = 1, \tag{2.63}$$

and are defined as

$$\begin{aligned} q_0 &= \cos\frac{\theta}{2}\cos\frac{\phi+\psi}{2} \\ q_1 &= \sin\frac{\theta}{2}\cos\frac{\phi-\psi}{2} \\ q_2 &= \sin\frac{\theta}{2}\sin\frac{\phi-\psi}{2} \\ q_3 &= \cos\frac{\theta}{2}\sin\frac{\phi+\psi}{2}. \end{aligned} \tag{2.64}$$

The rotation matrix then becomes

$$\mathbf{A} = \begin{bmatrix} {q_0}^2 + {q_1}^2 - {q_2}^2 - {q_3}^2 & 2\,(q_1q_2 + q_0q_3) & 2\,(q_1q_3 - q_0q_2) \\ 2\,(q_1q_2 - q_0q_3) & {q_0}^2 - {q_1}^2 + {q_2}^2 - {q_3}^2 & 2\,(q_2q_3 + q_0q_1) \\ 2\,(q_1q_3 + q_0q_2) & 2\,(q_2q_3 - q_0q_1) & {q_0}^2 - {q_1}^2 - {q_2}^2 + {q_3}^2 \end{bmatrix}. \tag{2.65}$$

The quaternions for each molecule satisfy the equations of motion

$$\begin{bmatrix} \dot{q}_0 \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{bmatrix} \begin{bmatrix} 0 \\ \omega_x^{\mathrm{b}} \\ \omega_y^{\mathrm{b}} \\ \omega_z^{\mathrm{b}} \end{bmatrix} \tag{2.66}$$

where $(0, \omega_x^{\mathrm{b}}, \omega_y^{\mathrm{b}}, \omega_z^{\mathrm{b}})$ is the angular velocity vector $\boldsymbol{\omega}$. A modified leap-frog formulation for quaternions has been proposed. The method is based on stored values of the angular momenta $\mathbf{I}^s(t - \frac{1}{2}\delta t)$, the quaternions $\mathbf{Q}(\mathbf{t})$ and the torques $\boldsymbol{\tau}^s(t)$ just computed from positions and orientations at time t. The first step is to bring all the angular momenta up to date

$$\mathbf{I}^s(t) = \mathbf{I}^s(t - \frac{1}{2}\delta t) + \frac{1}{2}\delta t\boldsymbol{\tau}^s(t). \tag{2.67}$$

in the space-fixed frame. These quantities are used to form the body-fixed angular velocity at time t, which in turn gives the time derivative of the quaternions $\dot{\mathbf{Q}}(t)$ through Equation 2.66. Then a guess at $\mathbf{Q}(t + \frac{1}{2}\delta t)$ is made

$$\mathbf{Q}(t + \frac{1}{2}\delta t) = \mathbf{Q}(t) + \frac{1}{2}\delta t\dot{\mathbf{Q}}(t). \tag{2.68}$$

Equations 2.67 and 2.68 are auxiliary equations used to estimate $\mathbf{Q}(t + \frac{1}{2}\delta t)$, so that transformations from space-fixed to body-fixed angular momentum, and the calculation of $\dot{\mathbf{Q}}$ can be implemented at the half-step time. The main algorithm equations are

$$\mathbf{l}^s(t + \frac{1}{2}\delta t) = \mathbf{l}^s(t - \frac{1}{2}\delta t) + \delta t \boldsymbol{\tau}^s(t) \tag{2.69}$$

and

$$\mathbf{Q}(t + \delta t) = \mathbf{Q}(t) + \delta t \dot{\mathbf{Q}}(t + \frac{1}{2}\delta t). \tag{2.70}$$

The constraint ${q_0}^2 + {q_1}^2 + {q_2}^2 + {q_3}^2 = 1$ should be preserved, it is common practice to 'renormalise' the quaternions , so as to guarantee that the sum of squares for each molecule is unity, at frequent intervals.

2.9.2 Linear Molecules

The motion of the Gay-Berne particles that were used in our lipid representation can be described by linear-molecule motion. A linear molecule can be defined by one moment of inertia, as one component vanishes and the two others are equal. The motion of such a molecule can be described by modifying the leap-frog technique used for translational motion. In the following section the appendix s refers to the space-fixed frame. The torque and the angular momentum are always perpendicular to the molecular axis (for a linear molecule). If $\mathbf{u}^s$ is the unit vector along the axis then the torque can be written:

$$\boldsymbol{\tau}^s = \mathbf{u}^s \times \mathbf{g}^s, \tag{2.71}$$

where $\mathbf{g}^s$ can be determined from the intermolecular forces (cf. Equation 2.59). The position of each atom in the molecule can be written

$$\mathbf{d}_a^s = d_a \mathbf{u}^s \tag{2.72}$$

where d_a defines the distance of the atoms from to the centre of mass of the molecule. The molecule being linear, all atoms are aligned along the molecular axis. The Gay-Berne potential is a single-site model, all forces and torques are resolved analytically. Nevertheless the inclusion of harmonic potentials to model bonds means that forces are acting not only at the centre of the particle but also on different points along the axis. $\mathbf{g}^s$ can then be written

$$\mathbf{g}^s = \sum_a d_a \mathbf{f}_a^s. \tag{2.73}$$

Only the perpendicular component $\mathbf{g}^{\perp}$ of $\mathbf{g}^s$ contributes to the torque

$$\boldsymbol{\tau}^s = \mathbf{u}^s \times \mathbf{g}^{\perp}, \tag{2.74}$$

where

$$\mathbf{g}^{\perp} = \mathbf{g}^s - (\mathbf{g}^s \cdot \mathbf{u}^s)\mathbf{u}^s. \tag{2.75}$$

The equation of rotational motion can now be written as the differential equation:[42]

$$\ddot{\mathbf{u}}^s = \mathbf{g}^{\perp}/I + \lambda \mathbf{u}^s \tag{2.76}$$

where I is the moment of inertia perpendicular to $\mathbf{u}^s$ and λ is a Lagrange multiplier which constrains $|\mathbf{u}|$ to be a constant of motion. A proposed solution to this algorithm uses the leap-frog algorithm.[43] Firstly an expression for λ is obtained by considering the advance of coordinates over half a time step

$$\dot{\mathbf{u}}^s(t) = \dot{\mathbf{u}}^s(t - \frac{1}{2}\delta t) + \frac{1}{2}\delta t \left[\mathbf{g}^{\perp}(t)/I + \lambda(t)\mathbf{u}^s(t)\right]. \tag{2.77}$$

Multiplying both sides of equation 2.77 by $\mathbf{u}^s(t)$, and considering that $\mathbf{u}^s(t).\dot{\mathbf{u}}^s(t) = 0$ and $\mathbf{u}^s(t).\mathbf{g}^{\perp}(t) = 0$ gives

$$\lambda(t)\delta t = -2\dot{\mathbf{u}}^s(t - \frac{1}{2}\delta t) \cdot \mathbf{u}_s(t) \tag{2.78}$$

and so

$$\delta t \ddot{\mathbf{u}}^s(t) = \delta t \mathbf{g}^{\perp}(t)/I - 2\left[\dot{\mathbf{u}}^s(t - \frac{1}{2}\delta t) \cdot \mathbf{u}^s(t)\right]\mathbf{u}^s(t). \tag{2.79}$$

This equation is then used to advance a full step in the integration algorithm

$$\dot{\mathbf{u}}^s(t + \frac{1}{2}\delta t) = \dot{\mathbf{u}}^s(t - \frac{1}{2}\delta t) + \delta t \ddot{\mathbf{u}}^s(t) \tag{2.80}$$

and the step is completed using

$$\mathbf{u}^s(t + \delta t) = \mathbf{u}^s(t) + \delta t \dot{\mathbf{u}}^s(t + \frac{1}{2}\delta t). \tag{2.81}$$

2.9.3 Temperature

While a spherical particle is only subject to translational motion, a rigid particle is also subject to rotational motion. Hence the kinetic energy is made up of two contributions a translational contribution $\mathcal{K}_{TRAN}$ and a rotational contribution $\mathcal{K}_{ROT}$

$$\mathcal{K}_{TRAN} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2, \tag{2.82}$$

and

$$\mathcal{K}_{ROT} = \frac{1}{2}\sum_{i=1}^{N} I_i w_i^2, \tag{2.83}$$

where m_i is the mass of particle i, v_i the velocity of particle i, I_i the moment of inertia of particle i and w_i the angular momentum of particle i among N particles. For linear particles there are $3N$ degrees of freedom for the translational motion (along the three axes) and $2N$ for the rotational motion (three rotations per particle minus the rotation along the long axis of the particle as this rotation leaves the particle unchanged) and therefore $5N$ for both motions. Each degree of freedom contributes to $kT/2$ to the kinetic energy, where k is the Boltzmann constant and T the temperature. The translational and rotational temperatures can be therefore linked to the associated kinetic energies:

$$\mathcal{K}_{TRAN} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{3N}{2} k T_{TRAN}, \tag{2.84}$$

and

$$\mathcal{K}_{ROT} = \frac{1}{2}\sum_{i=1}^{N} I_i w_i^2 = \frac{2N}{2} k T_{ROT}. \tag{2.85}$$

The translational and rotational temperatures can now be defined :

$$T_{TRAN} = \frac{1}{3kN}\sum_{i=1}^{N} m_i v_i^2, \tag{2.86}$$

and

$$T_{ROT} = \frac{1}{2kN}\sum_{i=1}^{N} I_i w_i^2. \tag{2.87}$$

The definition of the total kinetic energy $\mathcal{K}_{TOT}$ enables the overall temperature to be linked to the translational and rotational temperatures.

$$\mathcal{K}_{TOT} = \frac{5N}{2}kT = \frac{3N}{2}kT_{TRAN} + \frac{2N}{2}kT_{ROT} \tag{2.88}$$

$$T = \frac{1}{5}\left(3T_{TRAN} + 2T_{ROT}\right) \tag{2.89}$$

The temperature and kinetic energies for non-linear rigid particles can be defined in a similar fashion.

2.10 Summary

In this chapter, standard computer simulation techniques have been described. Special care was given to areas of methodology relevant to our work. The Ewald summation technique, which has been implemented for the treatment of electrostatic interactions, and rigid body motion, which was used for lipids, solvent and solute motion have therefore been described in more detail.

Chapter 3

Physical Background

In this chapter, the physical concepts related to liquid crystals and lipid bilayers are introduced.

3.1 Liquid Crystals

Matter is usually classified into three states, the solid, the liquid and the gaseous state.[24] In a gas, molecules are separated by large distances such that they do not see each other, and interact only weakly, apart from the occasional collision. In a liquid, the molecules are packed closer together and there is typically short range translational and rotational order between the molecules, persisting over a distance of up to a few molecular separations. Thermal motion disrupts this order over larger distances and so smooths out the average density. In a solid, the density is usually non-uniform over macroscopic distances; the molecules vibrate, on average about the sites of a crystal lattice.

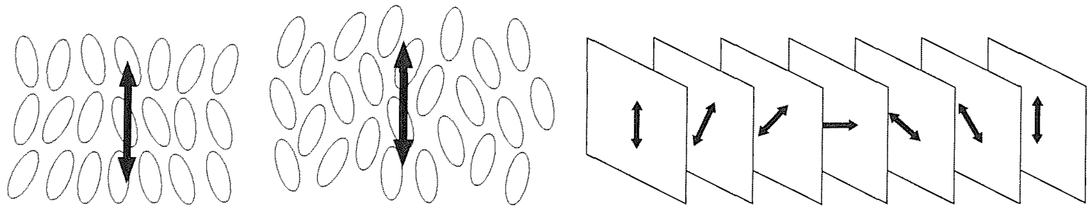
In addition to these three states, there are other intermediate ways that molecules can be assembled in the bulk. The liquid crystalline state of matter was first discovered by the Austrian botanist Reinitzer in 1888.[44] His work on the preparation of esters of cholesterol revealed unusual melting properties inconsistent with the three states of matter believed to exist. There was no concept of any possible overlap of any state of matter. Liquid crystals remained a scientific curiosity from their discovery in 1888, until renewed interest in their properties stimulated interest in the 1920s. In 1922 the term mesophase was advocated to describe correctly this new state of matter.[45] The original term liquid crystal had been a source of criticism and unpopularity in the scientific community, but both terms are still used today interchangeably. It was recognised in 1923 by Vorlander that the molecular shape of the molecule was an important factor determining whether a mesophase existed.[46] It had been observed that certain organic molecules did not show a single melting transition from the solid to a isotropic liquid, but instead exhibited a series of transitions through intervening mesophases. These molecules exhibit liquid crystalline phases by varying the temperature (thermotropic mesogens) and have highly anisotropic shapes, such as being long and thin, or disk-like. The intermediate mesophases were shown to flow like liquids, but also display the properties consistent with that of a solid, such as optical birefringence. The ability of a liquid crystal to flow suggests that there are only short range translational correlations between the positions of the molecules whereas the optical properties suggest that there are long range correlations between the orientations of the molecules.

3.1.1 Thermotropic Liquid Crystals

A mesophase is created when some of the long-range orientational and translational order characteristic of a solid crystalline lattice is retained after the crystal has appeared to melt, unlike in an isotropic liquid, where the constituent molecules are randomly aligned in all possible orientations and positions. A classification for the mesophases exhibited by elongated molecules was introduced by Friedel.[47] This scheme identified three different types of mesophases; the smectic, the nematic, and the chiral nematic mesophases as illustrated in Figure 3.1.

The Nematic Phase

The nematic phase (N) is characterised by an alignment of the molecules' molecular axes, but shows no long-range translational order. The orientational correlations



(a) Smectic A (b) Nematic (c) Chiral nematic

Figure 3.1: The different liquid crystalline phases.

typically extend over a correlation length ξ_0 of thousands of Å in an unperturbed system, whereas the translational correlation length ξ_T is typically of the order of only a few molecular separations, similar to an isotropic liquid in this respect. The direction in which the long axes of the molecules tend to align is known as the director and is represented by the unit vector $\mathbf{n}$. The nematic phase is uniaxial (there is rotational symmetry around the director axis).

The Chiral Nematic Phase

Optically active or chiral molecules can exhibit a twisted or chiral nematic phase (N^*). The molecules in this phase tend to align along a preferred direction as in the conventional nematic phase, but the director twists through the sample leading to a helical structure. The pitch of the helix is typically of the order of a few hundred nanometres, similar to the wavelength of visible light. As a consequence, samples of chiral nematics often appeared coloured, the colour depending on the pitch of the helix. Twisted nematic phases can also be formed when a chiral molecule, not necessarily a mesogen, is dissolved in a nematic phase.

The Smectic Phases

The smectic phases present long range orientational order, as in the nematic phase, but also possess long range correlations between the positions of the molecules in a single direction. The molecules, therefore, tend to arrange in equidistant layers, with only liquid-like translational order within each layer. Additional translational ordering of the molecules or tilting of the director within the layers leads to various smectic phases. The simplest of the smectic phases is the smectic A (S_A) in which the molecules tend to align parallel to the layer normal and exhibit no long range translational order within a layer. The smectic A phase can be viewed as an arrangement of liquid-like layers stacked on top of each other.

3.1.2 Lyotropic Liquid Crystals

Liquid crystalline phases can also be induced by varying the amount of solvent in a two or more component system; these mesophases are called lyotropic liquid crystals.[48,49] The classic example of a lyotropic system is that of an amphiphilic molecule, such as a soap, in water. The polar headgroup of the amphiphile prefers prefers a polar environment, the water in this case, whereas the alkyl chain prefers an organic environment. A single molecule in solution cannot satisfy both these conditions, but a group of molecules can. The aggregation of these molecules into a micelle is the driving force behind the formation of complex lyotropic liquid crystal phases.[48,49] At very low concentrations, the amphiphiles exist as single molecules, but, as the concentration is increased, various phase can be observed; these range from isotropic phases of spherical micelles to liquid crystalline phases consisting of, for example, infinite bimolecular layers or hexagonal columnar structures.[50] Lyotropic systems are important in systems as diverse as cell membranes to detergents. The phase diagram of DMPC is a perfect example of a lyotropic liquid crystal and is later illustrated in Figure 3.5.

3.1.3 Computer Simulation Studies of Liquid Crystals

It was not until the 1960s that research and development into liquid crystal mesophases grew considerably due to the discovery that biological membranes contain complex liquid crystalline structures, and the development of these materials in the new generation of optoelectronic display devices in the electronics and computing industries. The relationship between the intermolecular potential and the occurence of mesophases became the subject of numerous computer simulation experiments. Liquid crystals are complex molecules, they can have flexible or semi-flexible structures and complicated electronic charge distributions; they are therefore computationally expensive to model realistically. Moreover, liquid crystal phenomena occur over relatively large time and length scales, requiring, in some circumstances long simulations of large systems. Early attempts to simulate mesogens suffered from insufficient computing time, and prevented any definite conclusions being drawn. It was not until the 1980s that sufficient computer power was available to simulate mesogens unambiguously.

Lattice models were the first to be used significantly to simulate liquid crystals in the 1970s. Lebwohl and Lasher[51] developed a lattice version of the Maier-Saupe[52] model of a nematic liquid crystal and showed a first-order Nematic-Isotropic phase transition. This model consists of a cubic lattice of molecules represented by a headless vector, where each site only interacts with its nearest neighbours. The interaction energy can be written as $E_{ij} = -\epsilon P_2(cos\beta_{ij})$ where β_{ij} is the angle between the nearest neighbour's long axes and ϵ is an energy parameter. The interaction therefore only depends on the relative orientation of the molecules. As the interaction is simple, long runs of millions of Monte Carlo cycles on large systems are accessible, allowing the weak Nematic-Isotropic transition[53–55] to be investigated. Various systems have been studied using lattice models. However by their very nature, translational structures cannot be obtained.

The next historical step was the development of single site molecular models. Contrary to atomistic simulations, only the essential feature of the molecule is considered, reducing drastically the number of site-site interactions whilst retaining the fundamental information for bulk phase behaviour. Molecules can be modeled as simple rigid bodies interacting via a potential that depends on the relative orientations of the pair of molecules and on the intermolecular vector, in addition to their separation. The first attempt at creating a continuous single site potential was done by Corner,[56] his model being modified by Berne and Pechukas.[57] Both these models had flaws that were overcome by the model suggested by Gay and Berne.[1] These models will be described in depth in section 4.2.1. According to the parameterisation used, simulations performed with the Gay-Berne potential have been able to show the existence of nematic,[58,59] smectic A[59] and smectic B [60,61] phases. Other single site molecular models include hard potentials. Hard particles are idealised representations only taking shape into account. The molecules either overlap and their potential energy is infinite or they do not and their potential energy is zero. Frenkel et al.[62,63] studied the phase diagram of hard ellipsoids of revolution, and established the regions of nematic stability for ellipsoids of various elongations by Monte Carlo and free energy calculations. Only nematic liquid crystalline phases can be exhibited by hard ellipsoidal models. Phase diagrams have been studied for hard spherocylinders,[64] they can exhibit nematic and smectic liquid crystalline phases.[64,65]

The increase in computer power recently enabled the simulation of liquid crystal phases using atomistic potentials. McBride *et al.*[66] studied 4,4'-di-n-pentyl-bibicyclo-[2.2.2]octane with a united-atom approach where the central bicyclooctane group was maintained rigid. They were able to show the growth of a liquid crystalline phase. Other approaches use hybrid potentials. In their study, McBride and Wilson[67] used the Gay-Berne potential to model the rigid core of a liquid crystal molecule and an atomistic potential for the flexible hydrocarbon chains. This constituted a powerful approach to study the flexibility in mesogenic systems without the computational cost that would have been associated with a full atomistic simulation.

3.2 The Lipid Bilayer

A membrane is a complex molecular assembly containing lipids, membrane proteins and other molecules, that maintains the cell integrity. As a dynamic boundary the membrane is the location of many physiological processes such as cell signalling and transport. The experimental study of membranes is a difficult task and it is hence problematic to study dynamic properties at the atomic scale. Computer simulations can therefore be a crucial tool to predict properties not directly accessible through experiments. The basic structure of cellular membranes is the lipid bilayer, which also serves as a frame for membrane-associated proteins.

3.2.1 Phospholipid Properties

The occurrence of phospholipids as an essential membrane component is attributable to their ability to form bilayer vesicles spontaneously when dispersed in water. The bilayer self-organisation property results from the amphipathic character of membrane lipids, the hydrophobic region lying in the interior and hydrophilic region in the exterior of the bilayer. Most membrane lipids belong either to phospholipids, glycosphingolipids (GSLs) or cholesterol with a molar ratio in most cells of the order of 75%:5%:20%.[68] Membrane lipid diversity is increased by the variability within each of these three classes. As an example each phospholipid possesses one of some 10 different polar headgroups. The lipid composition of the cell membrane is related to the cell function, as a particular functionality corresponds to a unique lipid composition.[68] Phospholipids are divided in two main classes, the glycerophospholipids and the sphingophospholipids or glycosphingolipids. The glycerophospholipids are the predominant phospholipids in biological membranes. They are built on a glycerol frame, two fatty acid chains esterify two vicinal hydroxyl groups on the glycerol constituting the hydrophobic part of the lipid ; while a phosphate group esterifies the third hydroxyl group on the glycerol. At the same time, the phosphate group esterifies a choline, an ethalonamine or another polar group, thus bridging the hydrophilic and the hydrophobic region. The fatty acids are called lipid tails whereas the glycerol, the phosphate group together with the polar group bridged to the phosphate constitute the lipid headgroup. A dimyristoylphosphatidylcholine (DMPC) molecule is illustrated in Figure 3.2. One of the reasons why phospholipids are a key constituent of membranes is their ability to self-aggregate in water. The hydrophobic effect is the major driving force of the aggregation process. Only the headgroup of the phospholipid is able to form stable interactions with the aqueous environment; the tails disrupt the hydrogen bond network. The formation of an aggregate is entropy driven. According to their structure and to the environment; phospholipids are able to form various aggregates. One way of characterising a phospholipid is to examine the lateral stress profile of a monolayer.[69] The evolution of the pressure Π as a function of the depth z inside the monolayer can be split into three distinct zones: the aqueous phase, the headgroup and the tail regions. When two entities are attracted the pressure becomes negative whereas it becomes positive when they are repulsed. In Figure 3.3 the lateral stress profile of a type II phospholipid is illustrated.[69] In the headgroup region, repulsion is due to both strong electrostatic interactions and steric reasons, whereas it is mostly due to steric reasons in the tail region. The interface between the two regions is an attraction region. Although the lateral stress

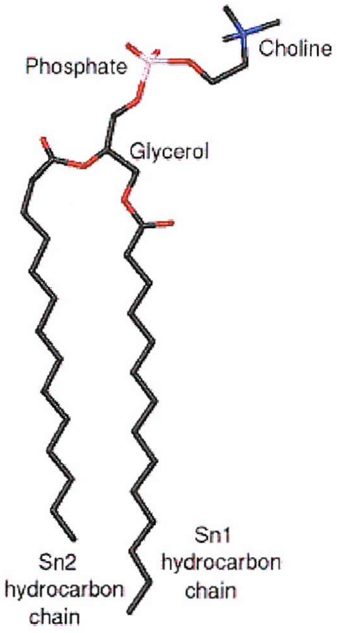


Figure 3.2: DMPC molecule (hydrogen atoms are omitted). Carbon atoms are in black, oxygen atoms are in red, phosphate atom is in pink and nitrogen atom in blue.

profile was initially a theory it has been measured experimentally and the theoretical profile[70–72] has been verified.[69]

A type II system will have a negative spontaneous curvature c_0[(a)] and will lead to an inverted micelle. A type I system will have a positive spontaneous curvature and is more likely to lead to a micelle. A bilayer is formed when the interactions in the tail and in the headgroup are balanced thus leading to spontaneous curvature close to zero. The three main aggregates' arrangements are described in Figure 3.4. For a type II phospolipid the interactions in the tail region are more repulsive than in the headgroup region. For a type I the repulsive interactions in the headgroup are higher than in the tail region. The lateral stress profile governs the spontaneous curvature of the assembly, and hence the shape of the aggregate. A phospholipid can be of Type II if for example it possesses two hydrocarbon chains and if the headgroup is relatively small, therefore leading to an increased pressure in the hydrocarbon domain. These arrangements are only some key phases; other phases can be accessed as indicated by the critical packing parameter N_S. The critical packing parameter is defined as

---

(a) The spontaneous curvature c_0 is the inverse of the curvature radius R_0 of the monolayer.

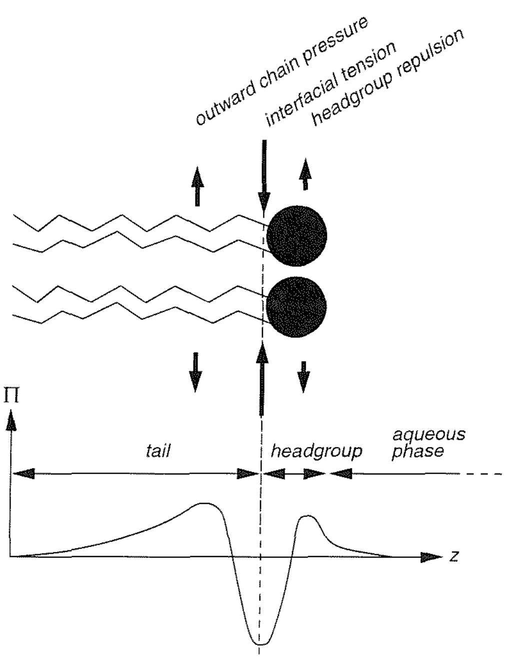
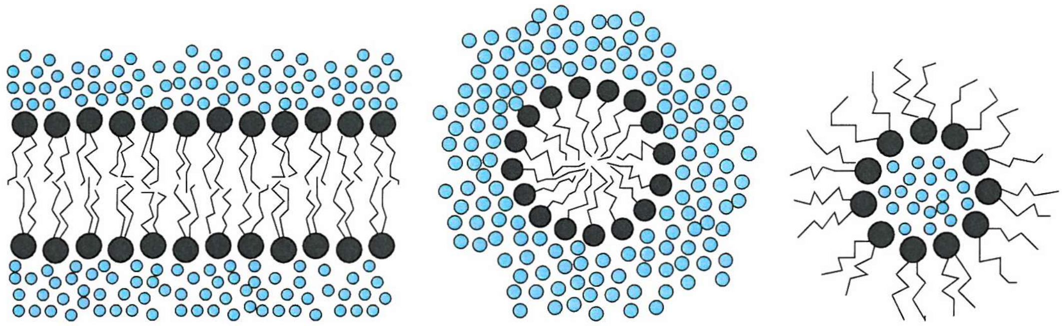


Figure 3.3: Lateral stress profile of a type II phospholipid.

$N_S = \frac{v}{l\,a_0}$ where v is the volume of the hydrophobic portion of the surfactant, l the length of the hydrocarbon chain and a_0 the area per head group.[73] For most two-chain phospholipids, bilayers are the favoured structure rather than micelles or inverted hexagonal phases.[73]

3.2.2 Bilayer Properties

In water phospholipids spontaneously form stable micelles or bilayers - depending on their structure. The bilayer can freeze or melt and exists in a range of states depending on temperature and level of hydration. When the temperature is higher than the phase transition temperature (from gel to liquid crystalline phase) the lipids are in a fluid state, they can rotate about their long axes and freely diffuse within the layer. Below the phase transition temperature they are in the gel state. The longer the hydrocarbon chains are and the lower the number of double bonds, the higher the phase transition temperature. A large number of phases are accessible to phosphatidylcholines. The phase diagram of DMPC is shown in Figure 3.5. In the $L_{\beta'}$ and $P_{\beta'}$ phases the hydrocarbon chains in the bilayer are ordered in a gel-state, and in the $P_{\beta'}$ case a rippling of the bilayer occurs. At temperatures in excess of



(a) bilayer $c_0 \sim 0$ (b) micelle $c_0 > 0$ (c) inverted cylinder $c_0 < 0$

Figure 3.4: States of aggregation of lipids in water.

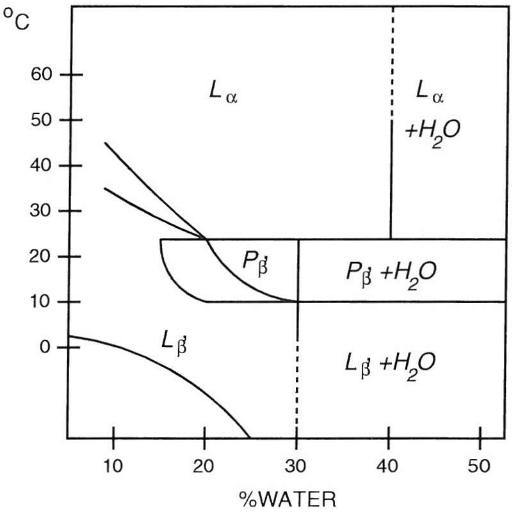
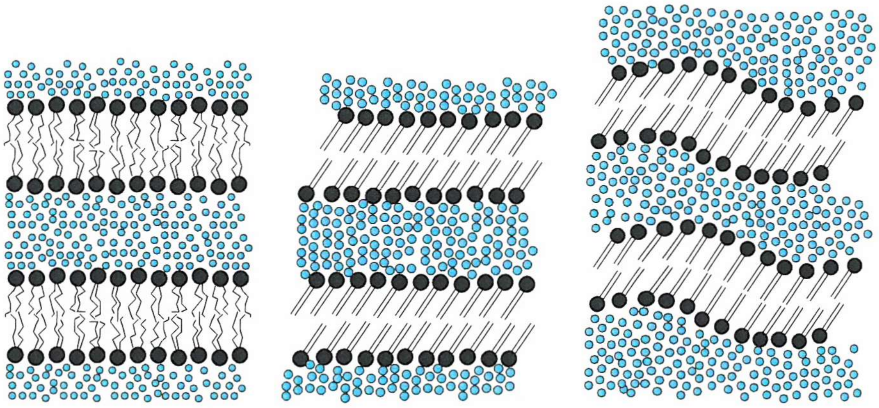


Figure 3.5: Temperature-composition phase diagram of hydrated DMPC.

24 °C (297 K) the L_α phase is adopted; this corresponds to the liquid crystalline state in which the hydrocarbon chains are molten and the lipids are free to diffuse in the plane of the bilayer. This is the phase present in physiological systems. The most common phosphoglycerides are dimyristoylphosphatidylcholine (DMPC) and dipalmitoylphosphatidylcholine (DPPC). They contain two hydrocarbon chains of 14 and 16 carbons respectively, and they share the same headgroup. The L_α phase is adopted in biological systems, and at temperatures higher than the phase transition temperature, which is 24°C for DMPC[68, 74] (see Figure 3.5) and 42°C for DPPC.[68, 75] When phosphatidylcholine dispersions are cooled for extended periods of time (hours



(a) L_α (b) $L_{\beta'}$ (c) $P_{\beta'}$

Figure 3.6: The different phase adopted by a bilayer

to days) at temperatures just above 0° C, a fourth lamellar phase is formed (not displayed in Figure 3.5), the L_c phase in which the chains are highly ordered in an orthorhombic lattice.[76] The L_α, L'_β and P'_β phase are illustrated in Figure 3.6.

3.2.3 Computer Simulations of Membranes

Since they allow access to properties that could not be derived by experiment, computer simulations of membranes are very popular, from the study of hydrated phospholipid bilayers on their own, to the diffusion of solutes through the membrane, and the inclusion of membrane bound proteins in bilayers. In 1988 Berendsen and Egberts[3] performed the first simulation of a realistic biological membrane. Over the last 15 years Berendsen *et al.* have performed numerous membrane simulations, using many different phospholipid species and system sizes.[5,7,12,23,77–83] The diffusion of solute molecules through lipid bilayer membranes have been studied by Stouch *et al.*[8,10] Essex *et al.*[4] performed one of the earliest DMPC simulation and extensively analysed it. Building on the simulations of Essex *et al.*, Robinson et al.[84,85] studied the influence of cholesterol on the bilayer structure. They showed that the inclusion of cholesterol increased the chain order. Berkowitz *et al.* [86–88] have produced several studies using a range of different phospholipid species. They paid particular attention to the structural effects at the lipid-water interface region and the understanding of hydration forces. These simulations were also among the first to incorporate the Ewald sum to treat long-range electrostatic interactions. The effect of surface tension has been a fierce subject of discussion over recent years.[89–93] Chiu et al.[89] used surface tension boundary conditions, a pressure of 1 atmosphere was applied normal to the plane of the bilayer while a negative pressure was applied parallel to the membrane plane to simulate surface tension effects. Tu et al.[92, 94–96] and Berger *et al.*[93] have argued that the most realistic boundary conditions is that of constant isotropic pressure with zero applied surface tension. Berendsen et al.[23] have performed simulations of hydrated DPPC bilayers with and without the inclusion of surface tension effects, and saw very little difference between both sets of results. Klein *et al.*[97] have presented a 1 ns simulation of a hydrated diphytanolphosphatidylcholine (DPhPC) lipid bilayer, comparing their results favourably with X-ray scattering and nuclear magnetic resonance spectroscopy results. The number of simulations of lipid bilayers has continuously grown since the first simulation by Berendsen *et al.*[3] in 1988. Simulation protocols are now well defined and results obtained can be directly linked to experimental data.

Chapter 4

Model Construction and Parameterization

4.1 Previous Reduced Representations

Computer simulations offer a powerful way of studying biological membranes at a molecular level, yielding detailed dynamic and thermodynamic data. However these simulations are computationally expensive, making long simulations or simulations of large assemblies difficult to perform. Coarse grain (CG) models are simplified models that still allow access to membrane behaviour but are far less computationally expensive. Although these models give less detailed information, they constitute an ideal compromise for the study of membranes on larger time or length scales.

4.1.1 Coarse-Grain Models for Phospholipids

Goetz *et al.*[98] developed a coarse grain model for the study of bilayer membranes. Their interaction potential is based upon three types of particles which can be hydrophilic or hydrophobic. The three types of particles are hydrophilic solvent particles (i.e. water), hydrophilic surfactant head group particles, and hydrophobic tail particles. The interaction between hydrophilic and hydrophobic particles is modeled by a repulsive soft core potential, while all other interactions are modeled using a 12-6 Lennard-Jones potential. Head group and tail particles are connected using harmonic bond potential to form surfactant molecules. In some of their simulations they added a bending potential along the two surfactant chains, thus creating a semi-flexible model, and for more complex surfactants a spontaneous or preferred tilt angle angle was added to the bending potential. Three surfactants were studied, two single chain molecules and one reminiscent of a lipid which contained two chains. They studied the assembly process of these particles. All three surfactants were capable of forming spherical micelles, cylindrical micelles, and bilayers. The assembly process took between 1 and 5×10^5 time steps, this time being greatly dependent upon the surfactant concentration. They also looked at stress and density profiles across the bilayer and were able to identify a tensionless bilayer.

Klein *et al.*[99–101] have also developed a coarse grain (CG) model for phospholipid simulations. Their model, which first provided a DMPC description, uses simplified representation for water, alkanes and phospholipids. Triplets of water molecules are represented by single spherical sites. Triplets of carbon atoms in the hydrocarbon chains and their accompanying hydrogen atoms are also represented by single spherical sites. These hydrocarbon sites are linked together to form chains using stretching and bending potentials. Non-bonded, non-electrostatic interactions are modelled using a Lennard-Jones 6-4 potential between water sites and a Lennard-Jones 9-6 potential between hydrocarbon sites. Lennard-Jones parameters were optimized to reproduce bulk water and both nonane and dodecane experimental densities. The model's water-hydrocarbon interaction parameters were selected to ensure phase separation and reasonable interface width. Single spherically symmetric sites were used to represent each of the choline ($(CH_3)_3N - CH_2 - CH_2$), phosphate ($O - PO_2 - O$), glycerol backbone ($CH_2 - CH - CH_2$) and ester groups ($O - CO - CH_2$). The choline and phosphate groups were assigned charges of $+e$ and $-e$ respectively, and a dielectric constant of 78 was used. "Tinfoil" Ewald periodic boundary conditions were used to treat the electrostatic interactions. They are no water-water or water-head group electrostatic interactions. Non-bonded, non-electrostatic interactions between water and hydrocarbon sites with head group sites was modelled using a Lennard-Jones 9-6 potential. Non-bonded, non-electrostatic interactions among the head group sites was modelled using tabulated potentials based on potentials of mean force derived from the atomistic radial distribution functions. Klein *et al.* carried out both Monte-Carlo and molecular dynamics studies. The MD simulations were conducted in the canonical ensemble using multiple time steps, with the shortest time step, 1 fs used for bond length and angle integration and the largest, 40 fs used for non bonded interactions in between 11 Å and the cutoff. Using a preassembled bilayer containing 50 lipids and 428 water sites (i.e. 1284 water molecules), their model was able to reproduce semiquantitatively the density profiles of an aqueous DMPC bilayer; the water sites do not penetrate into the bilayer as much as in the atomistic study, probably because the larger water sites cannot find sufficiently large spaces to be accommodated this far into the membrane. Klein *et al.* extended their study to a larger system containing 1024 lipids and 8768 water sites, and performed a 1 ns simulation with a 20 fs time step in the NPT ensemble. This study revealed a peak to peak distance of 32 Å, smaller than the all-atom and experimental value of 36 Å. They also studied the self-assembly of a phospholipid bilayer. Their results suggested that molecular dynamics was more efficient than Monte-Carlo for this purpose. An anæsthetic, halothane, was also introduced in the membrane; a single spherical site was used to model it. The system contained 50 lipids, 428 water sites and 25 halothane sites. They examined the halothane distribution along the bilayer normal and compared it to an all-atom simulation. After reparameterization the CG simulation exhibited a distribution in qualitatively good agreement with the all-atom simulation. The model developed by Klein *et al.* shows that it is possible to model realistically a phospholipid bilayer using a coarse grain approach.

Other applications include dissipative particle dynamics study of spontaneous vesicle formation of amphiphilic molecules by Yamamoto *et al.*[102] and study of amphiphilic bilayers as well as their stress profiles using dissipative particle dynamics by Lipowsky *et al.*[103]

4.1.2 Hybrid Models Using the Gay-Berne Potential

Apart from the model developed by Whitehead *et al.*[2] for the hydrocarbon region of a biomembrane (discussed in section 4.2.3), very few literature papers report the use of the Gay-Berne potential to represent biomembranes. A simplified membrane model has been developed by Ayton *et al.*[104, 105] Starting from the fact that atomistic molecular dynamics simulations are limited in time and space regimes due to the computational cost associated with them, and to the fact that continuum dynamics cannot explicitly model molecular-level interactions, they presented a new methodology for interfacing molecular dynamics simulations with continuum dynamics and applied it to a model lipid bilayer. They developed a feed-back mechanism in which microscopic-level molecular dynamics simulations are coupled to corresponding macro-scale continuum-level simulations. The techniques of non-equilibrium molecular dynamics were used to create the micro-to-macro interface, where transport coefficients that are required at the continuum level are calculated from detailed microscopic models. The reductionist model they used for the micro-level simulation cannot be considered as a stand-alone model but rather as a simple system to validate their technique. The membrane bilayer system was constructed with ellipsoids of revolution of length 20 Å and diameter 5 Å. A dipolar charge distribution embedded within a sphere of radius 5 Å was placed at one end of the Gay-Berne particle, making one end of the particle "polar" and one end non-polar. The charge distribution was constructed by placing a positive charge q_+=0.705 esu and a corresponding negative charge q_-=-0.705 esu at a distance of l=0.5 Å from the centre of the terminal sphere such that the resulting dipole $\mu=ql$ was perpendicular to the long axis of the ellipsoid. The solvent particles were modelled with a Lennard-Jones potential with a charge distribution identical to the membrane particles. The membrane-solvent interaction was calculated using the generalised Gay-Berne potential,[106] but without the attractive part of the potential to promote demixing. The electrostatic interactions were calculated using the Ewald sum. This model while being simple is not without drawbacks; any attempt to calculate order parameter will result in an artificially high-order system (high densities). Low order system will on the other hand lack the structural characteristics of a real membranes. Nevertheless, this model is interesting in the sense that it introduces the Gay-Berne potential as a tool to represent biomembranes.

To account for molecular flexibility in liquid crystals without having to use an atomistic potential, Wilson *et al.*[107] have developed a Gay-Berne/Lennard-Jones model. They describe molecular dynamics simulations of two Gay-Berne particles connected by an eight-site Lennard-Jones alkyl chain. Their model exhibits isotropic liquid, smectic-A and smectic-B phases. They were able to grow a smectic-A phase from an isotropic liquid and to identify odd-even effects in the alkyl chain, which could only be detected through atomistic simulations before.[66] Wilson *et al.*[67] have used a similar model, and performed molecular dynamics studies of a liquid crystal composed of a Gay-Berne particle core site with two alkyl chains of different length (C_7 and C_3) at either end of the molecule. They were able to grow a smectic-B phase from a pseudo-nematic phase.

Other additions to the Gay-Berne potential include the work of Zannoni *et al.*[108] who embedded dipoles at the end of a Gay-Berne particle to investigate the dimerization of particles and the dipole effect on the resulting phase.

4.2 Lipid Model

As highlighted by Whitehead *et al.*,[2] single site molecular potentials can be used as alternatives to atomistic potentials. Instead of considering each molecule as a sum of specific atoms, only the essential features of the molecule are considered. This option reduces drastically the number of site-site interactions to calculate and therefore the computational cost of the simulation, while keeping the fundamental interactions thought necessary to reproduce the bulk phase behaviour. In this section the construction of a simplified bilayer representation based on the popular single-site Gay-Berne potential is presented.

4.2.1 The Gay-Berne potential

Several single-site potentials have been suggested to represent aspherical molecules.[56,57,109] Corner[56] was the first to propose such a potential. His first idea was to calculate the second virial coefficients of molecules with aspherical geometries, and especially cylin-

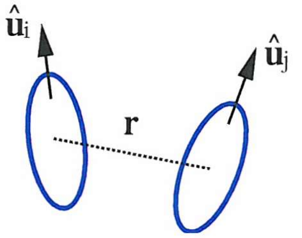


Figure 4.1: Schematic interaction between Gay-Berne particles. Gay-Berne particles with orientations $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_i$, linked by the intermolecular vector **r**

drically shaped molecules, to characterize their deviation from perfect gas behaviour. The calculation of the second virial coefficient requires the knowledge of the analytical form of the intermolecular potential. Corner assumed the molecular pair potential had the form

$$U_{corner}(\hat{u}_1, \hat{u}_2, \mathbf{r}) = 4\epsilon(\hat{u}_1, \hat{u}_2, \mathbf{r}) f\left(r/\sigma(\hat{u}_1, \hat{u}_2, \mathbf{r})\right) \tag{4.1}$$

where $\hat{u}_1$ and $\hat{u}_2$ define orientations of molecules one and two, and $\mathbf{r}$ is a unit vector defining the orientation of the intermolecular vector from molecule one to molecule two, as illustrated in Figure 4.1. To find the dependence of the well depth ϵ and of the range σ on $\hat{r}$, $\hat{u}_1$ and $\hat{u}_2$ Corner performed a numerical interpolation between the results obtained with a linear array of four Lennard-Jones centres per molecule (RLJ4) and his model, for a set of different relative orientations.

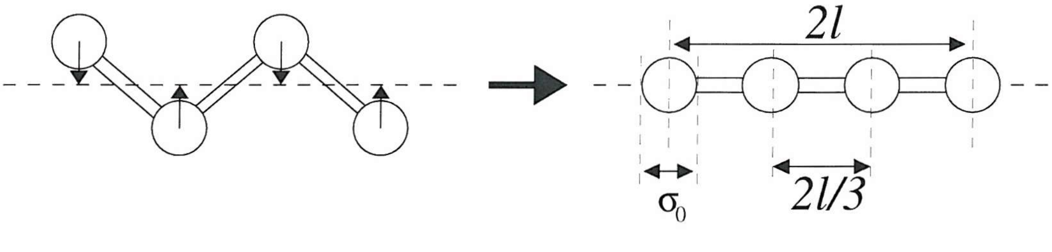


Figure 4.2: The RLJ4 model

To fit their model with an equivalent Lennard-Jones representation Corner used a linear array of four Lennard-Jones centres per molecule. This model is often referred to as RLJ4 model. It is obtained by projecting butane's atom positions along the long axis of the molecule (cf. Figure 4.2). The resulting model is axially symmetric, therefore enabling comparison with single-site models.

In the resulting fit, the shape anisotropy(a) is present in both the range and the strength functions. The virial coefficients show large errors for molecules with length to breadth ratio greater than 2.5, above this value the molecule is usually flexible in shape and the rigidity assumption is no longer appropriate.

The next step toward the Gay-Berne potential was made by Berne and Pechukas.[57] They give a new expression for the well depth and range of the potential by considering the overlap between two ellipsoidal Gaussian functions. The overlap was found to take the analytic form

$$S(\hat{u}_1, \hat{u}_2, \mathbf{r}) = \epsilon(\hat{u}_1, \hat{u}_2) exp\left(\frac{-r^2}{\sigma(\hat{u}_1, \hat{u}_2, \hat{r})}\right) \tag{4.2}$$

where $\hat{u}_1$ and $\hat{u}_2$ are unit vectors defining the orientations of the two molecules and $\hat{r}$ is a unit vector parallel to the intermolecular vector $\mathbf{r}$ ($\hat{r} = \mathbf{r}/r$). The explicit forms of the range and strength parameters are given by

$$\sigma(\hat{u}_1, \hat{u}_2, \hat{r}) = \sigma_\perp \left[1 - \frac{\chi}{2}\left(\frac{(\hat{u}_1.\hat{r} + \hat{u}_2.\hat{r})^2}{1 + \chi(\hat{u}_1.\hat{u}_2)} + \frac{(\hat{u}_1.\hat{r} - \hat{u}_2.\hat{r})^2}{1 - \chi(\hat{u}_1.\hat{u}_2)}\right)\right]^{-\frac{1}{2}} \tag{4.3}$$

and

$$\epsilon(\hat{u}_1, \hat{u}_2) = \epsilon_0\left[1 - \chi^2(\hat{u}_1.\hat{u}_2)^2\right]^{-\frac{1}{2}} \tag{4.4}$$

where the shape anisotropy parameter χ is defined by

$$\chi = \frac{(\sigma_\parallel/\sigma_\perp)^2 - 1}{(\sigma_\parallel/\sigma_\perp)^2 + 1}. \tag{4.5}$$

In this overlap model, the well depth $\epsilon(\hat{u}_1, \hat{u}_2)$ is independent of the intermolecular vector $\hat{r}$, which makes it unrealistic even if it brings more details than the model suggested by Corner. Moreover χ controls not only the shape anisotropy but also the anisotropic well depth potential for molecules with different orientations. Kushick and Berne[109] used the well depth $\epsilon(\hat{u}_1, \hat{u}_2)$ and the range of the potential $\sigma(\hat{u}_1, \hat{u}_2, \hat{r})$ defined by the overlap model in the Lennard-Jones 12-6 potential:

$$U_{BKP}(\hat{u}_1, \hat{u}_2, \mathbf{r}) = 4\epsilon(\hat{u}_1, \hat{u}_2)\left[\left(\frac{\sigma(\hat{u}_1, \hat{u}_2, \hat{r}}{r}\right)^{12} - \left(\frac{\sigma(\hat{u}_1, \hat{u}_2, \hat{r}}{r}\right)^{6}\right]. \tag{4.6}$$

(a) The shape anisotropy, which will be defined by the ratio $\sigma_\parallel/\sigma_\perp$ later in this chapter, can be described by the particle's length to width ratio

This model gives a continuous pair potential with ellipsoidal shape, but with flaws.[110] For example the strength parameter does not depend on the orientation of the intermolecular vector $\mathbf{r}$. All the problems associated with these potentials were overcome by Gay and Berne[1] who proposed a modified version of the overlap model, which is a shifted form of the Lennard-Jones 12-6 potential, given as:

$$U_{GB}(\hat{u}_1, \hat{u}_2, \mathbf{r}) = 4\epsilon(\hat{u}_1, \hat{u}_2, \hat{r}) \left[\left(\frac{\sigma_\perp}{r - \sigma(\hat{u}_1, \hat{u}_2, \hat{r}) + \sigma_\perp} \right)^{12} - \left(\frac{\sigma_\perp}{r - \sigma(\hat{u}_1, \hat{u}_2, \hat{r}) + \sigma_\perp} \right)^{6} \right]. \tag{4.7}$$

The well depth is now dependent on the molecule's orientations as well as on the intermolecular unit vector, and can be written as:

$$\epsilon(\hat{u}_1, \hat{u}_2, \hat{r}) = \epsilon_0 \epsilon^{\nu}(\hat{u}_1, \hat{u}_2) \epsilon'^{\mu}(\hat{u}_1, \hat{u}_2, \hat{r}) \tag{4.8}$$

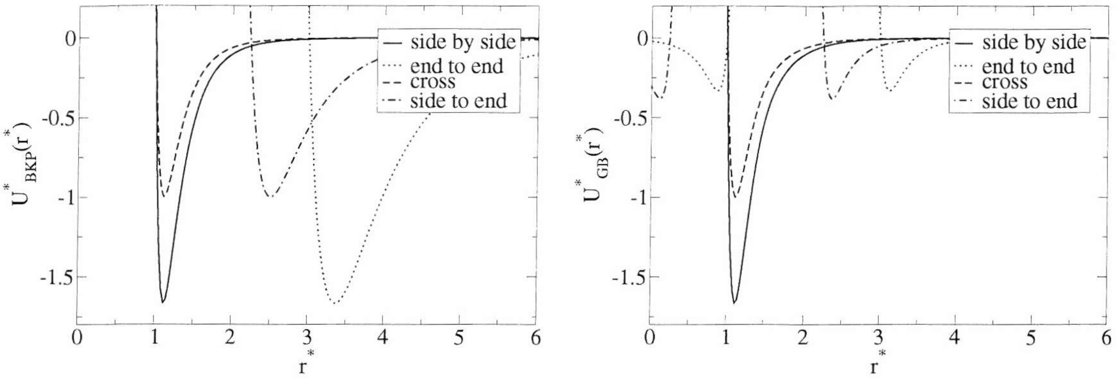
The original strength parameter $\epsilon^{\nu}(\hat{u}_1, \hat{u}_2)$, which depends only on the orientations of the molecules is identical to the one used in the overlap model without the ϵ_0 factor (cf. equation 4.4). The other term is defined as:

$$\epsilon'(\hat{u}_1, \hat{u}_2, \hat{r}) = 1 - \frac{\chi'}{2} \left(\frac{(\hat{u}_1.\hat{r} + \hat{u}_2.\hat{r})^2}{1 + \chi'(\hat{u}_1.\hat{u}_2)} + \frac{(\hat{u}_1.\hat{r} - \hat{u}_2.\hat{r})^2}{1 - \chi'(\hat{u}_1.\hat{u}_2)} \right), \tag{4.9}$$

where the anisotropy χ' depends on the ratio of the potential well depth for the side by side and end to end arrangements $(\epsilon_\perp/\epsilon_\parallel)$.

$$\chi' = \frac{(\epsilon_\perp/\epsilon_\parallel)^{1/\mu} - 1}{(\epsilon_\perp/\epsilon_\parallel)^{1/\mu} + 1} \tag{4.10}$$

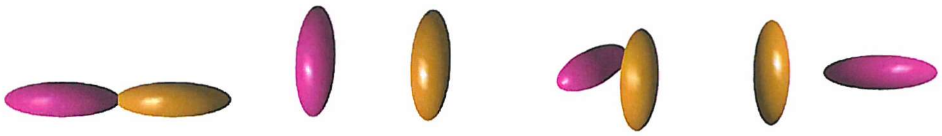
The other modification brought by Gay and Berne was to shift the potential along the r-axis by a distance of $\sigma(\hat{u}_1, \hat{u}_2, \hat{r}) - \sigma_\perp$ rather than to scale the r dependence by $\sigma(\hat{u}_1, \hat{u}_2, \hat{r})$, ensuring the potential well widths are independent of the relative orientations of the molecules. The effects of these modifications are shown in Figure 4.3 and 4.5. An approach to describing the Gay-Berne potential is to look at the distance dependence of the potential energy for particular orientations as illustrated in Figure 4.3 for the Berne-Kushick-Pechukas and the Gay-Berne potential. The four



(a) (b)

Figure 4.3: The distance dependence of (a) the Berne-Kushick-Pechukas and (b) the Gay-Berne intermolecular potentials for molecules of length-to-breadth ratio 3:1. The well depth anisotropy ratio in the Gay-Berne potential is 5:1 and the exponents μ and ν have the values 2 and 1, respectively.

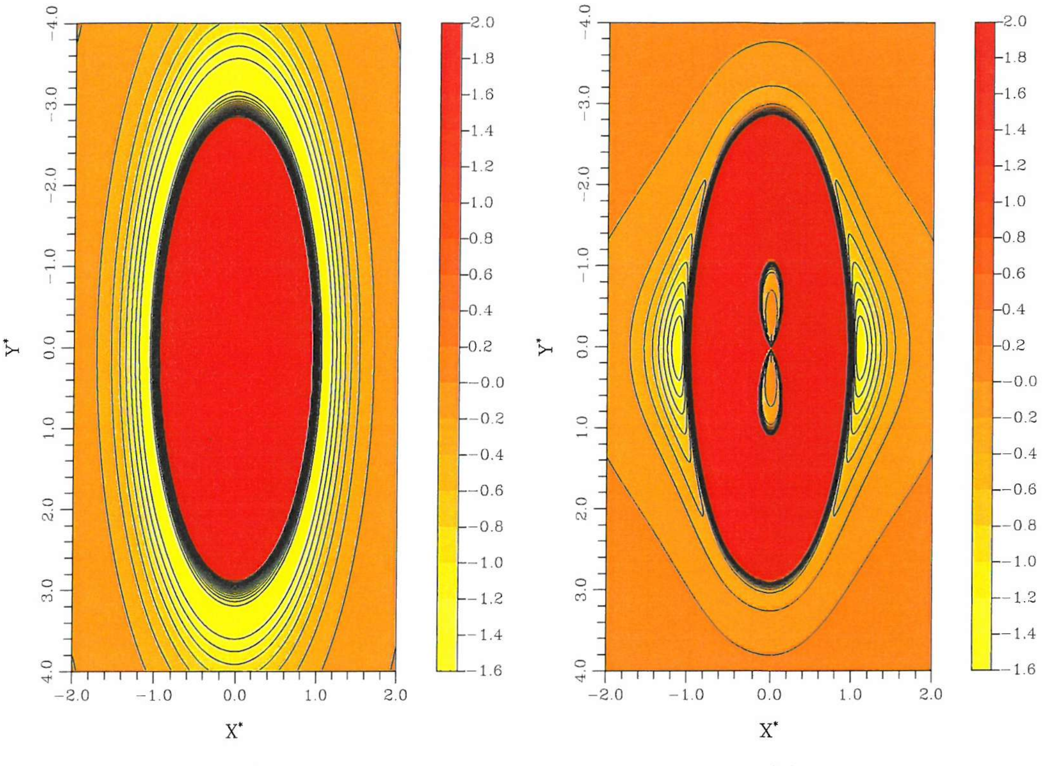
characteristic arrangements taken by two Gay-Berne particles are illustrated in Figure 4.4.



(a) **e** (b) **s** (c) **x** (d) **t**

Figure 4.4: The four characteristic arrangements taken by two Gay-Berne particles. (a) end to end conformation **e**: $\hat{u}_i \cdot \hat{u}_j = \hat{u}_i \cdot \hat{r} = \hat{u}_j \cdot \hat{r} = 1$; (b) side by side conformation **s**: $\hat{u}_i \cdot \hat{u}_j = 1$ and $\hat{u}_i \cdot \hat{r} = \hat{u}_j \cdot \hat{r} = 0$; (c) cross conformation **x**: $\hat{u}_i \cdot \hat{u}_j = \hat{u}_i \cdot \hat{r} = \hat{u}_j \cdot \hat{r} = 0$; (d) side to end conformation **t**: $\hat{u}_i \cdot \hat{u}_j = \hat{u}_i \cdot \hat{r} = 0$ and $\hat{u}_j \cdot \hat{r} = 1$ or $\hat{u}_i \cdot \hat{u}_j = \hat{u}_j \cdot \hat{r} = 0$ and $\hat{u}_i \cdot \hat{r} = 1$.

For each arrangement a well depth and a contact distance can be calculated, the well depth corresponds to the lowest value of the potential whereas the contact distance is the distance at which the potential is zero. For the Gay-Berne potential, the contact distance changes from a small value when the particles are side-by-side to a large value when they are end-to-end. Furthermore the well depth changes from a large value for side-by-side particles to a small value for end-to-end particles. Similarly the well depth is weak for the **t** conformation but becomes large for the **x** conformation. An interesting feature is the presence of a hole in the middle of the potential for the end-to-end and side-to-end conformations. This hole is due to the



(a)

(b)

Figure 4.5: The potential energy contours for a pair of parallel molecules of length to breadth ratio 3:1 interacting via (a) the Berne-Kushick-Pechukas and (b)the Gay-Berne intermolecular potentials. The contour are shown for values of the potential between -1.6 and +2.0, in steps of 0.2. The parameterization of the Gay-Berne potential is the same as in Figure 4.3. One particle is positioned at coordinates $(0, 0)$, the other at coordinates (X^*, Y^*), the axis Y^* is parallel to both particles' orientation vectors.

shifted nature of the Gay-Berne potential. When $\sigma(\hat{u}_i, \hat{u}_j, \hat{r}) < 2$ the potential has an "interior" well for $r < \sigma(\hat{u}_i, \hat{u}_j, \hat{r}) - 2$, but this does not pose a problem for molecular dynamics, since the inner and outer wells are separated by an infinite barrier.[1] This could cause problems if the Monte-Carlo method was used.

4.2.2 Simple Parameterization of the Gay-Berne Potential

The anisotropic shape of the Gay-Berne potential makes it dependent upon the orientation of the particles. To measure this dependence it is usual to look at certain conformations, namely, the side-by-side conformation (**s**), the end-to-end conformation (**e**), the T conformation or side-to-end (**t**) and the cross conformation (**x**), which are defined according to the particles orientation ($\hat{u}_i$ and $\hat{u}_j$) with respect to the intermolecular vector ($\hat{r}$). They are illustrated in Figure 4.4. In the cross arrangement (**x**) the two molecules are perpendicular to each other as well as to the intermolecular vector in the side-by-side (**s**) arrangement the molecules are parallel to each other but perpendicular to the intermolecular vector , in the end-to-end (**e**) arrangement the two molecules and the intermolecular vector are all aligned and in the T (**t**) arrangement the two molecules are perpendicular to each other, one of them being parallel to the intermolecular vector and the other perpendicular.

The well depth $\epsilon(\hat{u}_i, \hat{u}_j, \hat{r})$ and the intermolecular separation $\sigma(\hat{u}_i, \hat{u}_j, \hat{r})$, change dramatically with the orientations of the particles and the intermolecular vector. The value of σ and ϵ for particular conformations have been calculated by Luckhurst *et al.*[111] with the Gay-Berne potential and are displayed in Table 4.1.

Table 4.1: The well depth $\epsilon(\hat{u}_i, \hat{u}_j, \hat{r})$ and the intermolecular separation $\sigma(\hat{u}_i, \hat{u}_j, \hat{r})$ for particular orientational configurations.

Configuration	$\sigma(\hat{u}_i, \hat{u}_j, \hat{r})$	$\epsilon(\hat{u}_i, \hat{u}_j, \hat{r})$
e	$\sigma_0(\sigma_\parallel/\sigma_\perp)\ (= \sigma_\parallel)$	$\epsilon_0(\epsilon_\parallel/\epsilon_{perp})(1-\chi^2)^{\nu/2}$
s	$\sigma_0\ (= \sigma_\perp)$	$\epsilon_0(1-\chi^2)^{\nu/2}$
x	$\sigma_0\ (= \sigma_\perp)$	ϵ_0
t	$\sigma_0\sqrt{((\sigma_\parallel/\sigma_\perp)^2+1)/2}$	$\epsilon_0[2/[(\epsilon_\perp/\epsilon_\parallel)^{1/\mu}+1]]^\mu$

The parameterization of our model was mainly based on the Gay and Berne parameters that were described in their original paper.[1] Since in our model each tail particle mimics four CH_2 groups, and since Gay and Berne model was developed to fit a RLJ4 model, the use of their parameters seemed like a reasonable decision. The parameters they used were μ=2, ν=1, $\sigma_\parallel/\sigma_\perp$=3, $\epsilon_\perp/\epsilon_\parallel$=5. The parameter $\sigma_\parallel/\sigma_\perp$ defines the shape anisotropy of the repulsive core of the Gay-Berne potential. When $\sigma_\parallel/\sigma_\perp$=1 the core is spherical in shape and it becomes ellipsoidal when $\sigma_\parallel/\sigma_\perp$ is different from unity. Prolate cores correspond to $\sigma_\parallel/\sigma_\perp$ >1, and oblate cores to $\sigma_\parallel/\sigma_\perp$ <1. Similarly, ϵ_0 scales the potential energy term (it defines the unit of energy of the Gay-Berne potential), and $\epsilon_\perp/\epsilon_\parallel$ defines the well depth anisotropy for molecules with fixed relative orientations, but with varying orientations with respect to $\hat{r}$. Gay and Berne determined their parameters through a fit with the Lennard-Jones representation of the RLJ4 molecule.

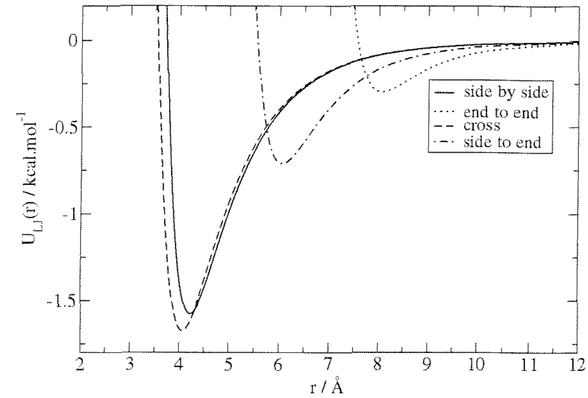
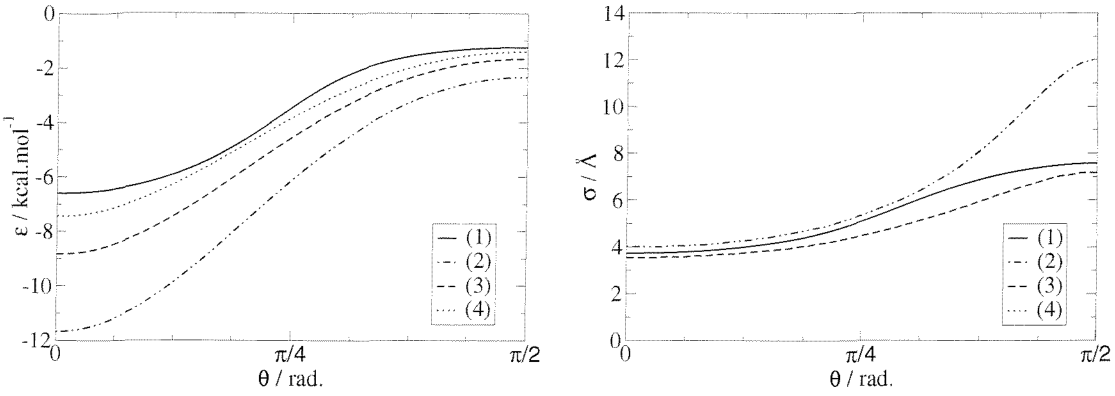


Figure 4.6: The distance dependence of the Lennard-Jones 12-6 intermolecular potential for the four representative arrangements of a butane molecule (RLJ4 model) : side by side (**s**), end to end (**e**), cross (**x**) and side to end (**t**) arrangements. Calculations were performed using the OPLS force field for united atoms : σ_{CH}=3.905 Å and $\epsilon_{CH} = 0.118\,\text{kcal.mol}^{-1}$.

These parameters can be evaluated directly by comparison with a site model possessing cylindrical symmetry. The parameters that are going to feed the Gay-Berne potential can be deduced from the interactions calculated with the Lennard-Jones potential between the molecules that one wishes to mimic, butane in our case. As will be described later in this chapter, each tail particle corresponds to four CH_2 groups. Since the majority of the Gay-Berne particles are located in the hydrocarbon region (six particles among ten), it was logical to base our parameterization on the RLJ4 model. To do so we calculated the Lennard-Jones interaction between two molecules of butane as a function of distance for different relative orientations. We adopted the OPLS united atom force field parameters for σ_{LJ} and ϵ_{LJ}, and used the geometry of the RLJ4 model described in Figure 4.2. The carbon atoms were aligned to remove the biaxiality of butane, and the distance between two carbon atoms was fixed to 1.3 Å, to take into account the projection onto the molecular axis (the optimal distance between two hydrocarbon carbon atoms is 1.526 Å). In Figure 4.6 the intermolecular potential for the different principal arrangements of two butane molecules are displayed.

Table 4.2: Well-depths and contact-distances for the different arrangements of butane calculated with the Lennard-Jones potential.

	x	s	e	t
Well depth (kJ.mol^{-1})	-7.0050	-6.5958	-1.2414	-2.9726
Contact distance (Å)	3.5450	3.7250	7.5650	5.5550



(a) (b)

Figure 4.7: Well depth (a) and contact distance (b) calculated as a function of the angle with the intermolecular vector when the two molecules are parallel ($\hat{u}_i.\hat{u}_j$=1) for (1) Lennard-Jones, (2) GB(3.0,5.0,2,1) with ϵ_0=7.005 kJ.mol^{-1} and σ_0=4Å, (3) GB(2.03,5.31,2,1) with ϵ_0=7.005 kJ.mol^{-1} and σ_0=3.545Å, and (4) GB(2.03,5.31,2,0.26) with ϵ_0=7.005 kJ.mol^{-1} and σ_0=3.545Å. In the contact distance plot (b) curves (3) and (4) are equivalent as the parameter ν has no influence on the contact distance.

The well depth of the cross configuration corresponds to ϵ_0 (cf. Table 4.1), the shape anisotropy is taken as the ratio of the contact distances in the end-to-end and side-by-side arrangements $\sigma_{\|}/\sigma_{\perp}= \sigma_E/\sigma_S$; the well depth anisotropy is derived the same way $\epsilon_{\perp}/\epsilon_{\|}=\epsilon_S/\epsilon_E$. The value of σ_0 is taken as the contact distance in the cross arrangement ($\sigma_S \approx \sigma_X = \sigma_0$). The values for ϵ_0, σ_0, $\sigma_{\|}/\sigma_{\perp}$, and $\epsilon_{\perp}/\epsilon_{\|}$ are therefore respectively : 7.005 kJ.mol^{-1}, 3.545 Å, 2.03, and 5.31. To describe these parameters in a more concise fashion, the notation GB($\sigma_{\|}/\sigma_{\perp}$,$\epsilon_{\perp}/\epsilon_{\|}$,$\mu$,$\nu$) is often used. These parameters would therefore be described by GB(2.03,5.31,μ,ν).

The well depth and the contact distance of the potential as a function of the angle with the intermolecular vector, when the two molecules are parallel, are displayed in Figure 4.7, for the Lennard-Jones potential (1) and the Gay-Berne potential with the Gay-Berne parameters[1] (2) as well as the parameters evaluated through comparison with the RLJ4 model (3). The use of the parameters obtained through a simple comparison with an RLJ4 model, allows better fits for the well depth and especially the contact distance. Nevertheless this fitting appears unsatisfactory; the parameters do not reproduce correctly the Lennard-Jones behaviour. The next step toward the complete parameterization concerns the parameter ν. It must be chosen such that

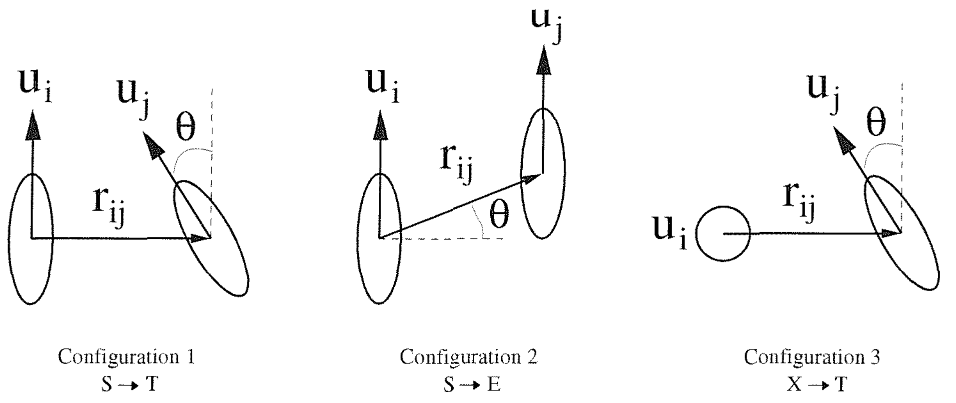


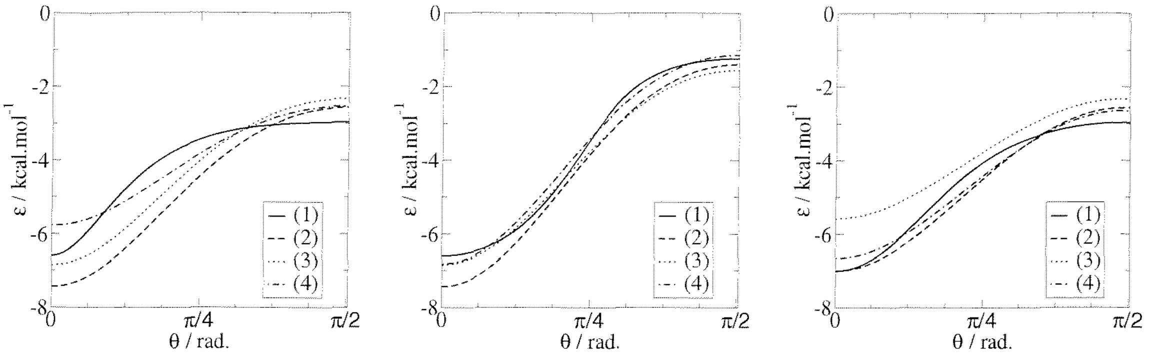
Figure 4.8: The three different arrangements chosen to sample the configurational space.

the well depth of the side-by-side arrangement is consistent with that for the RLJ4 model. For the side-by-side configuration the strength parameter is given by :

$$\epsilon_s = \epsilon_0 (1 - \chi^2)^{-\frac{\nu}{2}}. \tag{4.11}$$

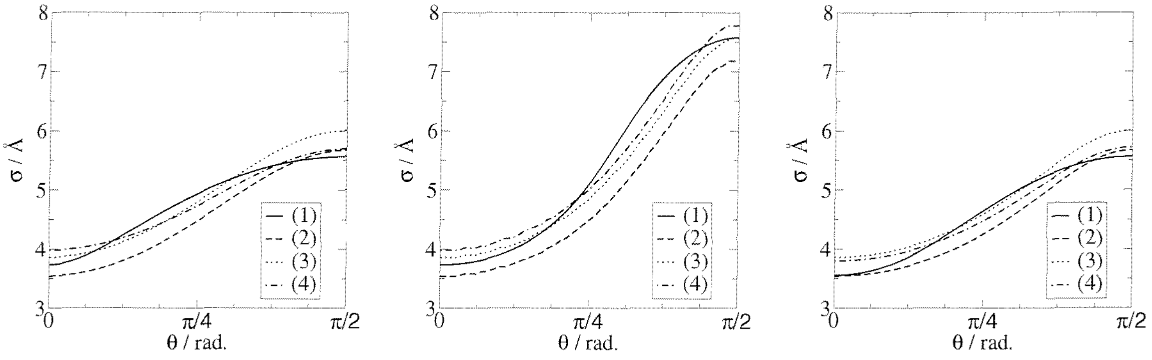
Using the values of ϵ_s,ϵ_0 and χ obtained previously we can determine ν. By comparison with the RLJ4 model a value of 0.26 was calculated for ν. This value clearly improves the fit to the Lennard-Jones curve (cf. Figure 4.7 - plot (4)), at least for the well depth. This way of characterising the different parameters allows an appropriate parameterization to be found for parallel particles, but the fitting is far from being perfect for other arrangements. The other arrangements that have been chosen to sample as efficiently as possible the configurational space are described in Figure 4.8. In each case the angle θ goes from 0 to $\pi/2$. These three arrangements, even if they do not take all the possible arrangements into account, sample most of the configurations two Gay-Berne particles could adopt, and seem therefore sensible choices. The well depth and contact distance of the potentials for the three configurations are displayed in Figure 4.9 and 4.10 respectively. Although the parameters based on comparison with the RLJ4 model bring an improvement, they clearly are not satisfactory for all three arrangements.

The parameterization of the Gay-Berne potential should ideally take into account all possible configurations that can be achieved by two Gay-Berne particles, rather than a few characteristic arrangements as described in this section. Such a parameterization technique has been developed and will be presented in section 4.2.4.



(a) Configuration 1 (b) Configuration 2 (c) Configuration 3

Figure 4.9: Well depth calculated as a function of the angle θ as described in Figure 4.8 for the three different arrangements. The notations LJ and GB refer to LJ: Lennard-Jones, and GB: GB(2.03,5.31,2.0,0.26) with ϵ_0=7.005 kJ.mol^{-1} and σ_0=3.545Å. (a) Configuration 1 goes from arrangement **s** to **t**, (b) configuration 2 from arrangement **s** to **e**, and (c) configuration 3 from **x** to **t**. Plot (1) calculated with the Lennard-Jones potential, plots (2) to (4) with the Gay-Berne potential and parameters described in Figure 4.7.



(a) Configuration 1 (b) Configuration 2 (c) Configuration 3

Figure 4.10: Contact distance calculated as a function of the angle θ as described in Figure 4.8 for the three different arrangements. The notations LJ and GB refer to LJ: Lennard-Jones, and GB: GB(2.03,5.31,2.0,0.26) with ϵ_0=7.005 kJ.mol^{-1} and σ_0=3.545Å. (a) Configuration 1 goes from arrangement **s** to **t**, (b) configuration 2 from arrangement **s** to **e**, and (c) configuration 3 from **x** to **t**. Plot (1) calculated with the Lennard-Jones potential, plots (2) to (4) with the Gay-Berne potential and parameters described in Figure 4.7.

4.2.3 The Gay-Berne Model of Biological Lipids

Starting from the work by Whitehead *et al.*[2] who simulated the hydrocarbon region of a biomembrane using a reduced representation model, we have extended the model to represent a complete biomembrane, and include solvation. This section will summarize the work carried out by Whitehead *et al.*, and will then describe the model presented here based on their approach.

The Gay-Berne Model of the Hydrocarbon Region of a Biological Lipid

The construction of a model of a biological lipid was based on the model previously developed by Whitehead *et al.*[2] for the hydrocarbon region of biological lipids. In the Whitehead parameterization, each Gay-Berne particle was used to model four methylene (CH_2) groups; a choice consistent with the original parameterization of the Gay-Berne potential. A $(CH_2)_{12}$ chain was therefore modelled using three linked Gay-Berne particles. The linking of the Gay-Berne was achieved using harmonic spring potentials anchored to the end of each. The length of the spring was adjusted to between 1.1 and 1.3 Å, and the force constant was set to 5.0 in reduced units. The anchoring points for the springs were taken as the positions of the terminal (CH_2) Lennard-Jones atoms subsumed into the potential. The absence of angular and torsional potentials enabled the Gay-Berne potential to contribute to the net translational motion of each particle but also their rotation. The choice of a weak force constant for the potential allowed the Gay-Berne chains to adjust their length, thereby mimicking the effect of *gauche* conformations in a "real" hydrocarbon chain. The two chains were linked together using a harmonic potential of length 7.9 Å anchored at the centre of the "head" beads with a force constant set to 5.0 in reduced units. The connectivities of this representation are illustrated in Figure 4.11. In the absence of solvation or head groups, the "head" Gay-Berne particles were anchored at their centre to a plane parallel to the membrane surface to maintain the integrity of the membrane. The force constant was again fixed to 5.0 in reduced units and the equilibrium distance was set to zero. The Gay-Berne particles are free to diffuse within this plane. Adjusting the separation distance between these two planes enabled control of the bilayer width.

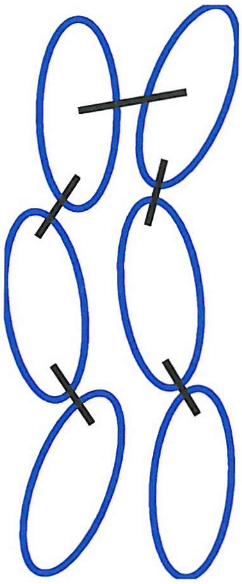


Figure 4.11: Schematic Gay-Berne representation of the hydrocarbon region of a biological lipid. Harmonic springs are shown linking Gay-Berne particles to form a "Gay-Berne" lipid.

The parameters used by Whitehead *et al.* were directly derived from the parameters originally presented by Gay and Berne[1] to mimic the potential between molecules comprising four centres equally spaced on a line at intervals of $2\sigma_0/3$ and interacting via the Lennard-Jones 12-6 potential. These parameters are described in Table 4.3 along with the parameters used by Whitehead *et al.* and parameters used by Luckhurst et al.[112, 113]

Table 4.3: Comparison of the different sets of parameters used for the Gay-Berne potential.

	$\sigma_{\parallel}/\sigma_{\perp}$	$(\epsilon_{\perp}/\epsilon_{\parallel})$	μ	ν
Gay and Berne[1]	3.0	5.0	2	1
Luckhurst *et al.*[112] (smectic)	4.4	20.0	1	1
Luckhurst *et al.*[113] (smectic and nematic)	3.0	5.0	1	2
Whitehead *et al.*[2] (non bonded)	2.0	3.0 / 5.0	2	1
Whitehead *et al.*[2] (bonded)	1.1 / 1.2 / 1.3	3.0 / 5.0	2	1

The first set of parameters were used by Luckhurst *et al.* to model the nematic liquid crystalline phase while the second were used to identify smectic and nematic mesophases and yielded a rich phase behaviour consisting of smectic, nematic, and isotropic liquid-crystalline phases. The parameters used for the study of the hydrocarbon region of a biological lipid were separated into *bonded* and *non bonded* parameters. Only the shape anisotropy differs between the bonded and non-bonded particles; for non bonded particles a ratio of 2 was used, consistent with the fact that each Gay-Berne particle is modelling four methylene groups. For bonded particles (with the exception of the two "head" particles which are considered to be non bonded with respect to the Gay-Berne interaction) ratios of 1.1, 1.2 and 1.3 were used. These smaller values reflects the fact that the bonded Gay-Berne particles are covalently linked. A bonded shape anisotropy of 1.3 proved to lead to better results. Similarly an energy anisotropy of 5.0 was determined to be the optimum value. Despite the simplicity of the model, the results clearly demonstrate the presence of a phase transition from an ordered low-temperature solid phase reminiscent of the $L_{\beta'}$, to a high-temperature disordered phase reminiscent of the L_α phase. Order parameters calculated for each layer of Whitehead's model are consistent with the experimental segmental order parameter reported for DMPC.

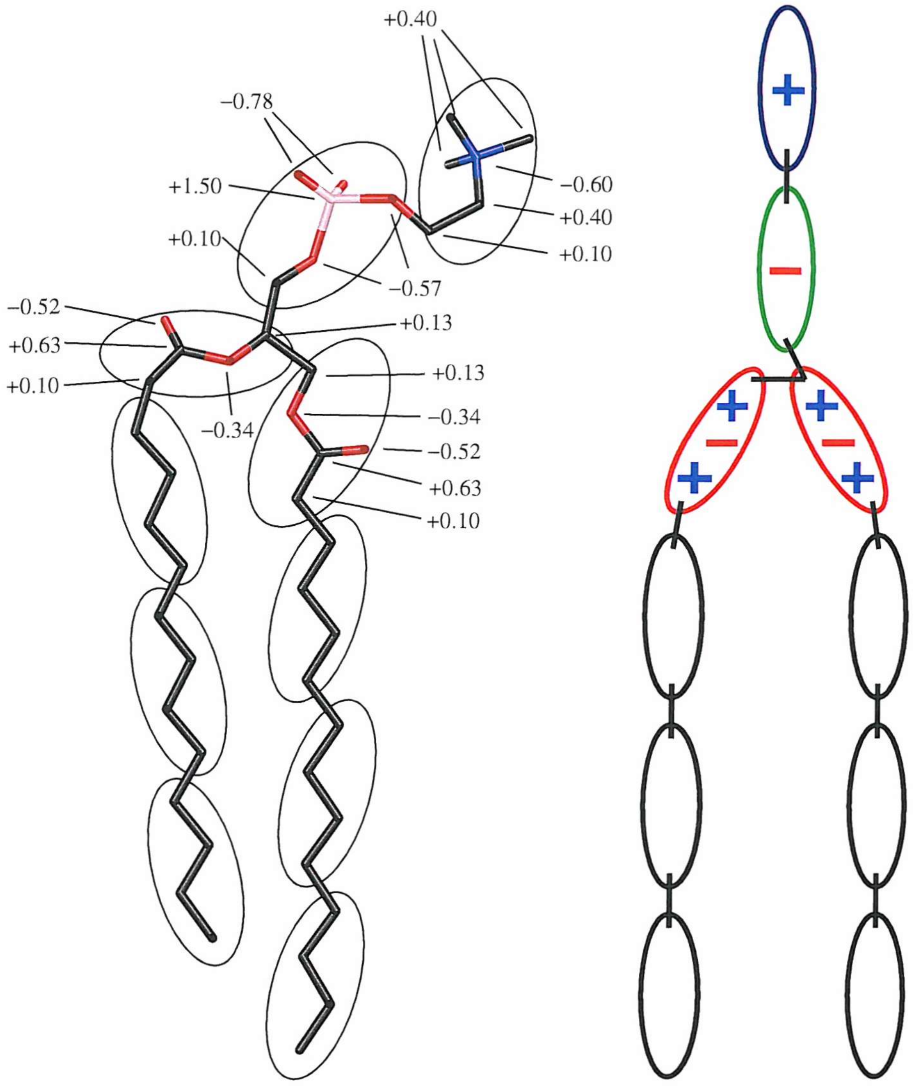
The Gay-Berne Model of a Biological Lipid Bilayer

Starting from the work previously reported by Whitehead *et al.*, the model has been extended to represent a complete biological lipid. The hydrocarbon region of the lipid was represented in the fashion outlined in the previous section. Only the parameterization changed, to fit more adequately to the phospholipid of choice (detailed later), and the harmonic potential linking the two "head" beads was removed, due to the presence of the head group. The addition of solvation also made the anchoring of the "head" beads to a plane redundant, and this has therefore been removed.

To be consistent, the representation of the head group had to reproduce the same *proportion* as the hydrocarbon region, *i.e.* approximately one bead or ellipsoid for four aligned atoms (with the exclusion of hydrogen atoms). The description also had to take into account the charge distribution. The charges on a united-atom DMPC are illustrated in Figure 4.12, together with the final Gay-Berne representation of DMPC. The charges are taken from the CHARMM27 all-atom force field[114] and are converted into united-atom charges. The charge distribution on the DMPC head group shows three distinctive regions; a positively charged region around the nitrogen atom, a negatively charged region around the phosphorus atom, and a neutral region around the glycerol moiety. The latter is globally neutral but dipolar. The

PhosphatidylCholine region can therefore, taking into account charge distribution and steric considerations, be split into two Gay-Berne particles. The first around the nitrogen atom includes two methylenes, the nitrogen atom and the three methyl groups attached to the nitrogen atom: $-CH_2 - CH_2 - N^+(CH_3)_3$. The total electronic charge on this cluster is +1.10 (taking into account CHARMM27 charges), and there are four aligned atoms, in agreement with the model used for the hydrocarbon region. The globally positive charge of the nitrogen bead makes the use of a single point positive charge at the centre of the particle a sensible representation of its electrostatic nature. The second cluster is centred around the phosphorus atom and contains the phosphorus atom, the four oxygen atoms that are attached to it and a methylene group: $-CH_2 - O - PO_2^- - O-$. The total CHARMM27 electronic charge on the phosphorus bead is -1.10 enabling the use of a single point negative charge. The phosphorus particle consists of four aligned atoms, once again in agreement with the hydrocarbon region representation. The rest of the head group was split into two ellipsoids, the first contained the CH group, a carboxyl group and a methylene; the second two methylene groups and a carboxyl group : $-CH - O - CO - CH_2-$ and $-CH_2 - O - CO - CH_2-$ respectively. Contrary to the nitrogen and phosphorus beads; the total electronic charge on these clusters is zero, making the use of a single point charge unrealistic. Instead three point charges were used for each particle to represent the electrostatic nature of these regions. The glycerol particles consist of four aligned atoms in agreement with the other parts of the lipid representation. The nature of the charges to be used was decided by examining the glycerol cluster CHARMM27 charge distribution. As can be seen in Figure 4.13, three groups of charges can be discerned. A small positive charge on the methylene linked to the rest of the head group, negative charges on the oxygen atoms, and a positive charge on the methylene linking the group to one of the hydrocarbon chains. The charges on the glycerol beads were therefore attributed as follows, two positive charges were located at each end of the particle and one negative charge was located at the centre.

The tail particles correspond to butane; they were linked in the fashion outlined by Whitehead *et al.* using harmonic spring potentials. For bonds between tail particles, the force constant of the potential was set to 5 in reduced units and the equilibrium



(a)

(b)

Figure 4.12: Two representations of a DMPC lipid : (a) united atom representation, the ellipsoids indicate the way the lipid was decomposed in order to establish the Gay-Berne representation; the electronic charges indicated are CHARMM27 all-atom charges converted to united atom charges. Carbon atoms are shown in black, oxygen atoms in red, phosphorus atom is shown in pink and nitrogen atom in blue. (b) Schematic Gay-Berne representation of a biological lipid. Nitrogen bead is shown in dark blue, phosphorus bead in green, glycerol beads in red and hydrocarbon beads in black. The plus and minus signs indicate the number of charges used as well as their nature (positive or negative).

distance to 1.3 Å, a distance that reflects the covalent nature of the link. The head group bead encompassed different atoms in different geometries, and because of this, we elected to invoke a simple geometry argument to justify the equilibrium distances in the head group region. All equilibrium distances are therefore based on an initial structure, that ensures the distance between the two tails is close to the experimen-

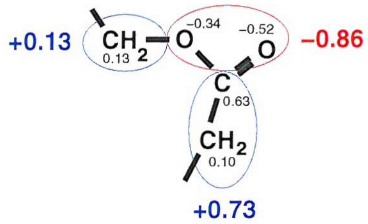


Figure 4.13: The CHARMM electronic charges on one of the glycerol cluster together with the three main charges cluster.

tal value (7.9 Å). The force constant being weak, the equilibrium distances and the initial structure have a limited influence on the overall membrane behaviour. Each of the glycerol beads was linked to a chain of three Gay-Berne tail particles through a harmonic potential of force constant set to 5.0 in reduced units and an equilibrium distance of 2.38 Å. The springs were anchored at the lower end of the particles, 1.95 Å away from the particle centre. This distance was chosen to correspond to the location of the terminal CH_2 groups. The equilibrium distance is larger than a conventional bond distance to reproduce the initial structure and ensure the width of the lipid remains close to the experimental value. The two glycerol particles were linked together at their top end using the same potential with a force constant of 40.0 in reduced units and an equilibrium distance of 4.0 Å. This equilibrium distance is, for the same reasons detailed before, larger than a conventional bond distance, an additional reason being to avoid close electrostatic interactions between opposed identical charges on the glycerol particles. The force constant was fixed at a larger value than in the tail region to account for the greater rigidity of the head group. Only one of the glycerol particles was linked to the phosphorus particle, the latter being linked to the nitrogen particle. Bonds were present in between the top end of the glycerol particle and the bottom end of the phosphorus particle, and between the top end of the phosphorus particle and the bottom end of the nitrogen particle. The harmonic potentials used for these links have a force constant of 40 in reduced units and an equilibrium distance of 2.38 Å for the glycerol-phosphorus bond and 1.3 Å for the nitrogen-phosphorus bond.

4.2.4 Parameterization

Gay-Berne Parameter Optimization

Even though the simple parameterization technique described in subsection 4.2.2 reproduces satisfactorily the results obtained with the RLJ4 model, it still only takes into account a few characteristic molecular arrangements. The parameterization of the Gay-Berne potential should ideally take into account all possible configurations that can be achieved by two Gay-Berne particles. A fitting procedure has therefore been developed that accounts for a larger set of conformations. To sample as efficiently as possible the configurational space, three sets of orientations have been chosen. They are described in Figure 4.8. Even though they do not sample all the configurational space, they include most of the arrangements two Gay-Berne particles can take. Configuration 1 goes from the **s** to the **t** arrangement, both orientation vectors are in the same plane and the angle between these two vectors is θ. Configuration 2 goes from the **s** to the **e** arrangement, both orientation vectors are parallel and the angle between these two vectors and the intermolecular vector is $\pi/2-\theta$. Configuration 3 goes from the **x** to the **t** arrangement, orientation vectors are perpendicular to each other, the first particle orientation vector is perpendicular to the intermolecular vector while the second makes an angle $\pi/2-\theta$ with it.

For each set of configurations, the angle θ is varied between 0 and $\pi/2$. The interaction between the two particles is calculated as a function of the intermolecular distance for the RLJ4 representation and as a function of the intermolecular distance and two of the parameters to be optimized for the Gay-Berne representation. From the interaction curve, the contact distance and the well depth of the potential are calculated. Two error functions are then derived, one depending on the contact distance, the other on the well depth.

The contact distance of the Gay-Berne potential (σ^{GB}) depends only on σ_0 and on the shape anisotropy of the particles $\sigma_{\|}/\sigma_{\perp}$; the well depth anisotropy $\epsilon_{\perp}/\epsilon_{\|}$ and ϵ_0 have no effect on the contact distance calculation. The well depth of the Gay-Berne potential (ϵ^{GB}) depends on ϵ_0, μ, ν, $\sigma_{\|}/\sigma_{\perp}$ and $\epsilon_{\perp}/\epsilon_{\|}$. The μ and ν parameters were not optimized and kept to their original values as described by Gay and Berne.[1]

The reason for this decision was to keep the analytical shape of the potential and of the forces as simple as possible. Changing the coefficients μ and ν would have meant changing the shape of the forces, and would have made the optimization process much more complicated. The first error function (ξ_{CD}), the contact distance error function, is therefore used to optimize the parameters σ_0 and $\sigma_{\|}/\sigma_{\perp}$.

$$\xi_{\mathrm{CD}}\left(\sigma_0, \sigma_{\|}/\sigma_{\perp}\right) = \sum_{\theta=0}^{\pi/2} \left[\sigma^{LJ}(\theta) - \sigma^{GB}(\theta, \sigma_0, \sigma_{\|}/\sigma_{\perp})\right]^2 \qquad (4.12)$$

For each value of the angle θ, σ_0 and $\sigma_{\|}/\sigma_{\perp}$ are systematically varied, and for each set the contact distance for the Gay-Berne potential $\sigma^{GB}(\theta, \sigma_0, \sigma_{\|}/\sigma_{\perp})$ is calculated. The contact distance for the Lennard-Jones potential $\sigma^{LJ}(\theta)$ is also calculated for each angle θ. Once the contact distances have been determined for all angles for the Lennard-Jones interaction and for all angles and all $(\sigma_0, \sigma_{\|}/\sigma_{\perp})$ sets for the Gay-Berne interaction, the error function $\xi_{\mathrm{CD}}\left(\sigma_0, \sigma_{\|}/\sigma_{\perp}\right)$ is calculated independently for each configuration. Three error functions are therefore calculated; one for each arrangement described in Figure 4.8. For each $(\sigma_0, \sigma_{\|}/\sigma_{\perp})$ set the difference between the contact distance calculated with the Lennard-Jones potential $\sigma^{LJ}(\theta)$ and that calculated with the Gay-Berne potential $\sigma^{GB}(\theta, \sigma_0, \sigma_{\|}/\sigma_{\perp})$ is calculated, squared and summed over all the θ angles. The minimum of this error function corresponds to the $(\sigma_0, \sigma_{\|}/\sigma_{\perp})$ pair that allows the best fitting between the Gay-Berne and the Lennard-Jones representations. The contact distance part of the parameterization process is summarised in Figure 4.14. The parameters σ_0 and $\sigma_{\|}/\sigma_{\perp}$ obtained from the contact distance parameterization are then used to optimize the remaining parameters ϵ_0 and $\epsilon_{\perp}/\epsilon_{\|}$ using the well depth error function (ξ_{WD}).

$$\xi_{\mathrm{WD}}\left(\epsilon_0, \epsilon_{\perp}/\epsilon_{\|}\right) = \sum_{\theta=0}^{\pi/2} \left[\epsilon^{LJ}(\theta) - \epsilon^{GB}(\theta, \epsilon_0, \epsilon_{\perp}/\epsilon_{\|})\right]^2 \qquad (4.13)$$

The same procedure applied to the contact distance parameterization is then applied to the well depth component with the well depth calculated using the Gay-Berne representation (ϵ^{GB}) and being compared to that obtained using the Lennard-Jones representation (ϵ^{LJ}). The minimum of this error function corresponds to the $(\epsilon_0, \epsilon_{\perp}/\epsilon_{\|})$

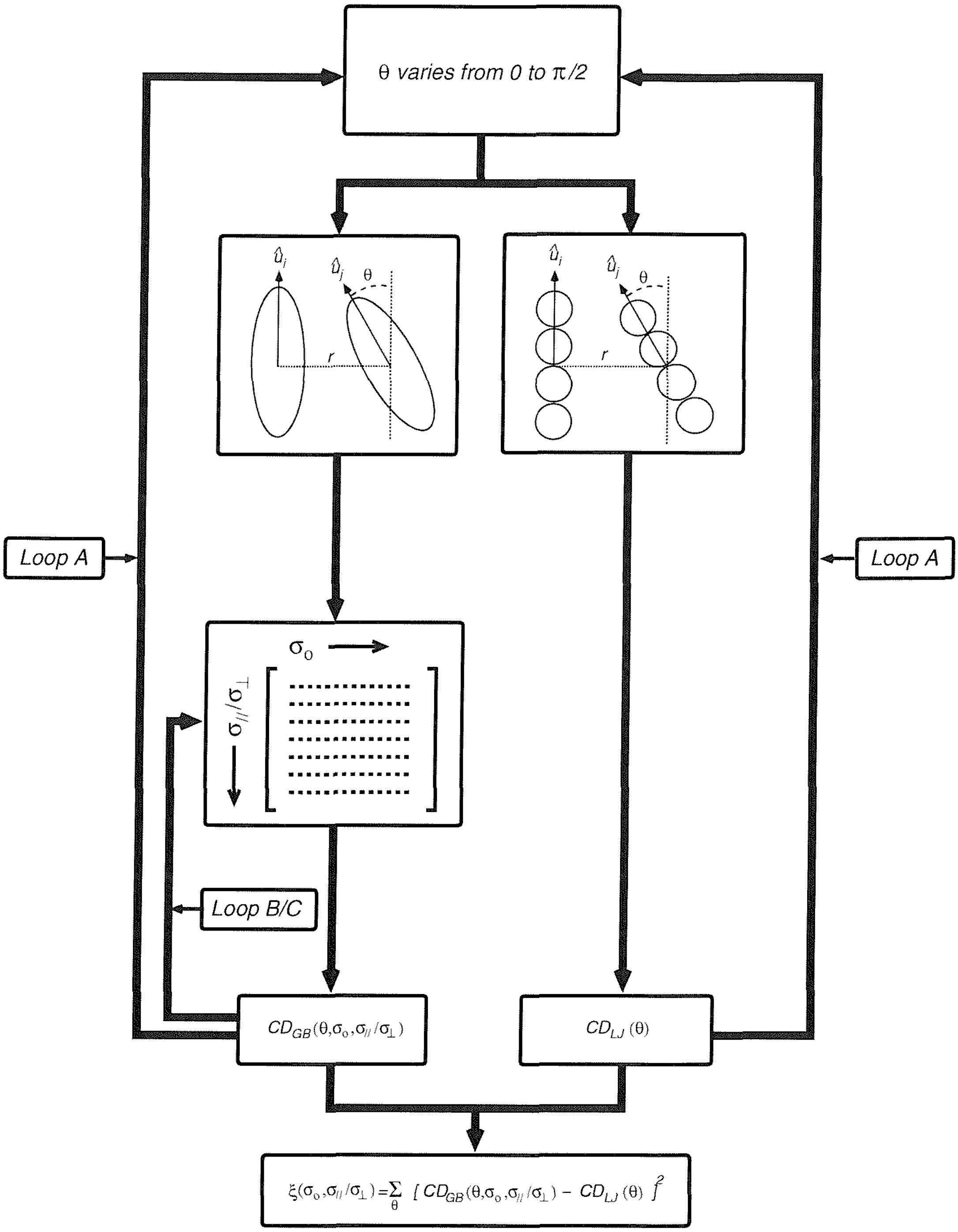


Figure 4.14: Schematic representation of the contact distance (CD) part of the parameterization process. Loop A concerns the angle θ that varies from 0 to $\pi/2$, loops B and C concern the parameters to optimize, σ_0 and $\sigma_{\|}/\sigma_{\perp}$. Loops B and C only apply to the Gay-Berne calculation.

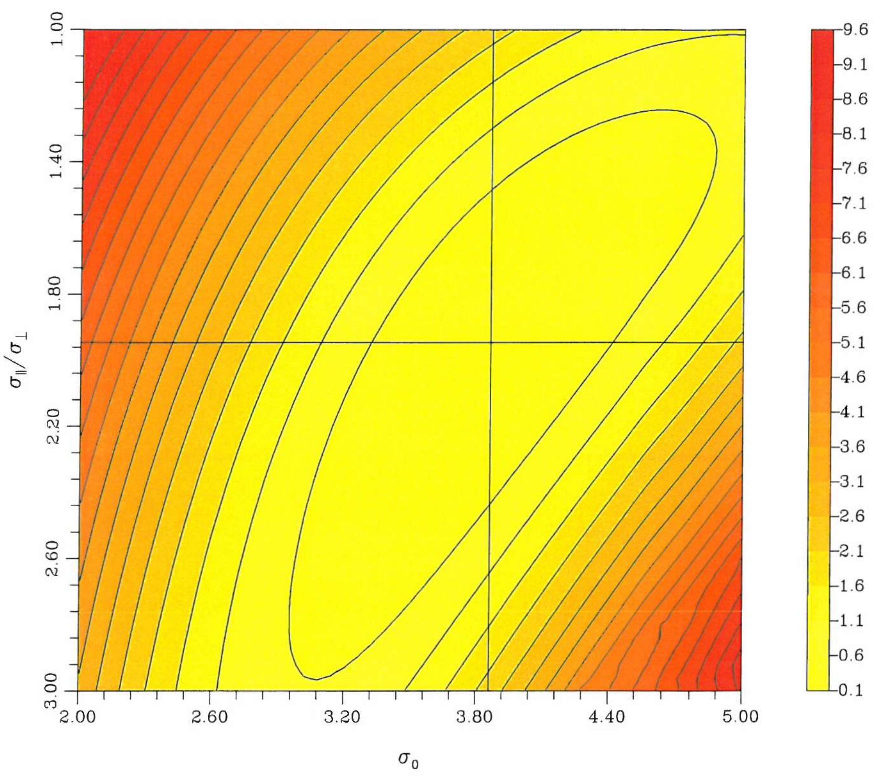
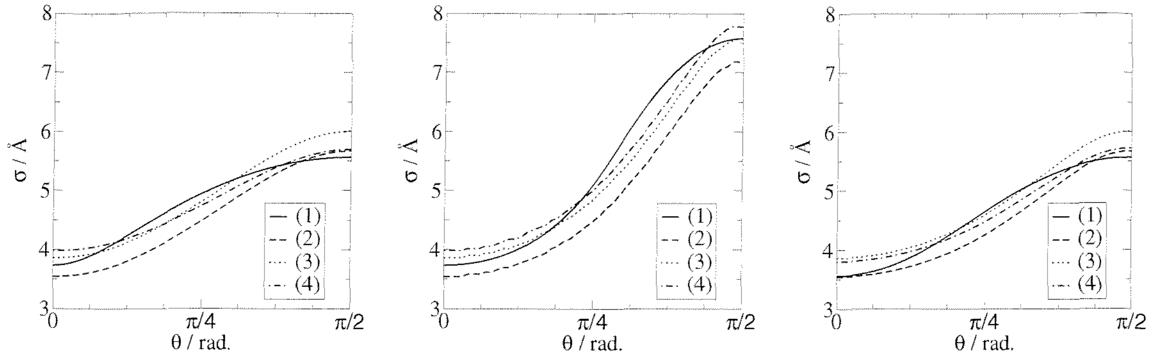


Figure 4.15: 2D contour for the contact distance error function for all three configurations as a function of σ_0 and $\sigma_{\|}/\sigma_{\perp}$. Minimum location ($\sigma_0$,$\sigma_{\|}/\sigma_{\perp}$) = (3.86, 1.96). The contour are shown for values of the error function in between 0.1 and 9.6 in steps of 0.5.

pair that allows the best fit between the Gay-Berne and Lennard-Jones representations.

The Lennard-Jones representation of butane (RLJ4) has been described in subsection 4.2.1. United atom OPLS[30] parameters were used, σ_{LJ} = 3.905 Å and ϵ_{LJ} = 0.118 kcal.mol^{-1}. The carbon atoms were aligned to remove the biaxiality of butane. The distance between adjacent atoms was fixed to 1.3 Å, instead of the optimal distance of 1.526 Å to account for the alignment of the sites (cf. Figure **??**).

The parameters σ_0 and $\sigma_{\|}/\sigma_{\perp}$ were optimized first with the contact distance error function ξ_{CD}. The parameters ϵ_0, $\epsilon_{\perp}/\epsilon_{\|}$, μ and ν were fixed and given the values predetermined based on the observation of characteristic conformations (**x**, **s** and **e**), respectively 7.005 kJ.mol^{-1}, 5.31, 2 and 1. In the Gay-Berne potential the contact distances are independent of this choice. The 2D contours for the contact distance error function for all three configurations are displayed in Figure 4.15. The error functions 2D contours for configurations 1, 2 and 3 are displayed separately in appendix B. The



(a) Configuration 1 (b) Configuration 2 (c) Configuration 3

Figure 4.16: Contact distance calculated as a function of the angle θ as described in Figure 4.8 for the three different arrangements. Plot (1) calculated with a Lennard-Jones potential, (2) to (4) with the Gay-Berne potential and (2) the parameters obtained through comparison with a RLJ4 model at certain configurations (cf. section 4.2.1), (3) the global parameters obtained from the error function calculation and (4) the parameters obtained from the error function calculation specific to each configuration. $\epsilon_0 = 7.005\,\mathrm{kJ.mol^{-1}}$ in plots (2) to (4). The set $(\sigma_{\|}/\sigma_{\perp},\epsilon_{\perp}/\epsilon_{\|},\mu,\nu)$ is equal to: in plot (2) (2.03,5.31,2,0.26) and $\sigma_0 = 3.545\,\text{Å}$, in plot (3) (1.96,5.31,2,1) and $\sigma_0 = 3.86\,\text{Å}$, in plot (4) (1.76,5.31,2,1) and $\sigma_0 = 3.98\,\text{Å}$ for configuration 1, (1.96,5.31,2,1) and $\sigma_0 = 3.98\,\text{Å}$ for configuration 2 and (1.88,5.31,2,1) and $\sigma_0 = 3.80\,\text{Å}$ for configuration 3.

overall error function is the sum of all three error function $\xi_{\mathrm{CD}}^{(\mathrm{all})} = \xi_{\mathrm{CD}}^{(1)} + \xi_{\mathrm{CD}}^{(2)} + \xi_{\mathrm{CD}}^{(3)}$, where $\xi_{\mathrm{CD}}^{(i)}$ corresponds to the error function of configuration i. The minimum of the overall error function $\xi_{\mathrm{CD}}^{(\mathrm{all})}$ is located at $(\sigma_0,\sigma_{\|}/\sigma_{\perp})$=(3.86,1.96). The minima of the individual error functions are displayed in Table 4.4.

Table 4.4: Parameters calculated from the contact distance fitting, based on all arrangements, arrangement 1, 2 and 3.

	1+2+3	1	2	3
σ_0 / Å	3.86	3.98	3.98	3.80
$\sigma_{\|}/\sigma_{\perp}$	1.96	1.76	1.96	1.88

The optimal values for the three arrangements taken together does not differ much from the optimal values from the individual arrangements. To check our parameterization method improved the fitting with the Lennard-Jones model, we plotted the contact distance as a function of the angle θ for the three different arrangements with different parameters. In Figure 4.16, Plot (1) corresponds to the Lennard-Jones model, plot (2) corresponds to the parameters obtained through comparison with a RLJ4 model at certain configurations, plot (3) to our set of optimized parameters taking into account all the orientations and plot (4) to the set of optimized parameters specific to each orientation. The set of parameters based on the three orientations (plot (3)) present a significant improvement compared to the simple parameterization (plot (2)) and enable a better fitting to the Lennard-Jones curve (plot (1)). The set of parameters specific to each orientation (plot (4)) unsurprisingly enables a better fitting. The best set of parameters therefore appears to be a value of σ_0 of 3.86 Å and a shape anisotropy $\sigma_{\|}/\sigma_{\perp} = 1.96$. These parameters are logically quite close to the one we found before. Some discrepancy can be found for certain configurations displayed in Appendix B,, for high values of σ_0 and $\sigma_{\|}/\sigma_{\perp}$. These artefacts come from the way the contact distance is calculated. For certain parameters, the Gay-Berne potential possesses a minimum in the centre of the ellipsoid. This flaw that could lead to problems if we were using Monte-Carlo simulations, makes the calculation of contact distance more difficult. Indeed, the potential can be zero at different points, leading to a false representation of the contact distance. These discrepancies only happen for the highest values of the parameters, far from the minimum. They therefore do not interfere with the determination of the error function minimum.

The remaining parameters ϵ_0 and $\epsilon_{\perp}/\epsilon_{\|}$ were then optimized using the well depth error function ξ_{WD}. The parameters σ_0 and $\sigma_{\|}/\sigma_{\perp}$ were fixed to the value determined using the contact distance error function ξ_{CD}, respectively 3.86 Å and 1.96. The 2D contour for the well depth error function of all three configurations combined is displayed in Figure 4.17. The error function 2D contours for configurations 1, 2 and 3 are displayed in appendix C.The overall error function is the sum of all three error function $\xi_{\mathrm{WD}}^{(\mathrm{all})} = \xi_{\mathrm{WD}}^{(1)} + \xi_{\mathrm{WD}}^{(2)} + \xi_{\mathrm{WD}}^{(3)}$, where $\xi_{\mathrm{WD}}^{(i)}$ corresponds to the error function of configuration i. The minimum of the overall error function $\xi_{\mathrm{WD}}^{(\mathrm{all})}$ is located at $(\epsilon_0, \epsilon_{\perp}/\epsilon_{\|}) = (5.58, 4.40)$. The minima of the individual error functions are displayed in Table 4.5. Contrary to the contact distance parameters, the values determined using the well depth error function show less uniformity among the configurations. This is particularly true for the well depth anisotropy which varies from 2.96 for configuration 1 to 5.96 for configuration 2. Even though configuration 2 is one of the most representative of our structures, its parameters cannot be used for all configurations

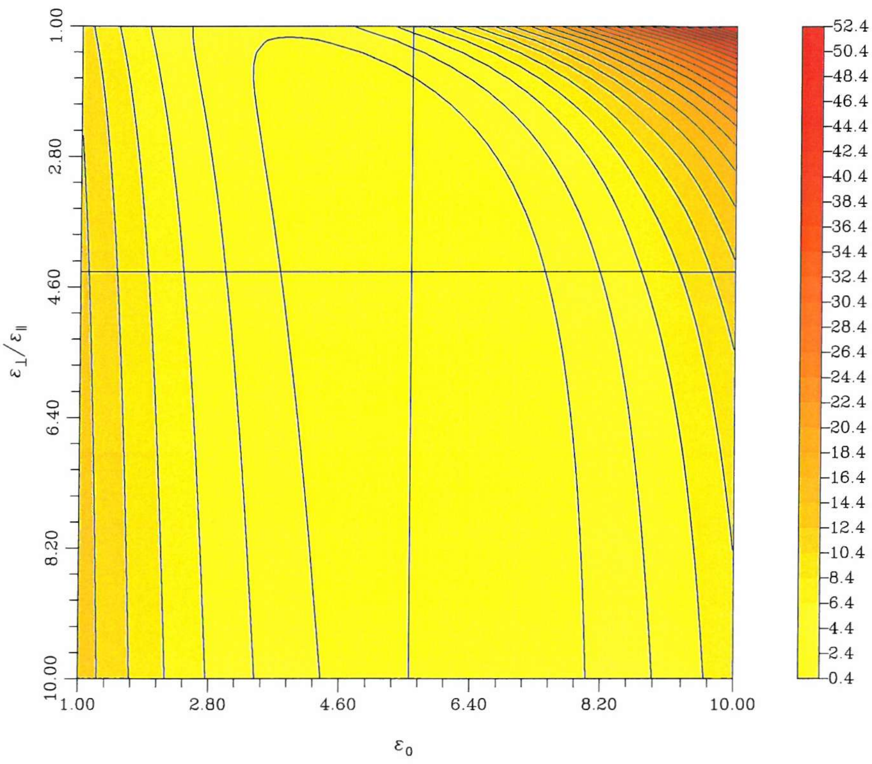


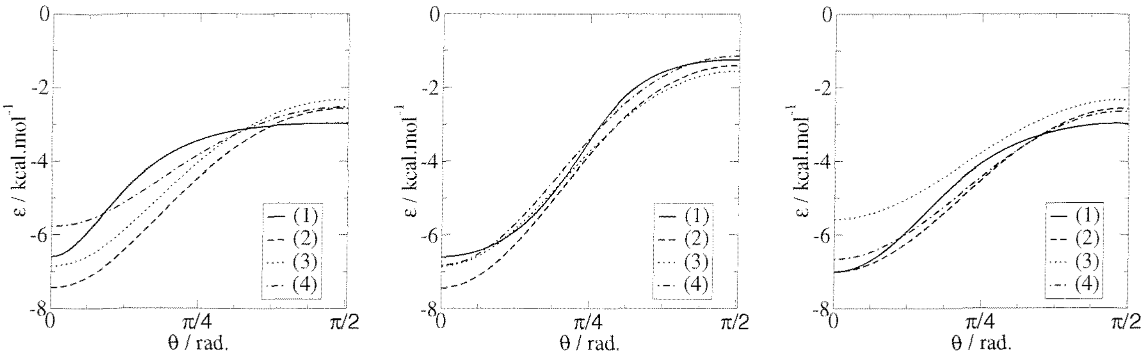
Figure 4.17: 2D contour for the well depth error function for all three configurations as a function of ϵ_0 and $\epsilon_\perp/\epsilon_\parallel$. Minimum location $(\epsilon_0, \epsilon_\perp/\epsilon_\parallel) = (5.58, 4.40)$.

as a value of $\epsilon_\perp/\epsilon_\parallel$ of 5.96 appears to be too high.

Table 4.5: Parameters calculated from the well depth fitting, based on all arrangements, arrangement 1, 2 and 3.

	1+2+3	1	2	3
ϵ_0 / kJ.mol^{-1}	5.58	4.70	5.56	6.66
$\epsilon_\perp/\epsilon_\parallel$	4.40	2.96	5.96	4.72

Even though different configurations give different results, the parameters obtained from the overall error function still give satisfactory results in each case, as can be seen in Figure 4.18. For each configuration, the well depth is plotted as a function of the angle θ with different parameters. For configuration 1 and 2 the overall parameters give better results than the parameters obtained from the simple parameterization. Configuration 3 is the only case where the overall parameters do not give better results. As the most representative conformations in our system are contained in configuration 1 and 2, it appears sensible to choose the overall parameters. The optimum values for ϵ_0 and $\epsilon_\perp/\epsilon_\parallel$ are therefore respectively 5.58 kJ.mol^{-1} and 4.40.



(a) Configuration 1 (b) Configuration 2 (c) Configuration 3

Figure 4.18: Well depth calculated as a function of the angle θ as described in Figure 4.8 for the three different arrangements. Plot (a) calculated with a Lennard-Jones potential, (b) to (d) with the Gay-Berne potential and (b) the parameters obtained through comparison with a RLJ4 model at certain configurations, (c) the global parameters obtained from the error function calculation and (d) the parameters obtained from the error function calculation specific to each configuration. The set $(\sigma_{\|}/\sigma_{\perp},\epsilon_{\perp}/\epsilon_{\|},\mu,\nu)$ is equal to: in plot (b) (2.03,5.31,2,0.26) with $\epsilon_0 = 7.005\,\mathrm{kJ.mol^{-1}}$ and $\sigma_0 = 3.545\,\text{Å}$, in plot (c) (1.94,4.40,2,1) with $\epsilon_0 = 5.58\,\mathrm{kJ.mol^{-1}}$ and $\sigma_0 = 3.86\,\text{Å}$, in plot (d) (1.94,2.98,2,1) with $\epsilon_0 = 4.70\,\mathrm{kJ.mol^{-1}}$ and $\sigma_0 = 3.86\,\text{Å}$ for configuration 1, (1.94,5.96,2,1) with $\epsilon_0 = 5.56\,\mathrm{kJ.mol^{-1}}$ and $\sigma_0 = 3.86\,\text{Å}$ for configuration 2 and (1.94,4.60,2,1) with $\epsilon_0 = 6.66\,\mathrm{kJ.mol^{-1}}$ and $\sigma_0 = 3.86\,\text{Å}$ for configuration 3.

Configuration 3 illustrates perfectly the limitation of the parameterization technique based on comparison with a RLJ4 model at certain configurations. The value found for ϵ_0 in configuration 3 is smaller than the one found using the simple parameterization technique , which only accounts for the **x** conformation. A better fitting through all values of θ is obtained with the smaller value of $6.66\,\mathrm{kJ.mol^{-1}}$ (compared to the previous value of $7.005\,\mathrm{kJ.mol^{-1}}$). On a larger scale, the overall ϵ_0 value is even smaller when taking into account all three configurations ($5.58\,\mathrm{kJ.mol^{-1}}$) but allows a better global correlation with the Lennard-Jones representation.

As can be seen in Figure 4.15 for the contact distance and in Figure 4.17 for the well depth; the contours are flat around the minima, leaving some freedom for the choice of the optimum parameters. For this reason, we decided to try different values around the optimized ones and to determine their effect on the simulation behaviour. The shape anisotropy $\sigma_{\|}/\sigma_{\perp}$ was close to the expected value, we therefore kept the optimized value of 1.94 and did not investigate it any further. Changing the energy anisotropy $\epsilon_{\perp}/\epsilon_{\|}$ did not have a significant effect on the membrane behaviour, the optimized value of 4.40 was therefore used. The well depth (ϵ_0) and contact distance (σ_0) had more influence on the overall structural and thermodynamic behaviour of the membrane, hence several values located around the minima were investigated. Several simulations were performed at temperatures ranging from 5 to 60 °C, with σ_0 taking values of 3.86, 4.0 and 4.5 Å and ϵ_0 values of 5.58, 6.0 and 6.5 kJ.mol^{-1}. These parameters are scaling values for the potential, their effect on the membrane behaviour is principally due to the balance with other interactions. The water-lipid van der Waals interaction, treated with the generalized Gay-Berne potential, which will be described later in this chapter, posses an intrinsic contact distance (σ_0^{GGB}) and well depth ($\epsilon_0^{\mathrm{GGB}}$). Therefore, σ_0^{GGB} and $\epsilon_0^{\mathrm{GGB}}$ being fixed, any change to σ_0 and ϵ_0 will affect the balance between the water-lipid and lipid-lipid van der Waals interactions and therefore affect the membrane structure. After consideration of the orientational ordering and of the electron density profiles, and comparison with experimental data, the best parameters were found to be $\sigma_0 = 4.0$ Å and $\epsilon_0 = 6.0$ kJ.mol^{-1}.

Charge Optimization

The next step toward the complete parameterization of the lipid model is the charge parameterization. The charges on the reduced representation model were optimized to reproduce selected all-atom dipoles and initially quadrupoles. The quadrupole is the third term in the multipole expansion, the first term being the charge and the second the dipole. The charges on our first lipid representation are described in Figure 4.19. These charges are based on the all-atom CHARMM27 charges[114] converted into united-atom charges. For the nitrogen and phosphorus particles the corresponding atomic charges are summed up, to give a single point charge placed at the centre of the ellipsoid. For the glycerol particles, the net charge being zero, three point charges have been placed on the axis. Their optimization will be described in this section. We were unable to optimize quadrupoles for geometry reasons. The quadrupole of any charge distribution is a symmetric and traceless tensor, which in the principal axis frame simplifies to a diagonal traceless tensor. In our glycerol representation, the charge distribution is cylindrically symmetric, the three charges being aligned on the principal axis of the bead as illustrated in Figure 4.19. Unfortunately the

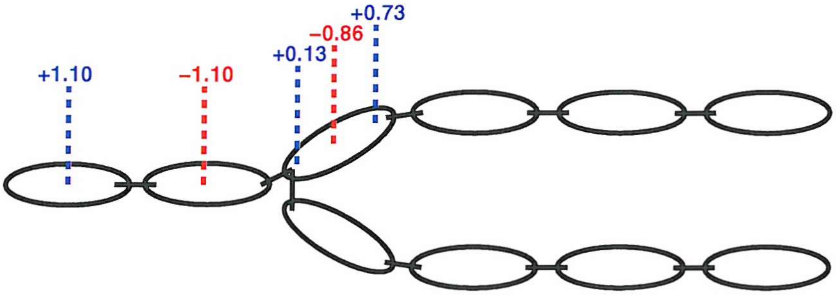
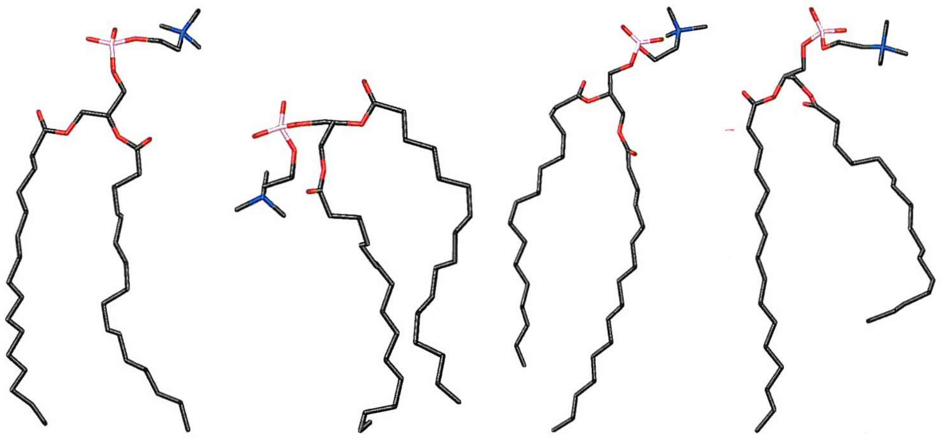


Figure 4.19: Original charge distribution, charges on the glycerol were originally placed at the centre of mass and 2Å apart from the centre of mass.

tensors of a cylindrically symmetric molecule and that of a non symmetric molecule cannot be compared. It is therefore impossible to optimize the charges in our reduced representation to reproduce the quadrupoles obtained from an all-atom model. Only the dipole moments were therefore used to optimize the charges.

The charges were optimized using four DPPC lipid structures. They are illustrated in Figure 4.20. Our reduced representation mimics a DMPC lipid. Since DMPC and DPPC only differ by different hydrocarbon chains lengths and since our parameterization only considers the head group charges, it is sensible to use DPPC. The four DPPC structures were taken from an equilibrated hydrated bilayer containing 72 lipids.[115] By design, the Gay-Berne charges on the two glycerol moieties are identical and consequently the dipoles are identical as well. Hence the four lipids were chosen such that they had similar all-atom dipoles on the two glycerol moieties. For each lipid, an equivalent reduced representation structure was generated. Only the atoms in the head group-and-glycerol region were taken into account ; the hydrocarbon chains were not considered. The centre of mass of each group of atoms (choline - or nitrogen -, phosphatidyl - or phosphorus - glycerol 1 and glycerol 2) was calculated and assumed to be the centre of geometry of each corresponding Gay-Berne bead in our reduced representations. The principal axis of inertia was also calculated and Gay-Berne beads were given the orientation of this axis. Dipoles were calculated for glycerol 1 (μ_{GL1}), glycerol 2 (μ_{GL2}), the entire head group and glycerol region (μ_{DMPC}) and for the phosphatidyl-choline (μ_{NP}) groups, for both all-atom and reduced Gay-Berne representation. Charges on the reduced representation (Figure 4.21) were optimized to minimize the difference between the all-atom dipoles (aa) and Gay-Berne dipoles (gb). Several conditions were applied to our set of charges :



(a) Lipid A (b) Lipid B (c) Lipid C (d) Lipid D

Figure 4.20: The four DPPC lipids structure used to optimize charges.

- the total charge on the glycerol bead should be zero, and hence only two charges are needed to represent glycerol beads : q_1 and q_2 as represented in Figure 4.21,
- the charges on the nitrogen and phosphorus beads should be equal and opposite in sign, hence only one charge is needed to represent nitrogen and phosphorus beads, $q_N = -q_P$.

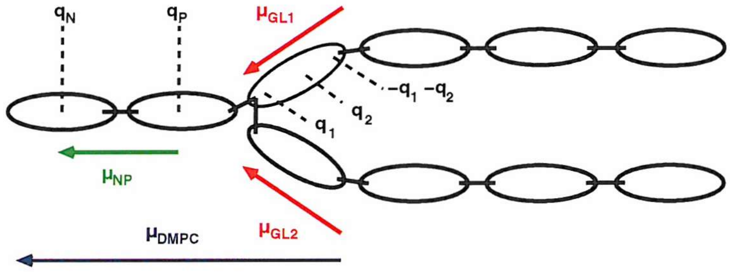


Figure 4.21: Charge and dipole description for the charge optimization process.

The first charges to be optimized are q_N and q_P ; they are manually modified until the all-atom and Gay-Berne phosphatidyl-choline dipole (μ_{NP}) become equivalent. The results are summarised in Table 4.6.

Table 4.6: q_N and q_P optimization with various lipids

	A (aa)[a]	A (gb)[b]	B (aa)[a]	B (gb)[b]	C (aa)[a]	C (gb)[b]	D (aa)[a]	D (gb)[b]
q_N	n.a.	1.100	n.a.	0.960	n.a.	0.920	n.a.	1.050
q_P	n.a.	-1.100	n.a.	-0.960	n.a.	-0.920	n.a.	-1.050
μ_{NP}[c]	23.967	23.971	21.477	21.422	21.778	21.699	23.995	24.089

[a] aa refers to all atom.
[b] gb refers to Gay-Berne and therefore to our reduced representation.
[c] dipoles in Debye.

The charges required to reproduce the all-atom dipoles are similar enabling the use of a simple arithmetic average. The average charge on the nitrogen and phosphorus particles is $q_N = -q_P = 1.0075$ which can be rounded to 1.0.

These charges are then fixed and the remaining charges q_1 and q_2 are optimized. Several error functions were examined. In the end. it was decided not to include the overall head group dipole μ_{DMPC} in the function used to optimize q_1 and q_2, as it made the choice of a single set of charges (q_1, q_2) too difficult. Indeed a small change in the nitrogen or phosphorus positions has a large influence on the head group dipole. The error function $\xi_1(q_1, q_2)$ (Equation 4.14), which only considers the two glycerol dipoles, makes the charge determination straightforward. We can justify the use of this simple function on the grounds that q_N and q_P have already been optimized.

$$\xi_1(q_1,q_2) = \left({\mu_{GL1}}^{\mathrm{gb}}(q_1,q_2) - {\mu_{GL1}}^{\mathrm{aa}}(q_1,q_2)\right)^2 + \left({\mu_{GL2}}^{\mathrm{gb}}(q_1,q_2) - {\mu_{GL2}}^{\mathrm{aa}}(q_1,q_2)\right)^2 \tag{4.14}$$

where, ${\mu_{GL1}}^{\mathrm{gb}}(q_1, q_2)$ is the dipole on the first glycerol bead in the Gay-Berne representation, and ${\mu_{GL1}}^{\mathrm{aa}}(q_1, q_2)$ is the dipole on the atomistic representation for the first glycerol group; ${\mu_{GL2}}^{\mathrm{gb}}(q_1, q_2)$ and ${\mu_{GL2}}^{\mathrm{aa}}(q_1, q_2)$ correspond to the second glycerol group. This error function is plotted in Figure 4.22, it corresponds to the sum of the $\xi_1(q_1, q_2)$ error functions of lipids A, B, C and D. The individual error functions of each lipid present the same shape. For all error functions (all lipids, lipid A, B, C and D), the same trend is observed, the minima are located in a valley along a straight line. In each case, the direction of this line remains the same, the overall error function hence does not help to find a unique solution for (q_1, q_2). To find a unique solution, another constraint needs to be added. The CHARMM27 charges for the glycerol moiety (Figure 4.23) can be used to link q_1 and q_2, by constraining

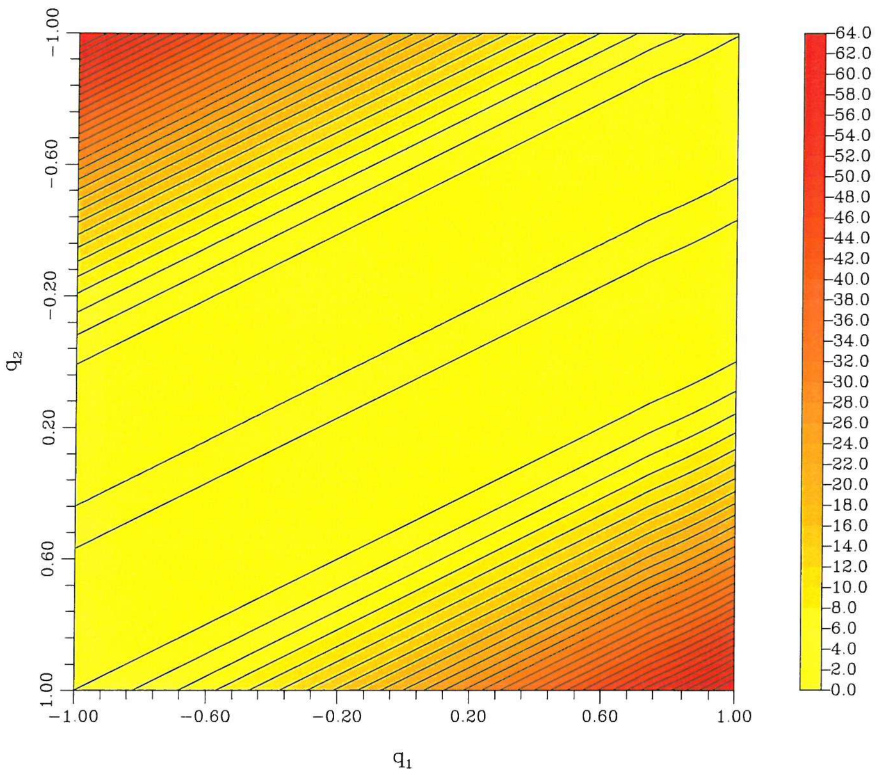


Figure 4.22: 2D contour of the ξ_1 error function as a function of q_1 and q_2.

q_2 (Equation 4.15) and keeping $q_1 > 0$, the original charge proportions assigned in Figure 4.23 can be reproduced.

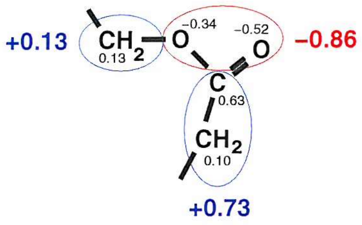


Figure 4.23: Charges initially used on the glycerol.

$$q_2 = -6.615 \times q_1 \tag{4.15}$$

Applying the constraint determined by equation 4.15 is effectively equivalent to tracing a line $q_2 = f(q_1)$, that will cross each valley (minima region) at one point, defining two overall minima. A new error function is defined which depends only on q_1 (Equation 4.16). This error function was examined for values of q_1 varying from -0.30 to

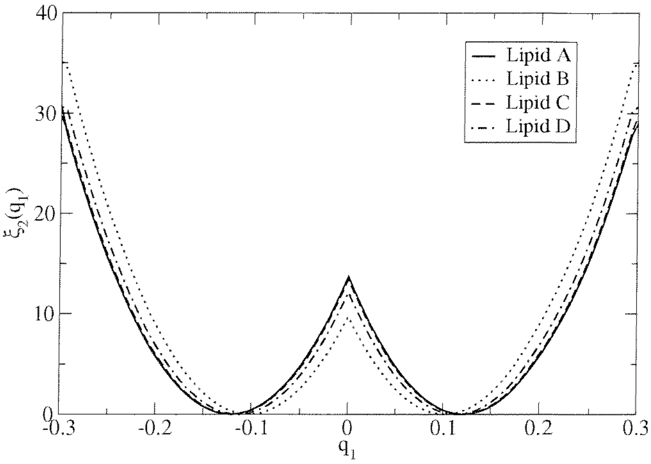


Figure 4.24: Error function $\xi_2(q_1)$, q_2 is determined by Equation 4.15.

0.30.

$$\xi_2(q_1) = \left(\mu_{GL1}{}^{\mathrm{gb}}(q_1) - \mu_{GL1}{}^{\mathrm{aa}}(q_1)\right)^2 + \left(\mu_{GL2}{}^{\mathrm{gb}}(q_1) - \mu_{GL2}{}^{\mathrm{aa}}(q_1)\right)^2 \tag{4.16}$$

This error function is plotted in Figure 4.24. Each lipid possess two minima, one for $q_1 < 0$ and one for $q_1 > 0$. The minima for the $q_1 > 0$ region are displayed in Table 4.7 as well as the average over the four lipids.

Table 4.7: Optimum charges found with $\xi_2(q_1)$.

	A (gb)[b]	B (gb)[b]	C (gb)[b]	D (gb)[b]	Average
q_1	0.121	0.102	0.120	0.114	**0.114**
q_2	-0.800	-0.675	-0.794	-0.754	**-0.756**
$-q_1 - q_2$	0.679	0.573	0.674	0.640	**0.642**

[b] gb refers to Gay-Berne and therefore to our reduced representation.

As the minima found for each lipid are close, a simple arithmetic average can be used. The average q_1 charge is 0.114, which implies a value of q_2 charge of -0.756. Because of the condition applied with Equation 4.15 this set is logically consistent with the first set of charges defined by eye. After rounding, the final set of charges is $q_N = 1.0$, $q_P = -1.0$, $q_1 = 0.11$, $q_2 = -0.76$ and $-q_1 - q_2 = 0.65$.

4.3 Water Model

4.3.1 Introduction

The presence of water plays a crucial role in the simulation of biophysical systems, especially in membranes where it is greatly contributing to the interface properties.[116,117] Despite its small size, water cannot be handled with transferable force fields, mostly because of its high density requiring a large number of molecules to represent realistically the system of choice and of the presence of hydrogen bonds. Therefore water simulation requires a specific force field. Most of these force fields use pairwise potentials without three-body terms or polarisation effects. They can be classified into three families. In the simple interaction site models the molecule is maintained rigid and Lennard-Jones and Coulombic interactions are calculated. Flexible models allow internal changes in the conformation of the molecule. Finally some models include polarisation and many body effects. Water models using rigid multiple interaction sites with partial charges can be represented by the three-site TIP3P,[118] SPC[119] and SPC/E,[120] the four-site TIP4P,[121] and the five-site ST2[122] and TIP5P[123] models. These models have been widely studied, giving similar structural and energetic properties but slightly different dielectric and dynamic properties.[124] Most of these models use an intermolecular potential consisting of coulombic interactions and a Lennard-Jones 12-6 interaction, the interaction between molecules m and n can be written as

$$U_{mn} = \sum_{i}^{onm} \sum_{j}^{onn} \frac{q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{C}{r_{OO}^{6}}. \tag{4.17}$$

Parameters q_i, q_j, A and C are optimized to yield reasonable structural and energetic results for the liquid phase water. There are also flexible multisite models that describe better the dynamic properties of water but these models are computationally more expensive. Given the choice of water models available, we decided to use a simplified model, to extend the simplification spirit of the membrane model to solvation. We therefore adopted the Soft Sticky Dipole (SSD) model of water.

4.3.2 The Soft Sticky Dipole Model of Water

The Soft Sticky Dipole model of water (SSD) has been recently introduced.[125] The SSD model is a modified soft-sphere version of the hard-sphere sticky dipole model of Bratko, Blum and others[126,127] and consists of a Lennard-Jones sphere embedded with a point dipole and a tetrahedral sticky potential. The SSD model yields thermodynamic, structural, and dielectric properties in agreement with experimental data. The dielectric constant of the SSD water has been determined to be 81 at 298 K,[125] in excellent agreement with experiment. Being a one site model, the SSD model is also computationally less expensive than multisite models. The dynamic properties of the SSD model have been studied by Chandra *et al.*[124] and have proved to be in good agreement with experimental results, and appear to be better than the TIP3P and SPC/E models in most cases.

Pure Water Potential

The water molecule of the SSD model is treated as a Lennard-Jones sphere with an embedded point dipole plus a tetrahedral "sticky" potential, all situated at the centre of mass (M) located on the H-O-H bisector, 0.0654 Å from the oxygen atom towards the hydrogens. The SSD water molecule is developed using the geometry of the TIP3P model, where distance between oxygen atom and hydrogen is 0.9572 Å and the H-O-H angle is 104.52 °. In SSD the centre of mass is the only interaction centre of the model and the explicit coordinates of the three atoms are never used in the calculation of the intermolecular potential energy. The coordinates can be derived from the position of the centre of mass and from the orientation of the dipole moment and of the molecular plane. The total interaction energy between two water molecules i and j is therefore

$$\mathcal{V}_{ij} = \mathcal{V}_{ij}^{\mathrm{LJ}}(r_{ij}) + \mathcal{V}_{ij}^{\mathrm{dp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) + \mathcal{V}_{ij}^{\mathrm{sp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) \tag{4.18}$$

where r_{ij} is the distance and $\mathbf{r}_{ij}$ the separation vector between the molecular centres, and Ω is the orientation of a water molecule defined by both the orientation of the dipole moment and of the molecular plane.

The first term in Equation 4.18 is the Lennard-Jones term.

$$\mathcal{V}_{ij}^{\mathrm{LJ}}(r_{ij}) = 4\epsilon_w \left[\left(\frac{\sigma_w}{r_{ij}} \right)^{12} - \left(\frac{\sigma_w}{r_{ij}} \right)^{6} \right] \tag{4.19}$$

The parameters used for the LJ interaction - well depth parameter ϵ_w and Lennard-Jones diameter σ_w - are shown in Table 4.8. The distance parameter σ_w is different from that used in TIP3P water (where $\sigma_w = 3.1506$ Å). This difference comes from the fact that in the SSD model the centre of mass M is the interaction site whereas it is the oxygen in TIP3P.

The second term is the point dipole-point dipole potential

$$\mathcal{V}_{ij}^{\mathrm{dp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = \frac{\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j}{r_{ij}^3} - \frac{3(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})}{r_{ij}^5} \tag{4.20}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are dipole moment vectors of water molecules i and j with an amplitude of 2.35 D.

The third term in Equation 4.18 is the tetrahedral sticky potential

$$\mathcal{V}_{ij}^{\mathrm{sp}}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = \frac{\mathcal{V}^\circ}{2} \left[s(r_{ij}) w_{ij}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) + s'(r_{ij}) w_{ij}^x(\theta_{ij}, \Omega_i, \Omega_j) \right] \tag{4.21}$$

where $\mathcal{V}^\circ = 3.7284$ kcal mol^{-1} determines the strength of the sticky potential. The function $w_{ij}(\mathbf{r}_{ij})$ which satisfies the tetrahedral coordination in the first shell is given by

$$w_{ij}(\mathbf{r}_{ij}) = \sin\theta_{ij}\sin 2\theta_{ij}\cos 2\phi_{ij} + \sin\theta_{ji}\sin 2\theta_{ji}\cos 2\phi_{ji} \tag{4.22}$$

in which (θ_{ij}, ϕ_{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 term $w_{ij}^x(\theta_{ij}, \Omega_i, \Omega_j)$ in Equation 4.21 is an empirical correction term and is given by

$$w_{ij}^x(\theta_{ij}, \Omega_i, \Omega_j) = (\cos\theta_{ij} - 0.6)^2 (\cos\theta_{ij} + 0.8)^2 + (\cos\theta_{ji} - 0.6)^2 (\cos\theta_{ji} + 0.8)^2 - 2w^\circ \tag{4.23}$$

where $w^\circ = 0.07715$. This additional function is necessary because $w_{ij}(\mathbf{r}_{ij}, \Omega_i, \Omega_j)$ vanishes at angles $\theta_{ij} = 0^\circ$ and $\theta_{ij} = 180^\circ$. Finally, $s(r_{ij})$ and $s'(r_{ij})$ are modulating functions given by

$$\begin{array}{lll} s(r_{ij}) = 1 & \text{if} & r_{ij} < r_L \\ s(r_{ij}) = \frac{(r_U - r_{ij})^2 (r_u + 2r_{ij} - 3r_L)}{(r_U - r_L)^3} & \text{if} & r_L \leqslant r_{ij} < r_U \\ s(r_{ij}) = 0 & \text{if} & r_{ij} \geqslant r_U \end{array} \tag{4.24}$$

where $r_L = 2.75$ Å, $r_U = 3.35$ Å for $s(r_{ij})$, and $r_U = 4.0$ Å for $s'(r_{ij})$.

Table 4.8: Parameters for the Ion-Water Potential

particle type	σ / Å	ϵ / kcal mol^{-1}	μ / D	Q_{xx} / b^a	Q_{yy} / b^a	Q_{zz} / b^a
water(M)	3.051	0.152	2.35	-1.682	1.762	-0.08
Na$^+$	2,579	0.118	-	-	-	-
Cl$^-$	4,445	0.100	-	-	-	-

[a] Calculated for the TIP3P monomer, Quadrupole moment unit $b = 10^{-26}$ esu cm^2.

Ion-Water Potential

The sticky potential of the SSD is present to take into account hydrogen bonds within the solvent only. To improve the generality of SSD Liu and Ichiye[125] have introduced a hybrid method in which ion-water interactions are modeled by ion-dipole plus ion-quadrupole interactions. The ion-water interaction is then given by

$$\mathcal{V}^{\mathrm{iw}}(r, \Omega) = \mathcal{V}^{\mathrm{LJ}}(r) + \mathcal{V}^{\mathrm{md}}(r, \Omega_w) + \mathcal{V}^{\mathrm{mq}}(r, \Omega_w) \tag{4.25}$$

in which the first term $\mathcal{V}^{\mathrm{LJ}}(r)$ is the Lennard-Jones potential (cf. Equation 4.19) between the ion and the centre of mass of the water molecule. Index i corresponds to ion and w to water. The ion-water cross parameters for energy are determined as follows :

$$\epsilon_{\mathrm{iw}} = \sqrt{\epsilon_{\mathrm{i}} \epsilon_{\mathrm{w}}} \tag{4.26}$$

and

$$\sigma_{\mathrm{iw}} = \frac{\sigma_{\mathrm{i}} + \sigma_{\mathrm{w}}}{2} \tag{4.27}$$

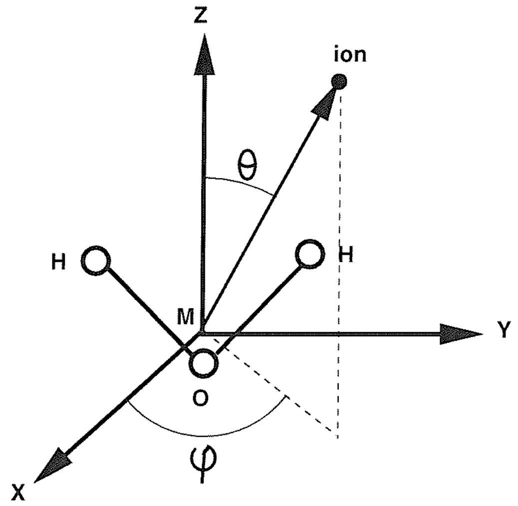


Figure 4.25: Schematic representation of the coordinate system used to describe the relative position and orientation of a water molecule with respect to an ion. The origin is located at the centre of mass of the water molecule.

The second term corresponds to the monopole-dipole interaction between the ion and the water molecule

$$\mathcal{V}^{\mathrm{md}}(r,\Omega_w)=\frac{q\mu\mathrm{cos}\theta}{r^2} \tag{4.28}$$

where q is the charge of the ion, μ is the dipole moment of the water molecule and θ is the angle formed by the bisector of the water molecule and the line joining the ion and the centre of mass of the water molecule, as shown in Figure 4.25. The last term is the monopole-quadrupole interaction,

$$\mathcal{V}^{\mathrm{mq}}(r,\Omega_w)=-\frac{q}{6r^3}\cdot\left[\mathcal{Q}_{xx}\left(3\mathrm{sin}^2\theta\mathrm{cos}^2\phi-1\right)+\mathcal{Q}_{yy}\left(3\mathrm{sin}^2\theta\mathrm{sin}^2\phi-1\right)+\mathcal{Q}_{zz}\left(3\mathrm{cos}^2\phi-1\right)\right] \tag{4.29}$$

where θ and ϕ are the angles shown in Figure 4.25 and $\mathcal{Q}_{xx}$, $\mathcal{Q}_{yy}$ and $\mathcal{Q}_{zz}$ are calculated by the centre of mass point quadrupole expansion of the charge distribution of the TIP3P monomer given the orientation shown in Figure 4.25. The quadrupole and dipole moments are given in Table 4.8.

4.4 Lipid-Water Interactions

4.4.1 Lipid-Water van der Waals Interactions

The interaction between a Gay-Berne fluid and a benzene molecule has already been described by Palke *et al.*[26] They resolved the positions of the four CH_2 group on the Gay-Berne particle axis (b), and calculated the solute - Gay-Berne interaction with a Lennard-Jones potential. In our study, to extend the simplification spirit used for modelling lipid-lipid interactions, we decided to model water-lipid van der Waals interactions using a single-point potential. The potential used would have to calculate interactions between a Gay-Berne particle and a Lennard-Jones atom, between an ellipsoid and a sphere.

The Gay-Berne potential allows only simulations of single component systems, as it intrinsically describes the interaction between two identical molecules. Without modifications it cannot be used to model the water-lipid interactions.

Cleaver *et al.*[106] created a general potential able to represent interactions between dissimilar molecules: *the generalized Gay-Berne potential.* They calculated the analytical overlap of two different uniaxial particles. In the following we consider two ellipsoidal particles i and j with cylindrical symmetry, with lengths l_i and l_j, and breadth d_i and d_j. The potential takes the general form:

$$U_{GGB}(\hat{u}_i, \hat{u}_j, \mathbf{r}) = 4\epsilon(\hat{u}_i, \hat{u}_j, \hat{r}_{ij}) \left[\left(\frac{\sigma_\perp}{r_{ij} - \sigma(\hat{u}_i, \hat{u}_j, \hat{r}_{ij}) + \sigma_\perp} \right)^{12} - \left(\frac{\sigma_\perp}{r_{ij} - \sigma(\hat{u}_i, \hat{u}_j, \hat{r}_{ij}) + \sigma_\perp} \right)^{6} \right]. \tag{4.30}$$

The range parameter can be expressed as:

$$\sigma(\hat{u}_i, \hat{u}_j, \hat{r}_{ij}) = \sigma_0 \left[1 - \frac{\chi}{2} \left(\frac{(\alpha \hat{r}_{ij}.\hat{u}_i + \alpha^{-1} \hat{r}_{ij}.\hat{u}_j)^2}{1 + \chi \hat{u}_i.\hat{u}_j} + \frac{(\alpha \hat{r}_{ij}.\hat{u}_i - \alpha^{-1} \hat{r}_{ij}.\hat{u}_j)^2}{1 - \chi \hat{u}_i.\hat{u}_j} \right) \right]^{-\frac{1}{2}}, \tag{4.31}$$

where

$$\sigma_0 = \sqrt{{d_i}^2 + {d_j}^2}, \tag{4.32}$$

(b) assuming Gay-Berne particles mimic an RLJ4 molecule

$$\chi = \left(\frac{(l_i^2 - d_i^2)(l_j^2 - d_j^2)}{(l_j^2 + d_i^2)(l_i^2 + d_j^2)} \right)^{\frac{1}{2}}, \tag{4.33}$$

and

$$\alpha^2 = \left(\frac{(l_i^2 - d_i^2)(l_j^2 + d_i^2)}{(l_j^2 - d_j^2)(l_i^2 + d_j^2)} \right)^{\frac{1}{2}}. \tag{4.34}$$

The well depth takes the same general form as the Gay-Berne potential:

$$\epsilon_C(\hat{u}_i, \hat{u}_j, \hat{r}_{ij}) = \epsilon_0 \epsilon^{\nu}(\hat{u}_i, \hat{u}_j) \epsilon_2^{\mu}(\hat{u}_i, \hat{u}_j, \hat{r}_{ij}) \tag{4.35}$$

where $\epsilon(\hat{u}_i, \hat{u}_j)$ is the original well depth strength described in Equation 4.4, the other term was suggested to take the form

$$\epsilon_2(\hat{u}_i, \hat{u}_j, \hat{r}_{ij}) = 1 - \chi' \left[\frac{\alpha'^2(\hat{r}_{ij}.\hat{u}_i)^2 + \alpha'^{-2}(\hat{r}_{ij}.\hat{u}_j)^2 - 2\chi'(\hat{r}_{ij}.\hat{u}_i)(\hat{r}_{ij}.\hat{u}_j)(\hat{u}_i.\hat{u}_j)}{1 - \chi'^2(\hat{u}_i.\hat{u}_j)^2} \right]. \tag{4.36}$$

by reference to the shape parameter. The parameters ϵ_0, μ, ν, χ' and α' must be related to the system of interest; this task is less straightforward than for the shape parameter, and requires the use of a fitting procedure, such as that used by Luckhurst and Simmonds,[59] and thus each system has its own exclusive set of parameters.

For the lipid-water interactions we are only interested in the shape taken by the potential for rod-sphere interactions. Cleaver *et al.*[106] defined a rod-sphere potential and derived the well depth and shape expression. If one of the particles is made spherical, for example $d_i = l_i = d$, then both χ and α goes to zero, nevertheless the range parameter remains finite in this limit and tends smoothly to

$$\sigma(\hat{u}_j, \hat{r}_{ij}) = \sigma_0 \left[1 - \chi \alpha^{-2} (\hat{r}_{ij}.\hat{u}_j)^2\right]^{-\frac{1}{2}} \tag{4.37}$$

where

$$\frac{\chi}{\alpha^2} = \frac{l_j^2 - d_j^2}{l_j^2 + d^2}. \tag{4.38}$$

Cleaver *et al.* also defined a well depth expression which is dependent upon orientation of molecule j with regard to the intermolecular vector, intermolecular distance, χ', α' and μ. The well depth was defined as:

$$\epsilon(\hat{u}_j, \hat{r}_{ij}) = \epsilon_0 \left[1 - \chi' \alpha'^2 (\hat{u}_j.\hat{r}_{ij})^2\right]^{\mu}, \tag{4.39}$$

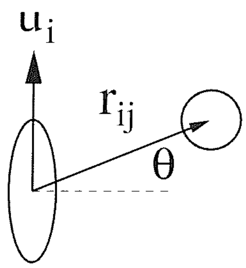


Figure 4.26: The relative orientations adopted for the generalized Gay-Berne parameters optimization.

where

$$\frac{\chi'}{\alpha'^2} = 1 - \left(\frac{\epsilon_{\|}}{\epsilon_{\perp}}\right)^{\frac{1}{\mu}}, \tag{4.40}$$

and $\epsilon_{\perp} = \epsilon_0$. In this particular case ν is zero and μ and $\chi'\alpha'^2$ may be found through a fit with the full potential. Forces and torques for the generalized Gay-Berne potential have been derived and are presented in Appendix A.

4.4.2 Parameterization

To optimize the parameters σ_0, ϵ_0, $\chi\alpha^{-2}$ and $\chi'\alpha'^{-2}$, the same protocol used to optimize the Gay-Berne parameters in section 4.2.4 was used. The generalized Gay-Berne interaction between a sphere and an ellipsoid is compared to the Lennard-Jones interaction between water and butane (RLJ4 representation). For each set of parameters - (σ_0,$\chi\alpha^{-2}$) for contact distance optimization, (ϵ_0,$\chi'\alpha'^{-2}$) for well depth optimization - an error function is calculated. This error function consists of the difference between the contact distance calculated with the generalized Gay-Berne representation and the Lennard-Jones representation, squared and summed over a range of relative orientations. These orientations, presented in Figure 4.26, consist in the Gay-Berne particle orientation vector $\hat{u}_i$ making an angle $\pi/2 - \theta$ with the intermolecular vector $\hat{r}_{ij}$, where θ is varied systematically between 0 and $\pi/2$.

$$\xi_{CD}(\sigma_0, \chi\alpha^{-2}) = \sum_{\theta} \left(\sigma^{\mathrm{LJ}}(\theta) - \sigma^{\mathrm{GGB}}(\sigma_0, \chi\alpha^{-2}, \theta)\right)^2 \tag{4.41}$$

$$\xi_{WD}(\epsilon_0, \chi'\alpha'^{-2}) = \sum_{\theta} \left(\epsilon^{\mathrm{LJ}}(\theta) - \epsilon^{\mathrm{GGB}}(\epsilon_0, \chi'\alpha'^{-2}, \theta)\right)^2 \tag{4.42}$$

The error functions are described in Equation 4.41 and 4.42, where σ refers to the contact distance and ϵ refers to the well depth. These error functions were used to optimize the values that were first estimated using approximations and which are summarised in Table 4.9. The parameter $\chi\alpha^{-2}$ was estimated using a value of d of 3.42 Å [c], a value of l_j/d_j of 2.0[d] and a value of l_j of 7.56 Å [e]. The parameter $\chi'\alpha'^{-2}$ was estimated using Gay-Berne parameter $\epsilon_{\|}/\epsilon_{\perp} = 3.0$ and $\mu = 2.0$. The parameter ϵ_0 was assumed to be a quarter of the Gay-Berne potential ϵ_0, as the number of interactions was four times smaller[f]. The contact distance error function (cf. Equation 4.41) is first calculated and plotted (Figure 4.27). The value of σ_0 and $\chi\alpha^{-2}$ vary, the values of ϵ_0 and $\chi'\alpha'^{-2}$ are the default values described above. The minimum is found for $\sigma_0 = 3.543$ Å and $\chi\alpha^{-2} = 0.586$. σ_0 was previously estimated to be 3.86 Å during the Gay-Berne parameters optimization, but a slightly higher value of 4.0 Å was chosen as it lead to better results. The estimated value for σ_0^{GGB} of 3.54 Å is lower and in the same range, in agreement with the smaller waist of the generalized Gay-Berne potential. The parameter $\chi\alpha^{-2}$ is close to the value originally assigned with only geometry considerations in mind. These values are then used in the well depth error function (Equation 4.42) calculation. The well depth error function is plotted in Figure 4.28. The well depth error function gives a minimum for $\epsilon_0 = 1.96\,\mathrm{kJ.mol^{-1}}$ and $\chi'\alpha'^{-2} = 0.401$. The optimal set of parameters is described in Table 4.9 together with the original set of parameters. Even though, the $\epsilon_0^{\mathrm{GGB}}$ value was roughly estimated as the quarter of the Gay-Berne potential ϵ_0, the value found through the optimization process is close to the estimate. The determination of ϵ_0 allows us to scale the generalized Gay-Berne potential with respect to the Gay-Berne potential, thus balancing the water-lipid and lipid-lipid interaction. This balance plays a key role in the interface structure. The optimized $\chi'\alpha'^{-2}$ differs from the original estimate, this difference can be explained by the fact that the original estimate only considered the energy anisotropy ($\epsilon_{\|}/\epsilon_{\perp}$) of the ellipsoid, and that no

---

[c] Obtained considering that $d = 2^{\frac{1}{6}}.\sigma_w$ where σ_w refers to the appropriate force field contact distance, here $\sigma_w = 3.051$ Å.

[d] Value used for the Gay-Berne potential.

[e] Contact distance of two Gay-Berne particles calculated in the end-to-end conformation.

[f] For the Gay-Berne interaction butane vs. butane interaction corresponded to 4 × 4 pairs while in the generalized Gay-Berne potential water vs. butane interaction corresponded to 4 × 1 pairs.

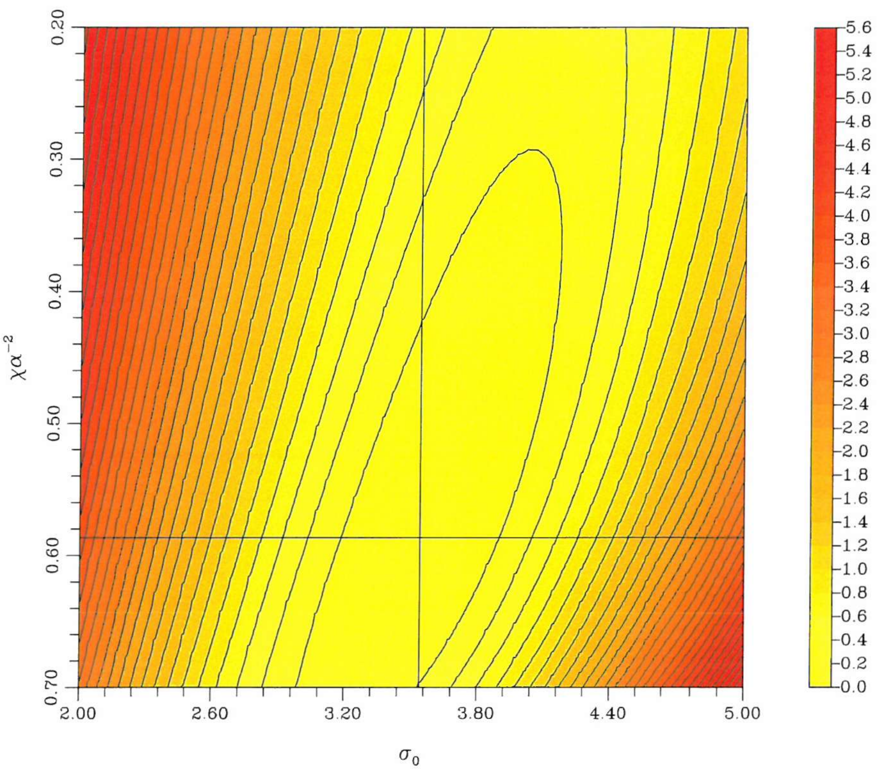


Figure 4.27: Contact distance error function. 2D contour.

energy parameter for water sphere was considered. The optimized value takes into account both the water and lipid energy parameters. The optimized $\chi\alpha^{-2}$ does not differ much from the first estimated value as the first estimation included both water and lipid geometry parameters. The contact distance of the generalized Gay-Berne potential (σ_0^{GGB}) plays an important role in the water-membrane interface structure. As the minimum of the contact distance error function is located in a flat valley (cf. Figure 4.27), several values around the minimum were investigated : 3.30, 3.35, 3.50 and 3.55 Å. The best bilayer structure was obtained with the smaller value of 3.30 Å. The main determinant for a better structure is the difference between the σ_0 and σ_0^{GGB}, rather than the absolute generalized Gay-Berne contact distance. Fixing σ_0 to be smaller than σ_0^{GGB} means a closer contact between water and lipid with respect to the lipid-lipid contact, thus affecting directly the interface structure.

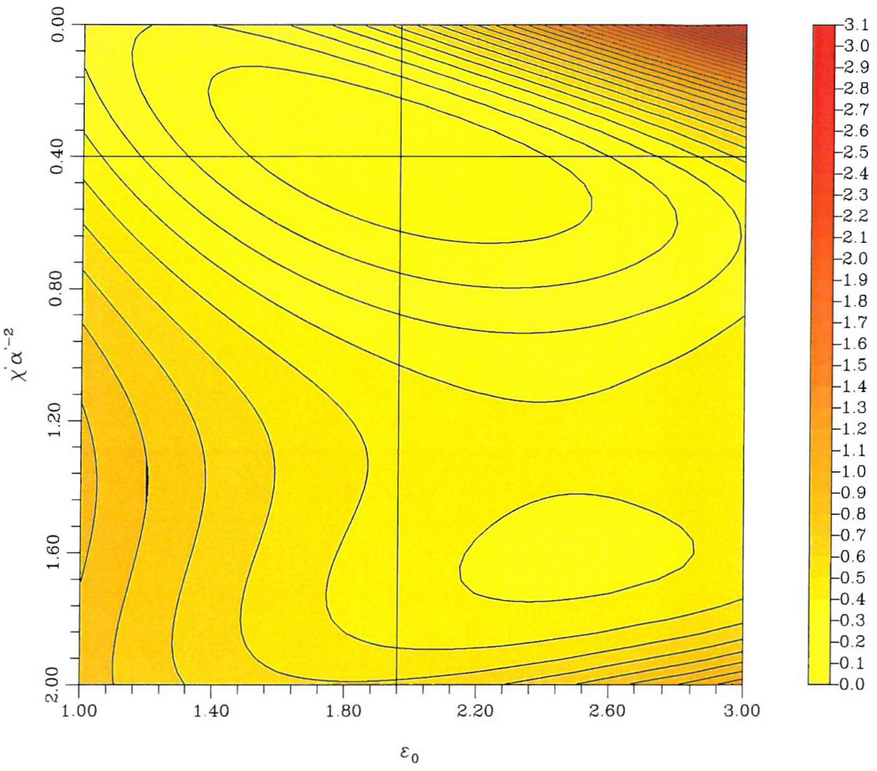


Figure 4.28: Well depth error function. 2D contour.

Table 4.9: Generalized Gay-Berne potential parameters

	σ_0	$\chi\alpha^{-2}$	ϵ_0	$\chi'\alpha'^{-2}$
simple parameterization	4.000	0.621	1.750	0.570
error function parameterization	3.543	0.586	1.960	0.401
final parameters	3.300	0.586	1.960	0.401

The final set of parameters is described in Table 4.9 along with the set determined through the simple parameterization method and the set determined through the use of an error function.

4.4.3 Electrostatic Interactions

Electrostatic interactions between the lipid head group and the water particles were simply taken into account by calculating point charge - dipole interactions. This calculation was performed using an Ewald summation. A simple and efficient way to increase the speed of electrostatic interactions calculation is to use Particle Mesh Ewald (PME)[128] instead of a simple Ewald summation. This method is based on interpolation of the reciprocal space Ewald sum. While being easily implemented for point charge - point charge interactions, this method is more difficult to implement for point charge - dipole interactions and has not been implemented in *CHAMPAGNE*. The use of a different water model, or further work towards the implementation of the PME method would therefore result in a greater efficiency of the code.

4.5 Summary

In this chapter, the development and parameterization of the Gay-Berne representation of a biological membrane have been described. The model is built upon the hydrocarbon region representation of Whitehead *et al.*,[2] and incorporates head group and solvent descriptions. Simple parameterization techniques, based on identification of simple configurations, have been presented. A more elaborate parameterization technique was introduced, which considers a larger set of configurations. This technique proved to give a better fit with an equivalent Lennard-Jones model (RLJ4) compared to the traditional parameters. These parameters were subsequently refined using short molecular dynamics simulations, and the results obtained with the optimum parameter set are presented in the next chapter.

Chapter 5

Model Validation

Extensive force field parameterization has been performed, and several parameter sets have been examined; this chapter only presents the results obtained for the optimum set. The parameters adopted in this section are $\sigma_0 = 4.0\,\text{Å}$, $\epsilon_0 = 6.0\,\text{kJ.mol}^{-1}$, $\sigma_{\|}/\sigma_{\perp} = 1.94$, $\epsilon_{\perp}/\epsilon_{\|} = 4.40$, $\mu = 2$ and $\nu = 1$ for the Gay-Berne potential. The charges on the head group are $q_N = 1.0$, $q_P = -1.0$, $q_1 = 0.11$, $q_2 = -0.76$ and $-q_1 - q_2 = 0.65$. The generalized Gay-Berne potential parameters for the water-lipid van der Waals interactions are $\sigma_0^{GGB} = 3.30$, $\epsilon_0^{GGB} = 1.96$, $\chi\alpha^{-2} = 0.586$ and $\chi'\alpha'^{-2} = 0.401$. The determination procedure for these parameters is explained in detail in Chapter 4.

5.1 Simulation Protocol

Molecular dynamics simulations using the microcanonical ensemble have been performed on a system consisting of 720 Gay-Berne particles, corresponding to two leaflets of a model bilayer each containing 36 lipid molecules, and 2094 water molecules. Periodic boundary conditions were applied in all three dimensions. A scaled time step, $\Delta t^* = 0.005$ was used ($\Delta t^* = \Delta t\sqrt{\epsilon_0/m\sigma_0^2}$) where m is the mass of the particles. This corresponds to a real time step of approximately 6 fs. For simplicity the mass of all particles have been set to unity in reduced units, corresponding to the mass of a particle in the hydrocarbon region (56 g.mol^{-1}). All inertia tensors have been set to 0.56, corresponding to the value obtained for butane (RLJ4 representation). The simulations were performed using *CHAMPAGNE*, a heavily modified version of a code written by Bill Palke.[26] Simulations were performed at the reduced temperatures ($T^* = k_B T/\epsilon_0$) 0.385, 0.392, 0.399, 0.406, 0.413, 0.420, 0.441 and 0.462 corresponding respectively to 5, 10, 15, 20, 25, 30, 45 and 60 °C. The simulations were performed in a tetragonal box of dimensions 46.32 Å along the plane bilayer, giving a head group surface area of 59.6 Å^2.lipid^{-1}, in accord with experimental data for the L_α phase of DMPC.[129] For the van der Waals interactions a 12 Å cutoff was employed and it was switched smoothly to zero over the region between 11.4 and 12 Å. The leap-frog algorithm for linear particles was employed to solve the equation of motion of Gay-Berne particles. The equations of motion of water particles were solved using the quaternion leap-frog algorithm. Electrostatic interactions were calculated using the Ewald sum. The convergence parameter α was set to 0.80, the number of reciprocal vectors k was 6, 10 and 6 (adapted to the box dimensions), and the real space part of the summation was truncated at 12 Å. The Verlet neighbour list was used to calculate van der Waals interactions and the real space portion of the Ewald sum. The Verlet cutoff was kept to 13.2 Å. The list was updated automatically by monitoring particles displacement.[130, 131] The simulations involved the generation of an initial structure at 5 °C, its hydration and the equilibration of this structure at all temperatures.

5.1.1 Initial Structure Generation

The initial configuration of the system was obtained by generating a hexagonally close-packed array of "Gay-Berne" lipid molecules at 5 °C. Each layer contains 36 lipids. The plane of the bilayer is therefore described by a grid of 6 × 6, each sub cell of the grid measuring $CELLX/6 \times CELLZ/6$. The first lipid coordinates are generated (lipid 1 in Figure 5.1). The lipid is placed in the sub cell of coordinates (1,1) and tilted such that the distances with future neighbour lipids are maximized. In Figure 5.2 the lipid drawn in blue is tilted by an angle θ such that the distances

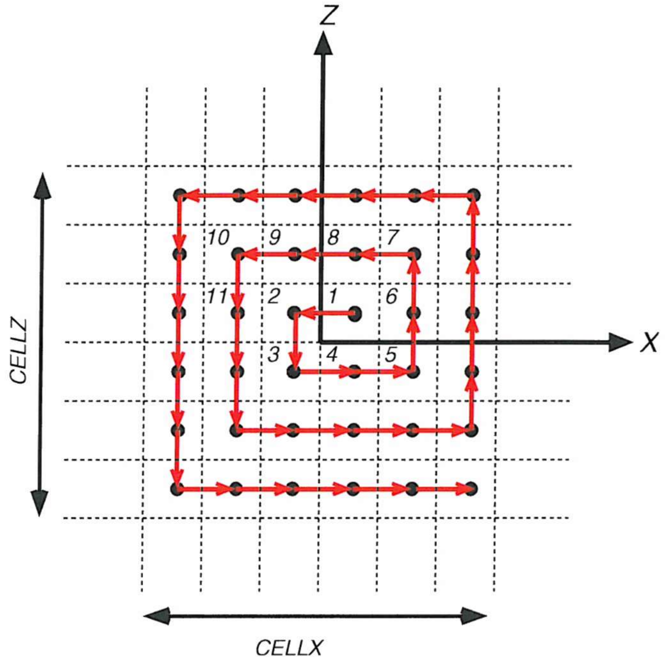


Figure 5.1: Lattice used for the initial membrane conformation generation.

a and b with its neighbours (in red) are maximized. This ensure there are no bad contacts when starting a simulation from scratch. The remaining lipid coordinates are obtained by translation of lipid 1 coordinates to the other sub-cells. The same process is repeated for the other layer. Along the y direction, lipids are placed such that the

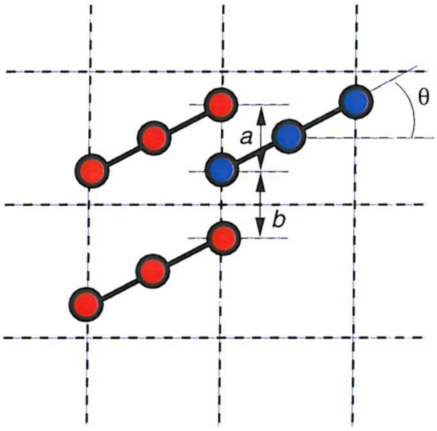


Figure 5.2: Placement of the lipid within each sub cell. Lipids are seen along the y direction in their initial configuration, hence the three circles.

hydrocarbon region thickness is 22.5 Å. This separation is based on estimates of the width of a $(CH_2)_{12}$ bilayer, derived from X-ray and neutron diffraction for DMPC and DPPC in the L_α and P'_β phases.[132–134] Once this initial structure has been generated,

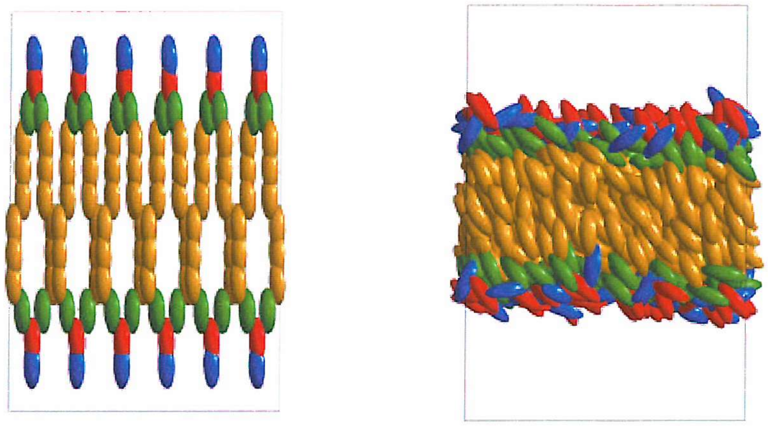

(a) (b)

Figure 5.3: membrane structure after initial 10 steps NVE simulation (a) and after 100 steps NVT simulation (b). Tail particles are drawn in orange, glycerol particles in green, phosphorus particles in red and nitrogen particles in blue.

a short NVE simulation is performed. The resulting structure, illustrated in Figure 5.3 (a), is then equilibrated. The equilibration run consisted of a 100 steps NVT simulation in which the simulation temperature must be stable to within 0.01. If this criterion is not met, the velocities are again rescaled and the equilibration phase continues until such time as a stable temperature is maintained. An obvious way to alter the temperature of the system is to scale the velocities.[32] If the temperature at time t is $T(t)$ and the velocities are multiplied by a factor λ, then the associated temperature change can be calculated as follows:

$$\Delta T = \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3}\frac{m_i(\lambda v_i)^2}{Nk_B} - \frac{1}{2}\sum_{i=1}^{N}\frac{2}{3}\frac{m_i {v_i}^2}{Nk_B} \tag{5.1}$$

$$\Delta T = (\lambda^2 - 1)T(t) \tag{5.2}$$

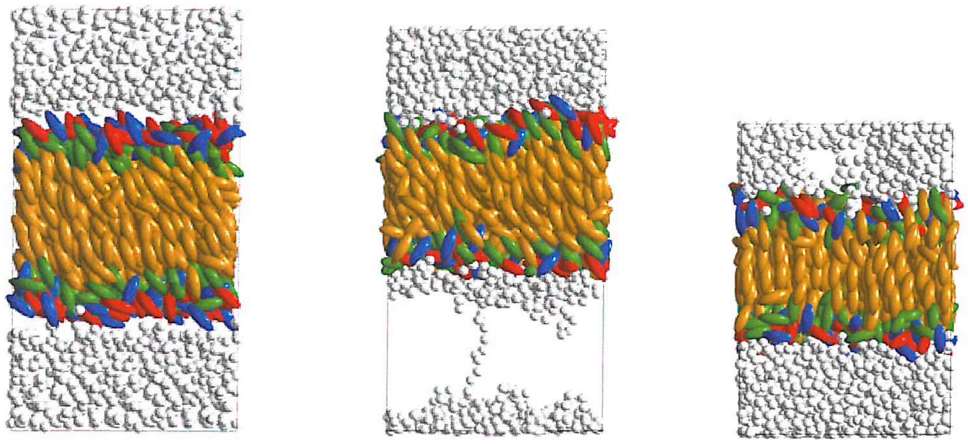
$$\lambda = \sqrt{\frac{T_{new}}{T(t)}} \tag{5.3}$$

The simplest way to control temperature is thus to multiply the velocities at each time step by the factor $\lambda = \sqrt{T_{req}/T_{curr}}$ where T_{curr} is the current temperature as calculated from the kinetic energy and T_{eq} is the desired temperature. More elaborate methods[33–36] exist to control the temperature. Since our production simulations were run in the microcanonical ensemble, and only the equilibration runs were run in the canonical ensemble, a crude scaling of the velocities was used. This equilibration phase was voluntarily kept short to avoid lipid bilayer organization in the absence of solvation. After this rapid equilibration phase the lipid bilayer density is higher, enabling more efficient hydration.

5.1.2 Hydration

The lipid bilayer is then hydrated. A pre-equilibrated box of water containing 256 molecules is replicated inside the lipid simulation box. The ideal system density could not be obtained from scratch; without the appropriate velocities, small distances between lipid and water particles lead to large forces that caused instability of the MD algorithm. When placing the water, the distances between water and lipids (d_{LW}) and the water-water (d_{WW}) distances were therefore checked to avoid close contacts that would lead to large forces. Placements were rejected for those water molecules for which $d_{LW} < 6.0\,\text{Å}$ or $d_{WW} < 2.8\,\text{Å}$. The initial configuration density is low, particularly at the lipid-water interface, as can be seen in Figure 5.4 (a). A structure of correct density is obtained through the succession of equilibration / hydration runs. The equilibration and hydration phases are as follows.

- Step 1: **Equilibration**. The structure is equilibrated in a 1000 steps simulation in which the simulation temperature must be stable to within 0.01. If this criterion is not met, the velocities are again rescaled and the equilibration phase continues until such time as a stable temperature is maintained.
- Step 2: **Hydration**. New water molecules are added and assigned random velocities with a Gaussian distribution. Distance checks are applied to newly inserted water molecules in the fashion outlined previously. The simulation box height is modified and water particles outside the new box are removed such that the box would contain 2094 (29.1 water per lipid) water molecules after the insertion procedure. Velocities of the lipid particles, old water particles and newly inserted particles are scaled independently to



(a) (b) (c)

Figure 5.4: Bilayer snapshots after the initial hydration (a), after the first hydration and equilibration sequence (b) and after the second sequence (c). Tail particles are drawn in orange, glycerol particles in green, phosphorus particles in red and nitrogen particles in blue.

reach the desired temperature.

During the equilibration phase (step 1), the interface water density increases, as water moves to solvate the lipids, while vacuum pockets form in the bulk water phase (cf. Figure 5.4 (b-c)). Water is inserted in these vacuum pockets in the hydration phases (step 2). The process of equilibration / hydration is repeated until a correct density is reached. This process was repeated five times.

A hybrid molecular dynamics-Monte Carlo algorithm was then applied to obtain the correct density. A normal molecular dynamics simulation was carried with a Monte Carlo move performed every 250 steps. The move involved modifying the box height and scaling all particle coordinates along this direction accordingly. A Boltzmann test was then performed. Velocities are rescaled after each Monte Carlo move to reproduce the desired temperature. The Monte Carlo moves ensure the right pressure and density are obtained. Monte Carlo moves were only performed along the bilayer normal, thus recreating a pseudo NP_NAT ensemble. Simulations with moves along all three axes have been performed; they did not show any significant drift of the surface area per lipid, and the surface area was therefore maintained constant. Two 50 000 steps simulations in the NP_NAT ensemble have been performed with an intermediate hydration (as described in step 2).

5.1.3 Equilibration

This structure was then equilibrated at all temperatures (5, 10, 15, 20, 25, 30, 45 and 60 °C). The equilibration consisted of two initial 100 time steps in which the simulation temperature must be stable to within 0.01 and a 500 000 time step simulation with a Monte Carlo move along the bilayer normal every 250 steps, and velocity rescaling after each move. The two 100 steps initial runs, ensure that each system's temperature is reached prior to the NP_NAT simulation. This protocol ensured the correct densities and temperatures for each simulation. At the end of this equilibration phase the dimension along the bilayer normal was 67.75 Å at 30 °C. This dimension varied between 64.40 and 70.79 Å across the entire temperature range, reflecting the change of density of the system with temperature.

5.1.4 Production

Eight production simulations of 1 200 000 time steps are then performed at temperatures 5, 10, 15, 20, 25, 30, 45 and 60 °C, with velocity rescaling every 250 steps. Each simulation was approximately 7.2 ns long and performed on a cluster of Linux PCs with 1400 MHz AMD Athlon processors. Coordinates from each of the simulations trajectories were stored for analysis every 500 time steps ($\approx$ 3 ps).

5.2 Simulation Results

Whereas a substantial amount of data is available on thermodynamic properties of lipid bilayers and biological membranes, the existing data for the mechanical and geometrical properties of membranes is more limited.[135]

5.2.1 Radial Distribution Functions

Radial distribution functions have been calculated for those Gay-Berne particles located in the hydrocarbon chains. Figure 5.5a shows the distribution functions generated from the simulations for temperatures ranging from 5 to 60 °C. Below 20 °C, the distribution clearly indicates the solid ordered nature of the hydrocarbon core of the membrane. Above 25 °C, the second peak (9.0 Å) is much less pronounced, while the third peak (12.5 Å) has almost disappeared and the fourth peak (17.0 Å) is no longer presen,t clearly showing the fluid nature of the membrane at this temperature. The examination of the distribution functions shows the transition from a highly ordered solid phase at low temperatures to a disordered liquid-crystalline phase at higher temperatures. This phase transition is only observed in the hydrocarbon region. The headgroup region does not show any loss of structure with increasing temperature. Radial distribution functions between identical Gay-Berne particles at 30 °C are presented in Figure 5.5b. The tail-tail function posses a short distance initial peak at 5.0 Å. Tail particles are uncharged, and their relative behaviour is therefore only governed by the Gay-Berne potential. The presence of identical negative charges on neighbour phosphorus particles creates a repulsion, the first peak in the phosphorus-phosphorus function is therefore slightly higher, located at around 6.0 Å. The nitrogen particles are located at the end of the lipid, they are therefore more mobile than the phosphorus particles, and the electrostatic repulsion has a greater effect on their distribution, the first peak of the nitrogen-nitrogen distribution is located at 9.0 Å. The charges on the glycerol particles have the opposite effect, the three point charges play an attractive role, the glycerol-glycerol function initial peak is shorter than for the tail-tail distribution, at just below 5.0 Å.

Distribution functions of Gay-Berne particles with the oxygen atom of the water particles are presented in Figure 5.5c. The reduced-representation RDFs of water with glycerol, nitrogen and phosphorus particles can be compared with atomistic RDFs of carbonyl oxygen, Choline methyl, and phosphate oxygen headgroup-water pairs respectively. Water-nitrogen and water-phosphorus RDFs have their initial peak located at 3.2 and 3.3 Å respectively. Those distances are much smaller than for the distributions between membrane particles. Indeed the σ_0 value for the generalized Gay-Berne potential (3.3 Å) is smaller than for the Gay-Berne potential (4.0 Å) thus implying shorter contact distances. Nitrogen RDFs show two hydration shells, in agreement with atomistic simulations for DPPC.[7] Our model mimics DMPC, but DMPC and DPPC have the same headgroup, and hence water-headgroup distribution functions should be similar. The first hydration shell is sharper and more defined in the reduced representation, whereas the second shell is less defined. The peak lo-

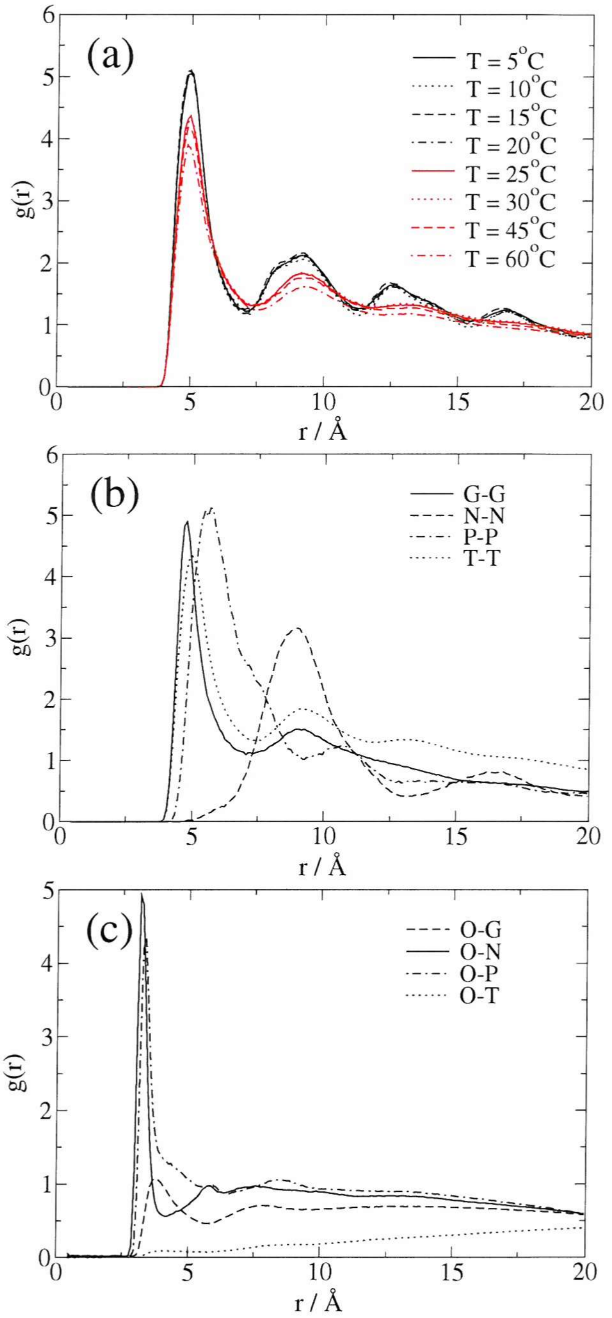
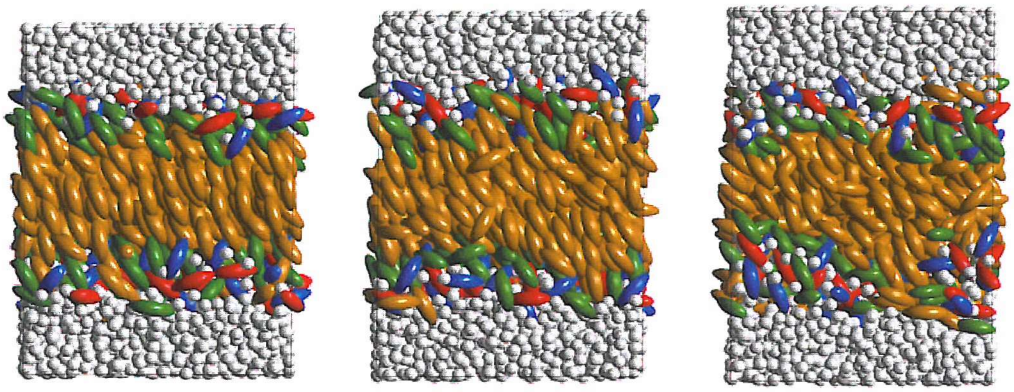


Figure 5.5: Radial distribution functions for (a) the tail Gay-Berne particles as a function of temperature; (b) for membrane glycerol (G), nitrogen (N), phosphorus (P) and tail particles (T) at 30 °C; (c) for water oxygen atom (O) against membrane glycerol, nitrogen, phosphorus and tail particles at 30 °C.

cations are surprisingly close, with the atomistic first and second hydration shells located at around 3.3 and 5.5 Å, and the reduced representation shells located at 3.2 and 5.9 Å. The phosphorus-water pair shows two hydration shells as well, the first shell is sharp and well defined while the second is hardly detectable but, being located at 5.9 Å, is in quite good agreement with the atomistic result located at around 5.5 Å. The reduced representation model is able to represent qualitatively the choline and phosphatidyl-water pair RDFs; the location of the first and second hydration shells is reproduced. Quantitatively; the intensity of the two first peaks is not in agreement with atomistic RDFs. The atomistic first and second hydration shells of the nitrogen-water pair contains 3 and 2 particles respectively, whereas the reduced representation contains 5 and 1 particles; suggesting that most water molecules are located in the first hydration shell, the second shell being hardly detectable. The atomistic and Gay-Berne representation phosphorus-water pairs both contain the same number of particles in the first shell (4.5), but show different hydration in the second shell with 3 water molecules for the atomistic model and only one for our model. These differences of hydration can explained by the larger size of Gay-Berne particles, when compared to Lennard-Jones spheres. The glycerol-oxygen pair first hydration shell is further apart (3.8 Å) in our model than in atomistic studies (3.1 Å). A second hydration shell can be noticed in our model (7.8 Å), which does not appear in atomistic RDFs. Quantitatively the reduced representation first hydration shell contains 1 molecule and the second 0.7, the total being comparable to the number of particle in the atomistic first hydration shell (1.5). Although the particles used in the reduced representation are larger than in atomistic representations, the hydration of the glycerol region is very similar to that observed in atomistic simulations. The agreement between the atomistic and reduced representation RDFs is globally correct, representing an achievement when considering the simplicity of the model used. Figure 5.6 presents structural snapshot images obtained from simulations at temperature of 5, 30 and 60 °C.

5.2.2 Electron Density Profile

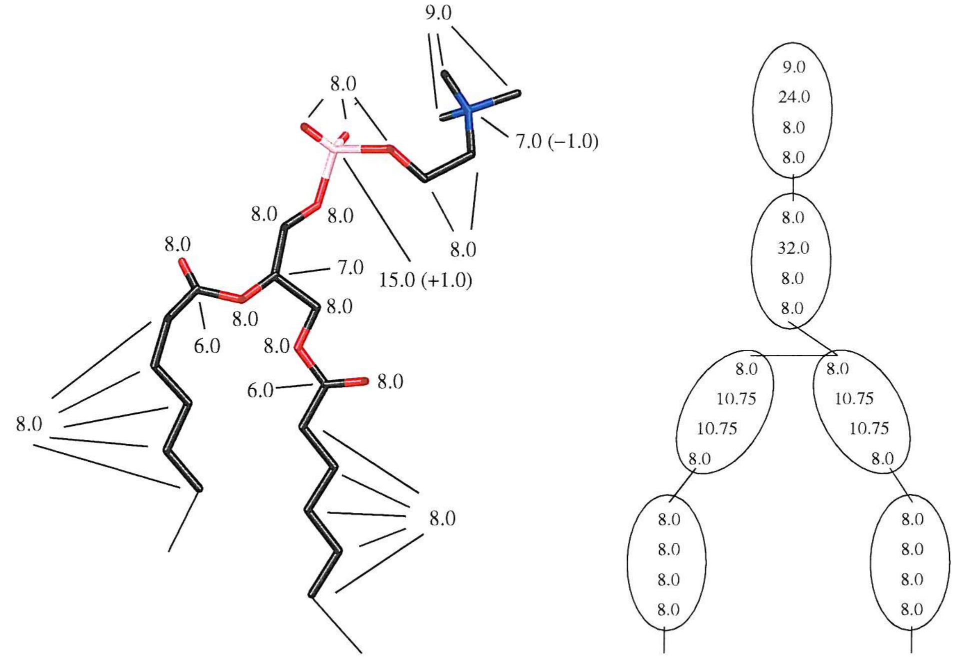
The organization of bilayers can be examined using a variety of techniques, including spectroscopic, X-ray and neutron diffraction techniques. Early diffraction work centred on determining the critical structural characteristics of bilayers : the width of the bilayer and the area per lipid molecule.[136] Early work by Levine *et al.*[137, 138] used X-ray diffraction data to calculate one-dimensional electron density profiles (EDP)



(a) T=5 °C (b) T=30 °C (c) T=60 °C

Figure 5.6: Structural snapshots.

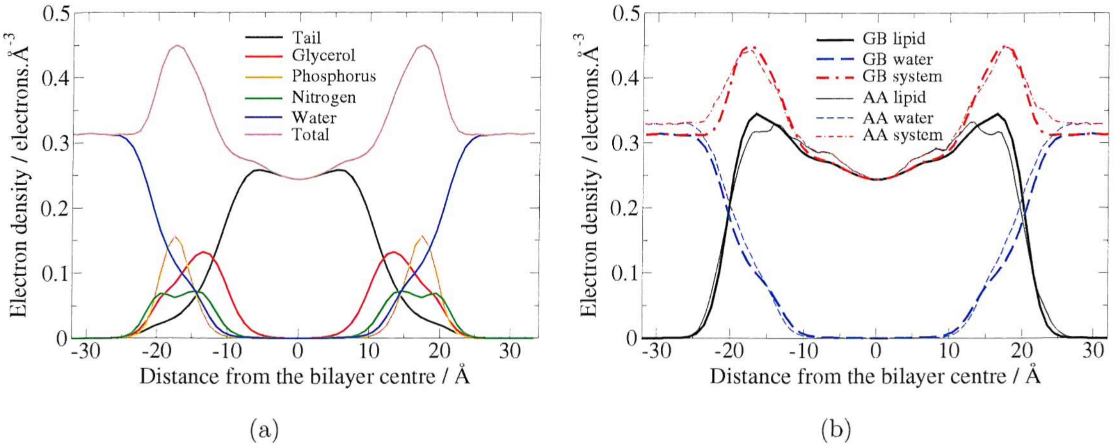
across bilayers of synthetic and naturally occurring phosphatidylcholines. These low resolution profiles gave quite accurate values for the separation distances of the phosphate groups. Higher resolution electron density profiles provide information on the organization of hydrated lipid bilayers, such as the variation of the bilayer thickness with the number of atoms in the lipid acyl chains[133, 139, 140] or the influence of double bonds in the lipid acyl chains on the area per lipid.[133, 141] Electron density profiles only provide a good measure for the location of the phosphate group. Information about the y-coordinates (along the bilayer normal) of other groups have been obtained using neutron diffraction. In atomistic simulations, electron density profile determination is quite straightforward, assuming atom centred electrons and the absence of electron reorganization, the number of electrons in each slab is counted and divided by the slab's volume. Computer simulations have the added advantage over experimental studies of being able to determine electron density profile contributions from the different groups, thereby allowing local organization to be studied. Electron density profiles were calculated from the reduced representation simulations. The reduced nature of our model meant electrons are not explicitly represented. One could count the number of electrons present in each ellipsoid and assign it to the centre of the particle, but for increased precision we decided to resolve their positions. Electrons were placed on four locations evenly spaced on the particle's axis. The electronic charge values were chosen to mimic as closely as possible the atomistic distribution,



(a) (b)

Figure 5.7: The electron distribution on the atomic representation (a), and the resolved electron distribution on the Gay-Berne representation (b). For reasons of clarity, two hydrocarbon Gay-Berne particles are omitted in each chain.

while the total number of electrons in each particle was identical to the corresponding atomistic representation. These values are illustrated in Figure 5.7. The nitrogen atom was assumed to have a net positive charge and the phosphorus atom to have a net negative charge. After calculation of the profile, symmetry was introduced through averaging over the two sides of the bilayer. Figure 5.8a presents the profile obtained with our reduced representation at 30 °C. The total electron density profile shows a minimum at the bilayer centre (about $0.24\,\mathrm{e.\AA^{-3}}$), a maximum located in the headgroup region (about $0.45\,\mathrm{e.\AA^{-3}}$) and a shoulder in between the headgroup and hydrocarbon regions. The bulk water electron density is $0.31\,\mathrm{e.\AA^{-3}}$, below the value reported experimentally (about $0.33\,\mathrm{e.\AA^{-3}}$). Figure 5.8b presents the comparison of the electron density profile obtained with our representation and one obtained with an atomistic representation. The atomistic profile was derived from a DPPC profile[142] by scaling the distance from the bilayer centre by the ratio of the peak to peak distances of DMPC and DPPC, respectively 36.0 and 38.3 Å.[129] Since these distances are



(a)

(b)

Figure 5.8: Electron density profiles, (a) the different contributions to the Gay-Berne representation and (b) comparison between the Gay-Berne (GB) representation (thick lines) and an all-atom (AA) representation for DPPC (thin lines). The all atom profile was scaled by the ratio of the peak-to-peak distances of DMPC and DPPC, respectively 36.0 and 38.3 Å[129] to mimic a DMPC profile.

close, and considering the reduced nature of our model, this scaled profile constitutes a decent comparison model. The atomistic profile was derived from a 1 ns simulation at 50 °C. The difference of temperature can be justified by the fact that DMPC and DPPC have different transition temperatures, respectively 24 and 42 °C. This choice of temperatures therefore ensures that both systems are located in the L_α phase and at similar distances from the transition temperature. The Gay-Berne electron density profile is in good agreement with the atomistic profile. The peak locations and intensities are reproduced, the shoulders at the hydrocarbon/headgroup interfaces have an identical shape and the water penetration inside the bilayer is identical. The only significant difference is in the bulk water electron density which is lower in our model. The most useful quantitative information from the electron density profile is the bilayer thickness or head-head separation D_{HH} taken as the peak to peak distance in the electron density profile. In our simulation the peak to peak distance is 35.0 Å at 30 °C in good agreement with the experimental value of 36.0 Å at 30 °C for DMPC.[129]

5.2.3 Atom Density Profile

Contrary to electron density profiles which only provide data on the phosphorus location, atom density profiles give more complete information on the bilayer structure.

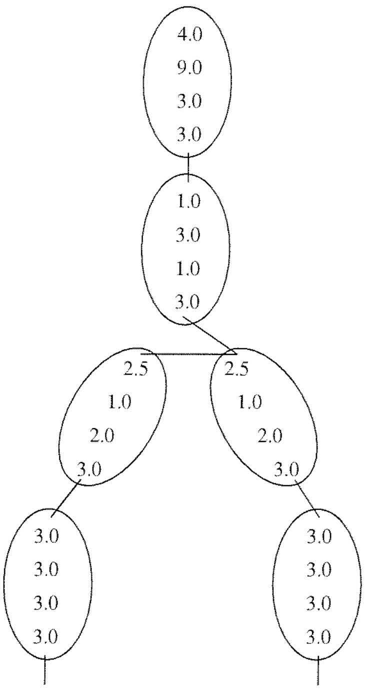
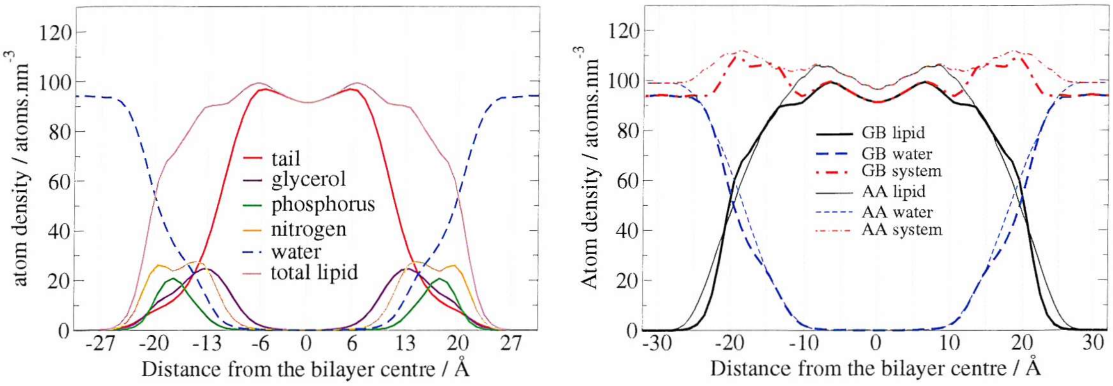


Figure 5.9: The resolved atom distribution on the Gay-Berne representation. For reasons of clarity, two hydrocarbon Gay-Berne particles are omitted in each chain.

The distribution of the position of certain lipid atoms along the bilayer normal can be obtained. The atom density is obtained by counting the number of atoms in slabs of width 1 Å along the bilayer normal and then dividing the number of occurrences by the slab volume. As has been detailed for the electron density profile (cf. section 5.2.2), in our model atoms are not explicitly represented. For increased precision we decided to resolve their positions. Atoms were placed on four locations evenly spaced on the particle's axis. The number of atoms were chosen to mimic as closely as possible the atomistic distribution, while the total number of atoms in each particle was identical to the corresponding atomistic representation. These values are illustrated in Figure 5.9. After calculation of the profile, symmetry was introduced through averaging over the two sides of the bilayer. The results for our simulation at 30 °C are presented in Figure 5.10 (a). The total lipid density increases above $y = \pm 25$ Å and reaches a maximum at $\approx \pm 7$ Å, where there are about 100 atoms.nm^{-3}. The density then decreases, there is a minimum at the centre of the bilayer with 90 atoms.nm^{-3}. The water has an inverse profile, with a density having a stable value of 94 atoms.nm^{-3} above $y = \pm 25$ Å. The water density then decreases steadily until $y = \pm 10$ Å, where



(a)

(b)

Figure 5.10: Atom density profiles, (a) the different contributions to the Gay-Berne representation and (b) comparison between the Gay-Berne (GB) representation (thick lines) and an all-atom (AA) representation for DPPC (thin lines). The all atom profile was scaled by the ratio of the peak-to-peak distances of DMPC and DPPC, respectively 36.0 and 38.3 Å[129] to mimic a DMPC profile.

it becomes zero. The bulk water atom density corresponds to a molecular density of 31.3 molecules.nm^{-3}, lower than the experimental value of 33 molecules.nm^{-3}.

Figure 5.10 (b) presents the comparison of the atom density profile obtained with our representation and one obtained with an atomistic representation. The atomistic profile was derived from a DPPC profile[142] by scaling the distance from the bilayer centre by the ratio of the peak to peak distances of DMPC and DPPC, as described in section 5.2.2. The atom density of the Gay-Berne representation is globally lower than that of the atomistic representation. Nevertheless, the agreement between the two profiles is correct. The depth of penetration of water particles inside the Gay-Berne bilayer is nearly identical to the atomistic depth, suggesting a correct reproduction of the bilayer-water interface. Water density is too low in the region $y = \pm 15$ to ± 22 Å. This fact can be explained by the high density of this region, and the larger size of Gay-Berne particles when compared to Lennard-Jones sphere, which leave less space for water. The Gay-Berne and atomistic lipid profiles are similar; in both cases, there is a minimum at $y = 0$ Å, and a maximum at $y \approx \pm 7$ Å. The Gay-Berne profile presents shoulders than are not present in the atomistic profile, they are probably caused by the larger size of the particles in our representation, and to a somehow rougher interface.

5.2.4 Lipid Volume V_L

Experimental lipid volume measurements have been performed using a variety of methods - volumetry[143] (V), neutral flotation combined with dilatometry[144] (NFD), absolute specific volume measurement[145] (ASVM) or using a differential vibrating tube densimeter[146] (DVTD). Most yield consistent results as shown in Table 5.1. Compared to other membrane characteristics (e.g. headgroup surface area) the lipid volume measurement is relatively precise with errors of the order 2 parts in 1000.[129]

Table 5.1: Comparison of literature values for DMPC volume/lipid.

Method	Temperature (^oC)	v_L (ml/g)	V_L (Å^3/lipid)
NFD[144]	30	0.977	1100
ASVM[145]	30	0.978	1101
V[143]	30	0.978	1101
DVTD[146]	30	0.972	1094

In simulations, the lipid volume can be accessed using the simple formula :

$$V_L = AD/2 - N_W V_W, \tag{5.4}$$

where A is the surface area per lipid, D is the repeat lamellar spacing (or the height of the simulation box in our case), N_W is the number of water molecules per lipid, and V_W is the volume of one water molecule calculated in the bulk water phase of our simulations. The lipid volume calculated from our simulation is $1101.23 \pm 2.07\,\text{Å}^3.\text{lipid}^{-1}$ at 30 °C in excellent agreement with experimental data.

5.2.5 Lamellar Repeat Spacings D

The lamellar repeat spacing D is determined using diffraction. Most studies have been performed on stacks of bilayers using multi-lamellar vesicles (MLVs). The experimental values show large fluctuations, in the case of DPPC ranging from 60 Å to 67.2 Å.[147,148] Nagle *et al.*[129] have suggested corrected values, with a proposed value for DMPC of 62.7 Å. In our simulation D corresponds to the height of the simulation box, corresponding to 67.75 Å at 30 °C. As lipid volume calculated in our simulations is in good agreement with experimental values, this higher value can therefore be explained by the low bulk water density, as reflected by the larger water volume - 31.55 $\mathring{A}^3$.molecule^{-1} compared with the experimental value of 30.4 $\mathring{A}^3$.molecule^{-1}.

5.2.6 Headgroup Surface Area A

The Area per lipid can be indirectly determined through the order parameter [135, 149, 150] using ^{2}H NMR, Petrache *et al.*[149] obtained values ranging from 60 Å^2.lipid^{-1} at 30°C to 68.5 Å^2.lipid^{-1} at 65°C. Other methods have been used such as gravimetric X-ray (GX) and corrected gravimetric X-ray (GXC). The former's reliability has been questioned, different studies often leading to different values. Corrections to the values obtained by different methods have been suggested by Nagle *et al.*,[129] enabling the different methods to be compared. The corrected values at 30 °C range from 59.5 Å^2.lipid^{-1} for GXC to 67.6 Å^2.lipid^{-1} for GX. According to Nagle *et al.* the best value is that obtained by EDP : 59.6 Å^2.lipid^{-1}. In NP_NAT simulations Feller *et al.*[90] showed that the larger the surface area, the smaller the peak-to-peak distance, consistent with a constant bilayer volume. In our case the balance is achieved as the head group surface area chosen enables the determination of the correct peak-to-peak distance. In our simulations, the head group surface area was fixed to 59.6 $\mathring{A}^2$.lipid^{-1}. The correct bilayer thickness D_{HH}, as well as lipid volume V_L were reproduced. The water density being too low, the lamellar repeat distance calculated from our simulations is logically too high.

5.2.7 Dipole

Owing to their zwitterionic nature, phosphatidylcholines posses a net dipole, often identified by the P-N vector. In the following section, the dipole values reported for our simulations correspond to the P-N dipole, involving phosphorus and nitrogen particles only. For each lipid l, each time-step t, the head group dipole $\mu(l,t)$ is defined as:

$$\mu(l,t) = \sqrt{\mu_x(l,t)^2 + \mu_y(l,t)^2 + \mu_z(l,t)^2} \qquad (5.5)$$

where the x, y and z dipole contributions are:

$$\begin{aligned}
\mu_x(l,t) &= q_N\left[x_N(l,t) - x_P(l,t)\right] = -q_P\left[x_N(l,t) - x_P(l,t)\right] \\
\mu_y(l,t) &= q_N\left[y_N(l,t) - y_P(l,t)\right] = -q_P\left[y_N(l,t) - y_P(l,t)\right] \\
\mu_z(l,t) &= q_N\left[z_N(l,t) - z_P(l,t)\right] = -q_P\left[z_N(l,t) - z_P(l,t)\right]
\end{aligned} \tag{5.6}$$

where, q_N and $q_P = -q_N$ are the charges on the nitrogen and phosphorus particles, and $x_N(l,t)$, $y_N(l,t)$ and $z_N(l,t)$ are the coordinates of the nitrogen particle of lipid l at time t, and $x_P(l,t)$, $y_P(l,t)$ and $z_P(l,t)$ the coordinates of the phosphorus particle.

Experimentally, the dipole moment of the phosphocholine group has been estimated to be 19 D by Sheperd *et al.*[151] In our simulation the average head group dipole μ_{NP} has been calculated, defined as:

$$\mu_{NP} = \frac{1}{N_{lip}}\frac{1}{N_{time}}\sum_{l,t}|\mu(l,t)| \tag{5.7}$$

where N_{lip} is the number of lipids, and N_{time} the number of time-steps. A value of 23.44 ± 1.43 D at 30 °C has been obtained; it does not fluctuate significantly with temperature with values varying between 23.30 and 23.68 D over the complete temperature range. This large value is in reasonable agreement with the value reported by Sheperd *et al.*, suggesting that the charge parameterization has been effective.

Various experimental methods have been used to estimate the average orientation angle of the head group. The main conclusion has been that, on average, the head-groups lie almost parallel to the bilayer surface.[151–156] Similar characteristics have been obtained for phospholipids with phosphatidyl glycerol, phosphatidyl choline or phosphatidyl ethanolamine head groups. On average, the head group dipoles lie approximately, to within 30 °, of the membrane plane.[152,157–160] In Figure 5.11 the dipole moment components along the x, y and z axes are presented for the top (layer 1) and bottom (layer 2) layers of the bilayer as a function of time. The value have been calculated using the formula:

$$\mu_{1,x}(t) = \frac{1}{N_{lip}/2}\sum_{l_1}\mu_x(l_1,t), \tag{5.8}$$

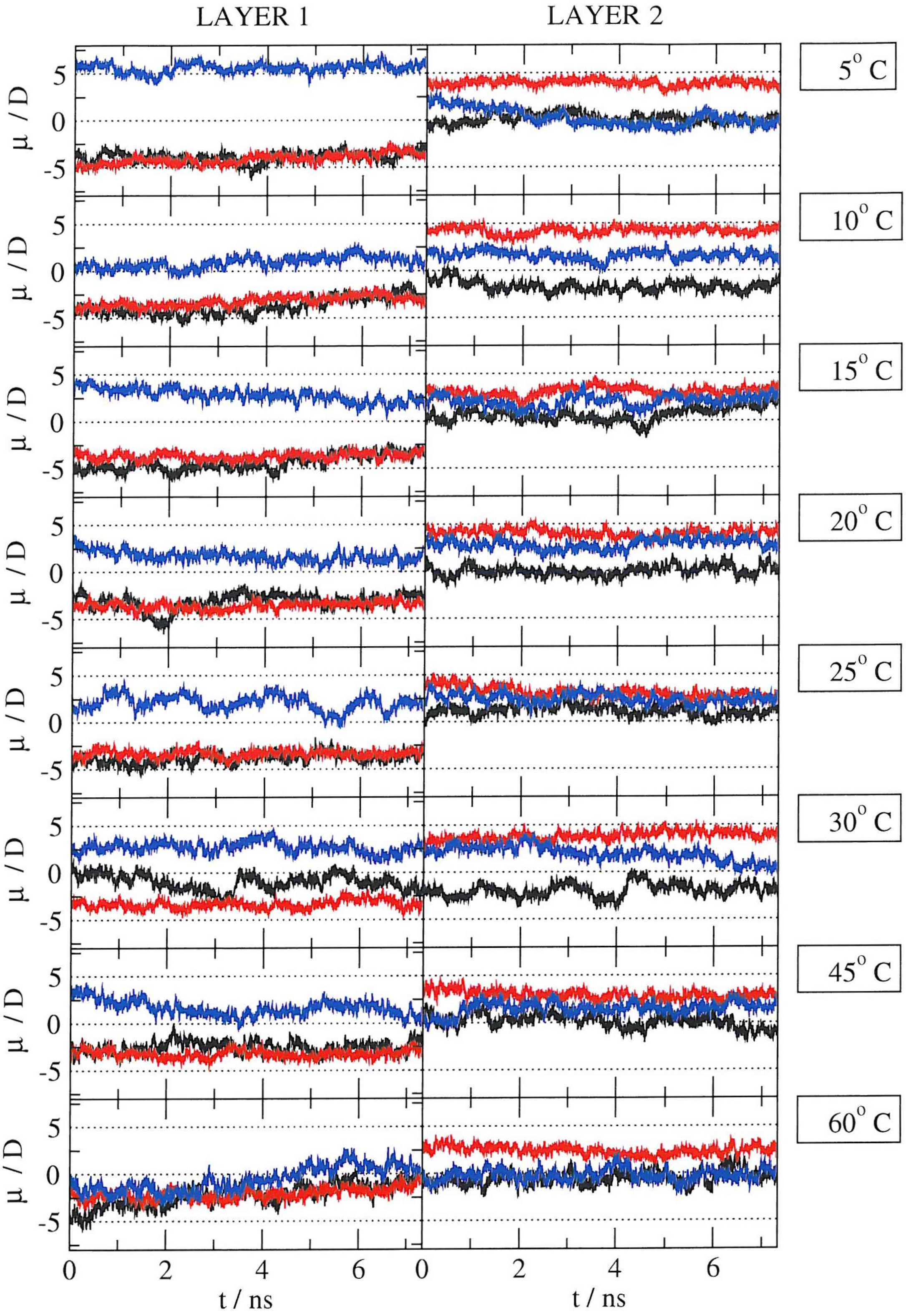


Figure 5.11: Temporal evolution of the average dipole components per lipid (averaged over all lipids within each layer). The x-component of the dipole is drawn in black, the y-component in red and the z-component in blue.

for the x component of layer 1. The sum is over all lipids in layer 1 (l_1). The y and z contributions, and the formula for layer 2 are simply derived by analogy with Equation 5.8.

For the dipole to lie in the plane, its y component (along the bilayer normal) should be virtually zero, or at least small with respect to the x and z components (in the plane of the bilayer). The examination of the x, y and z components of the total dipole of each of the membrane layers does not follow this trend. The y component of the dipole is non negligible, with a value of $\approx -3\,\text{to}\,-5\,\text{D}$ per lipid for the top layer and $\approx 3\,\text{to}\,5\,\text{D}$ per lipid for the bottom layer, suggesting that the layer dipole is pointing towards the membrane interior. While the two layers have opposite dipoles along the bilayer normal, the dipole components in the bilayer plane show no correlation between the two layers. This observation is confirmed when examining the temperature behaviour of the dipole; while the y component is systematically negative for the top layer, and positive for the bottom, the x and z components show no systematic trend and fluctuate more than their y counterpart, suggesting a random distribution of the dipole in the plane of the bilayer. The y dipole contribution behaviour can be easily explained by examining the average phosphorus and nitrogen bead positions along the y axis. The average y coordinate of the nitrogen bead is closer to the bilayer centre than the phosphorus bead y coordinate, although its height fluctuates more. The nitrogen particle mobility is greater than that of the phosphorus bead as it is at the end of the chain, allowing it to interact more freely with the glycerol beads, resulting in a lower height than expected from the initial structure.

To investigate further the dipole behaviour, the angle between the P-N vector and the bilayer normal was calculated. The angle $\phi(l,t)$ was calculated for each lipid l and time-step t. The bilayer normal considered was pointing up for those lipids in the top layer and pointing down for lipids in the bottom layer. An angle of 0° therefore corresponds to a dipole pointing towards the exterior of the membrane, while an angle of 180° correspond to a dipole pointing towards the membrane interior. In Figure 5.12 we present the average angle between the P-N vector and the bilayer normal

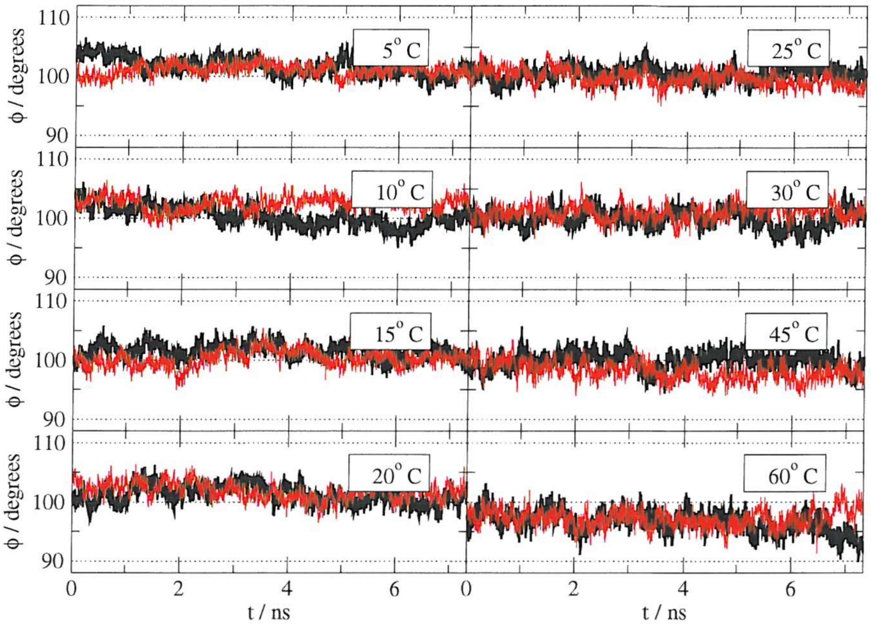


Figure 5.12: Temporal evolution of the angle ϕ between the P-N vector and the bilayer normal averaged over all lipids within each layer. The top layer angle is drawn in black, the bottom layer in red. For the sake of clarity the standard error is not represented; it spans between 32 and 64 °.

averaged over all lipids within each layer as a function of time:

$$\phi_1(t) = \frac{1}{N_{lip}/2} \sum_{l_1} \phi(l_1, t) \tag{5.9}$$

for layer 1, where the sum is over all lipids within layer 1 (l_1). $\phi_2(t)$ is defined similarly. For each temperature a similar trend is observed, the angle for both layers is close to $\approx 100\,°$ and is stable across the entire 7.2 ns simulated. Such an angle corresponds to a P-N dipole making an angle of 10 ° with the bilayer plane and pointing toward the membrane interior, as has been predicted by examining the dipole components. This result is a direct consequence of the relative average positions of the phosphorus and nitrogen beads. Nevertheless, considering only the average angle between the P-N vector and the bilayer normal would be a mistake, as the angle is distributed over a large range. This has been shown in molecular dynamics simulations of atomistic models[161,162] where the angle distribution spans a region between 0 ° and 150 °. In their simulations Tu *et al.*[163] identified a broad, bimodal distribution, indicating two dominant configurations. Skarjune *et al.*[164] performed the conformational analysis of the head group of phosphatidylethanolamine and came to the conclusion that one particular conformation was not adopted, but rather found three or four fami-

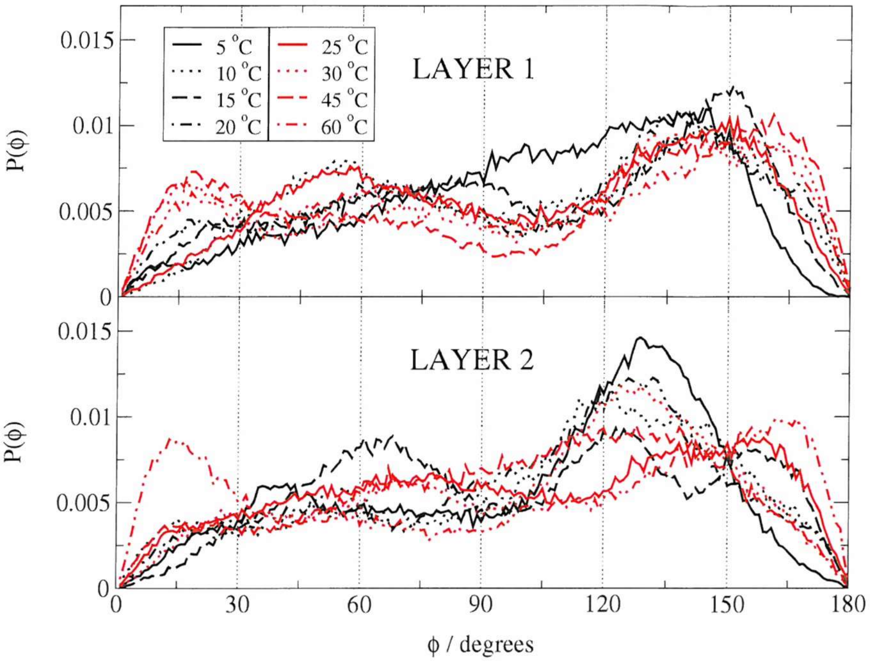


Figure 5.13: Distribution of the angle ϕ between the headgroup P-N vector and the bilayer normal for each of the two layers and the eight temperatures studied. Angles $\phi(l_1, t)$ and $\phi(l_2, t)$ are considered for the distribution, where l_1 is over lipids in layer 1 and l_2 is over lipids in layer 2.

lies of conformations. Using deuterium and phosphorus NMR combined with Raman spectroscopy, Akutsu *et al.*[152] could not isolate a unique solution for the head group conformation, but rather one or two families. These results suggest that different conformations are accessible to the head group, and hence that a description through an overall angle is only valid based on macroscopic grounds. These observations corroborate our analysis of the angle distribution, displayed in Figure 5.13 for the individual layer distributions - $p_1(\phi) = P(\phi(l_1, t))$ and $p_2(\phi) = P(\phi(l_2, t))$, and in Figure 5.14 for the distribution averaged over the two layers ($p(\phi) = (p_1(\phi) + p_2(\phi))/2$). Examining these distributions, it is not possible to identify a unique angle around which the distribution is centred. The distributions have several significant peaks, with different heights and widths.

To identify the characteristic angles, we performed a non linear curve fitting. In most cases there appears to be at least three significant angles; we therefore fitted the distributions with a sum of three Gaussian functions. The function used is :

$$\begin{aligned} p(\phi) &= A_0\, exp(-A_1(\phi - A_2)^2) \\ &+ A_3\, exp(-A_4(\phi - A_5)^2) \\ &+ A_6\, exp(-A_7(\phi - A_8)^2). \end{aligned} \tag{5.10}$$

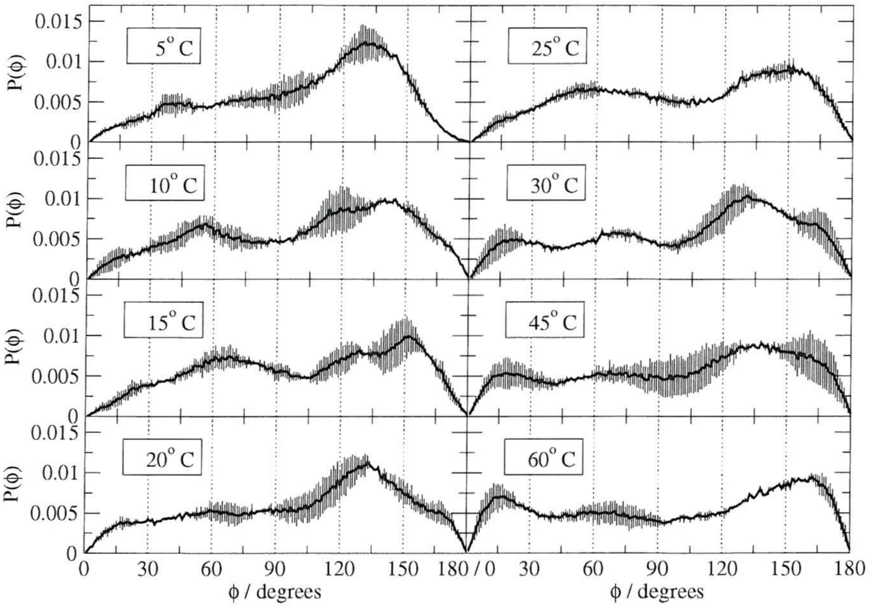


Figure 5.14: Distribution of the angle ϕ between the headgroup P-N vector and the bilayer normal averaged over the two layers for the eight temperatures studied, together with one standard error.

For a single normalized function, the coefficients A_0 and A_1 are linked :

$$A_0 = 1/\sigma\sqrt{2\pi} \tag{5.11}$$

and

$$A_1 = 1/2\sigma^2, \tag{5.12}$$

where σ is the standard error, and only two coefficients are needed; one for the standard error and one for the mean ($A_2 = \mu$ in this case). The same relation holds for A_3 and A_4, and A_6 and A_7. Since we are fitting three normal functions to a normalized distribution, the individual functions are no longer individually normalized, three coefficients are therefore required for each function. To rank the different peaks, the relative weights of each angle were resolved(a). They are displayed alongside each angle, as a percentage, thereby giving an indication of the relative importance of each peak. The evolution of the three main angles with temperature is illustrated in Figure 5.15.

As we were only interested in the main angles and their associated weights, the standard deviation from coefficients A_0 and A_1 (or A_3 and A_4, or A_6 and A_7) are not

---

(a) If we were to consider a single normal function, the relationship between A_0 and A_1 would be $A_0 = \sqrt{A_1/\pi}$. Since we are considering three Gaussian functions, the coefficients A_0, A_3 and A_6 incorporate the weights of each function, respectively α, β and γ: $p(\phi) = \alpha p_1(\phi) + \beta p_2(\phi) + \gamma p_3(\phi)$. The weights are therefore $\alpha = A_0/\sqrt{A_1/\pi}$, $\beta = A_3/\sqrt{A_4/\pi}$ and $\gamma = A_6/\sqrt{A_7/\pi}$.

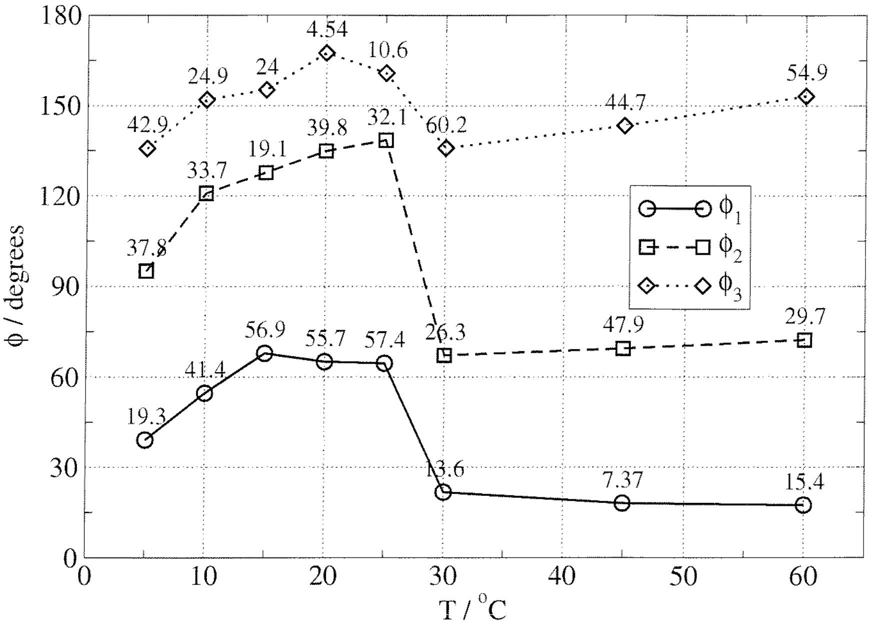


Figure 5.15: The three main angles ϕ between the headgroup P-N vector and the bilayer normal resolved from a non linear curve fitting of the average angle distribution (cf. Figure 5.14) with three normal functions. The percentage weights$^{(a)}$ are given alongside each angle.

presented. This fitting was performed on the average angular distribution (Figure 5.14). The correlation coefficient was at least 0.98 in each case.

This analysis enables extra conformations to be shown, that would not be accounted for by an average angle description. At 30 °C, the three angles are, in order of importance, 136 °, 67 ° and 22 °, corresponding to angles with the bilayer plane of respectively −46 °, 23 ° and 68 °. While the average angle is nearly identical for all temperatures (≈ 100 °), the three resolved angles are not distributed in the same fashion across the temperature range. The angles possess high values below 25 °C, but the lowest angles have the most significant weights (with the exception of 5 °C), above 30 °C, the angles are smaller but the highest angles' weights are more significant. The three angles have a similar temperature dependence. There is a plateau below 25 °C, a rapid decrease of the angle in between 25 and 30 °C, and a plateau above 30 °C. This rapid decrease is representative of a significant change in the organization of the head group and can be associated with a phase transition in the head group region. While the examination of the order parameter (described later in section 5.2.8) gives insight in to the hydrocarbon region behaviour, these angles enable the head group behaviour to be understood.

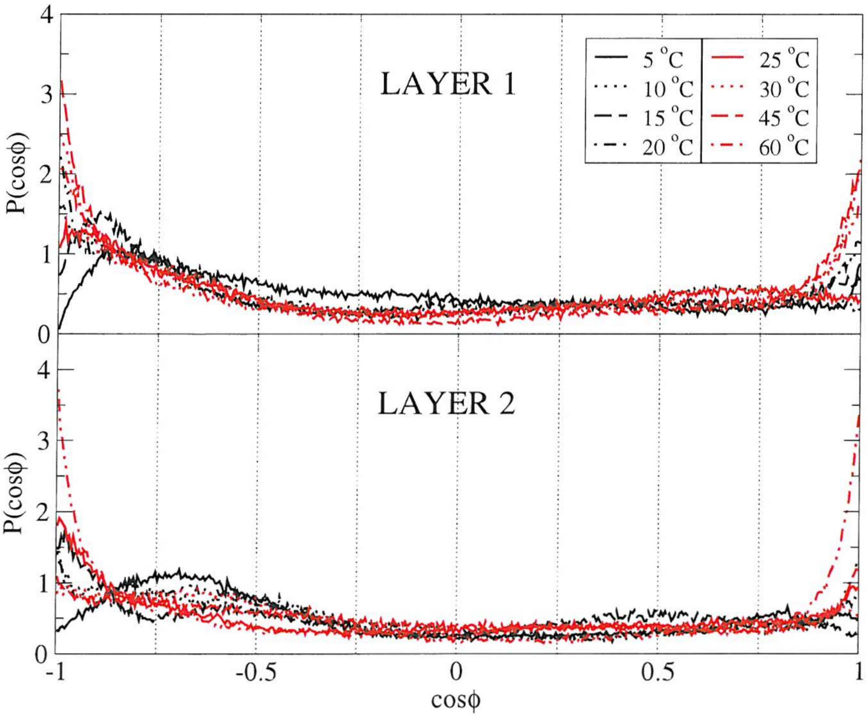


Figure 5.16: Distribution of the cosine of the angle ϕ between the headgroup P-N vector and the bilayer normal for each of the two layers and the eight temperatures studied.

To examine the distribution shape we investigated the distribution of the cosine of the angle ϕ with the bilayer normal. These distributions are shown in Figure 5.16 for the individual layers and in Figure 5.17 for the average over the two layers. A pure statistical distribution with no intrinsic preference arising from the model would give a probability $P(\cos\phi) = 1/2$ over the entire $\cos\phi$ range. This corresponds to the dipole moment lying predominantly in the plane of the bilayer, purely as a result of the way the system is defined. Figure 5.16 clearly indicates preferences for angles in the vicinity of 0 and 180 °. The results in these regions are less precise, as they contain less data. The analysis in this region is therefore complicated by the lower numerical precision of the data. Nevertheless, when looking at the way $P(\cos\phi)$ deviates from this ideal distribution, we obtain more values around $P(\cos\phi) = \pm 1$ than expected. This implies that in actuality the dipoles have a preference for pointing into and out of the bilayer. This trend is especially marked for angles in the $\cos\phi = -1$ region and comes as no surprise when the average position of the nitrogen and phosphorus particles is considered. The population for $\cos\phi = 1$ also shows deviation from a normal distribution, this deviation becomes more important as temperature

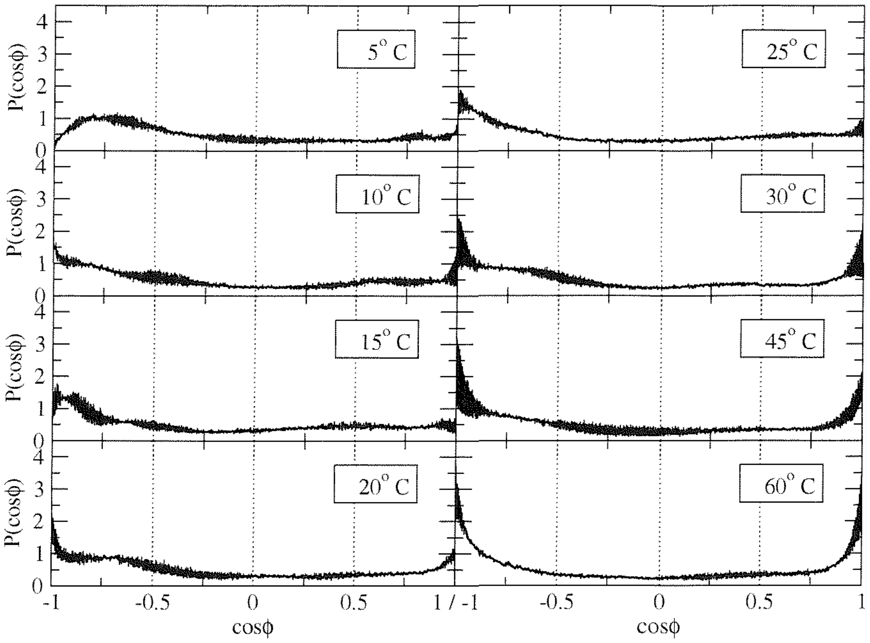


Figure 5.17: Distribution of the cosine of the angle ϕ between the headgroup P-N vector and the bilayer normal averaged over the two layers for the eight temperatures studied, together with one standard error.

increases. These results are consistent with experimental data which suggest that the P-N dipole lies mostly in the plane, as suggested by the sole examination of the average angle in the present study; but show much richer beyhaviour. These plots show that systematic deviations from this alignment in the plane are present.

5.2.8 Orientational Ordering

While the study of the P-N dipole gives insight into the head group behaviour, information about the bilayer's hydrocarbon region can be accessed through the orientational ordering examination.

Order Parameter

The concept of order parameter is now widely used in the world of liquid crystals. In biomembranes it has also been introduced[165–167] to provide a basis for understanding the concept of membrane fluidity. The second order Legendre polynomial P_2 is widely used as a measure of the order parameter.[21, 85, 89, 90, 93, 168, 169] Denoting the angle between a C-H bond on the ith carbon on the acyl chain and the bilayer director by β_i, the order parameter is defined by:

$$S_{CD}^{(i)} = \left\langle \frac{1}{2}\left(3cos^2\beta_i - 1\right)\right\rangle = \langle P_2(cos\beta_i)\rangle . \tag{5.13}$$

For a perfectly ordered system, $\beta_i = 0$ and $S_{CD}^{(i)} = 1$, for a system perfectly ordered perpendicularly to the director, $\beta_i = \pi/2$ and $S_{CD}^{(i)} = -1/2$. Order parameters can be accessed through the measurement of ^{2}H-NMR quadrupolar splittings; providing values of $|S_{CD}^{(i)}|$ rather than $S_{CD}^{(i)}$, through the specific deuteration of an acyl chain hydrogen atom. The quadrupolar splitting depends on the direction of the applied magnetic field

$$\Delta\nu(\gamma) = \Delta\nu(\gamma = 0)P_2(cos\gamma), \tag{5.14}$$

where γ is the angle made by the magnetic field with the director of the system. The quadrupolar splitting vanishes for $\gamma = 54.73°$, the so called magic angle. The quadrupolar splitting is linked to the CD bond vector order parameter along the chain.

$$\Delta\nu = \frac{3}{4}\frac{e^2qQ}{h}|S_{CD}| \tag{5.15}$$

where e^2qQ/h is the static quadrupole coupling constant that depends on the bond studied. The order parameter $S_{CD}(n)$=$S(n)$ can be measured for each carbon atom n on the chain. The "fluid bilayer signature" consists of a relatively constant "plateau region" of $S(n)$ for the first half of the chain (e.g. n=2-9 for a palmitate chain) and a more rapidly decreasing $S(n)$ for the end of the chain.[170] It has been extensively studied by Seelig *et al.*[156,166] Petrache *et al.*[149] reported bond order parameter profiles as a function of carbon number along the acyl chains for the homologous series of saturated phosphatidylcholines, and in particular for DMPC. The bond order parameter for the plateau ranges from 0.21 at 30°C to 0.18 at 50°C for DMPC. Other literature data[5,21,23,85,89,90,92,93,168,169] suggest similar values ; $S(n)$ have the value of approximately 0.18-0.22 for the methylenes 1 to 9, and then decrease to reach a value of 0.08 - 0.10 for the last methylene.

The second-rank Gay-Berne order parameter $\overline{P_2}^{GB}$ can be calculated and compared to the experimental S$_{mol}$ segmental order parameter, which corresponds to the C-C bond order parameter: S_{CC}. The Gay-Berne order parameter can be defined as $\overline{P_2}^{GB} = \langle P_2(\cos\beta)\rangle$, where β is the angle between the molecular axis and the director $\mathbf{n}$ and $\langle\rangle$ denotes a time average. The magnitude of $\overline{P_2}^{GB}$, and the location of the mesophase director $\mathbf{n}$ are obtained by first calculating $\overline{P_2}^{GB}(t)$, which is the order parameter for a configuration at a given time step, by forming the $\mathbf{Q}$ tensor with elements,

$$Q_{\alpha\beta} = \frac{1}{2N}\sum_i (3l_{\alpha i}l_{\beta i} - \delta_{\alpha\beta}), \tag{5.16}$$

where $l_{\alpha i}$ is the direction cosine of the ith Gay-Berne particle and the αth box axis. Diagonalising $\mathbf{Q}$ gives $\overline{P_2}^{GB}(t)$ as the largest eigenvalue and the director is the corresponding eigenvector.

$\mathrm{S_{mol}}$ corresponds to the segmental order parameter or to the average orientation of the chain segments to which the deuterium is attached. According to Seelig *et al.*[165] the $\mathrm{S_{mol}}$ order parameter can be derived from $\mathrm{S_{CD}}$ by the following relation

$$\mathrm{S_{mol}} = \mathrm{S_{CD}}\left[(3\cos^2 90° - 1)/2\right]^{-1} = -2\mathrm{S_{CD}}. \tag{5.17}$$

The equation for the terminal methyl is different:

$$\mathrm{S_{mol}} = \mathrm{S_{CD}}\left[(3\cos^2 109.5° - 1)/2\right]^{-1} \approx -3\mathrm{S_{CD}} \tag{5.18}$$

The dimyristoylphosphatidylcholine $\mathrm{S_{CD}}$ intermolecular order parameter has been determined experimentally by Douliez *et al.*[171] and can be used to determine $\mathrm{S_{mol}}$ using Equations 5.17 and 5.18.

Figure 5.18 presents the $\mathrm{S_{mol}}$ order parameter derived from $\mathrm{S_{CD}}$ along with the Gay-Berne order parameter $\overline{P_2}^{GB}$ of the entire hydrocarbon region and of the three layers of Gay-Berne particles that make the hydrocarbon chain, as a function of temperature. The error bars on the Gay-Berne order parameters are standard errors calculated by dividing the data collection phase into 6 equal batches, each containing 200 000 steps. The overall Gay-Berne order parameters show three distinctive area : a plateau at low temperatures, a fast decrease of the order parameter around 20 °C,

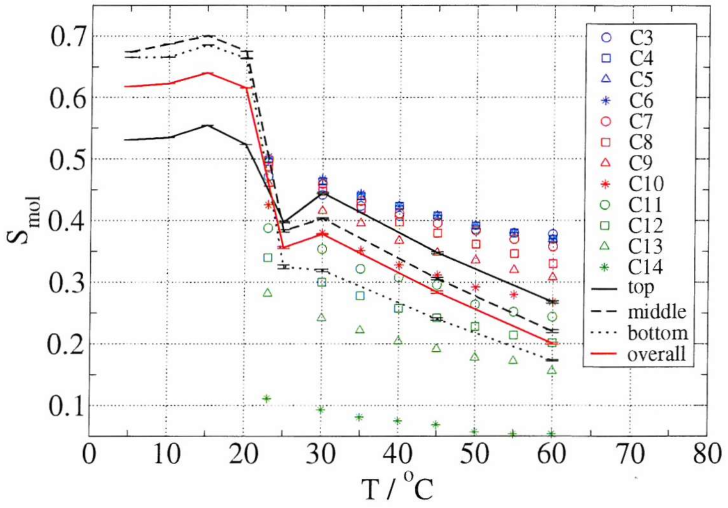


Figure 5.18: $\overline{P}_2^{GB}$ order parameter as a function of temperature for the entire hydrocarbon region and for the top, middle and bottom layers of the hydrocarbon region. S_{mol} order parameter as a function of temperature for DMPC.[165, 172, 173] The three colors, blue, red and green corresponds to the three layers used in the reduced representation, respectively top, middle and bottom. The C2 carbon atom is omitted as it is part of the glycerol backbone in our representation.

which can be associated with a phase transition and a slow decrease after the transition temperature. The system is clearly highly ordered below the transition temperature and highly disordered above it. It is therefore likely that at low temperatures the bilayer is in a solid phase, and at higher temperatures in a liquid-crystalline phase. The phase transition occurs in between 20 and 25 °C, in excellent agreement with the experimental phase transition of 24 °C.[129] In comparison, the phase transition identified in the head group region through examination of the P-N dipole orientation occured at a higher temperature (cf. Figure 5.15). The hydrocarbon and head group regions therefore appear to have different thermal behaviour, with an earlier phase transition in the tail region. At 5 and 10 °C, a higher order parameter was expected, this could be explained by the fact that the system is blocked in a local minimum and does not have enough energy to reorganize itself. Above 25 °C the order parameter for each layer is consistent with experimental data, where it is available. In the L_α phase the experimental segmental order parameter S_{mol} of the hydrocarbon region of DMPC declines from values of approximately 0.4 at the top of the chain to 0.2 at the chain terminus. This change is not uniform, however, the values over the top half of the chain are virtually identical with a marked decrease occurring in the second half. Above 25 °C, the simulation equivalent of S_{mol} shows the same behavior, with the

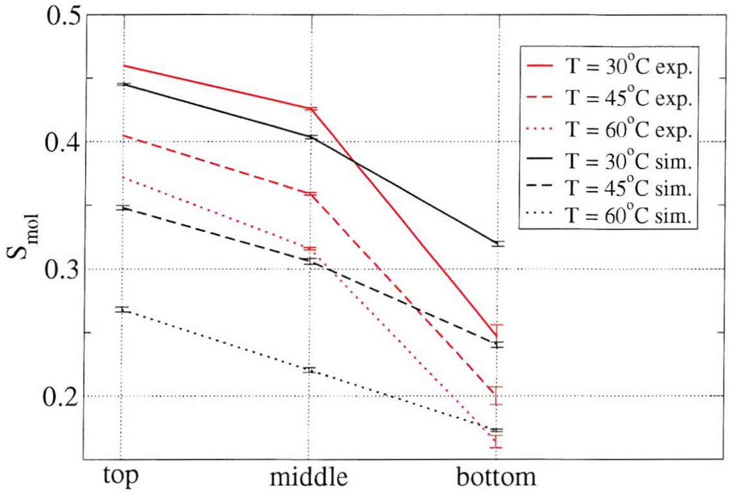


Figure 5.19: $\overline{P}_2^{GB}$ order parameter as a function of position along the hydrocarbon chain for three temperatures above the phase transition. The data calculated in our simulations are plotted in black, the data obtained from experimental values are plotted in red. The experimental value for the top region corresponds to the average of the order parameter over carbons C3 to C6, the middle region to average over carbons C7 to C10, and the bottom region to average over carbons C11 to C14.

top layer being the more ordered and the bottom layer the more disordered. Below 20 °C, the trend is inverted, with an almost identical order parameter for the bottom and middle layers at around 0.7, and a less ordered top layer at around 0.5. In the study of Whitehead *et al.*, even at low temperature the top layer was more ordered than the middle and bottom layers; the inversion observed in Figure 5.18 is probably due to the disturbance brought to the system by the inclusion of head groups and solvation.

Figure 5.19 presents the same data as a function of position along the hydrocarbon chain. Temperatures below the phase transition are not presented as there is no available experimental data enabling a comparison. The experimental data presented was obtained by averaging the S_{mol} order parameter of Douliez *et al*[172] for groups of four carbon atoms, therefore leading to three values for each temperature: top, middle and bottom. These can be compared to the values obtained in our simulations. The calculated order parameters show the same general trend as experimental values: high order parameter at the top of the chains which decreases when going down the chains. The discrepancy between the calculated and experimental values increases with temperature in the higher part of the chains and decrease in the lower part of the chains. The calculated values are systematically lower than the experimental

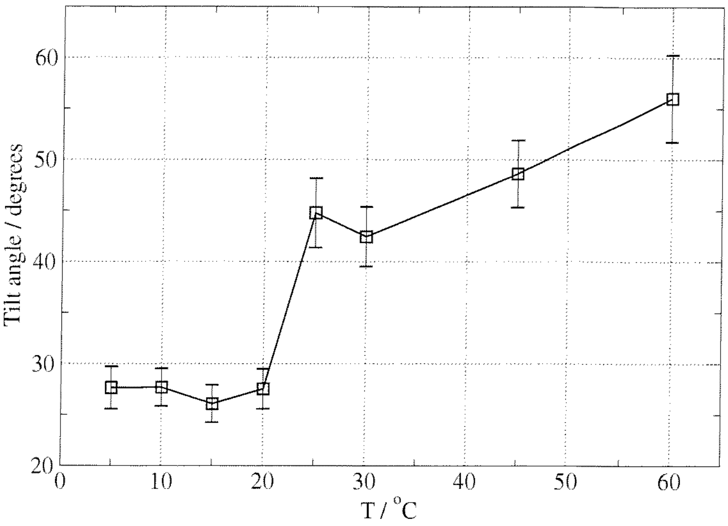


Figure 5.20: Tilt angle as a function of temperature. The Y direction was considered for the angle determination, rather than the director.

values in the top and middle layers and higher in the bottom layer. Considering the nature of the potential used, higher order parameter are expected in our model. This is the case in the unperturbed bottom layer, while disturbance brought to the higher layers by the head groups generates more disorder, and hence smaller values.

Chain Tilt

The chain tilt angle is defined as the angle between the bilayer normal and the vector connecting the first and the last carbon atom of the hydrocarbon chain. For simulations, the values reported in the literature range between 30 ° and 40 °.[4] The order parameter of the chain tilt can also be determined experimentally.[174,175] Using X-ray diffraction, Smith *et al.* reported values between 26 ° and 30 ° for DMPC at 23.5 °C, just below the phase transition. In our simulations, the tilt angle is the angle θ between the bilayer normal and the vector joining the last to the first tail particles of each chain. The tilt angles for temperatures ranging between 5 and 60 °C are displayed in Figure 5.20. Below 20 °C, the system has a tilt angle of 27 °, in perfect agreement with experimental data in the $L_{\beta'}$ and $P_{\beta'}$ phases, this value increases rapidly in between 20 and 25 °C, and continue to increase slowly after 25 °C. At 60 °C, the system reaches a tilt angle of 56 °. This angle is high when compared to experimental results, and can be explained by the slight loss of structure of the membrane at this temperature. The rapid change of tilt angle in between 20 and 25 °C can be associated with a phase transition and is consistent with the results previously obtained with the order parameter examination (cf. Figure 5.18). The transition is observed in the correct area - the transition temperature of DMPC from a $P_{\beta'}$ to a L_α phase is 24 °C.

5.2.9 Translational Diffusion

The average diffusion coefficient $\langle D \rangle$ of the Gay-Berne particles can be written in terms of the diffusion coefficients parallel and perpendicular to the director, $D_\parallel$ and $D_\perp$, respectively,

$$\langle D \rangle = \frac{1}{3}(D_\parallel + 2D_\perp). \tag{5.19}$$

The diffusion coefficients $D_\parallel$ and $D_\perp$ have been calculated using the following Einstein relationships

$$\langle (\Delta\mathbf{r}(t) \cdot \mathbf{n})^2 \rangle = 2D_\parallel t \tag{5.20}$$

$$\langle (\Delta\mathbf{r}(t) \cdot \mathbf{n}')^2 \rangle = 2D_\perp t, \tag{5.21}$$

where $\Delta\mathbf{r}(t) = \mathbf{r}(t) - \mathbf{r}(0)$, corresponds to the displacements of the centre of mass of each particle during the time, t, $\mathbf{n}$ is the unit vector along the director at t=0, and $\mathbf{n}'$ is a unit vector perpendicular to the director. The averages are accumulated over all Gay-Berne particles, hydrocarbon chain Gay-Berne particles and headgroup Gay-Berne particles for all time origins and consequently the statistical uncertainties on the mean square displacements increase with time.

Coordinates from each of the simulation trajectories were stored for analysis every 500 time steps, and the effects of periodic boundary conditions removed. Figure 5.21 presents both parallel-to and perpendicular-to the director mean-square-displacements for individual Gay-Berne particles and the lipid centre of geometry, for temperature ranging from 5 to 60 °C. The extent of diffusion within the plane of the bilayer (perpendicular to the director) clearly increases with temperature. At 60 °C, diffusion over $\approx 100\,\text{Å}^2$ is occurring for the Gay-Berne based mean squared displacements (Figure 5.21b), corresponding to almost a quarter of the simulation box. At an equivalent temperature, Whitehead *et al.* obtained a diffusion of $\approx 56\,\text{Å}^2$ over a

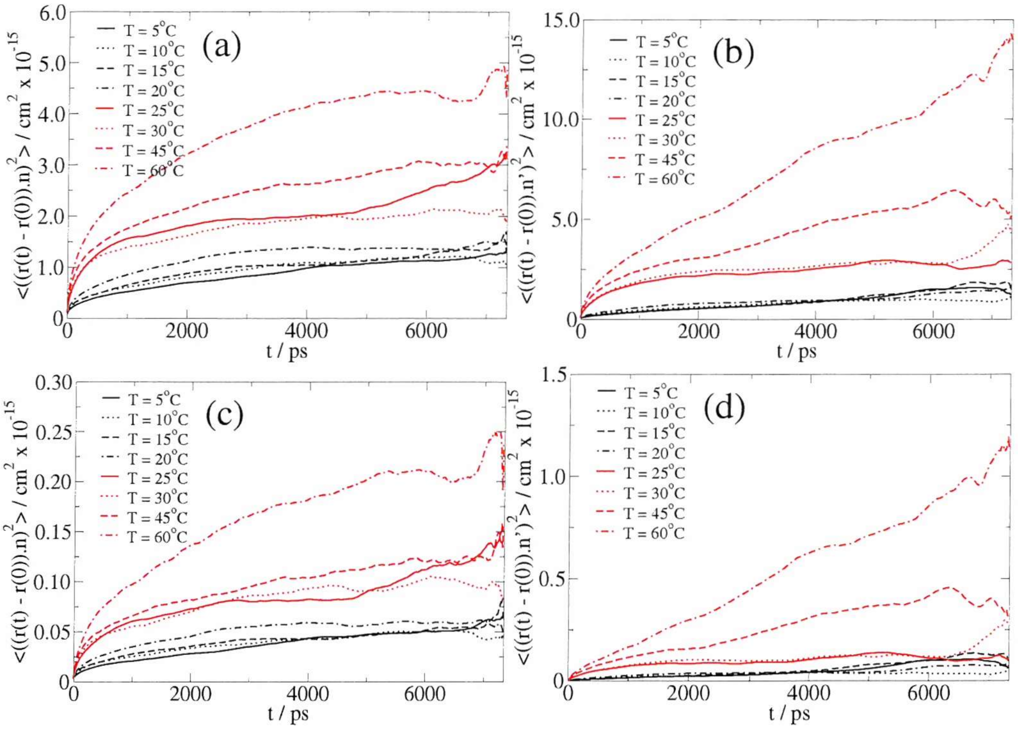


Figure 5.21: Mean square displacement plots: (a) using all particles, displacement parallel to the director, (b) using all particles, displacement perpendicular to the director, (c) using each lipid centre of geometry, displacement parallel to the director and (d) using each lipid centre of geometry, displacement perpendicular to the director.

period of 500 ps. Our model exhibits relatively less free diffusion in the plane, a logical result considering the absence of solvation in the original model and the overall lower density. The displacement along the director shows less free diffusion, consistent with the symmetry of the bilayer. The mean square displacement plots calculated with the lipid centre of geometry exhibit less diffusion, an order of magnitude smaller along the director and in the bilayer plane. In Figure 5.22, the average parallel, perpendicular and overall diffusion coefficients are presented for all Gay-Berne particles (plot (a)), for particles located in the hydrocarbon chains (plot (b)), for particles located in the headgroup (plot (c)) and for the lipid centre of geometry (plot (d)) as a function of temperature. The data collection phase was divided into four equal batches. The diffusion data was obtained by averaging the diffusion coefficients calculated using equations 5.20 and 5.21 over the two central batches : $t = 1800 - 5400$ ps. The error bars correspond to standard errors evaluated by dividing this data collection period into two equal batches. Data in the region $t = 0 - 1800$ ps were ignored, because pronounced librational motion was occurring, and data in excess of $t = 5400$ ps were

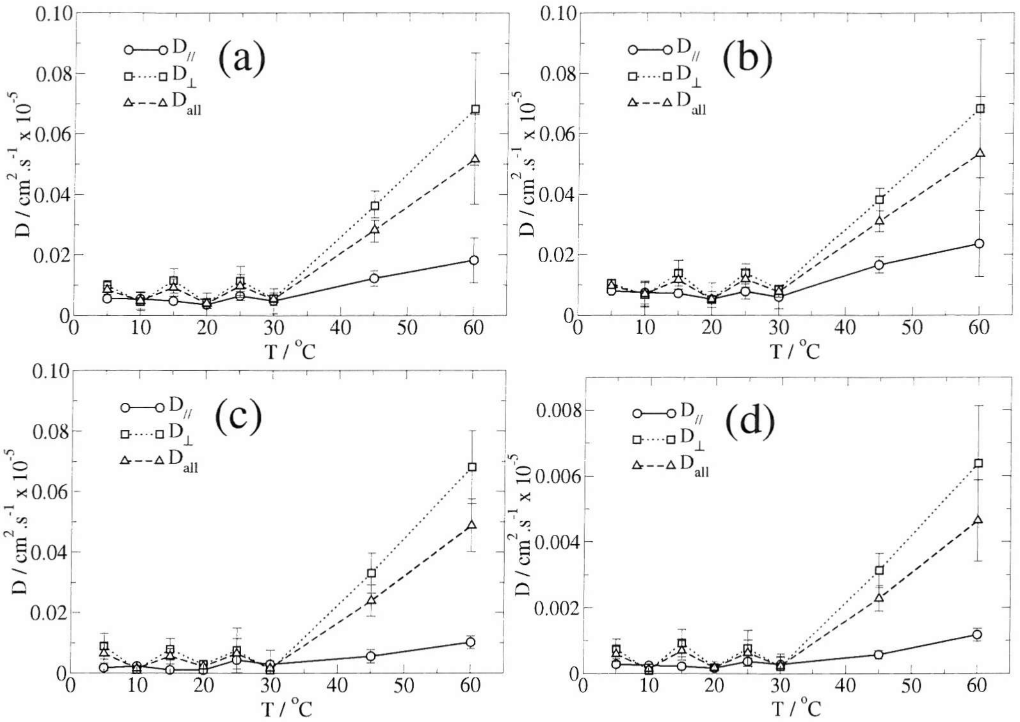


Figure 5.22: Average parallel, perpendicular and overall diffusion coefficients vs. temperature: (a) over all Gay-Berne particles, (b) over hydrocarbon chain Gay-Berne particles, (c) over headgroup Gay-Berne particles and (d) considering the lipid centre of geometry motion.

ignored because of decreasing precision in the calculated displacement.

From all plots, it is clear that diffusion along the director is small, in agreement with the geometry of the bilayer. All plots show the same general trend with two distinctive areas. Below 30 °C there is a plateau of low diffusion, with diffusion in the plane of the bilayer ($D_\perp$) and out of the plane ($D_\parallel$) being nearly identical. Above 30 °C, $D_\parallel$ remains nearly constant, while $D_\perp$ increases to reach $10^{-6}\,\mathrm{cm}^2.\mathrm{s}^{-1}$ for particle based plots (Figure 5.22a, b and c) and $10^{-7}\,\mathrm{cm}^2.\mathrm{s}^{-1}$ for the lipid based diffusion (Figure 5.22d) at 60 °C. The diffusion coefficients at 60 °C are summarized in Table 5.2.

Table 5.2: Diffusion coefficients D / $\times 10^{-7}\,\mathrm{cm}^2.\mathrm{s}^{-1}$ at 60 °C.

	$D_\parallel$	$D_\perp$	D_{all}
All GB particles	1.83 ± 0.74	6.83 ± 1.85	5.16 ± 1.48
Tail particles	2.36 ± 1.09	6.83 ± 2.29	5.34 ± 1.89
Head group particles	1.02 ± 0.21	6.81 ± 1.20	4.88 ± 0.87
Lipid based	0.118 ± 0.019	0.638 ± 0.175	0.464 ± 0.123

In our simulations a transition from an immobile level to a diffusive regime is detected in between 30 and 45 °C. Ordering and mobility transitions are not detected at the same temperature. At 30 °C the membrane is in a liquid crystalline phase as indicated by translational ordering data but is still in the non-diffusive regime. The examination of the temperature evolution for the head group angles detailed in Section 5.2.7 suggested a later structural phase transition for the head group particles. While the orientational ordering study suggests a phase transition of the hydrocarbon region between 20 and 25 °C, the angle examination suggests a head group transition between 25 and 30 °C. The delayed diffusive phase transition could therefore be largely due to the head groups. Nevertheless, the head groups only, cannot explain the delayed transition, as the structural head group transition is observed at a lower temperature (25-30 °C) than the diffusive transition (30-45 °C). The thermodynamic transition therefore appears to be more marked than the kinetic transition. A smaller increase of $D_{\perp}$ can be noticed for the tail region in between 20 and 25 °C, suggesting the possibility of an earlier phase transition in the tail region.

The lipid based diffusion plot is in good agreement with results obtained using fluorescence recovery after photobleaching (FRAP), where the diffusion increases at the phase transition to reach $9 \times 10^{-8}\,\mathrm{cm^2.s^{-1}}$ at 55 °C,[176] in good agreement with the value obtained for our simulations: $\approx 6 \times 10^{-8}\,\mathrm{cm^2.s^{-1}}$ at 60 °C. The time scale accessed by the FRAP method suggests than these diffusion coefficients correspond to long range lateral diffusion, rather than local diffusion.

König *et al.*[177] studied DPPC using backscattering. They interpreted their results in terms of superposition of local in-plane and out-of-plane diffusion and lateral diffusional jumps between adjacent sites as predicted by the free volume theory. In-plane diffusion coefficients (short time-scale diffusion) ranged from $D_{\perp}^{\mathrm{min}} = 1.5 \times 10^{-7}\,\mathrm{cm^2.s^{-1}}$ to $D_{\perp}^{\mathrm{max}} = 6.0 \times 10^{-6}\,\mathrm{cm^2.s^{-1}}$ at 60 °C. The lateral diffusion coefficient (long time-scale diffusion) is $D_{\perp}^{\mathrm{lat}} = 9.7 \times 10^{-8}\,\mathrm{cm^2.s^{-1}}$. These results suggest that particle based diffusion coefficients in the plane of the bilayer, correspond to local diffusion or *rattling in a cage*. In our simulations $D_{\perp}$ reaches $7.0 \times 10^{-7}\,\mathrm{cm^2.s^{-1}}$ at 60 °C (Figure 5.22a), inside the range described experimentaly by König *et al.* ($D_{\perp} = 1.5 \times 10^{-7}$ to $6.0 \times 10^{-6}\,\mathrm{cm^2.s^{-1}}$ at 60 °C). Lipid based coefficients correspond to lateral diffusion or *lipid swap*. The perpendicular diffusion coefficient based on the lipid centre of geometry reaches $6.0\times10^{-8}\,\mathrm{cm}^2.\mathrm{s}^{-1}$ at 60 °C (Figure 5.22d), in excellent agreement with the value described by König *et al.* for the lateral diffusion coefficient ($D_{\perp}^{\mathrm{lat}} = 9.7 \times 10^{-8}\,\mathrm{cm}^2.\mathrm{s}^{-1}$). Calculated and experimental diffusion coefficients are summarized in Table 5.3.

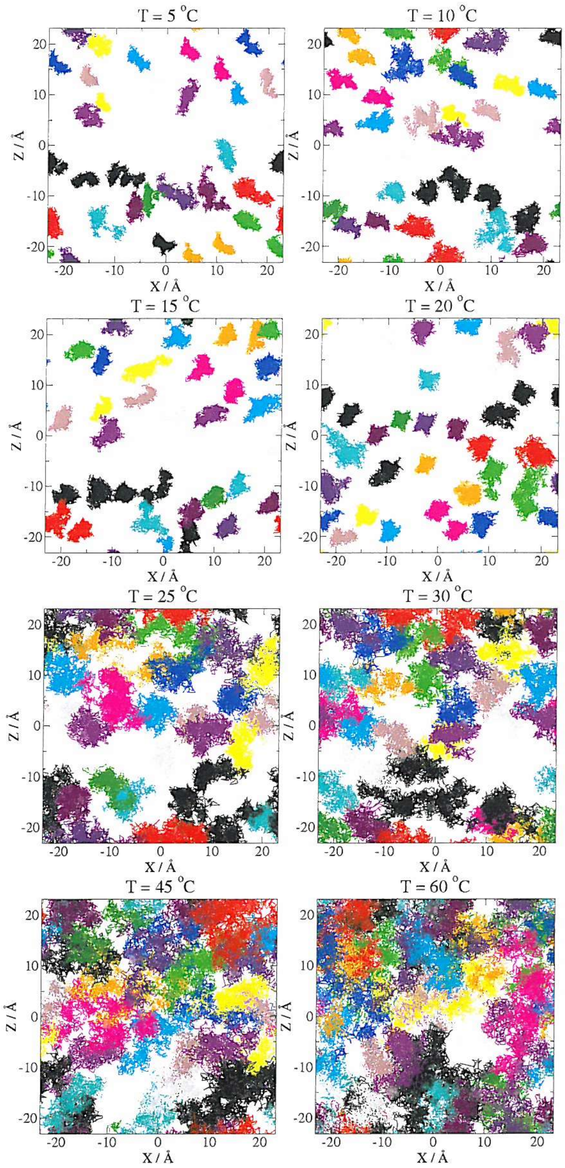


Figure 5.23: Lipid lateral motion projected on the xz plane (perpendicular to the bilayer normal) as a function of temperature. The trajectories shown are for the 7.2 ns simulation and the 32 lipids of the top layer.

Table 5.3: Comparison of calculated and experimental diffusion coefficients $D_\perp$ / cm^2.s^{-1}. The values reported for these MD simulations are the perpendicular diffusion coefficient for all Gay-Berne particles for the in plane motion and the lipid based perpendicular diffusion coefficient for lateral motion.

Time scale	These MD sim. at 60 °C	König *et al.*[177] at 60 °C	Lalchev *et al.*[176] at 55 °C
Short	$6.83 \pm 1.85 \times 10^{-7}$	$1.5\text{-}60.0 \times 10^{-7}$	-
Long	$6.38 \pm 1.75 \times 10^{-8}$	9.7×10^{-8}	9×10^{-8}

Lipid swap observation is further confirmed by examination of the centre of geometry motion in the plane of the bilayer illustrated in Figure 5.23. It is clear from the lipid displacement above 30 °C that lipid swaps are occurring. Tail and headgroup show similar diffusive behavior, the latter having only marginally smaller diffusion coefficients.

5.3 Simulation Efficiency

The main focus of this work was to propose a model that would be less computationally expensive than the equivalent atomistic representation. To measure the efficiency gain, the computation time of our model was compared to an atomistic simulation of DPPC.[115] The reduced representation of a 72 lipid biological membrane, containing 720 Gay-Berne particles and 2094 water molecules was run with *CHAMPAGNE* for 50 000 cycles and compared to the atomistic simulation of a 72 lipids system containing 2095 water molecules performed using CHARMM. The simulation of the Gay-Berne representation of a membrane took 34 hours on a single AMD Athlon 1400 MHz processor. Using CHARMM, the same number of cycles took 40 hours on the same processor. The proposed model is therefore only marginally faster than an atomistic model. The Gay-Berne model allows the use of bigger time steps (6 fs vs. 2 fs for CHARMM), therefore increasing the simulation length but that does not concern the intrinsic efficiency. No effort has been made to optimize the speed of *CHAMPAGNE*. The addition of water is largely responsible for the decrease in efficiency when compared to the efficiency of the hydrocarbon region representation.[2] The use of embeded dipoles in the water model means that the implementation of particle mesh Ewald is less straightforward than with point charges. Moreover the nature of the sticky potential does not allow the time step to be increased any further. However, the smoothness of the model's interaction potential results in rapid equilibration. Future work will concentrate on the implementation of the particle mesh Ewald method and should results in great improvement of the code efficiency. The water model used in our representation (SSD), gives satisfactory results and enables a good representation of the water-lipid interface. Nevertheless, this model is not without drawbacks, the bulk water density is too low. The implementation of a new hybrid water model, and its specific parameterization would lead significant improvements both in terms of structure and speed.

5.4 Self Assembly Properties

5.4.1 Introduction

Amphiphilic molecules possess polar head groups and nonpolar chains; to minimize unfavourable solvophobic interactions, they spontaneously aggregate to form a variety of microstructures, ranging from micelles, rodlike structures, and bilayers to more complex phases such as cubic phases (cf. Chapter 3). These aggregates can adopt different organizations depending on the temperature, pH, or composition of the mixture. If we consider amphiphilic molecules in a polar solvent, a compromise between complete phase separation and a isotropic mixture is reached. The latter is not achieved as it results in unfavourable water-hydrocarbon interactions. The formation of two distinct phases is also energetically less likely in biological conditions as it would remove the polar head groups from the polar solvent. The molecules therefore self-assemble to create a microphase in which the hydrocarbon chains sequester themselves inside the aggregate and the polar head groups orient themselves toward the polar solvent.[73] The aggregation process itself is the result of an equilibrium; transfer of the hydrocarbon chains into the aggregate interior is the driving force of the self-assembly, while electrostatic repulsion between head groups opposes it. [73] The self-aggregation process of surfactants is of fundamental importance to many biological and industrial processes. This phenomenon occurs on a very fast time scale (nanosecond) and on a very short length scale (nanometer), thus making experimental investigation difficult.[18] Theoretically, self-assembly has mainly been studied with coarse grain models using a few beads and molecular dynamics and/or Monte

Carlo,[98–101,178] dissipative particle dynamics,[102] stochastic dynamics,[179] or brownian dynamics.[180] One reason for this is that the time scales over which the aggregation occurs ,while being very short for experimental characterization, is long for computer simulations using atomistic models, which rather have to start from preassembled systems. In the present study, the main objective of the self aggregation study was to test the accuracy of our lipid representation. Atomistic simulations of the spontaneous aggregation of phospholipid into bilayers have been recently performed by Marrink *et al.*[19] They performed six simulations on systems containing 64 dipalmitoylphosphatidylcholine (DPPC) and 3000 water particles. Other phospholipids were investigated to test lipid dependence (POPC, DOPC, DOPE).[19] Larger DPPC systems were also examined.[19] Starting from a random lipid distribution and a cubic box, the formation of a perfect lipid bilayer was observed in all cases. Four steps can be identified in the aggregation process. There is an initial rapid separation into water and lipid phases ($\approx$ 200 ps). The rearrangement into a lipid bilayer is slower, requiring about 3 ns. This intermediate configuration contains a large water pore, which requires 20 ns to disappear. The system then relaxes to an equilibrium bilayer in about 5 ns. The latter accurately reproduces the available experimental data. Apart from the final bilayer structural analysis, their aggregation analysis was limited to visual inspection and to the time evolution of the density profile. This study gives a unique insight into the self-assembly of a bilayer at the atomic level. The observation of a hydrophilic pore, which is a basic feature of the bilayer formation mechanism is biologically important. It can explain observed ion permeation rates in the absence of translocation protein. Other bilayer self-assembly studies include work by Goetz *et al.*[98,178] and Shelley *et al.*[100] who both used coarse grain models. Analysis included density profiles, mobility and elasticity as well as stress profiles. Typically, self assembly of surfactants into micelle are analysed in more details,[98] with extension to kinetics[18] or diffusion[20] analysis.

In the present study, self-assembly is considered as a test for the robustness of the Gay-Berne representation of phospholipids.

5.4.2 Simulation set up

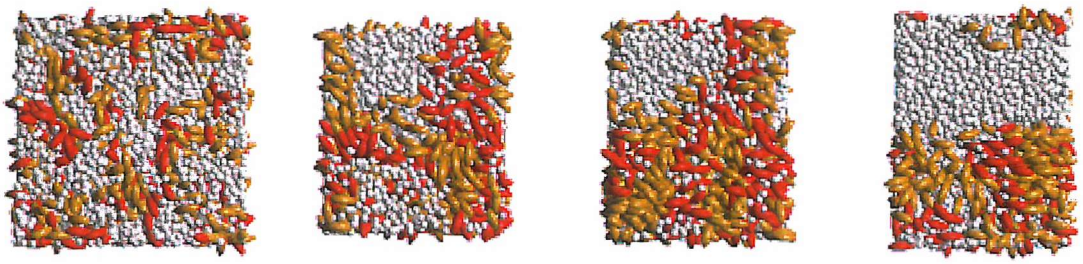
Two simulations were performed, both containing 64 lipids; the first labelled SLF1 contained 3000 water particles, the second labelled SLF2 contained 1863 water particles (i.e. 29.1 water/lipid). The study was conducted using molecular dynamics and a pseudo NPT ensemble in an orthorhombic unit cell at 45 °C. A hybrid molecular dynamics-Monte Carlo algorithm was applied to obtain the correct density. A normal molecular dynamics simulation was carried with a Monte Carlo move performed every 50 steps. The move involved modifying one of the three box dimensions (randomly chosen) and scaling all particle coordinates along this direction accordingly. A Boltzmann test was then performed. Velocities are rescaled after each Monte Carlo move to reproduce the desired temperature. The Monte Carlo moves ensure the right pressure and density are obtained. The box was initially cubic $|a| = |b| = |c| = 66.34\,\text{Å}$. The initial configuration was constructed by first randomly inserting lipids taken from a preequilibrated bilayer and then randomly inserting the water particles. Insertions that resulted in molecules that were within a certain distance to molecules already present were rejected. These distances were fixed at 6 Å for water-lipid and lipid-lipid contacts and at 2.8 Å for water-water contacts. The initial density was lower than the expected density; the initial box dimensions are on average 40% larger than the dimensions of a box containing 72 lipids and 2095 water particles. This condition was necessary for the lipid insertion procedure. Smaller initial values lead to clashes, or to the impossibility of inserting the 64 lipids correctly. The initial water-lipid random configurations were cubic; no bias was introduced in any direction. The MC moves (cf. section 5.1) on the box dimensions allowed the box to deform according to the forces acting within each system. Periodic boundary conditions were applied in all three directions. The initial structure was rapidly equilibrated. Equilibration runs consisted of two consecutive runs of 10 steps in which the simulation temperature must be stable to within 0.01. If this criterion is not met, the velocities are again rescaled and the equilibration phase continues until such time as a stable temperature is maintained. The length of these runs was voluntarily kept short to maintain the initial structure (randomly generated) for as long as possible prior to production runs. Production simulations of 2 400 000 and 3 300 000 time steps for respectively

SLF1 and SLF2 were then performed. Production runs were performed in the NVE ensemble, with a Monte Carlo move on the box dimension performed every 50 steps. After each Monte Carlo move, velocities were rescaled to control temperature. A scaled time step, $\Delta t^* = 0.005$ is used ($\Delta t^* = \Delta t\sqrt{\epsilon_0/m\sigma_0^2}$) where m is the mass of the particles. This corresponds to a real time step of approximately 6 fs. The protocol used is essentially identical to the one used for the preassembled bilayers simulations (cf. section 5.1).

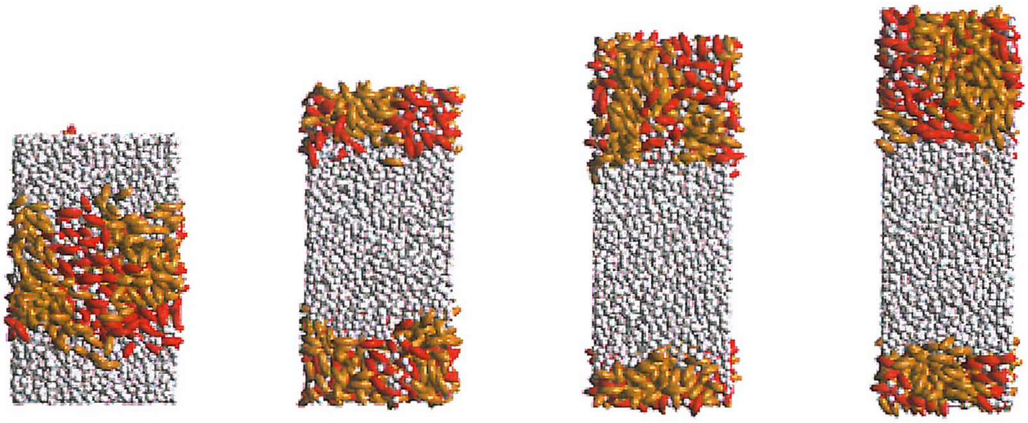
5.4.3 Results and Discussion

Neither of the two simulations show self-assembly of the system into a lipid bilayer structure, at least to one comparable to the structure of a preassembled bilayer. Nevertheless, local organization in a bilayer-like structure can be observed. Snapshots of simulation SLF1 and SLF2 are illustrated in Figure 5.24 and Figure 5.25 respectively. In both case there is an initial phase separation of water and lipid particles, occuring over a 3 ns period for the 3000 water molecules system and a 4 ns period for the 1863 water molecules system.

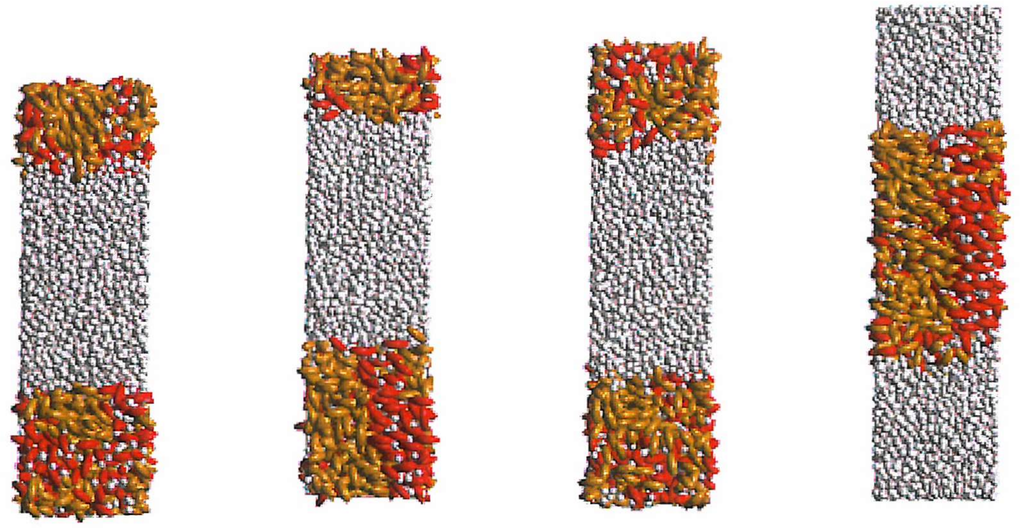
The system containing less water particles would have been expected to achieve phase separation faster. This contradictory result should be confirmed by further simulations with different starting structures. The next and final phase of the simulation consist in the reorganization of the tail and head group particles inside the lipid phase. This stage starts after the seggregation stage, continues until the end of the simulation, and would have probably continued if the simulations had not been stopped. Expected and final volumes are presented in Table 5.4. The expected volumes are based on previous simulations results. Final volumes are close to the calculated volumes, suggesting that equilibrium has been reached volume-wise.



(a) t= 0 ns (b) t= 1 ns (c) t= 2 ns (d) t= 3 ns

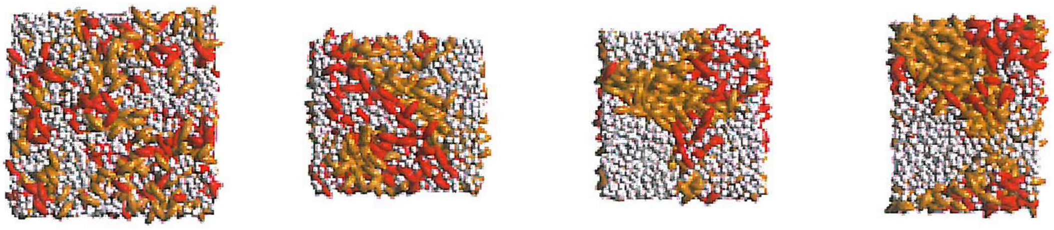


(e) t= 4 ns (f) t= 5 ns (g) t= 6 ns (h) t= 7 ns

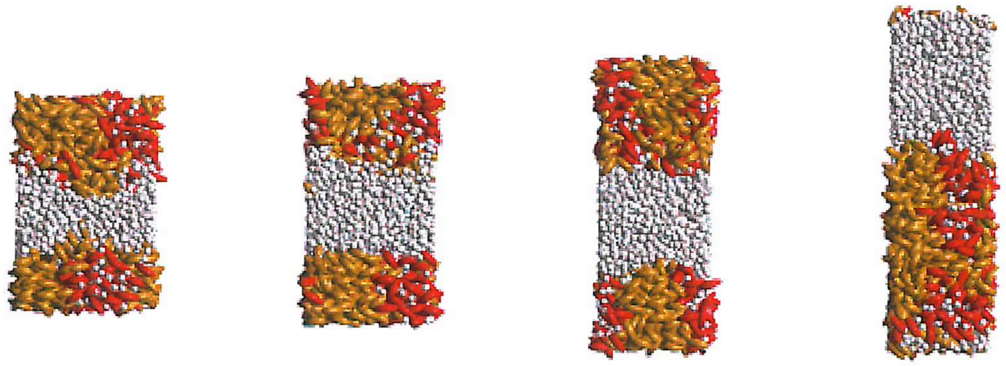


(i) t= 8 ns (j) t= 9 ns (k) t= 10 ns (l) t= 14.7 ns

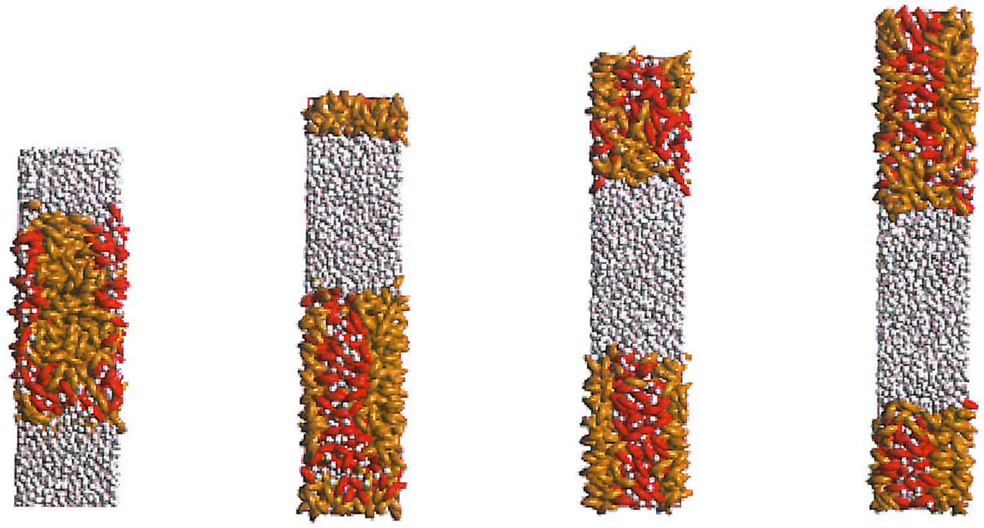
Figure 5.24: Snapshots of the simulation for system SLF1 (64 lipids and 3000 water particles) at start, intermediate and final stages. DMPC headgroups are shown in red, DMPC tails in orange and water particles in white.



(a) t= 0 ns (b) t= 1 ns (c) t= 2 ns (d) t= 3 ns



(e) t= 4 ns (f) t= 5 ns (g) t= 7 ns (h) t= 9 ns



(i) t= 10 ns (j) t= 12 ns (k) t= 15 ns (l) t= 20 ns

Figure 5.25: Snapshots of the simulation for system SLF2 (64 lipids and 1863 water particles) at start, intermediate and final stages. DMPC headgroups are shown in red, DMPC tails in orange and water particles in white.

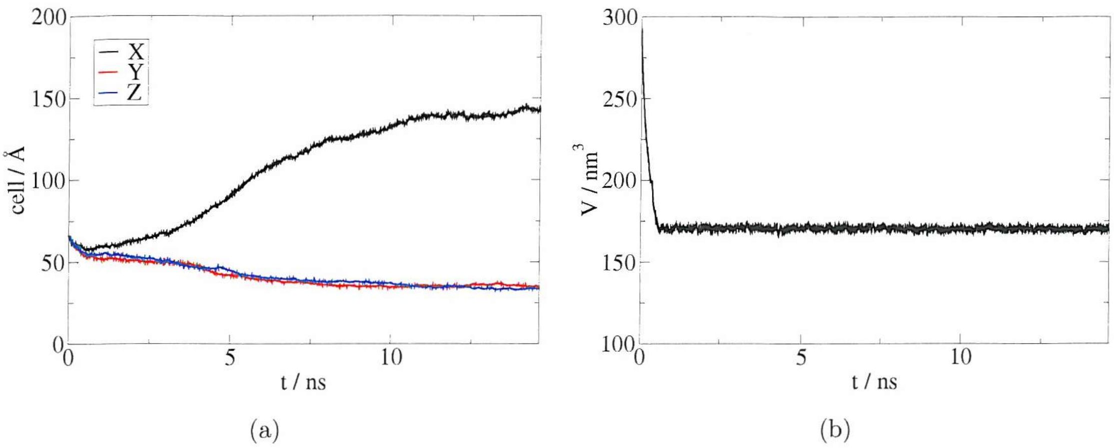


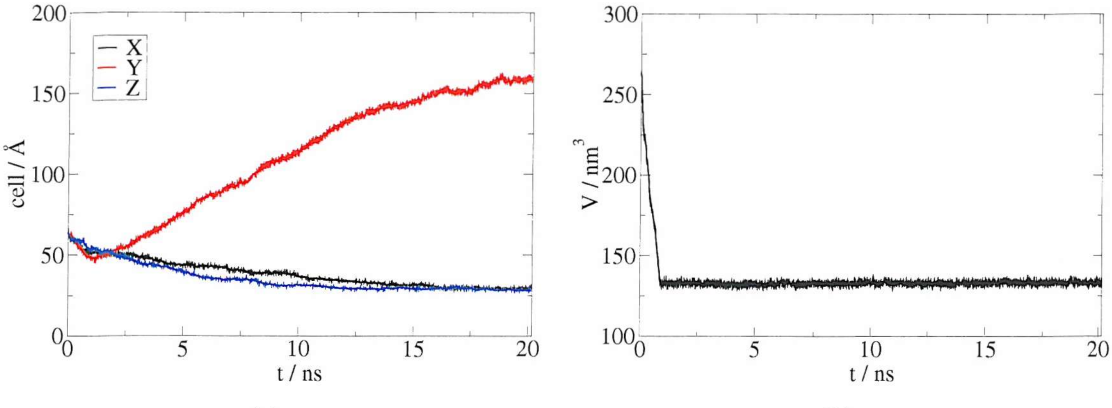
Figure 5.26: Box dimensions (a) and volume (b) time evolution for system SLF1, 64 lipids and 3000 water particles.

Table 5.4: Comparison of initial, expected and final box volumes / nm^3

	SLF1	SLF2
Initial volume	291.96	291.96
Expected volume[a]	167.39	131.52
Final volume	168.70	133.12

[a] The expected volume is calculated based on a volume of 147.96 nm^3 for a 72 lipids and 2095 water particles bilayer at 45 °C and a calculated water volume of 31.55 Å^3/molecule.

This fact is confirmed by the examination of the temporal evolution of the volume, illustrated in Figure 5.26 for SLF1 and in Figure 5.27 for SLF2. Although the box volume reaches equilibrium early in the simulation, the cell dimensions do not completly reach equilibrium at the end of the simulation. The equilibrium volume is reached in 0.6 ns for simulation SLF1 and in 0.9 ns for simulation SLF2. A consistent result considering the bigger difference between the expected and initial volumes for SLF2. That equilibrium volume is reached so quickly confirms the efficiency of the MC move routine for pressure and volume control. The cell evolution is characterized by the elongation of the box in one of the three direction, and the decrease of the remaining dimensions to compensate. This elongation happens along the x direction for simulation SLF1 and along the y direction for simulation SLF2. The head group particles and tail particles seggregate in both cases. Two layers forms, with head



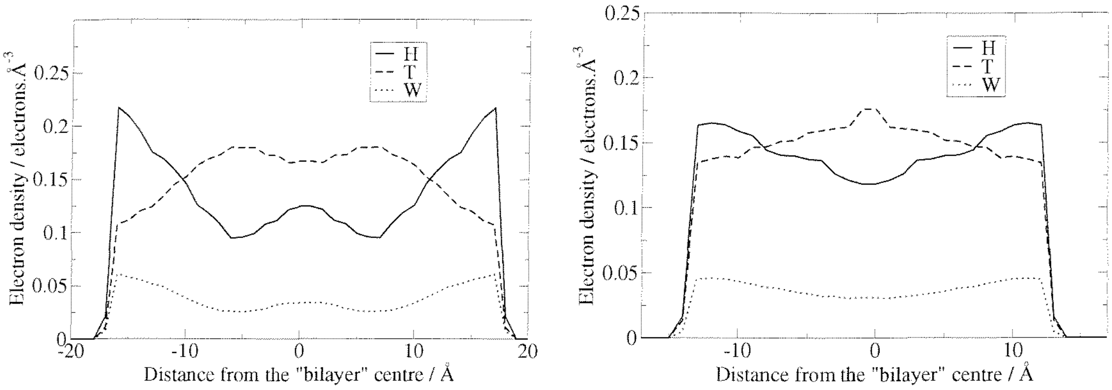
(a)

(b)

Figure 5.27: Box dimensions (a) and volume (b) time evolution for system SLF2, 64 lipids and 1863 water particles.

group of one layer strongly interacting with the head group of the opposite layer. This strong interaction is particularly visible when looking at an animation of the aggregation process. While tail particles posses a high mobility, head group beads hardly rattle. Some water particles also interacts with the head group and have similar mobility to the head group particles. The main defect of the final "bilayer" is the lack of water in between the two layers. A pseudo pore can also be observed at the interface with the water phase, where head group particles span across the entire width of the bilayer like structure. This can be observed in Figure 5.25h, and is present in both simulations. Thus although a perfect bilayer can not be obtained, a local bilayer-like structure can be identified. To characterize this structure we calculated electron density profiles around the location of the lipid phase. They are illustrated respectively in Figure 5.28a and 5.28b for system SLF1 and SLF2. The inclusion of the pores in the analysed region makes the separation of the structure less straighforward, increasing the head group and water density in the centre of the "bilayer". Nevertheless a local structure can be identified. The tail density is higher in the centre of the bilayer, while the water and head group densities are higher at the ends of the assembly. This trend is particularly marked for system SLF1, where even a drop in the tail density can be noted at the centre of the bilayer. The steep fall of density present at both ends is due to periodic boundary conditions and to the lack of hydration of the assembly.

Although a local bilayer-like structure has been identified, the self-assembly study



(a)

(b)

Figure 5.28: Local electron density profiles for system SLF1 (64 lipids and 3000 water particles) (a) and system SLF2 (64 lipids and 1863 water particles) (b).

presented here failed to isolate a correct phospholip bilayer structure. The aggregation process can be separated into two stages. A first rapid segregation of the order of three to four ns, where the water and lipid particles split into two distinct phases. The second stage consists of local reorganization of the lipid phase. At the end of the final stage, two phases are present, water and a low-hydration bilayer. The driving force of the reorganization seems to be the head group - head group interaction. These interactions are so large - in comparison to other interactions - that they prevent the correct hydration of the head group, and therefore the formation of a correct bilayer structure. To prevent this weakness of the model, the force field should ideally be reparameterized to balance more adequatly the interactions. The water - head group interactions should be as favourable as the head group - head group interactions and stronger than the water - tail interactions.

5.5 Conclusions

In this chapter the molecular dynamics simulation of a fully hydrated reduced representation of a DMPC phospholipid bilayer has been presented. These simulations have been performed at eight different temperatures ranging from 5 to 60 °C. The results obtained have been favourably compared to the available experimental data in most cases. Several physical properties of the bilayer have been analysed. The electron density profiles agrees with experimental data, the peak-to-peak distance in particular is in agreement with the value reported by Nagle *et al.*[129] The bulk water density is lower than it should be, a problem that could be solved by introducing a new water model. A phase transition in the hydrocarbon region from a highly ordered solid phase at low temperature to a disordered liquid-crystalline phase at higher temperature has been identified. The examination of radial distribution functions, orientational ordering as well as chain tilt indicate the presence of a phase transition in between 20 and 25 °C for the hydrocarbon region, in excellent agreement with the experimental temperature of 24 °C. Through the examination of the P-N dipole data, a phase transition was identified in the head group region, with the transition temperature located in between 25 and 30 °C, higher than the one identified in the hydrocarbon region. The diffusive behaviour of the membrane was also examined and proved to be in agreement with experimental data. Evidence of long-range lipid diffusion or "lipid swap" was observed. The self aggregation study failed to identify the formation of a bilayer of correct structure. Rather, the formation of a two phases system was observed, water and a low-hydration bilayer. Nevertheless, if our model is set up as a preassembled bilayer, its behaviour is in agreement with the available experimental data. The model's efficiency is sufficient to proceed to the study of the permeation of small solutes within this bilayer. Considering the simplicity of the model, the agreement with experimental data represents an achievement. The model developed while being simpler and potentially much faster than a conventional atomistic model still gives access to information at the molecular level.

Chapter 6

Small Molecule Permeation

This chapter presents the physico-chemical properties governing passive transportation and shows how permeability coefficients can be obtained experimentally. The techniques used to access permeation information through computer simulations are also reviewed. The method of choice is described in detail and its application to small molecules permeation across a *Gay-Berne membrane* is presented.

6.1 Solute Bilayer Interactions

6.1.1 Absorption

When considering oral absorption, the drug needs first to dissolve in the aqueous content of the gastrointestinal tract (dissolution) and then to cross the barrier of the gastrointestinal tract (membrane transfer) to reach the blood.[181] For a number of reasons membrane transfer can be limited and therefore absorption incomplete.

$$\boxed{\text{Solid Drug}} \xrightarrow{dissolution} \boxed{\text{Drug in Solution}} \xrightarrow[transfer]{membrane} \boxed{\text{Absorbed Drug}}$$

Dissolution depends on the surface area of the dissolving solid and the solubility of the drug at the surface of the dissolving solid. Processing and formulation of the compound play a key role in order to manipulate surface area. Solubility does not enter in the scope of this thesis, the drug will therefore now be considered to be in solution.

The barrier of the gastrointestinal tract is not different to any other that involves the crossing of a biological membrane. Biomembranes are made of lipid-bilayer with embedded proteins such as selective ion channels (Na^+, K^+, Ca^{2+}, Cl^-). Tight junctions between cells is insured by the interaction of membrane proteins at the contact surface. These pores have been estimated to have diameters in the range 3-10 Å, they only make up a small proportion of the whole surface, about 0.01 %. These tight junctions are in reality small aqueous filled pores.

The transport of drugs across the intestinal epithelium may occur by one or more of four routes: the carrier mediated route, by transcystosis, or the two passive routes. The first passive process by which a compound can cross a biological membrane is the transcellular mechanism where the compound has to diffuse through the core of the lipid bilayer. The second passive process is the paracellular mechanism, where the diffusion occurs through the aqueous filled pores. This process is slower due to the small surface area of the pores with respect to the membrane. Moreover this mechanism is very dependent on molecular size due the finite dimension of the aqueous pores.

The amount of drug absorbed (F_a) is dependent on the rate of absorption (k_a) and the rate of disappearance of the drug (k_d) from the absorption site which can be due to absorption or movement of the drug (k_m) away from the absorption site. The ratio absorbed can be expressed as:

$$F_a = k_a/k_d = k_a/(k_a + k_m). \tag{6.1}$$

The rates of absorption k_a and k_m depend on the nature of the drug. Lipinski[182] defined limiting values for good absorption of a compound by proving that poor permeability can be associated with:

- more than 5 H-bonds donors (expressed as the sum of OHs and NHs)
- molecular weight over 500
- Log P over 5

- more than 10 H-bonds acceptors (expressed as the sum of Ns and Os)

The 'rule of 5' is based on a distribution of properties among several thousands drugs. Some drugs will lie outside the parameter cutoffs in the rule; these drugs are believed to have structural features that enable them to act as substrates for naturally occurring transporters.

6.1.2 Models for Absorption Estimation

A variety of models have been suggested to estimate permeabilities in humans,[183, 184] varying from low throughput (*in situ* rat models) to high throughput (*in silico*) models. This section focuses on experimental techniques; *in silico* techniques will be detailed in the next section. The techniques used to estimate permeabilities depend on the size of the compounds considered. On one hand, the diffusion of small organics molecules occurs mostly through the paracellular route, and most techniques will therefore only give an estimate of the paracellular permeability. To access transcellular permeabilities of small organics special techniques are used. On the other hand, drugs diffusion because of their molecular size, is taking the transcellular route and most drug permeabilities will according to their size corresponds to transcellular permeabilities.

Small Organics

For small organic molecules, permeability coefficients can be measured across planar lipid bilayers formed on a $\approx 1\text{mm}R^2$ hole in a polyethylene or Teflon® partition separating two magnetically stirred chambers[185–192] , or across lipid bilayers forming large unilamellar vesicles.[191, 193–203] Solute fluxes from one chamber to the other can be measured using radioactive tracers, ^{1}H NMR measurements or fluorescence techniques, by collecting samples from the receiving compartment at regular interval of times.

The permeability of lipid bilayers is pH dependent, and hence for weak acids and bases, both the permeability of the neutral and ionized forms need to be considered. For electrolytes[186, 187, 196–199, 202–209] :

$$\mathcal{P}_{obs} = \mathcal{P}_n f_n + \mathcal{P}_i f_i \tag{6.2}$$

where $\mathcal{P}_{obs}$ is the observed permeability, $\mathcal{P}_i$ and $\mathcal{P}_n$ the respective permeabilities of the ionized and neutral forms, and f_i and f_n the respective fraction of the ionized and neutral species as governed by the solute pKa. Assuming that the ionized species do not cross the membrane $\mathcal{P}_i \approx 0$, and the permeability coefficient simplifies to the neutral species contribution.[186,187,204]

The overall membrane resistance to solute permeation, defined as the inverse of the permeability coefficient $\mathcal{P}$, can be expressed as the integral over the local resistances across the membrane :[202,205,207]

$$\mathcal{R} = \frac{1}{\mathcal{P}} = \int_0^d \frac{dz}{K(z)D(z)} \tag{6.3}$$

where $K(z)$ and $D(z)$ are the depth-dependent partition coefficient from water into the membrane and the diffusion coefficient in the membrane at depth z, respectively, and d is the membrane thickness.

When considering experimental results, a simplified model is used. The transport is considered to be governed directly by a distinct region within the lipid bilayer, and permeability is described by a simplified solubility-diffusion model:[186,187,197,202,205,207]

$$\mathcal{P} = \frac{K_{barrier} D_{barrier}}{d_{barrier}} \tag{6.4}$$

where $K_{barrier}$ and $D_{barrier}$ are the partition coefficient from water and the diffusion coefficient in the bilayer barrier, respectively, and $d_{barrier}$ is the barrier domain thickness.

For hydrophobic solutes K in the middle of the membrane is likely to be much greater than nearer to the water/lipid interfaces, while for hydrophilic solutes K is higher at the interfaces than in the membrane interior. The zone of maximum partition accounts only for a small fraction of the resistance to permeation and the zone of minimum partition accounts disproportionately for the resistance. Consequently, $K_{barrier}$ cannot be obtained from direct membrane/water partition experiments, as a solute tends to partition into that region with the lowest solvation energy rather than into the barrier domain where the solvation energy is the highest.[202,207] Therefore, to identify the barrier domain in biomembranes, one needs to search for a model solvent system which best describes the barrier domain for the permeant under study.

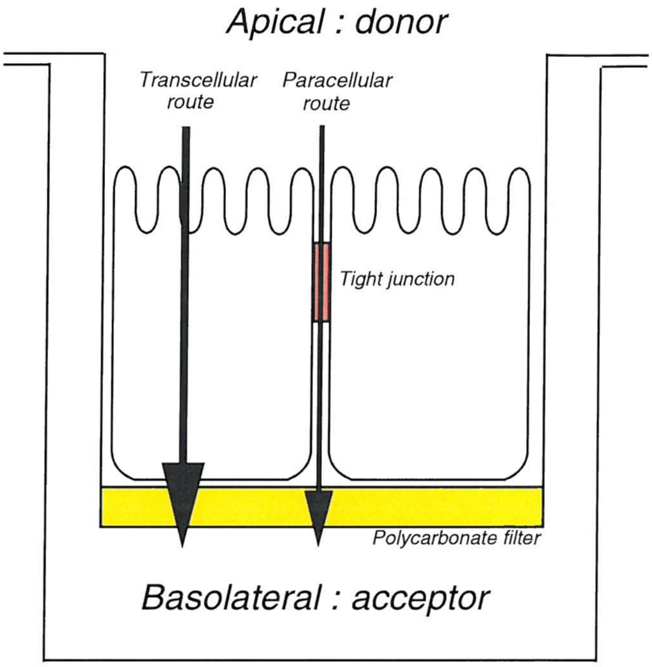


Figure 6.1: Measurement of permeability across Caco-2 cell monolayer.

Drugs

Drug permeability coefficients are not usually measured across isolated bilayers, but rather use more complex models that more closely correlate with *in vivo* absorption. The most common *in vitro* model to estimate absorption are Caco-2 cell monolayers.[210] Caco-2 cells originate from a human colorectal carcinoma that spontaneously differentiate into polarized monolayers with morphological and functional similarities to the small intestinal epithelium when seeded on a polycarbonate filter and allowed to grow on it for up to about 3-4 weeks. The basolateral side is toward the filter and the apical side toward the solution. The culture medium is usually maintained at 37°C and pH 7.4 to recreate physiological conditions. Permeability coefficients can be obtained using radiolabelled (^{14}C and ^{3}H) compounds, with ultraviolet spectroscopy, with liquid chromatography or with combinations of liquid chromatography and mass spectroscopy.

Rapidly and completely absorbed drugs, mainly hydrophobic drugs, distribute readily into the cell membrane. The transcellular surface area is >1000 fold larger than the tight junctions surface area. Hydrophobic drugs therefore diffuse mainly using the transcellular route. Slowly and incompletely absorbed drugs, mainly hydrophilic drugs, distribute poorly into the membrane, and it is generally assumed that they take the paracellular pathway. Nevertheless it has not been proven that they exclusively use this route. Indeed, the larger surface area of the transcellular route can compensate the poor partitioning of the drugs inside the membrane. Moreover the tight junctions marking the gate to the paracellular pathway restrict this transport to drugs with low molecular size. Compounds with molecular weights lower than 200 pass the intestinal mucosa using the paracellular route. The transcellular route can be used as well, but diffusion is mainly achieved through aqueous filled pores. Compounds with molecular weights between 200 and 500 cross the barrier by the transcellular route. With a molecular weight over 500 a decrease in membrane diffusion is observed. According to the molecular size and the hydrophilic/hydrophobic nature of the drug, one can predict the route of choice, this is summarized in Table 6.1.

Table 6.1: Possible passive routes

Molecular weight	< 200	200-500	> 500
Hydrophilic drugs	paracellular / transcellular	transcellular	small diffusion
Hydrophobic drugs	transcellular	transcellular	small diffusion

Assuming that only the neutral species can cross cell membranes, then:

$$\mathcal{P}_{cell} = f(\mathcal{P}^n_{trans} + \mathcal{P}^n_{para}) + (1 - f)\mathcal{P}^i_{para} \tag{6.5}$$

where $\mathcal{P}^n$ is the intrinsic permeability of the neutral species, $\mathcal{P}^i$ the intrinsic permeability of the ionized species, and f the fraction of the neutral species depending on solution pH and solute pK_a.

The first comparison of drug transport in cell monolayers and intestinal tissues[211] suggested that Caco-2 monolayers could be used to identify drugs with potential absorption problems. The effective permeabilities of lipophilic drugs differed 2- to 4-fold with Caco-2 monolayers. Hydrophilic drugs permeabilities are ranked properly using cell monolayers, but are transported at a 30- to 80-fold slower rate in Caco-2 monolayers. This large difference could be explained by fewer pores in the tight junctions in Caco-2 monolayers as compared with *in vivo* cells. Generally the permeability obtained using Caco-2 monolayers is comparable to *in vivo* permeability for high permeabilities ($> 1 \times 10^{-6}$) and lower than the *in vivo* data for low permeabilities($< 1 \times 10^{-7}$).[210]

Immobilized Artificial Membranes (IAMs) are also used as drug-membrane partitioning models. They do not give any information on permeability, since they only allow access to the water membrane partition coefficient $K_{water/membrane}$, and therefore do not enter in the scope of our study.

6.1.3 Computer Simulations of Drug Permeation

Stouch *et al.*[8] were the first to use MD simulations to study the diffusion of small solutes in lipid bilayer membranes. The first work concerned the simulation of benzene solutes within a bilayer of 36 DMPC lipid molecules at 320 K, above the phase transition (24 °C). Four simulations were performed, three of them contained a single solute, while the fourth contained four benzene molecules, to study the effect of increased benzene concentration. Simulation times varied between 500 ps and 1 ns. The benzene solutes were placed at various depths in the bilayer and were free to diffuse within the bilayer. The diffusion coefficients were calculated using the mean squared displacements of each molecule over all possible time origins ($\langle|\mathbf{r}(t)-\mathbf{r}(0)|^2\rangle$). The diffusion coefficients vary with the solute position. The benzene molecules placed in the centre of the bilayer diffuse 2-4 times faster than those close to the headgroup region, suggesting that membrane is not an homogeneous medium and cannot be treated as bulk phase. A more complete study of the diffusion data showed that benzene was experiencing two kinds of motion. Benzene rattles in a particular void for a relatively long period of time and takes infrequent jumps to another void. Experiments suggest that the voids through which the solute can move are more numerous in the centre of the bilayer. The frequency of the jumps is therefore greater in the centre of the bilayer where the voids are larger and more numerous, and correlates directly with higher diffusion coefficients. The bilayer structure was not significantly modified by the inclusion of benzene solutes; the bilayer thickness remain unchanged while the order parameter profile is only slightly modified. The benzene molecule rotated freely and experienced no preferred orientation. There was no indication of a preferred diffusion direction, either perpendicular or parallel to the bilayer normal. A further study of benzene diffusion at various temperatures[9] suggested that benzene prefers different locations at different temperatures. At 320 K, the benzene solutes show no regional preference, at 310 K they migrate to the centre of the bilayer, while at 340 K they reside mostly near the headgroup region. This behaviour is consistent with the distribution of free volume which concentrates at the centre of the bilayer at low temperature and becomes more diffuse at higher temperatures.

Stouch *et al.* also studied the orientation and diffusion of a drug analogue in biomembranes[10] - a nifepidine analogue which is a calcium channel antagonist and is based on a scaffold of two connected benzene rings. In contrast with the much smaller benzene, the rate of diffusion did not vary with location in the bilayer. Nifepidine cannot jump between voids due to its larger size, and hence its diffusion is not subject to the distribution of free volume inside the bilayer. Nifepidine motion was anisotropic with a preferred motion along the bilayer normal, whereas benzene motion was essentially isotropic. The drug-like molecule shows a preferred orientation when in the headgroup region. It is slightly tilted with repsect to the bilayer normal, in order to achieve hydrogen bonding with water.

The use of mean square displacements to calculate diffusion coefficients has been the subject of criticism. In the absence of constraints the solute changes depth and moves into a region where it will undergo different diffusive behaviour. The derivation of diffusion from mean square displacements is only valid at long times, and therefore it cannot be applied to small solutes without leading to poor statistics (the solute is not spending enough time at a particular location) or a false value (the solute is seeing different areas of the membrane).

Marrink and Berendsen[7, 12] studied the permeation of water, oxygen and ammonia across a united-atom bilayer. They deduced the permeation rate indirectly via computation of the free energy and diffusion rate profiles across the bilayer. Considering the homogeneous solubility-diffusion model was oversimplified, they derived an inhomogeneous solubility-diffusion model. The permeation resistance $\mathcal{R}$ (the permeability coefficient is defined as the inverse of the resistance) can be written:

$$\mathcal{R} = \int_{z_1}^{z_2} \frac{exp(\Delta G(z)/RT)}{D(z)} dz \tag{6.6}$$

where $D(z)$ is the local diffusion coefficient and $\Delta G(z)$ is the potential of mean force with respect to the bulk phase. The potential of mean force can be computed from simulations using different methods. Marrink and Berendsen used different methods in distinct regions of the membrane. The first method is *analysis of local densities*. This can only be applied to water, by counting the number of water particles per slice, thereby calculating a local partition coefficient $K(z)$ that allows the computation of the potential of mean force:

$$\Delta G(z) = -RTlnK(z). \tag{6.7}$$

This method, only applicable to water, is only valid for those regions of the membrane where there is enough water, it cannot for example, be used in the centre of the bilayer. The second method used is *particle insertion*, where a particle is inserted as a "ghost" without disturbing the configuration, randomly in the region of interest and determining its interaction energy E_{ins} with the "real" particles. From the insertion energy they calculated the insertion thermodynamic potential μ^{ins} which can be used to determine the potential of mean force. An advantage of this method is that it allows the determination of the enthalpic and entropic contributions. This method, because it uses small voids present in the membrane cannot be applied to large molecules. The last method, *average force on constrained particle*, is the most expensive, but the most general as well. The derivative of the potential of mean force can be determined by measuring the average force exerted on a particle constrained at a given depth z in the membrane. The computation of the local diffusion coefficients was determined with different methods according to the location in the bilayer. Computation from the *mean-square displacement* is straightforward but subject to criticism as the particle may not reside in the same region throughout the simulation. Using the *force autocorrelation method*, it is possible to relate the autocorrelation function of the random force acting on a molecule to the local time-dependent friction coefficient. Time integration gives the local static friction coefficient which can be related to the local diffusion coefficient through Einstein's relation.

The free energy profiles of water and ammonia calculated by the *average force on constrained particle method*, are very similar in shape. They increase in the interfacial region to reach a plateau in the centre of the bilayer. A dip is present in the centre of the profile for both molecules and corresponds to the larger free volume available. The height of the barrier is smaller for ammonia (≈ 10 kJ/mol), which is explained by the smaller dipole of ammonia ($\mu_{H_2O} = 2.27$ D and $\mu_{NH_3} = 1.47$ D). The slopes in the interfacial region are similar where one could have expected a steeper slope for water. This could be explained by the fact that water is able to form stronger hydrogen bonds with the headgroup. The excess free energy profile of oxygen is completely different, being a hydrophobic solute, it prefers dissolving in the membrane interior. The profile is complimentary to the accessible free volume profile of the membrane, with the potential of mean force decreasing as the solute goes towards the centre of the bilayer. The diffusion rate profiles of water, ammonia and oxygen show similar patterns. The highest diffusion rates are found in the centre of the bilayer, where the density is smaller whereas the lowest diffusion rates are in the interfacial region where the density is the highest. In the centre of the bilayer, ammonia is the penetrant that diffuses the fastest, and oxygen the slowest. This can be directly correlated with the permeant size, oxygen being the largest solute and ammonia the smallest. Knowing the excess free energy and the local diffusion rates, the local resistance to permeation can be calculated using equation 6.6. For all permeants, the shape of the resistance profiles is mainly dominated by the shape of the excess free energy profiles. Only the dip in the middle is enhanced by the fast diffusion in this region. For water and ammonia the resistance in the middle of the bilayer is greater than in the bulk water phase, since they are polar. The maximum resistance is obtained in the interface region where the density is the highest. In the case of oxygen, the maximum resistance to permeation is located in the water bulk phase. For this hydrophobic permeant, the membrane acts as an "accelerator" rather than as a barrier.

Pohorille *et al.*[13,212] used Molecular Dynamics to study the permeation of small solutes between water and simple membranes (a hexane slab or glycerol monooleate bilayer). The solute free energy of transfer from the water bulk phase to a depth z

inside the membrane is computed from the solute excess chemical potential $\Delta\mu^{\mathrm{exc}}$:

$$\Delta G(z) = \Delta\mu^{\mathrm{exc}}_{\mathrm{membrane}}(z) - \Delta\mu^{\mathrm{exc}}_{\mathrm{water}}. \tag{6.8}$$

The excess chemical potentials are calculated using Widom's particle insertion method.[213, 214] Pohorille *et al.* studied methane and fluoromethanes. For all of them they found that the excess chemical potential is lower in the membrane than in the aqueous phase.

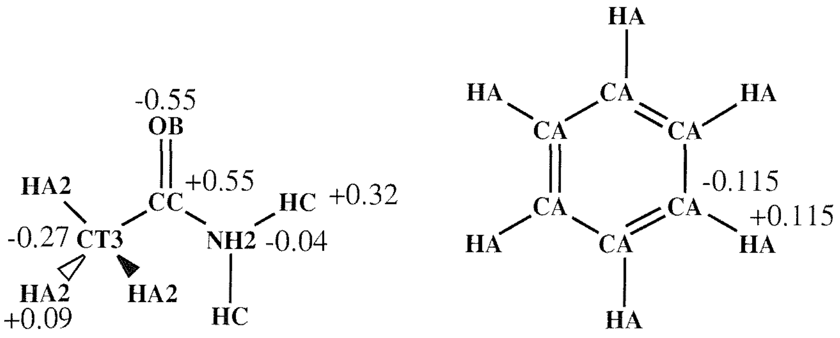
6.2 Solute Candidates

Seven solutes were initially chosen to represent most common functional groups. An identical study was performed by Daniele Bemporad[142] at the University of Southampton using an atomistic model and CHARMM. The same solutes were initially chosen to enable comparison between the results obtained with the simplified model and the atomistic. Parameterization problems have lead us to reduce this choice to four solutes. These are acetamide, benzene, ethane and methylacetate. This sets still covers a variety of functional groups, and contains two hydrophobic solutes (benzene and ethane), and two hydrophilic solutes (acetamide and methylacetate). These four solutes are therefore expected to have different behaviour inside the membrane. To facilitate comparisons with the atomistic simulations of Bemporad *et al.*, the same solute representation has been adopted. All of the solutes had tested parameters within the CHARMM27 force field, and these atomistic parameters have been used to derive our reduced representation parameters (cf. Section 6.3). In the atomistic simulations, the solutes proved not to undergo noticeable flexibility, they have therefore been held rigid in the present simulations. They are illustrated in Figure 6.2, together with labels and charges.

6.3 Solute Parameterization

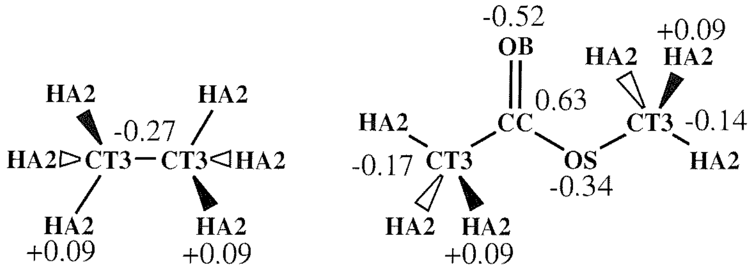
As the molecules were held rigid, intramolecular interactions were not considered. Interactions of the solutes with the water particles were represented with a *pseudo* atomistic potential, while interactions with the lipid were handled with the generalized Gay-Berne potential.

The solutes can interact with the water particles in a conventional fashion, using



(a) Acetamide

(b) Benzene



(c) Ethane

(d) Methylacetate

Figure 6.2: Solute candidates, labels, and charges taken from the CHARMM27 force field. All hydrogen atoms attached to the same atom have identical charges, hence only one charge is labeled in each case.

for example the non-bonded terms of the CHARMM27 force field. The nature of the water model meant that electrostatic interactions had to be treated slightly differently. The electrostatic nature of SSD water is accounted for by a dipole at the centre of mass of the water particle. Point charge - dipole interactions were therefore calculated using the Ewald sum in the fashion outlined in Section 4.4.3 for the lipid-water electrostatic interactions. van der Waals interactions were accounted for by a 12-6 Lennard-Jones potential. The water Lennard-Jones parameters are those described in the original SSD paper;[215] the solute Lennard-Jones parameters are taken from the CHARMM27 force field[114] and are described in Table 6.2. The electrostatic interactions between the solutes and the charged head group particles were calculated with an Ewald sum for point charge - point charge interactions. van der Waals interactions were handled by the generalized Gay-Berne potential - previously described in Section 4.4.1. The technique applied for the determination of the water generalized Gay-Berne parameters was applied identically to the determination of the solute parameters (cf. Section 4.4.2). For all solutes, each Lennard-Jones particle was used to determine the corresponding generalized Gay-Berne parameters. To optimize the parameters σ_0, ϵ_0, $\chi\alpha^{-2}$ and $\chi'\alpha'^{-2}$, the same protocol that was used to optimize the generalized Gay-Berne parameters in Section 4.2.4 is used. The generalized Gay-Berne interaction between a sphere and an ellipsoid is compared to the Lennard-Jones interaction between a Lennard-Jones particle and butane (RLJ4 representation). For each set of parameters - (σ_0,$\chi\alpha^{-2}$) for contact distance optimization, (ϵ_0,$\chi'\alpha'^{-2}$) for well depth optimization - an error function is calculated. This error function consists of the difference between the contact distance calculated with the generalized Gay-Berne representation and the Lennard-Jones representation, squared and summed over a range of relative orientations (the angle θ between the Gay-Berne particle orientation vector and the intermolecular vector is systematically varied between 0 and $\pi/2$). These error functions are described in detail in Section 4.4.2. The minima of the error functions indicate the best sets of parameters to use in the generalized Gay-Berne potential. The resulting parameters for all the atoms present in the studied solutes are detailed in Table 6.2.

Table 6.2: Lennard-Jones and resolved generalized Gay-Berne parameters for the solutes atoms

	ϵ / kcal.mol^{-1}	σ / Å	σ_0 / Å	ϵ_0 / kcal.mol^{-1}	$\chi\alpha^{-2}$	$\chi'\alpha'^{-2}$
	Lennard−Jones parameters		*Generalized Gay−Berne parameters*			
CA	0.0700	3.550	3.7714	1.3203	0.5714	0.3806
HA	0.0300	2.420	3.2000[a]	0.8406	0.6286	0.4005
CT3	0.0800	3.670	3.8286	1.4200	0.5643	0.3806
CC	0.0700	3.564	3.7714	1.3203	0.5714	0.3806
NH2	0.2000	3.296	3.6571	2.2400	0.5786	0.3924
OB	0.1200	3.029	3.5143	1.7200	0.5929	0.4006
OS	0.1521	3.154	3.5857	1.9600	0.5857	0.4006
HC	0.0460	0.400	2.2000[a]	0.9205	0.7400	0.4006
HA2	0.0220	2.352	3.1714[a]	0.7206	0.6286	0.4006

[a] The σ_0 values smaller than the one determined for the water lipid interaction ($\sigma_0^W = 3.30\,\text{Å}$) have been fixed to 3.30 Å to avoid overlapping charges.

The resulting parameters were examined for possible correlations against Lennard-Jones parameters. The parameterization was based on one Lennard-Jones sphere for the solute atom and on a butane molecule (RLJ4 representation) for the lipid particles. In all cases the lipid representation was identical and described in the same fashion, the only varying values were the Lennard-Jones parameters. The correlation were therefore performed between the generated generalized Gay-Berne parameters and the atom Lennard-Jones parameters. Contact distance and shape anisotropy (σ_0 and $\chi\alpha^{-2}$) were plotted against the Lennard-Jones contact distances (σ_{LJ}), while well depth and energy anisotropy (ϵ_0 and $\chi'\alpha'^{-2}$) were plotted against the Lennard-Jones well depths (ϵ_{LJ}). The results are displayed in Figure 6.3. Most generalized Gay-Berne parameters are strongly correlated to the atoms Lennard-Jones parameters. This is the case for the contact distance σ_0, the well depth ϵ_0, and the shape anisotropy $\chi\alpha^{-2}$. The correlation coefficients for these three fittings are in between 0.98 and 0.99, suggesting that parameters can be simply determined using linear relationships without the need for an error function calculation. The energy anisotropy $\chi'\alpha'^{-2}$ does not show any correlation with the Lennard-Jones well depth ($r^2 = 0.01$), a logical result considering that $\chi'\alpha'^{-2}$ only depends on $\epsilon_\perp/\epsilon_\parallel$ and μ.

The correlations obtained for parameters σ_0, ϵ_0, and $\chi\alpha^{-2}$, although only derived from a training set of 9 observables, are a clear indication that a simpler parameterization may be possible. These fits should be confirmed with a larger training set and a validation test. Owing to the nature of the generalized Gay-Berne potential the use of these parameters is nonetheless subject to caution. A small σ_0 value leads to simulation glitches when associated with a significant charges. This problem has been encountered in the water-lipid interaction parameterization. A small σ_0 leads to better structural results but caused the simulations to crash. The contact distance when set to a small value, enables the interacting particles to get in close contact, at a distance where electrostatic forces can become very large. These high forces when passed to the integrator, cause the integrator to fail. This phenomenon is further amplified in the solute case where point charges are used, in contrast to dipoles in the water model. In this study, three solutes that we planned to study have not been examined due to a combination of high charges and small contact distances which lead to simulations failing. For the remaining four solutes, atoms with σ_0 value lower than the distance established for water have seen their value fixed to $\sigma_0^W = 3.30\,\text{Å}$.

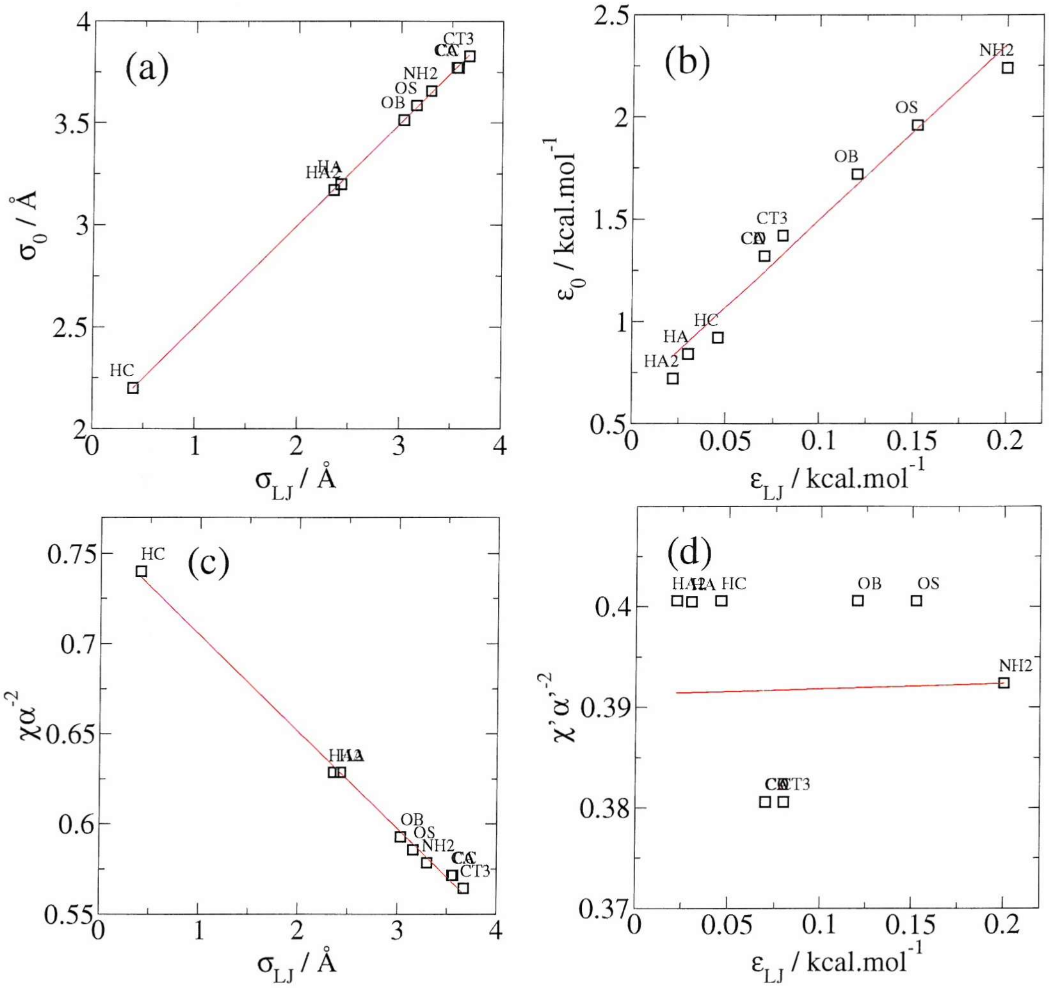


Figure 6.3: Linear regression between the generalized Gay-Berne (GGB) and the Lennard-Jones (LJ) parameters for the atoms present in the solutes. (a) GGB Contact distance vs. LJ contact distance ($\sigma_0 = 1.99 + 0.50\,\sigma_{LJ}, r^2 = 0.99$), (b) GGB well depth vs. LJ well depth ($\epsilon_0 = 0.63 + 8.62\,\epsilon_{LJ}, r^2 = 0.98$), (c) GGB shape anisotropy vs. LJ contact distance ($\chi\alpha^{-2} = 0.79 - 0.054\,\sigma_{LJ}, r^2 = -0.99$), and (d) GGB energy anisotropy vs. LJ well depth (no correlation). In the equations, contact distances are in Å and well depth in kcal.mol^{-1}

This tweak, although not strictly correct, avoids the systematic crash of the simulations. The use of the simple relationship derived from the regression, while being tempting, should be considered carefully as the parameterization technique does not account for charge effects. A possible solution to this problem would be to derive a parameterization technique accounting for electrostatic forces.

6.4 Permeation Calculations

Here we present the application of computer simulation techniques to the determination of the permeability coefficients of small solutes. The different techniques available to study permeability through computer simulations have been described in a previous section. Several methodologies are available for both the free energy and diffusion profile determination. To obtain a realistic description of the permeation process, different depths of the membrane have to be systematically sampled. The use of an unconstrained solute, diffusing freely in the membrane interior, is therefore excluded as the diffusion coefficients calculated are characteristic of different areas of the membrane. They cannot be related to a specific depth inside the membrane. The force autocorrelation method was therefore used for the diffusion profile determination. Considering the dimensions of the chosen solutes, the potential of mean force can in our case only be determined through the calculation of the average force on a constrained particle. The particle insertion method is limited to small solutes and the analysis of local densities is only applicable to water. The z-constraint method of Marrink *et al.*[7,12] was therefore used.

6.4.1 z-Constraint Algorithm

When studying solute permeation, the free energy difference associated with the process is a function of the depth z of the permeant inside the membrane. To determine this free energy change, a simple molecular dynamics simulation cannot be used. Classical molecular dynamics simulations only sample particular areas of the energy surface. The low energy regions are sampled more than the high energy regions, that account more importantly in the determination of the free energy barrier. To sample the entire membrane adequately, different characteristic depth of the membrane have to be screened. This is done by constraining the solute at different depths, ensuring that the information collected for a particular depth is characteristic of this depth. By measuring the average force exerted on the solute at each particular depth, the derivative of the potential of mean force can be determined. The method used to calculate the potential of mean force by constraining the solutes is referred to as the z-constraint method. This title is in our case misleading, as the bilayer normal lies along the y direction and not the z direction. The method should have been, in our case, more adequately named y-constraint.

When running a NPT simulation, four steps are required for the y-constraint algorithm, this reduces to two steps when running in the canonical or microcanonical ensembles. If we consider one solute located at depth $y(n)$ at a time step n, the two steps are as follows:

- Step 1: At time step n the solute centre of mass is located at depth $y(n)$ (with coordinates $x(n)$ and $z(n)$ in the plane of the bilayer). The integrator solves the equation of motion without any constraint and moves the solute to a new depth $y'(n)$ ($x'(n)$ and $z'(n)$ in the plane).
- Step 2: The y-constraint is applied. If the new depth of the solute is different from the previous depth i.e. if $y'(n) \neq y(n)$, the coordinates of the solutes are uniformly translated by $\Delta y = y(n) - y'(n)$ to reset the centre of mass at the desired depth $y(n)$. The permeant is free to rotate and its coordinates in the plane $x'(n)$ and $z'(n)$ are unmodified. The y component of the solute velocity is corrected.

As the permeants are, in this study, modelled as rigid bodies the y-constraint method is only applied to their centres of mass. The atom positions are resolved using a rotation matrix (quaternion leap-frog).

6.4.2 Computation of the Force

One single computation of the force acting on the solute enables both equilibrium and dynamic properties to be determined. The equilibrium properties consist of the solute partitioning behaviour and the dynamic properties are solute diffusion coefficients. The determination of the latter will be described in the following sections. In an unconstrained molecular dynamics simulation the total force acting on a body is equal to the opposite of derivative of the potential energy: $F_{tot} = -dV/d\mathbf{r}$. In dynamics with constraint, the total force is the sum of the term derived from the potential energy plus the constraint force:[28] $\mathbf{F}_{tot} = -dV/d\mathbf{r} + \mathbf{F}_c$. This can be incorporated into the integrator, leap-frog in the present study:

$$\begin{aligned}
\mathbf{r}_C(t+\Delta t) &= \mathbf{r}_C(t) + \mathbf{v}(t+\Delta t/2)\Delta t \\
&= \mathbf{r}_C(t) + [\mathbf{v}(t-\Delta t/2) + \mathbf{a}(t)\Delta t]\Delta t \\
&= \mathbf{r}_C(t) + \mathbf{v}(t-\Delta t/2)\Delta t + (\Delta t)^2\mathbf{F}_{tot}(t)/m \\
&= \mathbf{r}_C(t) + \mathbf{v}(t-\Delta t/2)\Delta t + (\Delta t)^2(-dV(t)/d\mathbf{r})/m + (\Delta t)^2\mathbf{F}_c(t)/m \\
&= \mathbf{r}(t+\Delta t) + (\Delta t)^2\mathbf{F}_c(t)/m
\end{aligned} \tag{6.9}$$

where t and $t+\Delta t$ are the old and new time steps, $\mathbf{r}(t+\Delta t)$ is the position the permeant would have assumed in the absence of constraint and $\mathbf{r}_C(t+\Delta t)$ is the position assumed by the solute after application of the constraint. If we define the displacement $\mathbf{S}(t) = \mathbf{r}_C(t) - \mathbf{r}(t)$ as the shift caused by the constraint, then the constraint force $\mathbf{F}_C(t)$ can be defined as:

$$\mathbf{F}_c(t) = m\frac{\mathbf{S}(t+\Delta t)}{(\Delta t)^2} = m\frac{\Delta y}{(\Delta t)^2}. \tag{6.10}$$

The y-constraint force is directly proportional to the solute displacement. It can be easily computed after the simulation, simply by writing the displacement of the solutes at each time step to a file. The same result is obtained with other integrators.

As previously described in the y-constraint methodology, the y component of the solute's velocity is set to zero after each time step. In the following $\mathbf{v}(t)$ and $\mathbf{v}_C(t)$ are the solute velocity before and after the application of the constraint. The constraint being only applied to the y direction:

$$\mathbf{r}^y_C(t+\Delta t) = \mathbf{r}^y(t+\Delta t) + \Delta y. \tag{6.11}$$

When a constraint is applied in a simulation - if we consider the constraint as being the distance between two particles r_{ij} - not only the square of the constraint must be constant, $\mathbf{r}^2_{ij} =$ constant, but its derivative as well:

$$\frac{d\mathbf{r}^2_{ij}}{dt} = 2\mathbf{r}_{ij}\mathbf{v}_{ij} = 0. \tag{6.12}$$

The relative velocities of the two constrained particles along the constraint vector must therefore be zero. The solutes centre of mass are in the present study constrained with respect to the membrane centre of geometry. The relative velocity of the solute with respect to the membrane centre of geometry must therefore be zero. The relationship found for the coordinates in equation 6.9 can be derived for the velocities. If the leap-frog algorithm is used:

$$\begin{aligned}\mathbf{v}_C(t+\Delta t/2) &= \mathbf{v}_C(t-\Delta t/2)+\mathbf{a}(t)\Delta t\\ &= \mathbf{v}_C(t-\Delta t/2)+\mathbf{F}_{tot}(t)\Delta t/m\\ &= \mathbf{v}_C(t-\Delta t/2)+(-dV(t)/d\mathbf{r})\Delta t/m+\mathbf{F}_C(t)\Delta t/m\\ &= \mathbf{v}(t+\Delta t/2)+\mathbf{F}_C(t)\Delta t/m\\ &= \mathbf{v}(t+\Delta t/2)+\mathbf{S}(t+\Delta t)/\Delta t\end{aligned} \tag{6.13}$$

In the y-constraint method, only the y-component of the velocity must be corrected:

$$\mathbf{v}_C^y(t+\Delta t/2)=\mathbf{v}^y(t+\Delta t/2)+\Delta\mathbf{y}/\Delta t \tag{6.14}$$

This correction ensures that the velocity of the solute centre of mass along y is zero relatively to the bilayer centre of geometry. This correction enables for the correct determination of the virial and kinetic energy.

6.4.3 Computation of the Potential of Mean Force

The computation of the average force on a constrained particle enables the potential of mean force to be determined. van Gunsteren[216] suggested a relationship between the average of the constraint force needed to keep the system on the hyperplane and the derivative of the free energy. He intuitively suggested that all that the constraint forces do during a constrained MD simulation is to counterbalance the thermodynamic force, and that therefore $dG(\xi)/d\xi$ should be equal to the constraint force; where ξ is the constrained variable, the depth y in our case, and G is the free energy. Even if this relationship does not hold true for all types of constraint, it is correct when geometrical constraints are used.[217–223] In our case, the derivative of the free energy difference is related to the average force required to constrain the solute at a particular depth y:

$$\frac{\partial G(y)}{\partial y}=-\left\langle\mathbf{F}(y)\right\rangle_t, \tag{6.15}$$

where $\left\langle\mathbf{F}(y)\right\rangle_t$ is the average of the instantaneous force $\mathbf{F}(y,t)$ acting on the solute at depth y and time t (the opposite of the constraint force), over time. The free energy of transfer from depth y_0 to depth y is defined as the potential of mean force:

$$\Delta G(y) = -\int_{y_0}^{y} \langle \mathbf{F}(y) \rangle_t \, dy. \tag{6.16}$$

The solute free energy is related to its partition coefficient $K(y)$, as described in Equation 6.7. In the present study, y_0 is located in the bulk water phase; the free energy of transfer from the bulk water to a depth y is therefore evaluated.

6.4.4 Computation of Local Diffusion Coefficients

From the computation of the force required to constrain the particle, one can also access dynamic properties, namely the diffusion profile of the permeant. It has been shown[224, 225] that the local time-dependent friction coefficient of the permeant molecule $\xi(t)$ is related to the time autocorrelation function of the fluctuation of the instantaneous force from the mean:

$$\xi(t) = \frac{\langle \Delta\mathbf{F}(y,t) \cdot \Delta\mathbf{F}(y,0) \rangle}{RT}, \tag{6.17}$$

where R is the universal gas constant and T the absolute temperature. The instantaneous fluctuation $\Delta\mathbf{F}(y,t)$ is defined as:

$$\Delta\mathbf{F}(y,t) = \mathbf{F}(y,t) - \langle \mathbf{F}(y) \rangle_t \tag{6.18}$$

The static friction coefficient $\tilde{\xi}$ is defined as the Laplace transform of the time-dependent friction $\tilde{\xi}(s)$ in the particular case where $s = 0$:

$$\tilde{\xi}(s) = \int_0^{\infty} \xi(t) e^{-st} dt \tag{6.19}$$

$$\tilde{\xi} = \int_0^{\infty} \xi(t) dt = \int_0^{\infty} \frac{\langle \Delta\mathbf{F}(y,t) \cdot \Delta\mathbf{F}(y,0) \rangle}{RT} dt \tag{6.20}$$

The static friction coefficient is a good measure of the dynamics of the system if non-Markovian behaviour and memory effects become negligible, i.e. if the time-dependent friction coefficient $\xi(t)$ is large and decays rapidly compared to other time scales in the system. This condition is met if the the slope of the free energy barrier over a distance covered by the particle during the decay time of its friction coefficient is less than the thermal fluctuation energy kT. In other words the static friction coefficient can be related to the local diffusion coefficient if the permeant remains in a region of constant free energy:

$$D(y) = \frac{RT}{\tilde{\xi}} = \frac{(RT)^2}{\int_0^\infty \langle \Delta\mathbf{F}(y,t) \cdot \Delta\mathbf{F}(y,0) \rangle \, dt}. \tag{6.21}$$

6.4.5 Computation of the Permeation Coefficient

Having computed the free energy $\Delta G(y)$ and diffusion profile $D(y)$, the overall permeation resistance $\mathcal{R}$ - defined in equation 6.6 - is obtained using the solubility-diffusion model by integrating over the local resistance $\mathcal{R}(y)$:

$$\mathcal{R} = \int_{y_1}^{y_2} \mathcal{R}(y) dy = \int_{y_1}^{y_2} \frac{\exp\left(\Delta G(y)/RT\right)}{D(y)} dy = \frac{1}{\mathcal{P}}. \tag{6.22}$$

The permeability coefficient $\mathcal{P}$ is defined as the inverse of the permeation resistance.

6.5 Simulation Protocol

In this study, $\Delta G(y)$, $D(y)$ and P were calculated for four small solutes: acetamide, benzene, ethane and methylacetate. For each of the four solutes, ten depths (y coordinates) were sampled from 0 to 30.5 Å from the bilayer centre. The solutes were constrained at each of these ten depths. Results from one leaflet were considered valid for the other layer too, for reasons of symmetry. For each depth y, four xz positions on the bilayer plane were sampled. Hence, $10 \times 4 = 40$ different positions are sampled for each solute. Five depths can be sampled in each simulation, and eight simulations are therefore required for each molecule. For each solute, eight simulations were therefore performed sampling ten depths and four xz positions on the bilayer plane. In each simulation five identical solutes were constrained at different depths. They were at least 6 Å apart from each other along the bilayer normal, but in most cases they were more than 12 Å apart. In each simulation, the five identical solutes had different xz coordinates on the plane of the bilayer. These precautions were taken to avoid solute clustering. This clustering, while possible, is unlikely to happen in our simulations as, there is no solute-solute interactions. All simulations have been run with a reduced time step $\Delta t^* = 0.003$ ($\Delta t^* = \Delta t\sqrt{\epsilon_0/m\sigma_0^2}$) at the reduced temperature of $T^* = 0.420$ ($T^* = k_B T/\epsilon_0$). This corresponds to a real time step of approximately 3.7 fs and a temperature of 30 °C. Correct solute masses and inertia tensors were used. The final structure of a 7.2 ns membrane simulation (described in Chapter 5) was used as the initial structure.

The solutes are inserted at the correct depths. Their Lennard-Jones well depths ϵ_i^{LJ}, generalized Gay-Berne well depths ϵ_i^{GGB}, charges q_i, as well as coordinates with respect to the centre of geometry $\mathbf{r}_i$ are initially scaled by a factor λ. Hence, if λ is fixed to 0.5, the solute will have half of its normal size, and all interactions with its environment will be scaled by 0.5. Three steps are required to generate the initial structure:

- Step 1: λ is fixed to 0.00001, and the system is equilibrated in a 1000 step simulation in which the simulation temperature must be stable to within 0.01. If this criterion is not met, the velocities are again rescaled and the equilibration phase continues until such time as a stable temperature is maintained.
- Step 2: λ is allowed to grow linearly in between 0.00001 and 1.0 in a 3000 steps simulation. No control is exerted on the temperature.
- Step 3: λ is fixed to 1.0 in a 10 000 steps simulation in which the simulation temperature must be stable to within 0.01. If this criterion is not met, the velocities are again rescaled and the equilibration phase continues until such time as a stable temperature is maintained.

The initial step allows the solute to be placed inside the membrane without using any energy minimization technique. The solute almost corresponds to a single-point; its width corresponds to the largest Lennard-Jones or generalized Gay-Berne contact distances (which have not been scaled). The solute is therefore easily inserted inside the membrane. The second step enables the solute to grow to its normal size, and in the mean time allows the membrane to deform to accommodate the inserted solute. The last steps ensures that the temperature of the initial structure is correct prior to production runs.

Production simulations consisting of 700 000 steps are then performed, with velocity rescaling every 100 steps. Membrane translation is also performed every 100 steps, to avoid any systematic drift. The centre of geometry of the membrane is calculated, and all particles are translated such as the new centre of geometry y coordinate is at the centre of the simulation box. The centre of mass motion is also reset. The protocol used is essentially identical to the one used for the preassembled bilayers simulations (cf. section 5.1). Each simulation was approximately 2.5 ns long and performed on a cluster of Linux PCs with 1400 MHz AMD Athlon processors. For each molecule, $\mathcal{P}$ was calculated from 20 ns of simulations and for the four molecules, a total of 80 ns of data collection was therefore performed.

6.6 Simulation Results

6.6.1 Free Energies

Free energy profiles $\Delta G(y)$ for the permeants studied are presented in Figure 6.4. They have been calculated using equation 6.16. The errors in $\Delta G(y)$ are calculated by propagating the errors in the force $\mathbf{F}(y)$. The latter are standard errors calculated from the difference of the mean force $\langle \mathbf{F}(y) \rangle_t$ to the force from each of the four individual simulations at each depth.

For all solutes, a general trend can be highlighted, $\Delta G(y)$ decreases moving from the water phase into the membrane. This systematic drop in free energy takes place between depths 30.5 Å and 16.5 Å. This region is dense and corresponds to the head group location. The associated free energy difference is logically higher for the most polar compound, acetamide with $-45\,\text{kcal.mol}^{-1}$, and less pronounced for the most hydrophobic compound, benzene with $-27\,\text{kcal.mol}^{-1}$. Ethane and Methylacetate have intermediate behaviours with free energy differences of respectively -40 and $-28\,\text{kcal.mol}^{-1}$. Bemporad *et al.*[142] have studied the permeation of small solutes across a DPPC bilayer using atomistic simulations. The solutes examined here entered the scope of their study. In their analysis, the permeants have a different behaviour when entering the membrane; their free energies increase systematically between 30.5 Å and 20.0 Å, with free energy differences in the range $\approx$ 2 to $5\,\text{kcal.mol}^{-1}$.

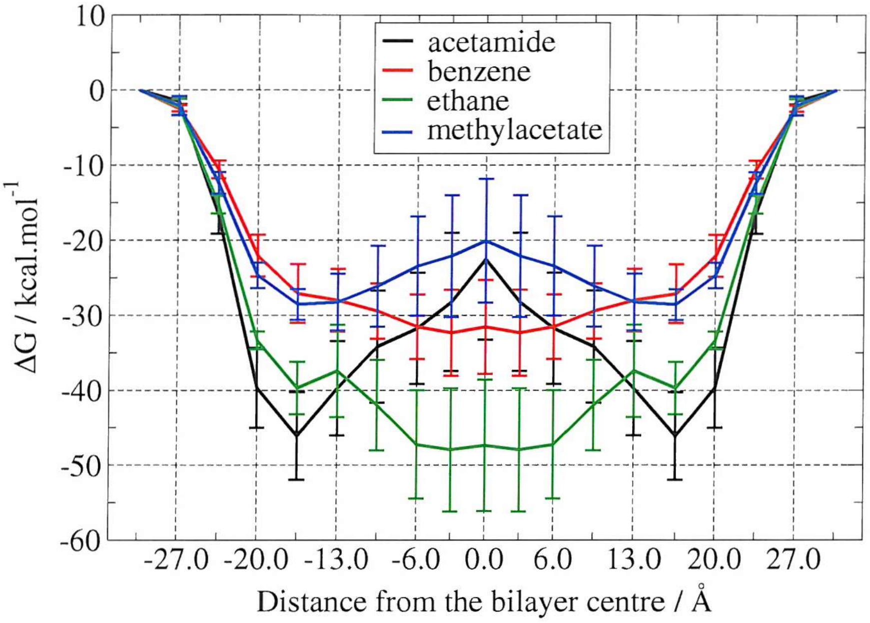


Figure 6.4: Free energy profiles. Error bars are standard errors calculated from the difference of the force in the four individual simulations from their average.

They explained this phenomenon by the increase in density in this region. Solute solubilization becomes more difficult as the free volume available to permeants becomes smaller. The presence of hydrogen bonds with water, that need to be displaced to allow solute permeation is another explanation put forward by Bemporad *et al.*

Different facts can explain the systematic $\Delta G(y)$ decrease of solutes entering the membrane in our study. One possibility is that the size of the head group particles is not sufficiently large. The free volume distribution has not been examined in this study, due to complications arising from the ellipsoidal shape of the Gay-Berne particles. In the absence of this data it is difficult to assert whether or not head group particles play any role as a "steric wall", as they should do. Nevertheless, the data presented in Figure 6.4 is consistent with the hypothesis advanced. In the presence of free volume, and due the charged nature of the head group particles, hydrophilic compounds (acetamide for example) are logically more attracted than hydrophobic compounds (benzene for example). If, as we suspect, the size of the head group particles is responsible for this anomalous behaviour, a new force field should be adopted. In our parameterization of van der Waals interactions all Gay-Berne particles are equivalent. Head group and tail particles have the same size, shape anisotropy ($\sigma_{\parallel}/\sigma_{\perp}$) and energy anisotropy ($\epsilon_{\perp}/\epsilon_{\parallel}$). To account for the differences in size of particles from different areas of the membrane, the generalized Gay-Berne potential should be implemented and parameterized for lipid-lipid van der Waals interactions. Another explanation would be the role played by the solvent. The SSD model of water has shown some limitations in the pure bilayer studies presented in Chapter 5. The bulk water density was lower than expected ($0.31\,\mathrm{e.\AA^{-3}}$ in our simulations vs. $0.33\,\mathrm{e.\AA^{-3}}$ experimentally). The role of hydrogen-bonding or an inaccurate representation of water-solute interactions would explain an higher affinity of the solute for the head group rather than for the solvent phase. It is impossible to tell if the initial drop in free energy is due to the size of the head group particles, the model used for water or a combination of these two effects.

Inside the bilayer (for those depths where $|y| \leqslant 16.5\,\text{Å}$), the four solutes have different behaviours. The free energies of acetamide and methylacetate, which are hydrophilic compounds, increases continuously from $-16.5\,\text{Å}$ to the centre of the bilayer, while it decreases continuously for benzene (with the exception of a small peak in the centre of the bilayer). Ethane, while being a hydrophobic compound, has a slightly different behaviour than benzene. The free energy increases by $\approx 2\,\mathrm{kcal.mol^{-1}}$ at $y = 13\,\text{Å}$, before decreasing to reach a minimum at $y = 3\,\text{Å}$. A small peak is present at the centre of the bilayer approximately $0.6\,\mathrm{kcal.mol^{-1}}$ higher than the minimum. This difference with benzene cannot be accounted for by size considerations as benzene is larger than ethane. The charges on ethane carbon atoms are larger than the one on the aromatic carbon atoms of benzene. Electrostatic repulsions could mean than ethane is less attracted to this intermediate region (location of the glycerol particles) than benzene. For benzene the free energy continues to decrease and reach $-32\,\mathrm{kcal.mol^{-1}}$ at the bilayer centre.

In the presence of a systematic free energy drop in the region $y = 30.5$ to $16.5\,\mathring{A}$, our profiles are difficult to compare to those obtained by Bemporad *et al.* Atomistic simulations[142] have shown that this region does not play a critical role for the differentiation of the solute behaviour; for all solutes there is a systematic increase of the free energy. The determinant region is the bilayer interior. We therefore decided to compare the free energy profile in the region $|y| \leqslant 16.5\,\text{Å}$ obtained in our simulations, to the atomistic profiles. This profile is presented in Figure 6.5. It was obtained by considering only forces for depths below $|y| = 16.5\,\text{Å}$ (included), and integrating them. The systematic decrease of free energy observed in our study is present between

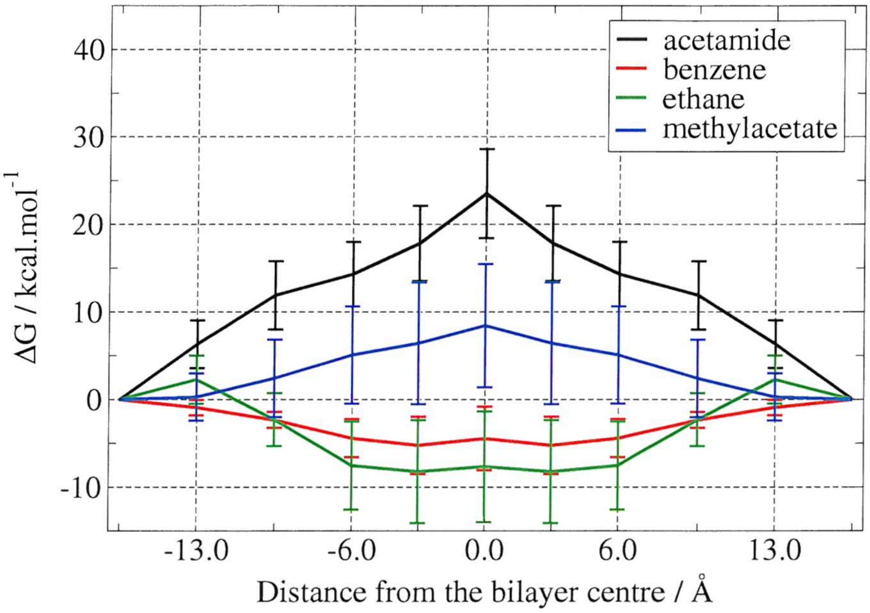


Figure 6.5: Free energy profiles in the region $|y| \leqslant 16.5\,\text{Å}$. Error bars are standard errors calculated from the difference of the force in the four individual simulations from their average.

bulk water and depth $y = 16.5\,\text{Å}$ inside the bilayer. This depth corresponds to the centre of the head group region. As our model mimics a DMPC bilayer, and as the membrane studied by Bemporad *et al.* was made of DPPC, the location of the head group centre was in their study at a different depth. We therefore compared the free energy profile between depths $y = -16.5$ and $16.5\,\text{Å}$ in our study to the free energy profile between depths $y = -20.0$ and $20.0\,\text{Å}$ in the study of Bemporad *et al.*

Acetamide behaviour in the membrane interior is similar in both studies. In the Gay-Berne membrane, acetamide's free energy increases continuously to reach a value of 23.5 kcal.mol^{-1} in the centre of the bilayer. In the atomistic study, the free energy increases from 5 kcal.mol^{-1} at $y = 20.0\,\text{Å}$ to 25 kcal.mol^{-1} at the centre of the bilayer, a net increase of 20 kcal.mol^{-1}. Both profiles are similar, the only difference being the absence of a slight free energy decrease at the bilayer centre in the reduced representation study. Acetamide's behaviour is logical, being a hydrophilic compound, the membrane hydrophobic core is the main barrier to permeation.

Benzene and ethane behaviours are similar in both the atomistic study of Bemporad *et al.*, and in our study. The free energy decreases continuously from the head group centre to the centre of the bilayer. Drops of ≈ 5 and $7.5\,\text{kcal.mol}^{-1}$ are observed in our study for respectively benzene and ethane, similar to the ≈ 7 and $11\,\text{kcal.mol}^{-1}$ decrease reported in the atomistic study. These two compounds are hydrophobic, and their partition inside the interior of the membrane is therefore favoured.

In the atomistic study, methylacetate's free energy initially decreases ($3\,\text{kcal.mol}^{-1}$), before increasing by $1\,\text{kcal.mol}^{-1}$, from $y = -9.5\,\text{Å}$ to the bilayer centre. In our study a completly different behaviour is observed, the energy increases continuously to reach $8.5\,\text{kcal.mol}^{-1}$ in the centre of the membrane. While being hydrophilic, methylacetate does not possess any hydrogen bonds donors and therefore cannot form any H-bonds with the surrounding lipids. This fact was put forward to explain methylacetate behaviour. In our study, methylacetate behaves similarly to the other hydrophilic solute. The nature of the lipid particles in our study makes the representation of H-bonds difficult to achieve; in this context acetamide and methylacetate behaviours are logically similar as no differentiation is brought by H-bonds.

The free energy behaviour is sensible in the centre of the bilayer, a region fre of short-ranged electrostatic interactions. This suggests that both van der Waals (generalized Gay-Berne potential) and long range electrostatic interactions have been parameterized appropriately. The problem encountered in bulk water and in the interface region might therefore arise from short range electrostatic interactions, with water and head group particles. The charge - dipole interaction might be an issue, not helped by the head group size. Indeed, solute parameterization highlighted that the association of a significant charge with a small generalized Gay-Berne contact distance could lead to simulation glitches. Moreover, the self aggregation study presented in Chapter 5 highlighted problems with the water - lipid interaction balance. Head group - head group interactions were essentially stronger than water - head group interactions, a weakness that may be associated with the water model, and possibly specifically to the handling of water electrostatic interactions with other particles. The free energy behaviour observed in the permeation study can therefore be partially explained by the way short range electrostatic interactions between water, head group and solute molecules are treated.

6.6.2 Diffusion Coefficients

The diffusion coefficients were obtained from the force time autocorrelation function, using equation 6.21. To improve the statistics, autocorrelation functions are generally averaged over all possible time origins. Even though this greatly improve the statistics, limitations in computer power and in the time scale simulated means that the final functions are still noisy. To derive diffusion coefficients, the time autocorrelation function of the force fluctuation needs to be integrated over time. A noisy curve would lead to large errors in the integral estimate. Standard autocorrelation functions are therefore often fitted to an exponential curve, the resulting fit is then integrated, giving an answer that is less subject to noise.

In our study, the force time autocorrelation time had to be fitted to a double exponential:

$$C(t) = C_1 \exp(-\mathrm{t}/\tau_1) + \mathrm{C}_2 \exp(-\mathrm{t}/\tau_2) \tag{6.23}$$

where $C_1 + C_2$ is the value of the force autocorrelation function at $t = 0$, and τ_1 and τ_2 are relaxation times, corresponding to two distinct motion time scales. The motions with a short decay time have typical τ_{short} between 0.04 and 0.07 ps depending on the solute and location in the bilayer. The smaller decay times are observed for acetamide and the largest for the remaining solutes. These decay times correspond to the response of the solutes to the local environment, i.e. to the friction with the surrounding lipids that keeps the permeants in their local cages. The long decay times are plotted in Figure 6.6. Errors in τ_{long} are standard errors calculated from the difference of the decay time in each of the four individual simulation at each depth from their average. Typical long decay times vary between 3 and 36 ps. They gradually increase when the solute enters the membrane. These decay times are associated with the overall permeant diffusion mechanism through the membrane. If the diffusion is assumed to be achieved through successive hops between free volume cages, τ_{long} corresponds to the residence time of the solutes inside their cage. Acetamide, benzene and methylacetate all have smaller long decay times at the centre of the bilayer than at $y = 3\,\text{Å}$. This result can be explained by the largest free volume available to permeants at the

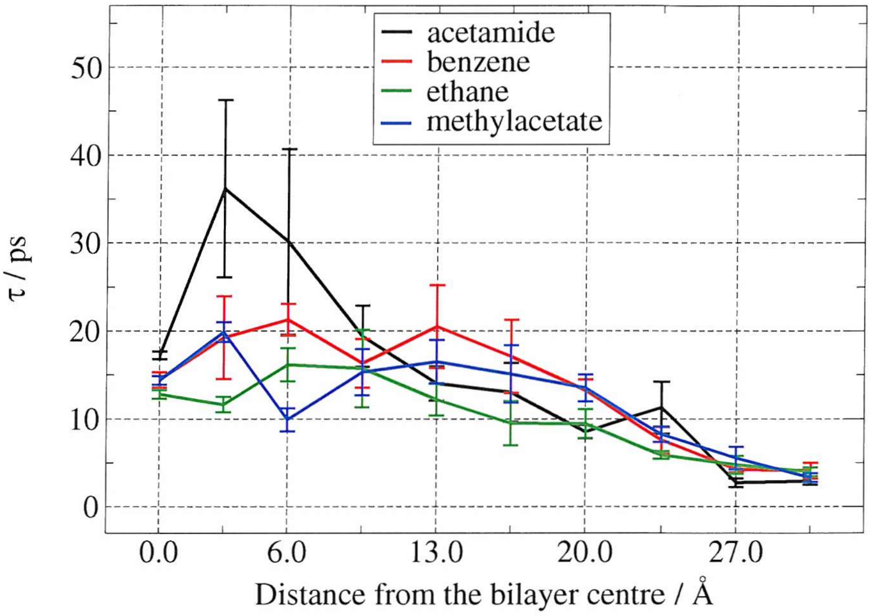


Figure 6.6: Long friction relaxation time τ_{long}.

centre of the bilayer. Residence times are smaller as more voids are present and jumps between cages therefore more frequent. The largest τ_{long} is obtained for acetamide at $y = 3$ Å, a logical result as acetamide is the most hydrophobic compound in this study.

Diffusion coefficients $D(y)$ for the four studied solutes, were calculated with equation 6.21 and the fitting procedure previously described. Errors in $D(y)$ are standard errors calculated from the difference of the diffusion coefficients in each of the four individual simulation at each depth from their average. The resulting diffusion coefficients are plotted in Figure 6.7. The observation of the four solute profiles does not allow to any systematic trend to be highlighted, as was the case in the atomistic model of Bemporad *et al.*[142] In their study, for all compounds, the diffusion coefficients decrease on entering the membrane, are constant below the head groups (20 Å), and slightly increase in the bilayer centre. In our study, most solute diffusion coefficients increase at the bilayer centre, as reported in the atomistic examination, nevertheless the diffusion coefficients in bulk water phase are systematically too low. To understand this anomalous diffusive behaviour, the solute diffusion coefficients in bulk water were investigated.

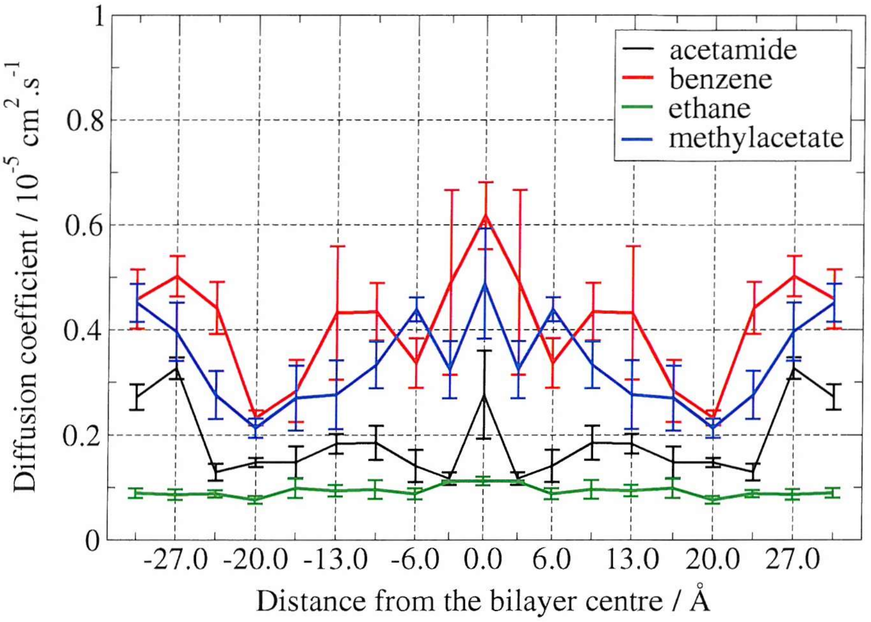
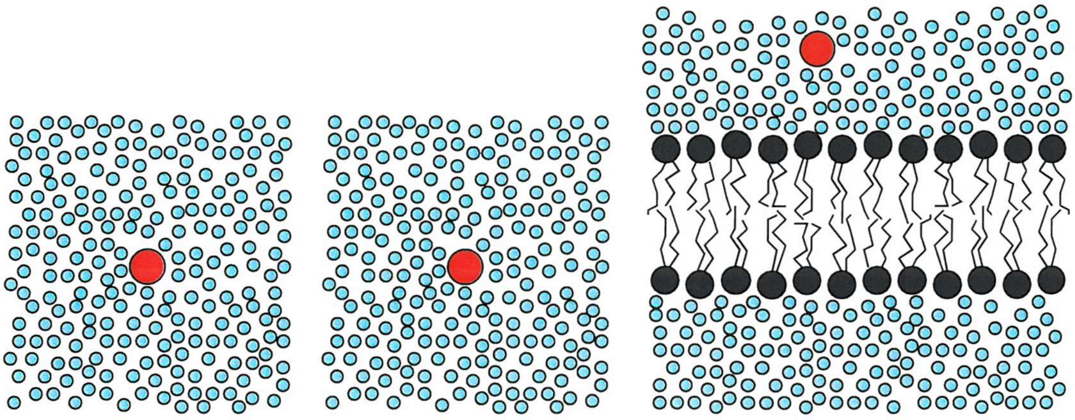


Figure 6.7: Diffusion coefficient profiles. Error bars are standard error calculated from the difference of the diffusion coefficients in the four individual simulations from their average.

Solute Diffusion in Bulk Water

For simplification purposes, the masses of the water particles have been set to unity, meaning that Gay-Berne and water particles have the same masses, an assumption that greatly overestimates water masses. Similarly, the water moment of inertia was given the same values as Gay-Berne particles. While these assumptions do not affect the thermodynamic behaviour of water, they greatly affect its dynamic and diffusive behaviour. These simplification have been implemented to control time step. To understand how this change can alter solute diffusion coefficients in bulk water, we calculated solutes diffusion coefficients in bulk water with different conditions. In the first sets, masses were set to unity, and inertia tensors were identical to those of Gay-Berne particles, while in the second sets water were given correct masses and inertia tensors. In both cases, the solute were constrained at $y = 0.0$ Å, and diffusion coefficients were calculated using Equation 6.21. Simulations were performed using *CHAMPAGNE* and a pre-equilibrated box of water containing 750 molecules, at the reduced temperature $T^* = 0.42$, corresponding to 30 °C. Simulations were performed in the microcanonical ensemble, with periodic velocity rescaling to control temperature (every 100 steps), for 500 000 steps. The time step was approximately 4.8 fs, leading to a simulation length exceeding 2.4 ns. The resulting diffusion coefficients



(a) Set 1 (b) Set 2 (c) Set 3

Figure 6.8: The three different cases used for the solutes' diffusion coefficient determination. In sets 1 and 2 the solutes (red circles) are constrained at depth $y = 0$ in a box of pure water. In set 1, water masses and inertia tensors are identical to those of Gay-Berne particles, in set 2 the correct masses and inertia tensors are used. In set 3 the solutes are constrained at depth $y = 30.5\,\text{Å}$ in the bilayer (corresponding to bulk water) and water masses and inertia tensors are identical to those of Gay-Berne particles.

are presented in Table 6.3, together with the diffusion coefficients obtained in the permeation study at $y = 30.5\,\text{Å}$ (bulk water). The three different cases used for the solutes' diffusion coefficient determination are summarized in Figure 6.8.

Table 6.3: Diffusion coefficients in bulk water / $10^{-5}\,\text{cm}^2.\text{s}^{-1}$. Set 1 corresponds to masses and inertia tensors used in permeability studies, set 2 to correct masses and inertia tensors and set 3 to the diffusion coefficients obtained for the solutes constrained at depth $y = 30.5\,\text{Å}$ (bulk water) in the membrane simulations. These sets are illustrated in Figure 6.8.

	Set 1	Set 2	Set 3	Experimental
Acetamide	0.518	1.823	0.272 ± 0.042	1.32 ± 0.03 (25 °C)[226]
Benzene	1.011	1.404	0.459 ± 0.097	≈ 1.1 (25 °C)[227]
Ethane	0.261	1.718	0.089 ± 0.016	1.52 ± 0.03 (25 °C)[228]
Methylacetate	0.314	1.392	0.451 ± 0.063	-

The modification of masses and inertia tensors leads to a systematic underestimate of diffusion coefficients. This is especially the case for ethane and hydrophilic compounds, whose values are three to six times smaller with the masses and inertia tensors used in permeability studies. Surprisingly, benzene diffusion coefficient is underestimated to a lesser extend (1.4 times smaller). The diffusion coefficients obtained with set 2 are close to the values determined experimentally. The ranking between ethane and acetamide is not reproduced, but values reported for set 2 and experimental values are of the same order of magnitude. Values obtained in the permeation studied for solutes constrained in bulk water are systematically underestimated.

Corrected Diffusion Coefficients

Water particles were given the same masses and inertia tensors as Gay-Berne particles. This simplification was essentially implemented to control time steps. While not affecting thermodynamic behaviour, this modification leads to a systematic underestimation of the diffusion coefficient in the presence of water. To balance this effect, empirical corrections have been introduced. It should be noted that solute molecules were given their correct masses and inertia tensors; the low values obtained for the diffusion coefficient in the presence of water are therefore only due to their environment and are not intrinsic.

Starting from the observation that the modification of water properties had an effect on the solutes' diffusion coefficients, we considered that this effect was proportional to the water density. Indeed, inside the bilayer, where no water is present, the solute behaviour is only the consequence of interactions with surrounding lipid particles, whereas in bulk water it is mainly affected by water molecules. In the interface its diffusion coefficient is the consequence of interactions with both lipid and water particles.

Table 6.4: Correcting factors used in the corrected diffusion profile determination

y / Å	Water density / atoms.nm^{-3}	Water density / %	Weight w
0.0	0.01	0.01	0.00
3.0	0.03	0.03	0.00
6.0	0.26	0.3	0.00
9.5	1.35	1.4	0.00
13.0	12.73	13.5	0.15
16.5	30.06	31.9	0.30
20.0	54.50	57.9	0.60
23.5	87.92	93.4	0.95
27.0	93.51	99.4	1.00
30.5	94.09	100.0	1.00

We therefore considered that solute diffusion is the result of two contributions, one from the solvent, one for the lipids, and that these contributions are directly proportional to the atom densities. Corrected diffusion coefficients $D_C(y)$ were derived from the calculated diffusion coefficients $D(y)$. Diffusion coefficients were considered to be the sum of two contributions:

$$D(y) = D_{lipids}(y) + w(y).D(y = 30.5), \tag{6.24}$$

the contribution from the lipids depends on the location inside the bilayer, the contribution from the water is considered to be the bulk water solute diffusion coefficient weighted by a factor proportional to the water atom density ($w(y)$). The weights are described in Table 6.4; they are directly proportional to the water atom density presented in Chapter 5. As bulk water diffusion coefficients are underestimated, as illustrated in the previous section, the water contribution is removed and replaced by a contribution involving the correct bulk water diffusion coefficient D_0. This coefficient was taken as the bulk water diffusion coefficient obtained with correct masses and inertia tensors (cf. Table 6.3 - set 2). The final formula for the corrected diffusion coefficients determination is therefore:

$$D_C(y) = D(y) - w(y).D(y = 30.5) + w(y).D_0. \tag{6.25}$$

This simple correction supposes equal contributions from solvent and lipid particles. Nevertheless, considering the simplicity of our model, this protocol constituted a simple and efficient way of obtaining a corrected profile. The corrected diffusion profiles for the four solutes are presented in Figure 6.9. Errors in $D_C(y)$ are standard errors calculated from the difference of the diffusion coefficients in each of the four individual simulation at each depth from their average.

The corrected profiles are much closer to those observed in the atomistic study of Bemporad *et al.* The bulk water coefficients are generally smaller, but are closer to experimental values; a logical result, considering that the diffusion coefficients used for the correction are closer to experimental values, as described in Table 6.3. For all solutes, diffusion coefficients decrease when entering the membrane, largely due to the design of the correction. Apart from ethane, all permeants see their diffusion coefficient increase at the bilayer centre. Assuming a hopping diffusive mechanism, the presence of larger and more frequent voids in the centre of the membrane explain these increases in diffusion coefficients. Ethane is subject to very limited diffusion in the interior of the membrane.

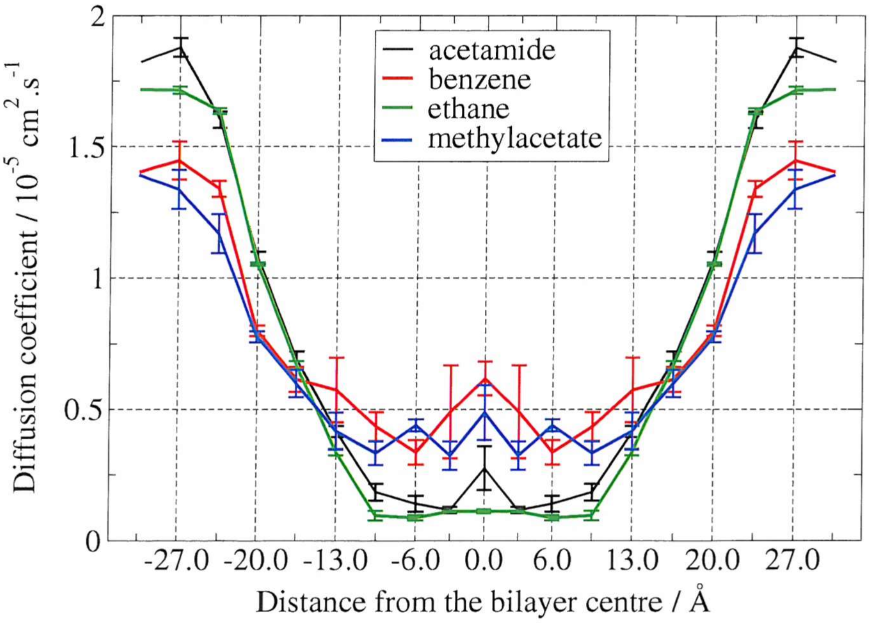


Figure 6.9: Corrected diffusion coefficient profiles. Error bars are standard error calculated from the difference of the diffusion coefficients in the four individual simulations from their average. No error bars are present at depth 30.5 Å as all four positions have the same corrected diffusion coefficient.

6.6.3 Local Resistances and Permeability Coefficients

For each permeant, four different resistances were calculated using equation 6.22. The first two resistance profiles were determined using the complete free energy profile (from -30.5 to 30.5 Å) and both the diffusion profile and the corrected diffusion profile described in section 6.6.2. The other two resistances were calculated using the free energy profile between -16.5 and 16.5 Å, and both the diffusion profile and the corrected diffusion profile. Error bars are not plotted; they are large, since a small error in $\Delta G(y)$ is amplified in $\mathcal{R}(y)$, owing to the exponential relationship between resistance and free energy.

Local resistances for the region -30.5 to 30.5 Å are plotted in Figure 6.10, both in their uncorrected and corrected form. Both corrected and uncorrected profiles show the same trend, they only differ through their magnitude. Bulk water resistances vary between 0.05 and 1×10^6 s.cm^{-1}. They decrease drastically when approaching the membrane (between -30.5 and -23.5 Å) to reach a value close to zero. In the region -23.5 to 23.5 Å, the resistance is stable for hydrophobic compounds, while a small peak is present for acetamide and methylacetate at the centre of the bilayer. This

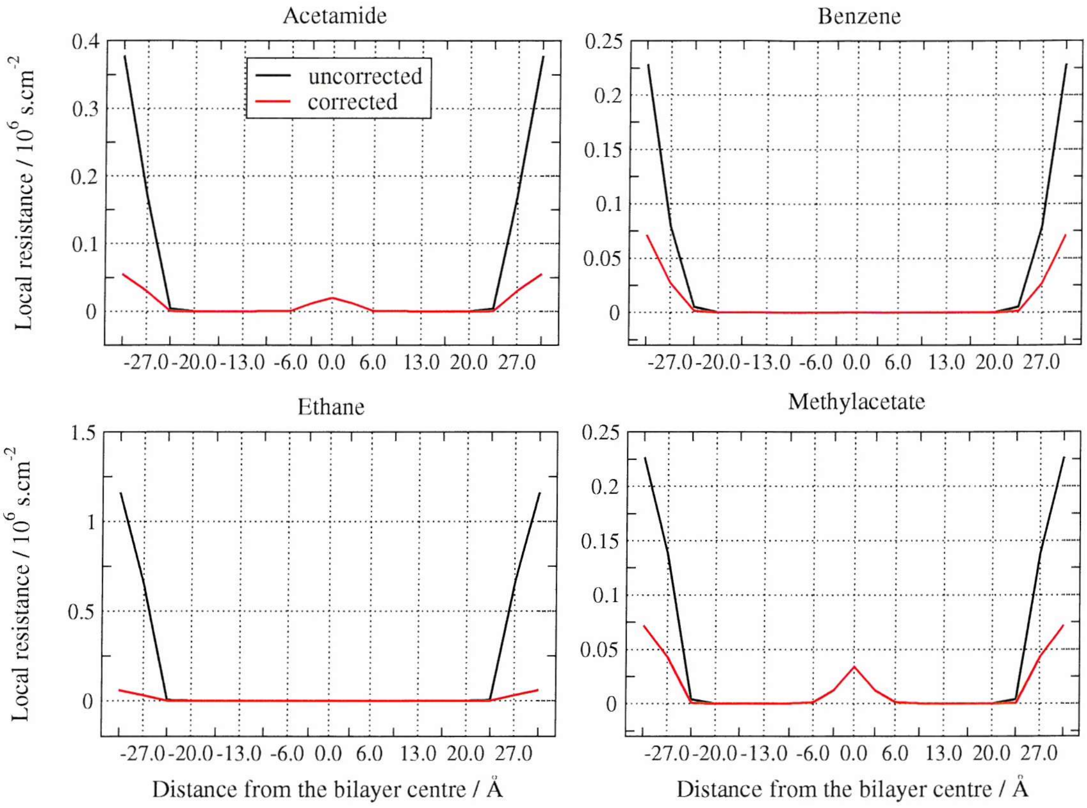


Figure 6.10: Local resistances profiles for the region −30.5 to 30.5 Å. For clarity error bars are not plotted. Uncorrected profiles are plotted in blacks, corrected profiles using corrected diffusion coefficients are plotted in red.

peak is due to the increase in free energy observed in this region for these compounds. Where they are significant (water phase), corrected resistances are smaller than the uncorrected resistances, owing to the larger corrected diffusion coefficients.

The complete free energy profiles showed a systematic decrease of the solutes $\Delta G(y)$ when entering the membrane, reflecting a general preference for the membrane interior. It is therefore not surprising to see the solutes resistances decrease when entering the membrane. The systematic defect observed in the free energy profiles is propagated to the resistance profiles. As could have been guessed by examining $\Delta G(y)$ values, the membrane does not constitute a barrier to permeation. This is reflected in the permeability coefficients presented in Table 6.5.

Table 6.5: Permeability coefficients / cm.s^{-1}. The free energy profile considered is for $|y| \leqslant 30.5$ Å

	uncorrected	corrected	Bemporad *et al.*[142]
Acetamide	37.74 ± 2.99	218.92 ± 43.73	0.00657 ± 0.00192)
Benzene	74.70 ± 9.28	225.00 ± 16.83	9.91 ± 1.11
Ethane	14.01 ± 3.41	265.09 ± 47.87	6.75 ± 0.82
Methylacetate	53.47 ± 7.37	147.66 ± 26.70	9.50 ± 1.11

Both corrected and uncorrected permeabilities are large when compared to the values reported by Bemporad *et al.*[142] No significant differences can be observed between hydrophilic and hydrophobic compounds. The membrane does not act as a barrier for any of the permeants, regardless of their polarity. These results are greatly affected by the defect observed in the free energy profiles, the examination of the resistance profiles computed with the modified free energy profile (cf. Figure 6.5) gives more insight into the solutes behaviour in the membrane interior. These profiles (both corrected for diffusion and uncorrected) are presented in Figure 6.11. For these plots, the free energy differences between bulk water and depths up to 16.5 Å (included) were considered to be zero.

The examination of these profiles enables a distinction between hydrophilic and hydrophobic compounds to be made. Acetamide and Methylacetate have similar profiles; the resistance is low in bulk water and increases in the middle of the bilayer. Acetamide reaches a value of $68\,000 \times 10^6$ s.cm^{-2}, while methylacetate reaches 2000×10^6 s.cm^{-2}. These trends are in agreement with the fact that the main barrier to permeation of these solutes is the hydrophobic core of the membrane. In the Bemporad *et al.* study, the acetamide resistance amplitude is smaller ($\approx 1200 \times 10^6$ s.cm^{-2}) and a drop of resistance is present at the centre of the bilayer, owing to the increased presence of free volume, but the general trend is similar to that observed in our study.

Methylacetate, while being hydrophilic, cannot form H-bonds with the surrounding lipids and therefore had a resistance profile typical of hydrophobic compounds in the Bemporad study. In our study, where H-bonds are not represented adequately, methylacetate behaves as a hydrophilic compound. Its profile is therefore different to the profile obtained from atomistic simulations.

Both hydrophobic compounds have similar behaviour. Unlike acetamide and

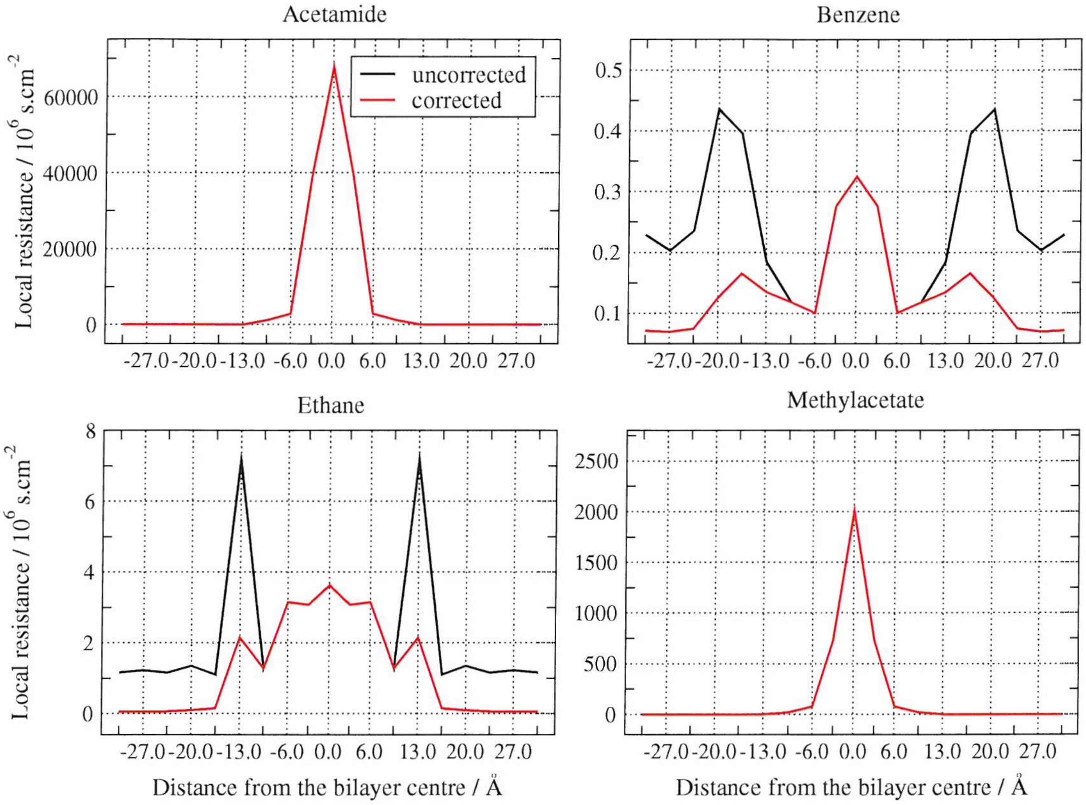


Figure 6.11: Local resistances profiles for the region -16.5 to 16.5 Å. For clarity error bars are not plotted. Uncorrected profiles are plotted in blacks, corrected profiles using corrected diffusion coefficients are plotted in red.

methylacetate, the diffusion correction modifies the shape of their profiles, systematically lowering resistances in the region 9.5 to 30.5 Å. The profiles show smaller amplitudes when compared to hydrophilic compounds, respectively ≈ 0.5 and $\approx 7 \times 10^6\,\text{s.cm}^{-2}$ for benzene and ethane. Both compounds undergo a resistance increase when they cross the interface, located at 20 Å for benzene and at 13 Å for ethane. In the atomistic study these peaks were located at the same depth (20 Å), with similar magnitudes for benzene, and one order of magnitude smaller for ethane. One feature, which is not present in the atomistic profiles of Bemporad *et al.*, is the increase of resistance at the bilayer centre detected in our study. This could be explained by the slight increase of free energy observed in this region, amplified by the use of the exponential in Equation 6.22. Interestingly, benzene and ethane profiles are closer to atomistic studies when they are not corrected.

Permeability coefficients were calculated based on these resistance profiles, and are presented in Table 6.6.

Table 6.6: Permeability coefficients / cm.s^{-1}. The free energy profile considered is for $|y| \leqslant 16.5$ Å

	uncorrected	corrected	Bemporad *et al.*[142]
Acetamide	0.0195 ± 0.0171	0.0205 ± 0.0182	0.00657 ± 0.00192
Benzene	7.37 ± 1.38	15.40 ± 3.59	9.91 ± 1.11
Ethane	1.32 ± 0.65	9.82 ± 7.53	6.75 ± 0.82
Methylacetate	2.71 ± 1.81	6.82 ± 5.64	9.50 ± 1.11

Corrected and uncorrected permeability coefficients and values reported by Bemporad *et al.* are of the same order of magnitude, with the exception of acetamide which is an order of magnitude larger in our study. Owing to large errors in the estimate of the permeability coefficients, a ranking of the permeants is difficult to achieve. Nevertheless the study succeeds in separating acetamide from the other compounds. A permeability of $\approx 2 \times 10^{-2}$cm.s^{-1} is obtained, two orders of magnitude larger than the value reported experimentally,[187, 190] (1.7×10^{-4}cm.s^{-1}), but still smaller than the values obtained for the other permeants (for which no experimental data is available).

While methylacetate showed profiles consistent with an hydrophobic behaviour, the permeability coefficients calculated in this study are of the same order of magnitude as those of of hydrophobic compounds, benzene and ethane. This is in excellent agreement with the results of Bemporad *et al.*, where these three permeants all had permeability coefficients of the same order of magnitude.

6.6.4 Solute Behaviour

Solute Orientation

The solutes behaviour inside the bilayer was monitored by examining the orientation of their principal axes of inertia. When determining the inertia tensor of a solute, after diagonalization, the vector with the largest eigenvalue is the principal axis of inertia, while the vector with the smallest eigen value corresponds to the molecular axis. The principal axis of inertia is, in our case, the normal to the molecular plane. We examined the angle θ between the principal axis of inertia of the solute and the bilayer normal (y direction). A measure of the degree of order of the system can be obtained through the calculation of an order parameter S (second Legendre

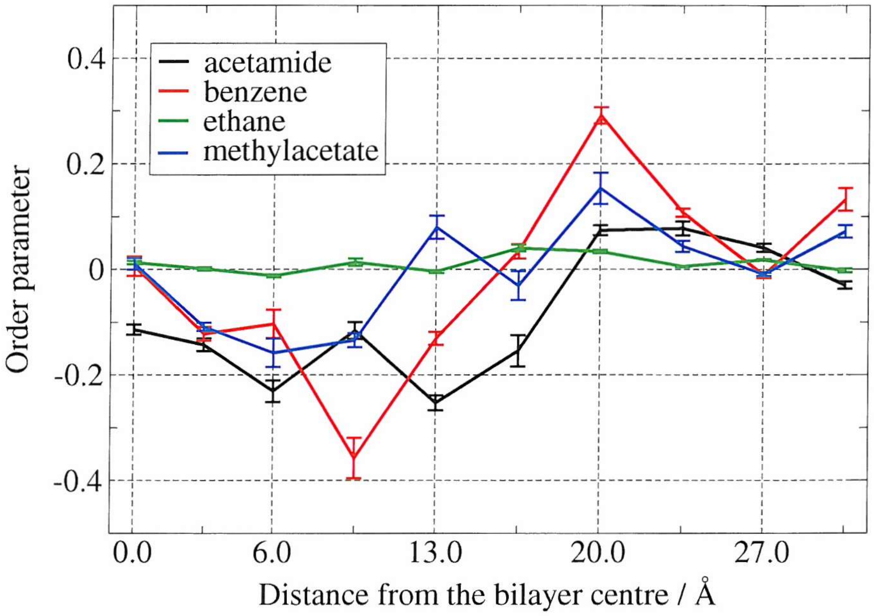


Figure 6.12: Solute order parameter as a function of depth, for the angle θ between the bilayer normal and the principal axis of inertia of each solute.

polynomial):

$$S = \frac{1}{2}\langle 3\cos^2\theta - 1\rangle \tag{6.26}$$

where the brackets denote an ensemble average. If the system is fully ordered, the order parameter takes a value of $-1/2$ (principal axis of inertia aligned with the bilayer normal), while it takes a value of 1 for full order perpendicular to the bilayer normal. A value of zero is associated with full isotropic motion (no ordering). The evolution of the solute order parameter as a function of depth in the bilayer for the four solutes studied is presented in Figure 6.12. Standard errors are calculated from the difference of the mean $\langle S\rangle$ in each of the four simulations from their overall average value for each depth.

The order parameter of ethane is close to zero at all depth, indicating the absence of a preferential orientation. Ethane nearly reaches full isotropic motion across the entire bilayer. This fact can be explained by the small size of ethane, which can be easily accommodated at every depth inside the membrane. The order parameter calculated using the molecular axis was also calculated and showed a similar behaviour, a value close to zero across the entire membrane, confirming the fact that ethane achieves full isotropic motion.

For all three remaining solutes, a general trend can be seen. Permeants tend to be aligned parallel to the lipid chains ($S < 0$). In the hydrocarbon region, the available free volume is along the chains, and solutes are tightly packed due to interfacial constraints. This trend is less marked at the bilayer centre, probably because of the greater free volume available to solutes, which can therefore adopt more conformations. It can be noticed that the most negative order parameter (larger organization of the molecular plane along the bilayer normal) is obtained for benzene. This can easily be explained by its size, it is the largest solute, and by its geometry, it is flat.

The smallest S values are obtained above $y = 27.0\,\text{Å}$, a logical result, as solutes are expected to undergo nearly free isotropic motion in the bulk water phase.

Interestingly, at 20.0 Å, i.e. at the end of the membrane-water interface, solutes tend to be aligned with the bilayer plane. This is especially the case for benzene, which order parameter steadily increases from 9.0 to 20.0 Å to reach $S \approx 0.3$ at $y = 20.0\,\text{Å}$. As revealed by the P-N dipole examination in Chapter 5, the head group dipole does not lie parallel to the plane of the bilayer. The explanation for benzene behaviour at $y = 20.0\,\text{Å}$ must therefore be found elsewhere. One possible justification, would be the shape of the water-lipid interface and simple geometric considerations. The interface being flat, the simplest arrangement for benzene would be to lie on this interface (parallel to the bilayer plane).

Solute Orientational Correlation Times

The ability of the solute to rotate around its centre of mass was evaluated using the time autocorrelation function of the principal axis of inertia vector orientation. The following function was used:

$$C(t) = \langle P_2\left[\boldsymbol{\mu}(t)\cdot\boldsymbol{\mu}(0)\right]\rangle \tag{6.27}$$

where P_2 is the Legendre polynomial of rank 2 and $\boldsymbol{\mu}$ is the principal axis of inertia of the solute (of norm unity), i.e. the unit vector along the normal to the molecular plane, as defined previously to calculate the angle θ with the bilayer normal. As the vectors $\boldsymbol{\mu}(t)$ and $\boldsymbol{\mu}(0)$ are normalized the equation transforms to:

$$C(t) = \frac{1}{2}\left\langle 3\cos^2\alpha - 1\right\rangle \tag{6.28}$$

where α is the angle between $\boldsymbol{\mu}$ at time t and at time 0. The resulting time autocorrelation functions were then fitted with single exponentials of form:

$$C_F(t) = C_0\exp(-\mathrm{t}/\tau). \tag{6.29}$$

The relaxation times, τ, obtained from these fittings are detailed in Table 6.7. The first 500 000 steps of the simulations were used, corresponding to approximately 1.83 ns. For those solutes with large relaxation times, autocorrelation functions were very noisy especially for times greater than 200 ps. Since the most characteristic part of the curve is the initial drop, and since data in excess of 200 ps prevents a correct fit of the initial drop to be obtained, we decided to restrict the fit to the region 0-100 ps. To enable a better fitting with the initial part of the curve, a weighting function was introduced: $w(t) = C(t)^2$. As a consequence, the results obtained are only valid for relaxation times lower than 100 ps. Some relaxation times were found to be greater than this limit; they were considered as being greater than 100 ps without any indication of their value. Fitting was performed for each solute, except ethane, each depth and each initial x-z position. Averages over the four initial x-z positions were performed. Ethane relaxation times were not examined as ethane does not adopt any particular orientation, as revealed in the previous section. When calculating the average, positions with relaxation times greater than 100 ps were given a value of 100 ps. The resulting average is indicated as being greater than the calculated value. Average values are plotted in Figure 6.13 together with standard errors where available. Standard errors were calculated from the differences of the four independent simulations from their average. Points where only an estimate is available (average greater than $\cdots$) are indicated by an asterisk.

For all solutes, a general trend can be observed, τ increases entering the membrane and reaches its highest value between 6 and 20 Å from the bilayer centre. The highest value is obtained for benzene at $y = 9.5$ Å, with $\tau \geqslant 81.25$ ps, a logical result as benzene is the largest permeant in this study, and its rotation in the densest part of the membrane are the most difficult.

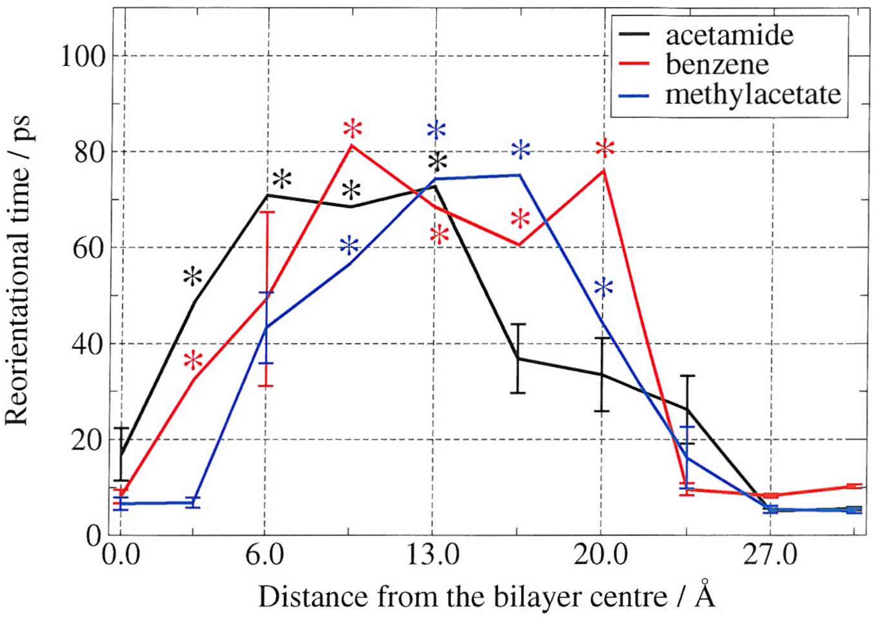


Figure 6.13: Solute reorientational correlation times as a function of depth in the membrane. For points marked with an asterisk, at least one of the four positions lead to a correlation time greater than 100 ps. The overall average is therefore at least the indicated value. For clarity purposes, those points are marked in black for acetamide, in red for benzene and in blue for methylacetate.

Table 6.7: Average reorientational time as a function of depth in ps.

y / Å	Acetamide	Benzene	Methylacetate
0.0	16.84 ± 9.46	8.06 ± 2.44	6.58 ± 2.33
3.0	$\geqslant 48.70$	$\geqslant 32.53$	6.82 ± 1.77
6.0	$\geqslant 70.88$	49.29 ± 31.41	43.30 ± 12.77
9.5	$\geqslant 68.47$	$\geqslant 81.25$	$\geqslant 56.70$
13.0	$\geqslant 72.74$	$\geqslant 68.51$	$\geqslant 74.30$
16.5	36.84 ± 12.51	$\geqslant 60.58$	$\geqslant 75.10$
20.0	33.47 ± 13.25	$\geqslant 75.99$	$\geqslant 44.37$
23.5	26.23 ± 12.23	9.61 ± 2.20	16.19 ± 11.09
27.0	5.11 ± 0.79	8.27 ± 0.67	5.42 ± 1.24
30.5	5.52 ± 0.58	10.22 ± 0.67	5.07 ± 0.81

The largest relaxation times are obtained between $y = 3.0$ and 13.0 Å for acetamide, and between $y = 9.5$ and 20.0 Å for benzene and methylacetate, i.e. in the hydrocarbon region for the most polar permeant (acetamide), and in the lipid-water interface for the most hydrophobic compound (benzene).

Previous simulations of benzene diffusion in the hydrocarbon region of a biomembrane[8] have lead to reorientational correlation times of $\approx$ 25 ps. In our simulations, a value of the same magnitude is obtained at $y = 0.0$ Å (8 ps), higher values are obtained between $y = 3.0$ and $9.5\,\text{Å}$ ($\geqslant 81.25$ ps at $9.5\,\text{Å}$). The atomistic simulations did not used any constraint on the solute, even though benzene resided mostly in the centre of the bilayer, the relaxation time reported corresponds to different regions of the membrane. The more restricted sampling of locations inside the bilayer of these studies explains the difference with our results.

6.7 Conclusions

In this chapter, permeability coefficients for four small organic molecules have been calculated using molecular dynamics simulations and the Gay-Berne representation of a DMPC membrane. The solubility-diffusion model has been used, this model incorporates informations about both equilibrium and dynamic properties. Equilibrium properties are accounted for by the free energy profile of transfer across the membrane, while the dynamic information is represented by the diffusion coefficient profile.

The results reported here show different defects. The computation of the free energy profiles enabled issues with the head group parameterization to be highlighted. The solutes' free energies systematically decrease when entering the membrane, instead of a small free energy increase that would have been expected. This was suggested to be a consequence of the size of the head group particles. Larger particles could reduce this defect. The computation of diffusion coefficients showed the influence of the masses and inertia tensors atributed to water particles. These were given the same values as Gay-Berne particles. This choice lead to a systematic underestimate of diffusion coefficients in bulk water. A correct attribution of water masses and inertia tensors would solve this issue, although this will inevitably result in a compromise in terms of time step.

Corrections have been brought to both diffusion coefficients and free energy profiles to model an *unbiased* estimate of permeability coefficients. Even though our model possesses several flaws, the corrected permeability coefficients presented are suprisingly close to the values reported by Bemporad *et al.*[142] in atomistic simulations. For acetamide, which was the only truly hydrophilic compound in this study, the permeability coefficient is far smaller than those of the remaining solutes, as in the Bemporad *et al.* study. Considering the simplicity of the model, reproducing values of the correct order of magnitude (using corrections), as well as the correct separation of acetamide from the other solutes, is an achievement. Leaving aside the systematic decrease of free energy observed when the permeants approach the bilayer, the solutes' behaviour inside the membrane met our expectations. While the bilayer core is essentially a barrier for hydrophilic molecules, it acts as an accelerator for hydrophobic solutes. In the Bemporad *et al.* study, methylacetate, while being a hydrophilic solute, was classified as a hydrophobic molecule, owing to the fact it does not possess hydrogen bonds donors. In the present study, where H-bonds are not well represented, methylacetate behaved in the same fashion as acetamide, and was therefore considered as a hydrophilic molecule.

Examination of solute orientation and orientational correlation times showed results in agreement with atomistic studies, both qualitatively and quantitatively. Thus although further parameterization is needed to improve solute representation, these results show that the generalized Gay-Berne potential can be used to model realistically solute interactions with the bilayer.

Further parameterization will be required for the model to yield an accurate description of the permeation process. Nevertheless, considering the simplicity of the model and the potential gain in computational time, the results presented in this chapter are very promising. We have few doubts that minor changes would lead to a better description of the permeation process and would make *CHAMPAGNE* a useful tool for the determination of permeability coefficients.

Chapter 7

Conclusions and Future Work

7.1 Summary

This thesis describes the development of a reduced representation of a biological membrane, and the study of small molecules permeation across this bilayer. The main goal of this study was the development of a membrane model that will be faster than an atomistic model and yet still yield correct membrane behaviour. Ultimately, this model was to be applied to the determination of the permeabilties of small solutes.

The development of a simplified biological membrane based on DMPC, described in Chapter 4 required extensive parameterization. A new technique for the determination of Gay-Berne parameters has been created. This technique which, to our knowledge, has never been used before, considers multiple comformations, rather than certain characteristic conformations. These parameters were subsequently refined in a number of molecular dynamics simulations before being used in production molecular dynamics simulations of phospholipid bilayers.

These simulations, presented in Chapter 5 have been performed at eight different temperatures ranging from 5 to 60 °C, with simulation times exceeding 7 ns for each temperature. The results obtained have been favourably compared to the available experimental data in most cases. Phase transitions have been identified both in the hydrocarbon region between 20 and 25 °C, through an examination of the orientational ordering; and in the head group region between 25 and 30 °C, through an examination of the P-N dipole. Evidence of long-range lipid diffusion or "lipid swap" was observed. The self aggregation study failed to identify the formation of a bilayer of correct structure. Rather, the formation of a two phase system was observed - water and a low-hydration bilayer. Nevertheless, if our model is set up as a preassembled bilayer, its behaviour is in agreement with the available experimental data. Considering the simplicity of the model, the agreement with experimental data represents an achievement. The model developed while being simpler and potentially much faster than a conventional atomistic model still gives access to information at the molecular level.

This model was therefore employed to model small solute permeation. The permeation study, described in chapter 6 concerned four small solutes, acetamide, benzene, ethane and methylacetate. A simulation protocol was implemented to obtain equilibrium and dynamic properties associated with the permeation process. Equilibrium properties consisted of the solute partitioning behaviour between water and different areas of the membrane, i.e. the free energy profile for membrane crossing. Dynamic properties consisted of the solute diffusive behaviour at different depths in the membrane interior. Permeability coefficients were calculated using free energy and diffusion coefficients. Both free energy and diffusion profiles showed defects, they have therefore been corrected in an attempt to give an *unbiased* estimate of permeability coefficients. The corrected permeability coefficients presented are suprisingly close to the values reported by Bemporad *et al.*[142] in atomistic simulations. Considering the simplicity of the model, the reproduction of values of the correct order of magnitude (using corrections), as well as the correct separation of acetamide from the other solutes, is an achievement. The solutes behaviour inside the membrane met our expectations. While the bilayer core is essentially a barrier for hydrophilic molecules, it acts as an accelerator for hydrophobic solutes. Solute conformations were also examined through the study of their orientation and orientational correlation times. Results are in agreement with the atomistic studies of Bemporad *et al.* and show that the generalized Gay-Berne potential can be used to model reallistically solutes interactions with the bilayer. Considering the simplicity of the model and the potential gain in computational time, the results presented in this chapter are very promising. We have few doubts that minor changes would lead to a better description of the permeation process and would make *CHAMPAGNE* a useful tool for the determination of permeability coefficients.

7.2 Further Work

Although the membrane model is satisfactory in a number of areas, several improvements are required. While being faster than an atomistic model, essentially because of the use of larger time steps, the intrinsic speed of the proposed model is only marginally faster than an atomistic model. The scope of this thesis was mainly the development of a simplified model; little effort has been dedicated to the improvement of the code's efficiency. The addition of water is largely responsible for the decrease in efficiency when compared to the efficiency of the hydrocarbon region representation.[2] The use of embedded dipoles in the water model means that the implementation of particle mesh Ewald is less straightforward than with point charges. Moreover the nature of the sticky potential does not allow the timestep to be increased any further. However, the smoothness of the model's interaction potential results in rapid equilibration. Future work will concentrate on the implementation of the particle mesh Ewald method and should result in a significant improvement of the code's efficiency. The implementation of a new hybrid water model, and its specific parameterization would lead to significant improvements both in terms of structure and speed.

The solute permeation study highlighted problems in the head group region. The size of the head group particles appeared to be too small. The introduction of different size particles interacting through the generalized Gay-Berne potential and a specific parameterization would lead to a better representation of the water-membrane interface and a more accurate description of the partitioning behaviour of solutes inside the membrane. The attribution of identical masses and inertia tensors to lipid and water particles leads to a systematic underestimate of solute diffusion coefficients in bulk water. Future modifications to *CHAMPAGNE* will therefore include the correct attribution of masses and inertia tensors to all particles.

References

[1] J. G. Gay and B. J. Berne, *J. Chem. Phys.*, **74**, 3316, (1981).

[2] L. Whitehead, C. M. Edge and J. W. Essex, *J. Comp. Chem.*, **22**, 1622, (2001).

[3] E. Egberts, *Molecular Dynamics Simulations of Multibilayer Membranes*, Ph.D. thesis, University of Groningen, (1988).

[4] J. W. Essex, M. M. Hann and W. G. Richards, *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.*, **344**, 239, (1994).

[5] D. P. Tieleman, S. J. Marrink and H. J. C. Berendsen, *Biochim. Biophys. Acta-Rev. Biomembr.*, **1331**, 235, (1997).

[6] S. E. Feller, *Curr. Opin. Colloid Interface Sci.*, **5**, 217, (2000).

[7] S. J. Marrink and H. J. C. Berendsen, *J. Phys. Chem.*, **98**, 4155, (1994).

[8] D. Bassolino-Klimas, H. E. Alper and T. R. Stouch, *Biochemistry*, **32**, 12624, (1993).

[9] D. Bassolino-Klimas, H. E. Alper and T. R. Stouch, *J. Am. Chem. Soc.*, **117**, 4118, (1995).

[10] H. E. Alper and T. R. Stouch, *J. Phys. Chem.*, **99**, 5724, (1995).

[11] T. R. Stouch, H. E. Alper and D. Bassolino, *J. Comput. Aided Mol. Des.*, **589**, 127, (1995).

[12] S. J. Marrink and H. J. C. Berendsen, *J. Phys. Chem.*, **100**, 16729, (1996).

[13] A. Pohorille, M. H. new, K. Schweighofer and M. A. Wilson, *Curr. Top. Membr.*, **48**, 49, (1999).

[14] M. S. P. Sansom, *Curr. Opin. Struc. Biol.*, **8**, 237, (1998).

[15] L. R. Forrest and M. S. P. Sansom, *Curr. Opin. Struc. Biol.*, **10**, 174, (2000).

[16] M. Ulmschneider and M. S. P. Sansom, *Biochim. Biophys. Acta*, **1512**, 1, (2001).

[17] S. Bandyopadhyay, M. Tarek and M. L. Klein, *J. Phys. Chem. B*, **103**, 10075, (1999).

[18] S. J. Marrink, D. P. Tieleman and A. E. Mark, *J. Phys. Chem. B*, **104**, 12165, (2000).

[19] S. J. Marrink, E. Lindahl, O. Edholm and A. E. Mark, *J. Am. Chem. Soc.*, **123**, 8638, (2001).

[20] S. J. Marrink and A. E. Mark, *Biochemistry*, **41**, 5375, (2002).

[21] S. E. Feller and R. W. Pastor, *J. Chem. Phys.*, **111**, 1281, (1999).

[22] R. S. Armen, O. D. Uitto and S. E. Feller, *Biophys. J.*, **75**, 734, (1998).

[23] D. P. Tieleman and H. J. C. Berendsen, *J. Chem. Phys.*, **105**, 4871, (1996).

[24] P. Atkins, *Physical Chemistry* - 4^{th} *ed*, Oxford University Press, (1990).

[25] M. A. Bates and G. R. Luckhurst, *In Press, Structure and Bonding*, volume 94, D.M.P. Mingos.

[26] W. E. Palke, J. W. Emsley and D. J. Tildesley, *Mol. Phys.*, **82**, 177, (1994).

[27] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford Science Publications, (1987).

[28] A. R. Leach, *Molecular modelling: principles and applications.*, Pearson Education Limited, Harlow, (1996).

[29] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179, (1995).

[30] W. L. Jorgensen and J. Tiradorives, *J. Am. Chem. Soc.*, **110**, 1666, (1988).

[31] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comp. Chem.*, **4**, 187, (1983).

[32] L. Woodcock, *Chem. Phys. Lett.*, **10**, 257, (1971).

[33] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola and J. R. Haak, *J. Chem. Phys.*, **81**, 3684, (1984).

[34] H. Anderson, *J. Chem. Phys.*, **72**, 2384, (1980).

[35] S. Nosé, *Mol. Phys.*, **53**, 255, (1984).

[36] W. Hoover, *Phys. Rev. A*, **31**, 1695, (1985).

[37] J. W. Essex, *Mol. Sim.*, **20**, 159, (1998).

[38] P. Ewald, *Ann. Phys.*, **64**, (1921).

[39] L. Verlet, *Phys. Rev.*, **159**, 98, (1967).

[40] H. Goldstein, *Classical Mechanics*, Addison-Wesley, (1980).

[41] D. J. Evans, *Mol. Phys.*, **34**, 317, (1977).

[42] K. Singer, A. Taylor and J. V. L. Singer, *Mol. Phys.*, **33**, 1757, (1977).

[43] D. Fincham, *CCP5 Quarterly*, **8**, 47, (1984).

[44] F. Reinitzer, *Monat. für Chem.*, **9**, 421, (1888).

[45] G. Freidel, *Ann. Phys.*, **18**, 273, (1922).

[46] D. Vorlander, *Zeit. für Phys. Chem.*, **105**, 211, (1923).

[47] G. Friedel, *Ann. Phys.*, **18**, 273, (1922).

[48] G. H. Brown, J. W. Doane and V. D. Neff, *A Review of the Structure and Physical Properties of Liquid Crystals*, CRC Press, (1971).

[49] P. S. Pershan, *Physics Today*, **35**, 34, (1982).

[50] M. A. Bates, *Computer Simulation of Liquid Crystals*, Ph.D. thesis, University of Southampton, (1996).

[51] P. Lebwohl and G. Lasher, *Phys. Rev. A*, **6**, 426, (1972).

[52] W. Maier and Z. Saupe, *Naturf. A*, **13**, 564, (1958).

[53] U. Fabbri and C. Zannoni, *Mol. Phys.*, **58**, 763, (1986).

[54] G. R. Luckhurst and P. Simpson, *Mol. Phys.*, **47**, 251, (1982).

[55] C. Zannoni, *J. Chem. Phys.*, **84**, 424, (1986).

[56] J. Corner, *Proc. R. Soc. Lond. A*, **192**, (1948).

[57] B. J. Berne and P. Pechukas, *J. Chem. Phys.*, **56**, (1972).

[58] D. J. Adams, G. R. Luckhurst and R. W. Phippen, *Mol. Phys.*, **61**, 1575, (1987).

[59] G. R. Luckhurst and P. S. J. Simmonds, *Mol. Phys.*, **80**, 233, (1993).

[60] E. Demiguel, L. F. Rull, M. K. Chalam and K. E. Gubbins, *Mol. Phys.*, **74**, 405, (1991).

[61] R. Hashim, G. R. Luckhurst and S. Romano, *J. Chem. Soc., Fararay Trans.*, **91**, 2141, (1995).

[62] D. Frenkel, B. M. Mulder and J. P. Mctague, *Phys. Rev. Lett.*, **52**, 287, (1984).

[63] D. Frenkel and B. M. Mulder, *Mol. Phys.*, **55**, 1171, (1985).

[64] P. Bolhuis and D. Frenkel, *J. Chem. Phys.*, **106**, 666, (1997).

[65] A. Stroobants, H. N. W. Lekkerkerker and D. Frenkel, *Phys. Rev. Lett.*, **57**, 1452, (1986).

[66] C. Mcbride, M. R. Wilson and J. A. K. Howard, *Mol. Phys.*, **93**, 955, (1998).

[67] C. Mcbride and M. R. Wilson, *Mol. Phys.*, **97**, 511, (1999).

[68] G. Cevc and D. Marsh, *Phospholipid Bilayers*, John Willey and Sons, (1987).

[69] R. H. Templer, S. J. Castle, A. R. Curran, G. Rumbles and D. R. Klug, *Faraday Discussions*, **111**, 41, (1999).

[70] I. Szleifer, A. Benshaul and W. M. Gelbart, *J. Phys. Chem.*, **94**, 5081, (1990).

[71] I. Szleifer, D. Kramer, A. Benshaul, W. M. Gelbart and S. A. Safran, *J. Chem. Phys.*, **92**, 6800, (1990).

[72] I. Szleifer, D. Kramer, A. Benshaul, D. Roux and W. M. Gelbart, *Phys. Rev. Lett.*, **60**, 1966, (1988).

[73] D. Evans and H. Wennerstrom, *The Colloidal Domain: Where Physics, Chemistry, Biology, and Technology Meet*, Wiley, New York, (1999).

[74] J. R. Silvius, *Lipid-Protein Interactions*, John Wiley and Sons, New York, (1982).

[75] V. Racansky, D. Valachovic and P. Balgavy, *Acta Physica Slovaca*, **37**, 166, (1987).

[76] A. Blume, *Thermochim. Acta*, **193**, 299, (1991).

[77] S. J. Marrink, M. Berkowitz and H. J. C. Berendsen, *Langmuir*, **9**, 3122, (1993).

[78] S. J. Marrink, R. M. Sok and H. J. C. Berendsen, *J. Chem. Phys.*, **104**, 9090, (1996).

[79] S. J. Marrink, F. Jahnig and H. J. C. Berendsen, *Biophys. J.*, **71**, 632, (1996).

[80] J. J. L. Cascales, J. G. de la Torre, S. J. Marrink and H. J. C. Berendsen, *J. Chem. Phys.*, **104**, 2713, (1996).

[81] S. J. Marrink, F. Jahnig and H. J. C. Berendsen, *Biophys. J.*, **71**, 632, (1996).

[82] J. J. L. Cascales, H. J. C. Berendsen and J. G. Delatorre, *J. Phys. Chem.*, **100**, 8621, (1996).

[83] P. Tieleman, H. J. C. Berendsen and M. S. P. Sansom, *Biophys. J.*, **76**, 3186, (1999).

[84] A. J. Robinson, W. G. Richards, P. J. Thomas and M. M. Hann, *Biophys. J.*, **67**, 2345, (1994).

[85] A. J. Robinson, W. G. Richards, P. J. Thomas and M. M. Hann, *Biophys. J.*, **68**, 164, (1995).

[86] K. Raghavan, M. R. Reddy and M. L. Berkowitz, *Langmuir*, **8**, 233, (1992).

[87] U. Essmann, L. Perera and M. L. Berkowitz, *Langmuir*, **11**, 4519, (1995).

[88] L. Perera, U. Essmann and M. L. Berkowitz, *Langmuir*, **12**, 2625, (1996).

[89] S. W. Chiu, M. Clark, V. Balaji, S. Subramaniam, H. L. Scott and E. Jakobsson, *Biophys. J.*, **69**, 1230, (1995).

[90] S. E. Feller, R. M. Venable and R. W. Pastor, *Langmuir*, **13**, 6555, (1997).

[91] S. E. Feller and R. W. Pastor, *Biophys. J.*, **71**, 1350, (1996).

[92] K. Tu, D. J. Tobias and M. L. Klein, *Biophys. J.*, **69**, 2558, (1995).

[93] O. Berger, O. Edholm and F. Jahnig, *Biophys. J.*, **72**, 2002, (1997).

[94] K. Tu, D. J. Tobias and M. L. Klein, *Biophys. J.*, **70**, 595, (1996).

[95] D. J. Tobias, K. Tu and M. L. Klein, *J. Chim. Phys. Phys.- Chim. Biol.*, **94**, 1482, (1997).

[96] K. C. Tu, M. L. Klein and D. J. Tobias, *Biophys. J.*, **75**, 2147, (1998).

[97] T. Husslein, D. M. Newns, P. C. Pattnaik, Q. F. Zhong, P. B. Moore and M. L. Klein, *J. Chem. Phys.*, **109**, 2826, (1998).

[98] R. Goetz and R. Lipowsky, *J. Chem. Phys.*, **108**, 7397, (1998).

[99] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay and M. L. Klein, *J. Phys. Chem. B*, **105**, 4464, (2001).

[100] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, P. B. Moore and M. L. Klein, *J. Phys. Chem. B*, **105**, 9785, (2001).

[101] C. F. Lopez, P. B. Moore, J. C. Shelley, M. Y. Shelley and M. L. Klein, *Comput. Phys. Commun.*, **147**, 1, (2002).

[102] S. Yamamoto, Y. Maruyama and S. Hyodo, *J. Chem. Phys.*, **116**, 5842, (2002).

[103] J. C. Shillcock and R. Lipowsky, *J. Chem. Phys.*, **117**, 5048, (2002).

[104] G. S. D. Ayton, S. Bardenhagen, P. Mcmurtry, D. Sulsky and G. A. Voth, *IBM J. Res. Dev.*, **45**, 417, (2001).

[105] G. Ayton, S. G. Bardenhagen, P. Mcmurtry, D. Sulsky and G. A. Voth, *J. Chem. Phys.*, **114**, 6913, (2001).

[106] D. J. Cleaver, C. M. Care, M. P. Allen and M. P. Neal, *Phys. Rev. E*, **54**, 559, (1996).

[107] M. R. Wilson, *J. Chem. Phys.*, **107**, 8654, (1997).

[108] R. Berardi, M. Fehervari and C. Zannoni, *Mol. Phys.*, **97**, 1173, (1999).

[109] J. Kushick and B. J. Berne, *J. Chem. Phys.*, **64**, (1976).

[110] A. Stone, *The Molecular Physics of Liquid Crystals*, Academic Press, London, (1979).

[111] G. R. Luckhurst, R. A. Stephens and R. W. Phippen, *Liq. Cryst.*, **8**, 451, (1990).

[112] M. A. Bates and G. R. Luckhurst, *Chem. Phys. Lett.*, **281**, 193, (1997).

[113] G. R. Luckhurst, R. A. Stephens and R. W. Phippen, *Liq. Cryst.*, **8**, 451, (1990).

[114] S. E. Feller and A. D. Mackerell, *J. Phys. Chem. B*, **104**, 7510, (2000).

[115] D. Bemporad, *Private Communication*, (2001).

[116] A. R. van Buuren, S. J. Marrink and H. J. C. Berendsen, *J. Phys. Chem.*, **97**, 9206, (1993).

[117] P. Ahlstrom and H. J. C. Berendsen, *J. Phys. Chem.*, **97**, 13691, (1993).

[118] W. L. Jorgensen, *J. Am. Chem. Soc.*, **103**, 335, (1981).

[119] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Intermolecular Forces*, Pullman. B. Ed. Reidel, (1981).

[120] H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, **91**, 6269, (1987).

[121] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, **79**, 926, (1983).

[122] F. H. , Stillinger and A. Rahman, *J. Chem. Phys.*, **60**, 1545, (1974).

[123] M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, **112**, 8910, (2000).

[124] A. Chandra and T. Ichiye, *J. Chem. Phys.*, **111**, 2701, (1999).

[125] Y. Liu and T. Ichiye, *J. Phys. Chem.*, **100**, 2723, (1996).

[126] L. Blum, F. Vericat and D. Bratko, *J. Chem. Phys.*, **102**, 1461, (1995).

[127] D. Bratko, L. Blum and A. Luzar, *J. Chem. Phys.*, **83**, 6367, (1985).

[128] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, **98**, 10089, (1993).

[129] J. F. Nagle and S. Tristram-Nagle, *Biochim. Biophys. Acta-Rev. Biomembr.*, **1469**, 159, (2000).

[130] D. Fincham and B. Ralston, *Comput. Phys. Commun.*, **23**, 127, (1981).

[131] S. Thompson, *CCP5 Quarterly*, **8**, 20, (1983).

[132] M. J. Janiak, D. M. Small and G. G. Shipley, *J. Biol. Chem.*, **254**, (1979).

[133] B. A. Lewis and D. M. Engelman, *J. Mol. Biol.*, **166**, 203, (1983).

[134] G. Buldt, H. Gally, A. Seelig and J. Seelig, *Nature*, **271**, (1978).

[135] J. H. Ipsen, O. G. Mouritsen and M. Bloom, *Biophys. J.*, **57**, 405, (1990).

[136] T. J. Mcintosh, *Membrane Permeability*, **48**, 23, (1999).

[137] Y. K. Levine, A. Bailey and M. Wilkins, *Nature (London)*, **220**, 577, (1968).

[138] Y. K. Levine, A. Bailey and M. Wilkins, *Nature (London) New Biol.*, **230**, 69, (1971).

[139] T. McIntosh, *Biochim. Biophys. Acta*, **513**, 43, (1978).

[140] W. Sun, S. Tristam-Nagle, R. Suter and J. Nagle, *Biophys. J.*, **71**, 885, (1996).

[141] T. McIntosh, S. Advani, R. Burton, D. Zhelev, D. Needham and S. Simon, *Biochemistry*, **34**, 8520, (1995).

[142] D. Bemporad, *Computer Simulation of Biological Membranes and Small Molecule Permeation*, Ph.D. thesis, University of Southampton, (2003).

[143] P. Laggner and H. Stabinger, Colloid and Interface Science, Academic Press, New York, volume V, pp. 91–96, (1976).

[144] J. F. Nagle and D. Wilkinson, *Biophys. J.*, **23**, 159, (1978).

[145] H. I. Petrache, S. Tristram-Nagle and J. F. Nagle, *Chem. Phys. Lipids*, **95**, 83, (1998).

[146] G. Schmidt and W. Knoll, *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics*, **89**, 36, (1985).

[147] R. Zhang, S. Tristam-Nagle, W. Sun, R. Headrick, T. Irving, R. Suter and J. Nagle, *Biophys. J.*, **70**, 349, (1996).

[148] M. Janiak, D. Small and G. Shipley, *Biochemistry*, **15**, 5475, (1976).

[149] H. I. Petrache, S. W. Dodd and M. F. Brown, *Biophys. J.*, **79**, 3172, (2000).

[150] R. L. Thurmond, S. W. Dodd and M. F. Brown, *Biophys. J.*, **59**, 108, (1991).

[151] J. Sheperd and G. Büldt, *Biochim. Biophys. Acta*, **514**, 8394, (1978).

[152] H. Akutsu and T. Nagamori, *Biochemistry*, **30**, 4510, (1991).

[153] H. Hauser, I. Pascher, R. Pearson and S. Sundell, *Biochim. Biophys. Acta*, **650**, 21, (1981).

[154] G. Büldt and R. Wohlgemuth, *J. Membr. Biol.*, **58**, 81, (1981).

[155] A. Seelig and J. Seelig, *Biochemistry*, **16**, 45, (1977).

[156] J. Seelig and A. Seelig, *Q. Rev. Biophys.*, **133**, 19, (1980).

[157] K. Gawrisch, D. Ruston, J. Zimmerberg, V. A. Parsegian, R. P. Rand and N. Fuller, *Biophys. J.*, **61**, 1213, (1992).

[158] J. Seelig, P. M. Macdonald and P. G. Scherer, *Biochemistry*, **26**, 7535, (1987).

[159] H. Gally, W. Niederberger and S. J., *Biochemistry*, **14**, 3647, (1975).

[160] P. G. Scherer and J. Seelig, *Biochemistry*, **28**, 7720, (1989).

[161] M. T. Hyvonen, T. T. Rantala and M. Alakorpela, *Biophys. J.*, **73**, 2907, (1997).

[162] L. Saiz and M. L. Klein, *J. Chem. Phys.*, **116**, 3052, (2002).

[163] K. Tu, D. J. Tobias, J. K. Blasie and M. L. Klein, *Biophys. J.*, **70**, 595, (1996).

[164] R. Skarjune and E. Oldfield, *Biochemistry*, **18**, 5903, (1979).

[165] J. Seelig and N. W., *J. Am. Chem. Soc.*, **96**, 2069, (1974).

[166] J. Seelig, *Q. Rev. Biophys.*, **10**, 353, (1977).

[167] J. H. Davis, *Biochim. Biophys. Acta*, **737**, 117, (1983).

[168] H. L. Scott, E. Jakobsson and S. Subramaniam, *Comput. Phys.*, **12**, 328, (1998).

[169] S. W. Chiu, M. M. Clark, E. Jakobsson, S. Subramaniam and H. L. Scott, *J. Comput. Chem.*, **20**, 1153, (1999).

[170] M. Bloom, E. Evans and O. G. Mouritsen, *Quarterly Reviews of Biophysics*, **24**, 293, (1991).

[171] J. P. Douliez, A. Leonard and E. J. Dufourc, *Biophys. J.*, **68**, 1727, (1995).

[172] J. P. Douliez, A. Leonard and E. J. Dufourc, *Biophys. J.*, **68**, 1727, (1995).

[173] E. J. Dufourc, C. Mayer, J. Stohrer, G. Althoff and G. Kothe, *Biophys. J.*, **63**, 1081, (1992).

[174] J. Katsaras, *J. Phys. Chem.*, **99**, 4141, (1995).

[175] G. S. Smith, E. B. Sirota, C. R. Safinya, R. J. Plano and N. A. Clark, *J. Chem. Phys.*, **92**, 4519, (1990).

[176] Z. I. Lalchev, P. J. Wilde and D. C. Clark, *J. Colloid Interface Sci.*, **167**, 80, (1994).

[177] S. Konig, W. Pfeiffer, T. Bayerl, D. Richter and E. Sackmann, *J. Phys. II*, **2**, 1589, (1992).

[178] R. Goetz, G. Gompper and R. Lipowsky, *Phys. Rev. Lett.*, **82**, 221, (1999).

[179] F. K. Von Gottberg, K. A. Smith and T. A. Hatton, *J. Chem. Phys.*, **106**, 9850, (1997).

[180] H. Noguchi and M. Takasu, *Phys. Rev. E*, **6404**, art. no., (2001).

[181] *Pharmacokinetics and Metabolism in Drug Design*, Wiley-VCH Verlag GmbH, (2001).

[182] C. A. Lipinski, F. Lombardo, B. W. Domony and P. J. Feeney, *Adv. Drug Del. Rev.*, **23**, 3, (1997).

[183] R. Borchardt, P. Smith and G. Wilson, *Models for Assesing Drug Absorption and Metabolism*, Plenum Press, New York, (1996).

[184] L. Barthe, J. Woodley and G. Houin, *Fund. Clin. Pharmacol.*, **13**, 154, (1999).

[185] R. C. Bean, W. C. Shepherd and H. Chan, *J. Gen. Physiol.*, **52**, 495, (1968).

[186] A. Walter and J. Gutknecht, *J. Membr. Biol.*, **77**, 255, (1984).

[187] A. Walter and J. Gutknecht, *J. Membr. Biol.*, **90**, 207, (1986).

[188] J. M. Wolosin and H. Ginsburg, *Biochimica et Biophysica Acta*, **389**, 20, (1975).

[189] A. Finkelstein, *J. Gen. Physiol.*, **68**, 127, (1976).

[190] E. Orbach and A. Finkelstein, *J. Gen. Physiol.*, **75**, 427, (1980).

[191] J. Brunner, D. E. Graham, H. Hauser and G. Semenza, *J. Membr. Biol.*, **57**, 133, (1980).

[192] J. Gutknecht and A. Walter, *Biochimica et Biophysica Acta*, **649**, 149, (1981).

[193] Z. Bar-On and H. Degani, *Biochimica et Biophysica Acta*, **813**, 207, (1985).

[194] A. Bochain, L. Estey, G. Haronian, C. Reale, M. andRojas and J. Cramer, *J. Membr. Biol.*, **60**, 73, (1981).

[195] M. B. Lande, J. M. Donovan and M. L. Zeidel, *J. Gen. Physiol.*, **106**, 67, (1995).

[196] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **148**, 157, (1995).

[197] T. X. Xiang and B. D. Anderson, *Biophys. J.*, **72**, 223, (1997).

[198] T. X. Xiang and B. D. Anderson, *Biophys. J.*, **75**, 2658, (1998).

[199] T. X. Xiang and B. D. Anderson, *Biochimica et Biophysica Acta*, **1370**, 64, (1998).

[200] G. J. M. Bresseleers, H. L. Goderis and P. P. Tobback, *Biochimica et Biophysica Acta*, **772**, 374, (1984).

[201] J. A. Dix, D. Kivelson and J. M. Diamond, *J. Membr. Biol.*, **40**, 315, (1978).

[202] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **173**, 187, (2000).

[203] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **177**, 137, (2000).

[204] T. X. Xiang, X. Chen and B. D. Anderson, *Biophys. J.*, **63**, 78, (1992).

[205] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **140**, 111, (1994).

[206] T. X. Xiang and B. D. Anderson, *J. Chem. Phys.*, **103**, 8666, (1995).

[207] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **165**, 77, (1998).

[208] T. X. Xiang, *J. Chem. Phys.*, **109**, 7876, (1998).

[209] T. X. Xiang, *J. Phys. Chem. B*, **103**, 385, (1999).

[210] P. Artursson, K. Palm and K. Luthman, *Adv. Drug Del. Rev.*, **22**, 67, (1996).

[211] P. Artursson and J. Karlsson, *Biochem. Biophys. Res. Commun.*, **175**, 880, (1991).

[212] A. Pohorille and M. A. Wilson, *J. Chem. Phys.*, **104**, 3760, (1996).

[213] B. Widom, *J. Chem. Phys.*, **39**, 2808, (1963).

[214] B. Widom, *J. Chem. Phys.*, **86**, 869, (1982).

[215] Y. Liu and T. Ichiye, *J. Phys. Chem.*, **100**, 2723, (1996).

[216] W. van Gunsteren, *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications*, volume 1, ESCOM, Leiden, The Netherlands, (1989).

[217] E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, *Chem. Phys. Lett.*, **156**, 472, (1989).

[218] E. Paci, G. Ciccoti, M. Ferrario and R. Kapral, *Chem. Phys. Lett.*, **176**, 581, (1991).

[219] M. Sprik and G. Ciccotti, *J. Chem. Phys.*, **109**, 7737, (1998).

[220] D. A. Pearlman, *J. Chem. Phys.*, **98**, 8946, (1993).

[221] W. K. Den Otter and W. J. Briels, *J. Chem. Phys.*, **109**, 4139, (1998).

[222] W. K. Den Otter and W. J. Briels, *Mol. Phys.*, **98**, 773, (2000).

[223] W. K. Den Otter, *J. Chem. Phys.*, **112**, 7283, (2000).

[224] B. Roux and M. Karplus, *J. Phys. Chem.*, **95**, 4856, (1991).

[225] R. Kubo, *Reports on Progress in Physics*, **29**, 255, (1966).

[226] C. M. Gary-Bobo and H. W. Weber, *J. Phys. Chem.*, **73**, 1155, (1969).

[227] R. Niesner and H. A., *J. Chem. Eng. Data*, **45**, 1121, (2000).

[228] P. Witherspoon and D. Saraf, *J. Phys. Chem.*, **69**, 3752, (1965).

Appendix A

Forces and Torques

A.1 Determining the Generalized Gay-Berne Potential Forces and Torques

Luckhurst et al.[111] described how to derive forces and torques for the Gay-Berne potential, the same methodology can be applied to derive the forces and torques for the generalized Gay-Berne potential. We just developed forces and torques for the particular case of the generalized Gay-Berne potential where one particle is made spherical. In the following we consider two particles i and j as defined in paragraph 4.4.1, particle i is considered to be spherical. The force experienced by a particle subject to a potential U can be defined from the differential of the pair potential with respect to movement along each of the three space-fixed axes,

$$\hat{F}_i = -\begin{pmatrix} \partial U/\partial x_i \\ \partial U/\partial y_i \\ \partial U/\partial z_i \end{pmatrix}. \tag{A.1}$$

The torque experienced by a particle under the influence of a potential U can be calculated from the differential of the potential with respect to the particle's direction cosines,$\hat{u}_{xi}$,$\hat{u}_{yi}$,$\hat{u}_{zi}$, if we define the vector

$$\hat{E}_i = -\begin{pmatrix} \partial U/\partial \hat{u}_{xi} \\ \partial U/\partial \hat{u}_{yi} \\ \partial U/\partial \hat{u}_{zi} \end{pmatrix}. \tag{A.2}$$

then the torque experienced by a paticle i is

$$\hat{\tau} = \hat{u}_i \wedge \hat{E}_i \tag{A.3}$$

where $\wedge$ represents a vector product. The intermolecular potential can take the shape :

$$U(\hat{u}_j, \mathbf{r}) = 4\epsilon(\hat{u}_j, \hat{r}_{ij}) \left[R^{-12} - R^{-6}\right], \tag{A.4}$$

where R is a shifted distance and is defined as :

$$R = \frac{r - \sigma(\hat{u}_j, \hat{r}_{ij}) + \sigma_\perp}{\sigma_\perp}. \tag{A.5}$$

The interaction strength parameter $\epsilon(\hat{u}_j, \hat{r}_{ij})$ is defined in equations 4.39 and 4.40 and $\sigma(\hat{u}_j, \hat{r}_{ij})$ in equations 4.37 and 4.38.

A.1.1 Derivatives with respect to translation along a space-fixed axis

Here we form the derivative along the x-axis. The derivatives with respect to other axes can be obtained by permutation of the subscripts.

$$\frac{\partial U}{\partial x} = 4\epsilon(\hat{u}_j, \hat{r}_{ij}) \left[-12R^{-13} + 6R^{-7}\right] \frac{\partial R}{\partial x} + 4\left[R^{-12} - R^{-6}\right] \frac{\partial \epsilon(\hat{u}_j, \hat{r}_{ij})}{\partial x} \tag{A.6}$$

$$\frac{\partial R}{\partial x} = \frac{1}{\sigma_\perp}\left[\frac{\partial r}{\partial x} - \sigma_0 \chi \alpha^{-2}(\hat{r}_{ij}.\hat{u}_j)\frac{\partial(\hat{r}_{ij}.\hat{u}_j)}{\partial x}\left(1 - \chi\alpha^{-2}(\hat{r}_{ij}.\hat{u}_j)^2\right)^{-\frac{3}{2}}\right] \tag{A.7}$$

where

$$\frac{\partial r}{\partial x} = x/r \tag{A.8}$$

and

$$\frac{\partial(\hat{r}_{ij}.\hat{u}_j)}{\partial x} = \hat{u}_{xj} \tag{A.9}$$

$$\frac{\partial \epsilon(\hat{u}_j,\hat{r}_{ij})}{\partial x} = -2\epsilon_0 \mu \chi' \alpha'^{-2} (\hat{r}_{ij}.\hat{u}_j) \frac{\partial(\hat{r}_{ij}.\hat{u}_j)}{\partial x} \left[1 - \chi' \alpha'^{-2} (\hat{r}_{ij}.\hat{u}_j)^2\right]^{\mu-1} \tag{A.10}$$

The substitution of A.7 and A.10 into A.6 gives the complete explicit formula for $\partial U/\partial x$.

A.1.2 Derivatives with respect to direction cosines

Here we form the derivatives with respect to u_{xj}, the component parallel to x of the unit vector defining the orientation of particle j.

$$\frac{\partial U}{\partial u_{xj}} = 4\epsilon(\hat{u}_j,\hat{r}_{ij}) \left[-12R^{-13} + 6R^{-7}\right] \frac{\partial R}{\partial u_{xj}} + 4\left[R^{-12} - R^{-6}\right] \frac{\partial \epsilon(\hat{u}_j,\hat{r}_{ij})}{\partial u_{xj}} \tag{A.11}$$

$$\frac{\partial R}{\partial u_{xj}} = -\frac{1}{\sigma_\perp} \frac{\partial \sigma(\hat{u}_j,\hat{r}_{ij})}{\partial u_{xj}} = -\frac{1}{\sigma_\perp} \sigma_0 \chi \alpha^{-2} (\hat{r}_{ij}.\hat{u}_j) \frac{\partial(\hat{r}_{ij}.\hat{u}_j)}{\partial u_{xj}} \left[1 - \chi\alpha^{-2} (\hat{r}_{ij}.\hat{u}_j)^2\right]^{-\frac{3}{2}} \tag{A.12}$$

where

$$\frac{\partial(\hat{r}_{ij}.\hat{u}_j)}{\partial u_{xj}} = \hat{r}_{xij} \tag{A.13}$$

$$\frac{\partial \epsilon(\hat{u}_j,\hat{r}_{ij})}{\partial u_{xj}} = -2\epsilon_0 \mu \chi' \alpha'^{-2} (\hat{r}_{ij}.\hat{u}_j) \frac{\partial(\hat{r}_{ij}.\hat{u}_j)}{\partial u_{xj}} \left[1 - \chi' \alpha'^{-2} (\hat{r}_{ij}.\hat{u}_j)^2\right]^{\mu-1} \tag{A.14}$$

Appendix B

Contact Distance Error Functions

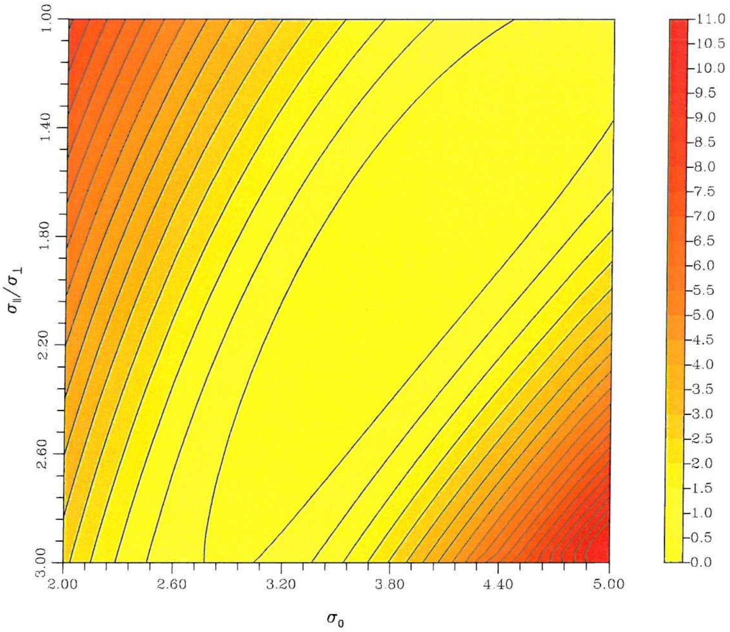


Figure B.1: 2D contour for the contact distance error function for configuration 1 as a function of σ_0 (X axis) and $\sigma_{\parallel}/\sigma_{\perp}$ (Y axis). All other parameters are fixed and have the values : ϵ_0=7.005 kJ.mol^{-1}, $\epsilon_{\perp}/\epsilon_{\parallel}$=5.31, μ=2 and ν=1. Minimum location $(\sigma_0,\sigma_{\parallel}/\sigma_{\perp})$=(3.98,1.76). The contour are shown for values of the error function in between 0.0 and 11.0 in steps of 0.5.

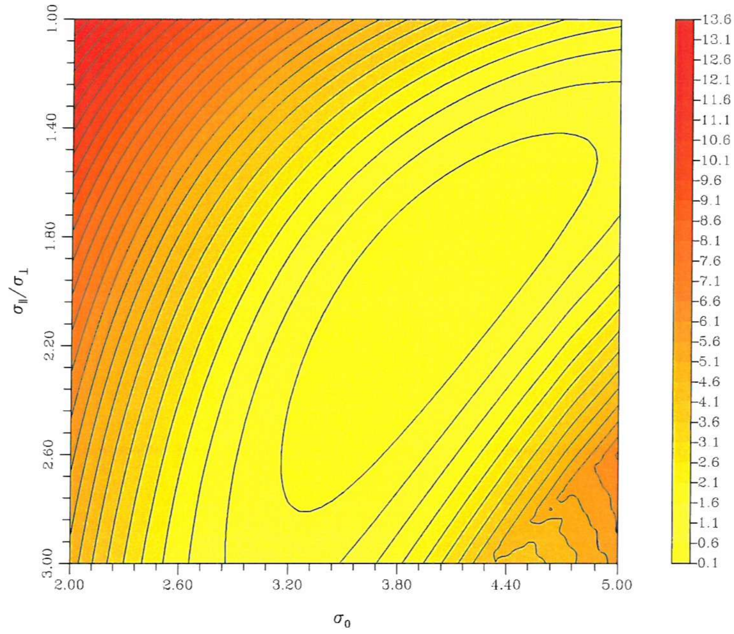


Figure B.2: 2D contour for the contact distance error function for configuration 2 as a function of σ_0 (X axis) and $\sigma_{\parallel}/\sigma_{\perp}$ (Y axis). All other parameters are fixed and have the values : ϵ_0=7.005 kJ.mol^{-1}, $\epsilon_{\perp}/\epsilon_{\parallel}$=5.31, μ=2 and ν=1. Minimum location (σ_0,$\sigma_{\parallel}/\sigma_{\perp}$)=(3.98,1.96). The contour are shown for values of the error function in between 0.1 and 13.6 in steps of 0.5.

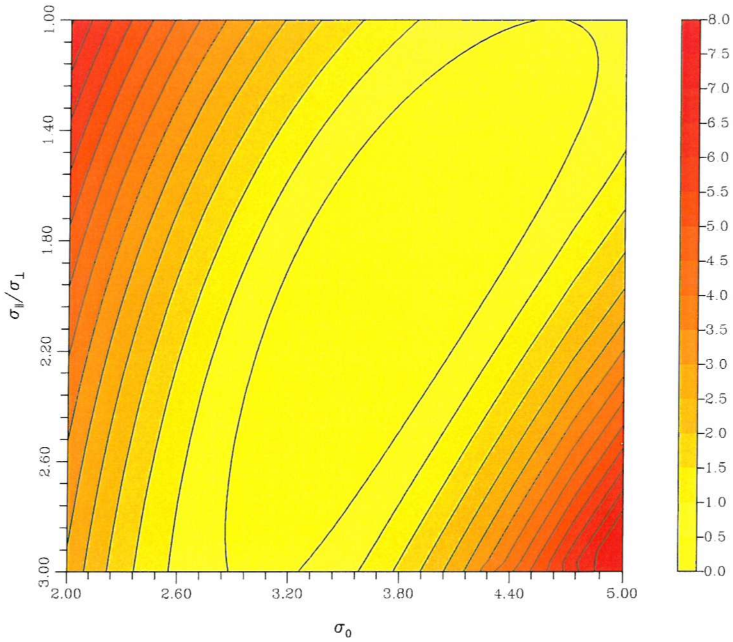


Figure B.3: 2D contour for the contact distance error function for configuration 3 as a function of σ_0 (X axis) and $\sigma_{\parallel}/\sigma_{\perp}$ (Y axis). All other parameters are fixed and have the values : ϵ_0=7.005 kJ.mol^{-1}, $\epsilon_{\perp}/\epsilon_{\parallel}$=5.31, μ=2 and ν=1. Minimum location (σ_0,$\sigma_{\parallel}/\sigma_{\perp}$)=(3.80,1.88). The contour are shown for values of the error function in between 0.0 and 8.0 in steps of 0.5.

Appendix C

Well Depth Error Functions

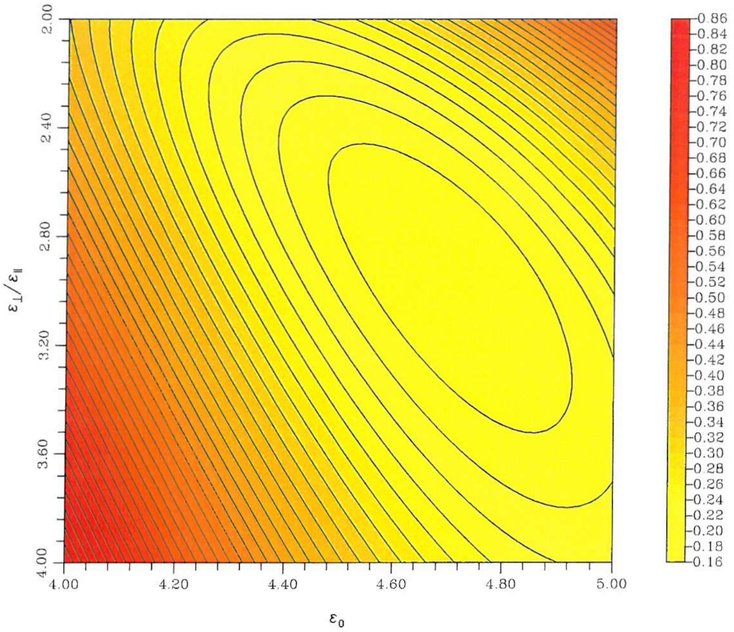


Figure C.1: 2D contour for the well depth error function for configuration 1 as a function of ϵ_0 (X axis) and $\epsilon_\perp/\epsilon_\parallel$ (Y axis). All other parameters are fixed and have the values : σ_0=3.86 Å, $\sigma_\parallel/\sigma_\perp$=1.96, μ=2 and ν=1. Minimum location (ϵ_0,$\epsilon_\perp/\epsilon_\parallel$)=(4.70,2.96). The contour are shown for values of the error function in between 0.16 and 0.86 in steps of 0.02.

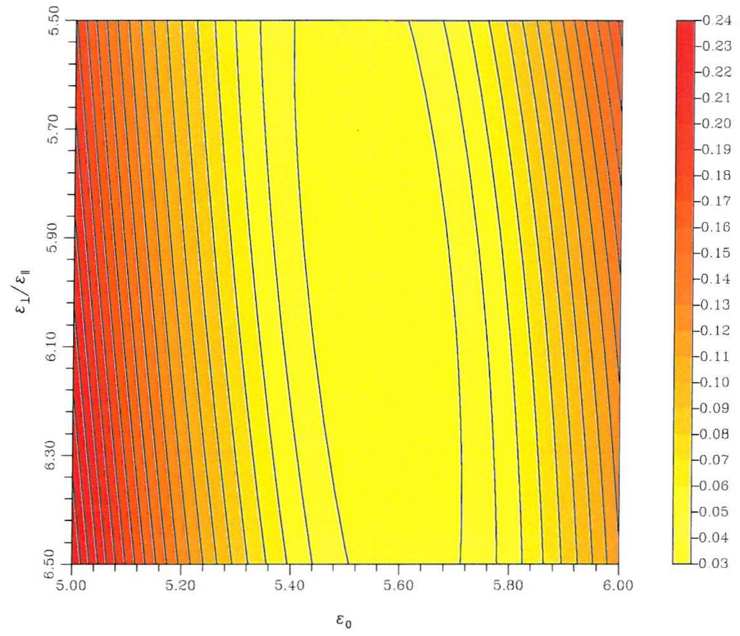


Figure C.2: 2D contour for the well depth error function for configuration 2 as a function of ϵ_0 (X axis) and $\epsilon_\perp/\epsilon_\parallel$ (Y axis). All other parameters are fixed and have the values : σ_0=3.86 Å, $\sigma_\parallel/\sigma_\perp$=1.96, μ=2 and ν=1. Minimum location $(\epsilon_0,\epsilon_\perp/\epsilon_\parallel)$=(5.56,5.96). The contour are shown for values of the error function in between 0.03 and 0.24 in steps of 0.01.

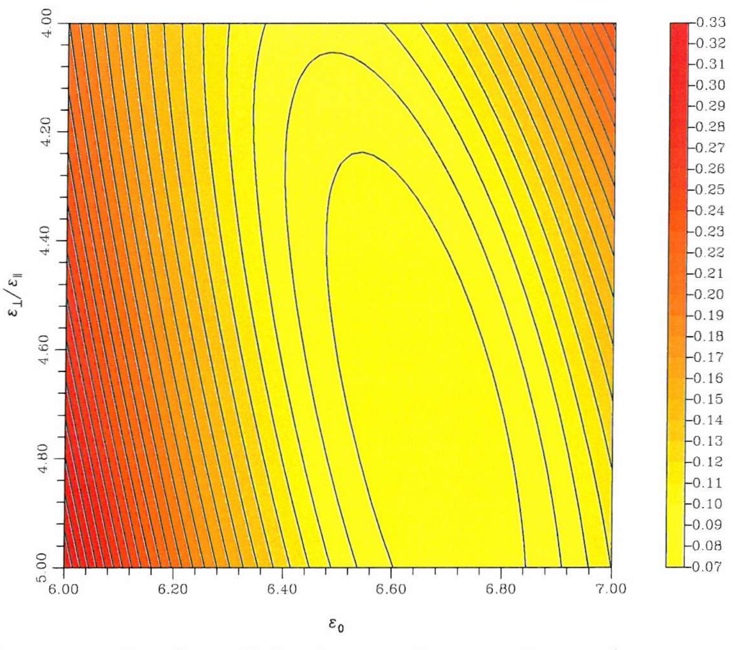


Figure C.3: 2D contour for the well depth error function for configuration 3 as a function of ϵ_0 (X axis) and $\epsilon_\perp/\epsilon_\parallel$ (Y axis). All other parameters are fixed and have the values : σ_0=3.86 Å, $\sigma_\parallel/\sigma_\perp$=1.96, μ=2 and ν=1. Minimum location $(\epsilon_0,\epsilon_\perp/\epsilon_\parallel)$=(6.66,4.72). The contour are shown for values of the error function in between 0.07 and 0.33 in steps of 0.01.