UNIVERSITY OF SOUTHAMPTON

Computer Simulation of Biological Membranes and Small Molecule Permeation

Daniele BEMPORAD

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Southampton.

Department of Chemistry

April 2003

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE

CHEMISTRY

Doctor of Philisophy

COMPUTER SIMULATION OF BIOLOGICAL MEMBRANES AND SMALL MOLECULE PERMEATION

by Daniele Bemporad

In this thesis, the results of a series of computer simulation studies into the thermodynamics and the kinetics of biological membranes and permeation processes are presented. The first chapter contains a brief overview of the project and its aim, and the second introduces the techniques of molecular dynamics computer simulations. In chapter 3, the biological background of cell membranes is presented and experimental techniques to study properties of pure lipid bilayers are described. Previous simulation studies are also reviewed. The membrane computer model is the subject of chapter 4, containing the simulation protocol and results. Membrane physical properties as they result from simulations are compared with available experimental data to validate the computer model proposed and the latter is shown to be physically realistic. Chapter 5 describes how permeability coefficients across lipid bilayers and cell monolayers can be measured in experiments. A technique which allows for the permeability coefficients to be obtained from molecular dynamics simulations is also described in detail. In chapter 6, the latter technique is employed to study the permeation of small organic solutes across the membrane computer model previously described and is validated by comparison with experimental data, which are available for most of the chosen small molecules. This yields new insights into the mechanism of membrane permeation. Chapter 7 describes the choice, the parameterization and the conformational study of candidate drug molecules, for which permeability coefficients across the membrane model are calculated in chapter 8 with the same methodolgy used for the small organic solutes in chapter 6. Results are compared with permeability coefficients through Caco-2 cells, which are commonly employed to measure the ability of drugs to permeate human tissues, and help to elucidate the process of drug absorption after administration. The final chapter contains conclusions and future work. The appendix reports the parameters developed and employed for simulating the drug molecules.

Acknowledgements

I wish to greatly thank Dr. J. W. Essex for all his advice, encouragement and enthusiasm these years. Also thanks to Claude Luttmann at Aventis for funding the project as my industrial supervisor.

I would also like to thank all the members of the computing suite for their vital help to my work and their very good company. Namely (in alphabetical order): Adrian, Chris, David, Donna, Francesca, Ian, Julen, Julien, Lewis, Loredana, Luca, Martin, Phil, Richard H., Richard M., Richard T., Rob, Seb, Steve, Stuart, Tim. A special thought goes to David, whose desk was next to mine for more than two years. He is a kind friend, and shared with me a similar project and identical "sufferings". Finally, I will never forget the enjoyable time I spent with the italian mates Benedetta, Daniela, Francesca, Giuseppina, Luca, Paolo e Roberta.

Contents

List of Tables

List of Figures

CHAPTER 1

Overall aim of the project

The ease with which small molecule pharmaceutical agents can pass through biological membranes is a critical factor influencing their *in vivo* potency, since drug disposition and rate of drug appearance in the receptor surroundings are decisive aspects of drug action. Understanding the kinetics and the thermodynamics of this process is then a prerequisite for rational drug development.[7] Means exist for measuring these effects either directly with artificial membranes or indirectly by consideration of partition coefficient measurements. Using these methods[8,9] a quantitative relationship can be estimated between the structure of a molecule and its behaviour within the cell membranes. The structure of a molecule can be expressed in terms of its physico-chemical properties and these properties can be quantified using mathematical descriptors, e.g. the octanol/water distribution coefficient for the lipophilicity, the molecular weight for the molecular size, the number of H-bond donors and acceptors for the molecule hydrogen-bonding ability, the polar surface area (PSA)[10] for the hydrophilicity. Then, mathematical equations can quantitatively correlate these descriptors with the known bioavailability data for this molecule to develop a method to predict drug behaviour. This kind of approach is the basis for Quantitative Structure-Activity Relationship (QSAR) studies.[8,9]

However, empirical approaches employing statistical techniques to select the decisive drug properties and easy-to-fit models are often too crude to reveal crucial details in the strongly non-

linear disposition-property dependencies. As a consequence, the subcellular pharmacokinetic approach has the potential to reveal more details about the structure of drugs with suitable dispositions more accurately than either traditional QSAR studies or simple rules for identification of compounds with potential permeability problems.

The most used and easily acceptable route of administration is the oral one.[7] Gastrointestinal absorption requires crossing the epithelial cell membranes. Even drugs able to diffuse via the paracellular route in the extracellular space soon encounter cell membranes to be crossed in order to reach their biological target. Therefore, the entire process of absorption from the site of administration into the bloodstream and of subsequent distribution from the blood into tissues can be seen as the diffusion of drug molecules in a series of membranes and aqueous compartments.[7] Most of the drugs cross cell membranes by passive diffusion without the help of protein carriers, unless they are analogues of physiological molecules.[7] An understanding of passive diffusion processes through phospholipid bilayer membranes is therefore crucial for subcellular pharmacokinetics and rational drug design.[7]

In any case, there is still a need to investigate these diffusional processes at a molecular level, and computer simulations offer a powerful way of studying diffusion at the single molecule level yielding detailed dynamic and thermodynamic data. From such data it may be possible to build model systems with predictive ability to measure the diffusion processes of potential drug molecules. The application of computer simulations to biological membranes as a method of studying drug transport and drug partitioning is therefore proposed. The project is summarized as follows:

1. Construction of a computer model of a biological membrane and validation of the simulations through comparisons with the available experimental data;

2. Simulation of, first, a set of small molecules and, then, real drug molecules within this membrane to determine depths of partitioning, rates of diffusion and profiles of permeation resistance;

3. Correlation of these results with known bioavailability data.

CHAPTER 2

Computer Simulation Methodology

The following chapter will describe how the dynamic evolution of a system can be simulated, how thermodynamic properties can be collected and evaluated, how simulations can be constrained to reproduce experimental conditions, and what time saving techniques exist to reduce the computational expense.

2.1 Statistical Mechanics

Let us consider the instantaneous mechanical states of a chemical system consisting of N interacting particles to be characterised by their positions in space and their momenta (assuming a classical approximation). That is, let us consider the set of $3N$ coordinates and $3N$ momenta $(\boldsymbol{r}^N, \boldsymbol{p}^N)$. Each set is a snapshot of the chemical system and is described as a location in phase space. An ensemble can be considered to be a collection of points in phase space, i.e a collection of snapshots.

The value of a particularly interesting system 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 measured experimentally can be considered as an average of $A(\boldsymbol{p}^N, \boldsymbol{r}^N)$ over time, and is therefore referred to as a time

average:

$$A_{obs} = \langle A \rangle_{time} = \lim_{t_{obs} \to \infty} \frac{1}{t_{obs}} \int_0^{t_{obs}} A(\boldsymbol{p}^N(t), \boldsymbol{r}^N(t)) dt \tag{2.1}$$

To calculate the average values of the properties of this system the simulation of its dynamic behaviour is required. This can be achieved[11,12] by describing an energy function that calculates the forces within the system due to interatomic interactions, and from the force the acceleration can be determined. Integrating Newton's laws of motion then yields a trajectory which describes the positions, velocities and accelerations of the particles in the system over time, enabling the calculation of system properties directly using equation 2.1. This is exactly what is done in Molecular Dynamics (MD) simulations.[11,12] A major problem is that in simulations the integral cannot be calculated to infinity. The value of A is then averaged over a sufficiently long time t_{MD}. In fact, the motion of the system is described on a step-by-step basis, i.e. a large finite number τ_{MD} of time steps of length $\Delta t = t_{MD}/\tau_{MD}$ are taken. In this case equation 2.1 may be rewritten in the form:

$$A_{MD} = \langle A \rangle_{time} = \frac{1}{\tau_{MD}} \sum_{\tau=1}^{\tau_{MD}} A(\boldsymbol{p}^N(\tau), \boldsymbol{r}^N(\tau)) \tag{2.2}$$

Because of the complexity of the time evolution of A, a different approach is to replace the single system evolving in time by a large number of replications that are considered simultaneously. The time average calculation in equation 2.1 can then be replaced by an ensemble average, where the average value of the property A is calculated over all the copies of the system:

$$A_{obs} = \langle A \rangle_{ensemble} = \int \int d\boldsymbol{p}^N d\boldsymbol{r}^N A(\boldsymbol{p}^N, \boldsymbol{r}^N) \rho_{ens}(\boldsymbol{p}^N, \boldsymbol{r}^N) \tag{2.3}$$

The ergodic hypothesis is a fundamental axiom of statistical mechanics and assumes that the ensemble average is equal to the time average, allowing the transformation of equation 2.1 to 2.3,[11,12] with $\rho_{ens}(\boldsymbol{p}^N, \boldsymbol{r}^N)$ being the probability of finding a certain value of $A(\boldsymbol{p}^N, \boldsymbol{r}^N)$ in that ensemble. This is what is done in Monte Carlo (MC) simulations.[11,12] A succession of a large finite number τ_{MC} of state points are generated to sample in accordance with the desired ρ_{ens} and the value of A is obtained as:

$$A_{MC} = \langle A \rangle_{ensemble} = \frac{1}{\tau_{MC}} \sum_{\tau=1}^{\tau_{MC}} A(\boldsymbol{p}^N(\tau), \boldsymbol{r}^N(\tau)) \tag{2.4}$$

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, another arrives to replace it. Such a distribution of states is called a *Boltzmann distribution.* The probability density of each 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, and is calculated differently in different ensembles.

In the canonical ensemble (NVT), that is when the number of particles, the volume and the temperature of the system remain constant, this ratio is equal to:

$$\rho_{NVT}\left(\boldsymbol{p}^N, \boldsymbol{r}^N\right) = \frac{n_i}{N} = \frac{\exp(-\mathscr{H}_i/k_BT)}{Q_{NVT}} \tag{2.5}$$

where $\mathscr{H}_i$ is the Hamiltonian of state i, k_B the Boltzmann constant, T the temperature, and Q_{NVT} the so called partition function. The latter plays a fundamental role in statistical mechanics, and it is from this property that all thermodynamic functions can be derived. The molecular partition function in the canonical ensemble is defined as:

$$Q_{NVT} = \frac{1}{h^{3N}N!}\int\int d\boldsymbol{p}^N d\boldsymbol{r}^N \exp(-\mathscr{H}_i/k_BT) \tag{2.6}$$

where $\mathscr{H}_i$ is $\mathscr{H}(\boldsymbol{p}^N, \boldsymbol{r}^N)$. Since $\mathscr{H}$ is always expressible as a sum of kinetic $\mathscr{K}(\boldsymbol{p}^N)$ and potential $\mathscr{V}(\boldsymbol{r}^N)$ contributions, then:

$$\begin{aligned} Q_{NVT} &= \frac{1}{N!}\frac{1}{h^{3N}}\int d\boldsymbol{p}^N \exp(-\mathscr{K}(\boldsymbol{p}^N)/k_BT)\int d\boldsymbol{r}^N \exp(-\mathscr{V}(\boldsymbol{r}^N)/k_BT) = \\ &= Q^{id}_{NVT}Q^{exc}_{NVT} \end{aligned} \tag{2.7}$$

For an atomic system:

$$Q^{id}_{NVT} = \frac{V^N}{(h^2/2\pi m k_BT)^{3N/2}N!} \quad , \quad Q^{exc}_{NVT} = \frac{1}{V^N}\int d\boldsymbol{r}^N \exp(-\mathscr{V}(\boldsymbol{r}^N)/k_BT) \tag{2.8}$$

In Monte Carlo (MC) simulations it therefore prooves possible to probe just the configurational part of phase space, with the ideal gas properties being added afterwards. Equation 2.4 can then be rewritten as:

$$A_{MC} = \frac{1}{\tau_{MC}}\sum_{\tau=1}^{\tau_{MC}} A(\boldsymbol{r}^N(\tau)) \tag{2.9}$$

In the isothermal-isobaric ensemble (NPT), that is when the number of particles, the pressure and the temperature of the system remain constant, the probability density is equal to:

$$\rho_{NPT}\left(\boldsymbol{p}^N, \boldsymbol{r}^N\right) = \frac{n_i}{N} = \frac{\exp(-(\mathscr{H}_i + PV)/k_B T)}{Q_{NPT}} \tag{2.10}$$

where P and V are the pressure and the volume of the system, respectively. Then, the partition function is:

$$Q_{NPT} = \frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{V_0}\int\int\int d\boldsymbol{p}^N d\boldsymbol{r}^N dV \exp(-(\mathscr{H}_i + PV)/k_B T) \tag{2.11}$$

where V_0 is the unit of volume, required to render Q_{NPT} dimensionless. Again, it is possible to separate the configurational properties from the kinetic ones and to devise a Monte Carlo procedure to probe configurational space only.

2.2 Force-field and Potential Energy Function

To sample the positions $\boldsymbol{r}$ and momenta $\boldsymbol{p}$ of the N particles comprising the simulation system, a functional form for the calculation of the potential energy is required. This functional form is called the force-field and it assigns energetic penalties when the particles deviate from their equilibrium positions. In this approach, called molecular-mechanics,[11,12] the molecular properties are considered to be dependent on the nuclei positions only and electron motion is ignored. A classical description is also employed and the Hamiltonian $\mathscr{H}$ of the simulation system is written as the sum of the particle kinetic $\mathscr{K}$ and potential $\mathscr{V}$ energies:

$$\mathscr{H}(\boldsymbol{r}, \boldsymbol{p}) = \mathscr{K}(\boldsymbol{p}) + \mathscr{V}(\boldsymbol{r}) \tag{2.12}$$

In real systems the interaction of a particle with another is influenced by the presence of all the other particles, but in most computer simulations a pairwise approximation gives a good description of liquid, gas and solid phase properties.

In CHARMM,[13] one of the most common force-fields, the potential energy $\mathscr{V}(\boldsymbol{r}^N)$ of a system of N particles as a function of the positions $\boldsymbol{r}^N$ is computed as the sum of the contributions from bond stretching, angle bending, dihedral torsions and non-bonded interactions (electrostatic and van der Waals interactions):

$$\mathscr{V}(\boldsymbol{r}_N) = \sum \mathscr{V}_{bonds} + \sum \mathscr{V}_{angles} + \sum \mathscr{V}_{dihedrals} + \sum \mathscr{V}_{electrostatic} + \sum \mathscr{V}_{van-der-Waals} \tag{2.13}$$

The bond and angle contributions are described by harmonic potentials that give the increase in energy as the bond length b_i and the valence angle θ_i deviate from their reference values $b_{i,0}$ and $\theta_{i,0}$, by means of a force constant $k_{i,b}$ and $k_{i,\theta}$:

$$\sum \mathscr{V}_{bonds} = \sum_{i=1}^{N_{bonds}} k_{i,b} \left(b_i - b_{i,0}\right)^2 \tag{2.14}$$

$$\sum \mathscr{V}_{angles-harmonic} = \sum_{i=1}^{N_{angles}} k_{i,\theta} \left(\theta_i - \theta_{i,0}\right)^2 \tag{2.15}$$

The CHARMM force-field also includes a cross term to describe the angle terms. The angle bending, which involves the atoms 1, 2 and 3, is modelled using the Urey-Bradley potential, which considers the van der Waals interactions between the atoms 1 and 3 along their distance r_{1-3}:

$$\sum \mathscr{V}_{angles-Urey-Bradley} = \sum_{i=1}^{N_{angles}} k_{i,UB} \left(r_{i,1-3} - r^0_{i,1-3}\right)^2 \tag{2.16}$$

Here $r^0_{i,1-3}$ is the reference distance and $k_{i,UB}$ the force constant. The torsional terms are computed as:

$$\sum \mathscr{V}_{dihedrals} = \sum_{i=1}^{N_{dihedrals}} k_{i,\phi} \left[1 + \cos\left(n_i \phi_i - \delta_i\right)\right] \tag{2.17}$$

Here ϕ_i is the dihedral angle, $k_{i,\phi}$ the force constant, n_i is the multiplicity (which gives the number of minimum points in the function as the torsion changes from 0 to 360 degrees) and δ_i is the phase angle (which determines the location of the minimum values).

The non-bonded interactions are calculated between all pairs of atoms i and j that are in different molecules or in the same molecule but separated by at least three bonds (1,4 relationship). The electrostatic interactions are usually modelled using a Coulombic potential term:

$$\sum \mathscr{V}_{electrostatic} = \sum^{pairs(ij)} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.18}$$

where q_i and q_j are the charge on atom i and atom j respectively, r_{ij} their separation distance, and ϵ_0 the relative permittivity of free space. The van der Waals interactions are modelled

using a Lennard-Jones (LJ) potential:

$$\sum \mathscr{V}_{van-der-Waals} = \sum^{pairs(ij)} \epsilon \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \tag{2.19}$$

where ϵ is the well depth at the energy minimum and $R_{min,ij}$ is the i-j distance at the energy minimum. In common force fields, these parameters are derived for pairs of identical atoms,[12] for instance ii and jj. When interactions between different types of atoms are required, it is common to apply mixing rules, e.g. in CHARMM:

$$R_{min,ij} = \frac{R_{min,ii}}{2} + \frac{R_{min,jj}}{2} \quad , \quad \epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2} \tag{2.20}$$

The i-j distance at which the LJ potential energy is zero is usually referred to as σ_{ij} and is equal to $2^{-1/6} R_{min,ij}$.

Out of plane bending, such as that in carbonyl groups, is incorporated by treating the atoms involved using an "improper" torsion angle ψ_i and incorporating an additional energy term in the force-field. In the CHARMM force-field these "improper" torsions are described using a harmonic potential:

$$\sum \mathscr{V}_{improper} = \sum_{i=1}^{N_{improper}} k_{i,\psi} \left(\psi_i - \psi_{i,0} \right)^2 \tag{2.21}$$

Force-fields are empirically determined, that is there are no rules that allow one to define *a priori* equilibrium values, force constants, charges and van der Waals parameters, but rather several values are employed in an iterative fashion until the simulated system has a static and dynamic behaviour as close as possible to that of the real system. At the beginning of a force-field parameterization, equilibrium values, force constants, charges and van der Waals parameters are usually obtained from both quantum-mechanical calculations and spectroscopic data, such as X-ray diffraction, IR and NMR. They are then modified and refined to reproduce, during a simulation, available experimental data for the system of interest.[12]

Force-fields describe in a classical formalism the potential energy of a system of any kind of particle. In computational chemistry each atom may be explicitly represented in the model, but also a group of two or more atoms may be modelled as a single "pseudo-atom". In the first case the force-field is called "all-atom", while in the second case the force-field is called "united-atom". Usually, just the hydrogen atoms bonded to non-polar atoms such as carbon

are modelled as united-atoms, while hydrogens bonded to polar atoms such as nitrogen or oxygen that are able to participate in hydrogen bonding interactions are modelled explicitly. With united-atom force-fields a considerable computational saving is possible, since the number of particles simulated is reduced and consequently the number of interactions to be calculated is smaller. However, united-atom force-fields have some drawbacks. With this kind of representation the level of detail and accuracy is lower.[12] Moreover, chiral centres may be able to invert during the simulation because of the absence of an explicit hydrogen atom on the chiral carbon, and so an additional improper torsion must be added into the force-field to prevent the inversion.[12]

2.3 Energy Minimisation Methods

Energy minimisation methods are typically applied to simulation systems prior to modelling studies, but are also useful for insights into molecular structure and, occasionally, sufficient to predict the properties of a system accurately.[12] The procedure of minimisation optimises the atomic positions subject to the forces generated by the molecular force field. Minimisation algorithms are designed to move atoms in such a way as to reduce the energy of the molecule until a minimum value is reached or an agreed energy convergence criterion has been met. Two common minimisation algorithms are the steepest descent and the conjugate gradient methods.[12]

The steepest descent method makes moves parallel to the direction of the force on a particle, so a move is taken directly downhill on the potential energy surface, with either a step of arbitrary size, or, from the result of a line search, to the minimum energy position along the new direction. Then the process is repeated without using any information from the previous move. Conversely, in the conjugate gradient method the previous step directions are used to refine the current direction of move.

More efficient minimisation algorithms include second derivative information and perform better along flat areas of the potential energy surface. The Newton-Raphson algorithm is an example.[12]

2.4 Molecular Dynamics Simulations or Monte Carlo Simulations

The Monte Carlo (MC) method produces a set of configurations based on the biased sampling of phase space. The method generates configurations randomly, using a special set of criteria to decide whether a MC move can be accepted or rejected. This procedure involves choosing a particle i, randomly or sequentially, followed by a trial translational or rotational displacement of this particle to generate a new configuration. The potential energy difference between the new configuration and its predecessor is then calculated:

$$\Delta\mathcal{V}(\boldsymbol{r}^N) = \mathcal{V}(\boldsymbol{r}^N_{new}) - \mathcal{V}(\boldsymbol{r}^N_{old}) \tag{2.22}$$

If the difference is negative, a move downhill on the energy surface has been performed and this trial move is accepted. In contrast, an uphill move on the energy surface is not immediately accepted. A random number between 0 and 1 is generated and compared to $\exp(-\Delta H/k_BT)$. In the NVT ensemble:

$$\Delta H = \Delta\mathcal{V}(\boldsymbol{r}^N) \tag{2.23}$$

in the NPT ensemble:

$$\Delta H = \Delta\mathcal{V}(\boldsymbol{r}^N) + P(V_{new} - V_{old}) - Nk_BT\ln(V_{new}/V_{old}) \tag{2.24}$$

Here k_B is the Boltzmann constant, T the temperature, P the pressure, V the volume. If the random number is less than $\exp(-\Delta H/k_BT)$, the move is accepted, otherwise the move is rejected and the original position is retained and recounted.

This acceptance/rejection criterion ensures that the configurations collected during the MC simulation are distributed according to the Boltzmann distribution. For this reason, the calculation of interesting system properties, for instance A, can be achieved at the end of the simulation by simply averaging over the number τ_{MC} of values for that property obtained from those configurations produced, according to equation 2.9.

The MC method does not provide a straight-forward method of obtaining time dependent information , since the configurations generated during the simulation are obtained by random cartesian moves and are not connected in time. To obtain time dependent properties such as those this thesis deals with, namely diffusion coefficients, Molecular Dynamics (MD) methods are required. These methods calculate the "real" dynamics of the system by solving an equation

of motion that allows for the calculation of new particles' positions, velocities and accelerations based on the previous values of these properties, the time interval between two subsequent steps and the force acting on each particle. At the end of the simulation a *trajectory* is generated. MD simulations are generally run for hundreds of picoseconds or some nanoseconds, depending on the computational expense, and thermodynamic properties are obtained averaging over the number τ_{MD} of steps, according to equation 2.2. The subsequent section describes in more detail how a trajectory is generated.

2.5 Equation of Motion

For any arrangement of the particles in the simulation system the force $\boldsymbol{F}_i$ acting on each particle i due to its interactions with the other particles can be calculated by taking the first derivative of the energy function $\mathscr{V}_i$ with respect to the particle positions:

$$\boldsymbol{F}_i = -\nabla \mathscr{V}_i \tag{2.25}$$

These forces allow the system to change by collective motions of particles over time in a way described by Newton's second law of motion:

$$\boldsymbol{F}_i = m_i \boldsymbol{a}_i \tag{2.26}$$

where m_i is the particle mass and $\boldsymbol{a}_i$ the particle acceleration. Since it is impossible to solve the complex coupled second order differential equations required for all the particles in the system, solutions are calculated using a numerical treatment and evaluating the particle displacements over small increments of time. From position $\boldsymbol{r}(t)$, the force $\boldsymbol{F}(t)$ acting at time t is calculated for all the pairs of particles of the system, and from the force $\boldsymbol{F}(t)$ the acceleration $\boldsymbol{a}(t)$ is computed. By knowing the velocity $\boldsymbol{v}(t)$ and the acceleration $\boldsymbol{a}(t)$ the new position $\boldsymbol{r}(t+\Delta t)$ and the new velocity $\boldsymbol{v}(t+\Delta t)$ are calculated after a time interval Δt which is called the time step. The force is assumed to be constant during Δt, and then the calculation is repeated again. Initial velocities are obtained from Gaussian, Maxwell-Boltzmann or uniform distributions which allow the system to be at the desired temperature. There are several algorithms for integrating the equations of motion and they all assume that the positions, velocities and accelerations of the system can be approximated as Taylor series expansions. An example of one such algorithm is the leap-frog algorithm,[14] which solves the expansion in the following way:

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \Delta t \boldsymbol{v}\left(t + \frac{1}{2}\Delta t\right) \tag{2.27}$$

$$v\left(t+\frac{1}{2}\Delta t\right) = v\left(t-\frac{1}{2}\Delta t\right) + \Delta t a(t) \tag{2.28}$$

$$v(t) = \frac{1}{2}\left[v\left(t+\frac{1}{2}\Delta t\right) + v\left(t-\frac{1}{2}\Delta t\right)\right] \tag{2.29}$$

The choice of time step depends on the time scale of the motions in the system: it should be one order of magnitude smaller than the shortest (=fastest) motion.[12] In the case of an atomic simulation, where the simulated particles are atoms, this motion is represented by the bond stretching and so a time step of 0.5 fs is required.

Before a simulation starts to produce useful data, the system usually needs to be minimized. This is required since the initial structure, which may be derived either from X-ray crystallography or quantum-mechanical calculations or both, generally does not correspond to the most populated structure under the simulation condition of temperature, pressure and solvation. Moreover, bad contacts may occur very easily. Minimization algorithms find energy arrangements of the particles which correspond to stable low-energy states of the system.

After minimization, before starting a production simulation, the system is allowed to undergo a period of equilibration. That is a period after which all memory of the initial configuration is lost and after which thermodynamic properties have ceased to show a systematic drift and have started to oscillate about steady mean values.

2.6 Application of Constraints in Simulations

The MD procedure described so far samples the *NVE* ensemble, where the number of particles N, the volume V and the energy E of the system are constant and the temperature T and the pressure P are allowed to fluctuate. Therefore the Hamiltonian in equation 2.12 can be referred to as $\mathscr{H}_{NVE}$. This is not always a useful ensemble; it would be nice to study systems of interest under conditions of constant T and P, that represent the situation found in an experimental laboratory. Algorithms are available to constrain these degrees of freedom and *NVT* and *NPT* ensembles can be simulated.

2.6.1 Constant Temperature

Temperature is directly related to the kinetic energy of the system:[11,12]

$$\mathscr{K} = \sum_{i=1}^{N_{particles}} \frac{\boldsymbol{p}_i^2}{2m_i} = \frac{k_B T}{2}(3N - N_c) \tag{2.30}$$

Here $\mathscr{K}$ is the kinetic energy, $\boldsymbol{p}_i$ the momentum of particle i, m_i its mass, k_B the Boltzmann constant, T the current kinetic temperature, N the number of particles simulated in the system and N_c the number of constraints. The velocities are generated during the simulation and hence the momenta and the temperature can be calculated.

One of the most rigorous algorithms to constraint the temperature is the Nosé-Hoover thermostat.[15–17] A thermal reservoir is considered to be part of the system and it is represented as an additional degree of freedom s. The Hamiltonian $\mathscr{H}$ of the extended system is equal to the sum of the kinetic $\mathscr{K}$ and potential $\mathscr{V}$ energies of the particles with the kinetic $\mathscr{K}_s$ and potential $\mathscr{V}_s$ energies of the reservoir:

$$\mathscr{H}_{NVT} = \mathscr{K} + \mathscr{V} + \mathscr{K}_s + \mathscr{V}_s = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i'^2}{2m_i s^2} + \mathscr{V}\left(\boldsymbol{r}'^N\right) + \frac{p_s'^2}{2M} + (n+1)k_B T_0 \ln s \qquad (2.31)$$

Here T_0 is the desired temperature, n is the number of degrees of freedom, p_s is the momentum conjugate to s, and M is the thermal inertia parameter, which can be considered the (fictitious) mass of the extra degree of freedom (it has units of energy time^2). The prime indicates "extended-system" variables or virtual variables, which are related to the real variable of the physical system by:

$$\text{coordinates: } \boldsymbol{r}' = \boldsymbol{r} \quad , \quad \text{momenta: } \boldsymbol{p}' = \boldsymbol{p}/s \quad , \quad \text{time: } t' = \int_0^t dt/s \qquad (2.32)$$

The equation of motion for the real variables then modifies to:

$$\frac{d^2\boldsymbol{r}_i}{dt^2} = -\frac{1}{m_i}\frac{d\mathscr{V}(\boldsymbol{r}^N)}{d\boldsymbol{r}} - \frac{sp_s}{M}\frac{d\boldsymbol{r}_i}{dt} \qquad (2.33)$$

The factor sp_s/M, also indicated by ξ, acts as a thermodynamic friction coefficient and is controlled by:

$$\frac{d(sp_s/M)}{dt} = \frac{1}{M}\left(\sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{m_i} - nk_B T_0\right) \qquad (2.34)$$

The choice of M is arbitrary but is critical to the success of a run. Too high a value results in slow energy flow between the system and the reservoir and the thermalization process is not efficient. On the other hand, if M is too low, strong energy oscillations occur, resulting in equilibration problems.

2.6.2 Constant Pressure

In simulations the pressure is related to the virial W.[11] The total virial function has an internal contribution, related to the forces acting between the simulated particles, and an external contribution, related to the externally applied forces:

$$W_{internal} = \frac{1}{3}\langle \sum_{i=1}^{N} \boldsymbol{x}_i \cdot \boldsymbol{F}_{internal} \rangle = -\frac{1}{3}\langle \sum_{i=1}^{N} \boldsymbol{x}_i \cdot \nabla \mathscr{V}_i \rangle \tag{2.35}$$

$$W_{external} = \frac{1}{3}\langle \sum_{i=1}^{N} \boldsymbol{x}_i \cdot \boldsymbol{F}_{external} \rangle = -PV \tag{2.36}$$

$$W_{total} = \frac{1}{3}\langle \sum_{i=1}^{N} \boldsymbol{x}_i \cdot \boldsymbol{F}_{total} \rangle = -Nk_bT \tag{2.37}$$

Here $\boldsymbol{x}_i$ is the position of the particle i, $\boldsymbol{F}_{internal}$ is the force acting on that particle according to the force-field and due to the interactions with the other particles of the system, $\boldsymbol{F}_{external}$ is an externally applied force acting on i, which is related to the external pressure, and $\boldsymbol{F}_{total}$ is the sum of intermolecular and external forces. The pressure P is consequently easily obtained in the following way:

$$PV = Nk_bT + \frac{1}{3}\langle \sum_{i=1}^{N} \boldsymbol{x}_i \cdot \boldsymbol{F}_{internal} \rangle \tag{2.38}$$

One of the most rigorous algorithms ensuring the simulation pressure is constant is Andersen's extended system method.[18] This method is based on including an extra degree of freedom mimicking the action of a piston and corresponding to the volume V of the simulation box which adjusts itself to equalize the internal and the applied pressures, respectively P_{int} and P_{ext}. The Hamiltonian of the extended system is:

$$\mathscr{H}_{NPH} = \mathscr{K} + \mathscr{V} + \mathscr{K}_\eta + \mathscr{V}_\eta = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i'^2}{2m_i V^{2/3}} + \mathscr{V}\left(V^{1/3}\boldsymbol{r}'^N\right) \frac{p_\eta^2}{2M} + P_0 V \tag{2.39}$$

Here M is the fictitious mass of the piston (with units mass length^{-4}), and p_η the momentum associated with V. The extended system variables are indicated with primes and relate to the real system variables by:

$$\text{coordinates: } \boldsymbol{r}' = \boldsymbol{r}/V^{1/3} \quad , \quad \text{momenta: } \boldsymbol{p}' = V^{1/3}\boldsymbol{p} \tag{2.40}$$

The equation of motion is modified in the following way:[18,19]

$$\frac{d^2\boldsymbol{r}(t)}{dt^2} = -\frac{1}{m_i}\frac{d\mathscr{V}(\boldsymbol{r}^N)}{d\boldsymbol{r}} - \frac{1}{3}\frac{dV(t)}{dt}\frac{1}{V(t)}\frac{d\boldsymbol{r}(t)}{dt} \tag{2.41}$$

Here the factor $\frac{1}{3}\frac{dV(t)}{dt}\frac{1}{V(t)}$ is the barostat friction coefficient. It is controlled by:[18,19]

$$\frac{d^2V(t)}{dt^2} = \frac{\boldsymbol{P}_{int}(t) - \boldsymbol{P}_{ext}}{M} \tag{2.42}$$

Here $\boldsymbol{P}_{int}(t)$ is the instantaneous internal pressure (calculated with equation 2.38) and $\boldsymbol{P}_{ext}$ the applied (desired) external pressure.

Nosé and Klein extended this method to the case of non-cubic simulation cells:[15]

$$\mathscr{H}_{NPH} = \mathscr{K} + \mathscr{V} + \mathscr{K}_\eta + \mathscr{V}_\eta = \sum_{i=1}^{N}\frac{\boldsymbol{p}'^2_i}{2m_i\boldsymbol{h}^t\boldsymbol{h}} + \mathscr{V}\left(\boldsymbol{h}^{-1}\boldsymbol{r}'^N\right)\frac{p_\eta^2}{2M} + P_0V \tag{2.43}$$

Here $\boldsymbol{h}$ is a matrix with three columns representing the sizes of the three cell sides, $\boldsymbol{h}^t$ is its transpose, and:

$$\text{coordinates: } \boldsymbol{r}' = \boldsymbol{r}/\boldsymbol{h} \quad , \quad \text{momenta: } \boldsymbol{p}' = \boldsymbol{h}\boldsymbol{p} \tag{2.44}$$

The equation of motion for the real system variables becomes:

$$\frac{d^2\boldsymbol{r}(t)}{dt^2} = -\frac{1}{m_i}\frac{d\mathscr{V}(\boldsymbol{r}^N)}{d\boldsymbol{r}} - \frac{d\boldsymbol{h}(t)}{dt}\frac{1}{\boldsymbol{h}(t)}\frac{d\boldsymbol{r}(t)}{dt} \tag{2.45}$$

The factor $\frac{d\boldsymbol{h}(t)}{dt}\frac{1}{\boldsymbol{h}(t)}$ is controlled by:

$$\frac{d^2\boldsymbol{h}(t)}{dt^2} = \frac{\boldsymbol{P}_{int}(t) - \boldsymbol{P}_{ext}}{MV\boldsymbol{h}^t} \tag{2.46}$$

In this case M has the unit of a true mass.

Andersen's method has been shown to produce strong oscillations in the volume of the system. It has been improved by Feller,[19] who introduced the Langevin piston method. In this case:

$$\frac{d^2V(t)}{dt^2} = \frac{\boldsymbol{P}_{int}(t) - \boldsymbol{P}_{ext}}{M} - \gamma\frac{dV(t)}{dt} + R(t) \tag{2.47}$$

where $R(t)$ is a random number taken from a Gaussian distribution and γ is the collision frequency. This algorithm was also implemented in the Nosé-Klein version for non-cubic cells and has been proved to minimize the effect that the choice of piston mass has on the dynamics.

2.7 Special Techniques

It has been already mentioned that the size of the time step depends on the fastest motion of the system, and in the case of atomic simulations this motion corresponds to the bond stretching.[12] Using a larger time step would speed up the calculation, since with the same number of steps a greater system evolution in time would be sampled. A solution is to freeze out some vibrations, for instance bonds, by constraining them to their equilibrium value, while the rest of the degrees of freedom are allowed to vary. The normal forces $\boldsymbol{f}_i$ from inter- and intra-molecular interactions defined in the force-field act on all the particles of the system, and some constraining forces $\boldsymbol{g}_i$ act on the selected constrained degrees of freedom only. Consequently, the equation of motion for the constrained degrees of freedom, for instance bond lengths, can be rewritten as: $m_i\boldsymbol{a}_i = \boldsymbol{f}_i + \boldsymbol{g}_i$. Within the SHAKE[20] algorithm, the technique is to solve the equation of motion for one time step in the absence of the constraint forces and subsequently determine their magnitude by correcting the atomic positions to ensure that the difference between the square of the new atomic distance and the square of the old atomic distance is zero at all times, within a chosen tolerance, for all the bond lengths selected to be constrained. Each constraint is considered in turn and solved. This may violate another constraint, and so it is necessary to iterate until all the constraints are satisfied.

To reduce the number of particles in the simulation system, periodic boundary conditions are employed; they are also used to overcome the problem of surface effects. The simulation box is replicated throughout space in all directions in an infinite lattice and consequently is completely surrounded by its image boxes. There are no empty spaces between the simulation central box and its images, where the first terminates the image starts, and for this reason the particles experience forces as if they were in bulk fluid. If a particle leaves the box, it is replaced by its image particle entering from the opposite side coming from an image box, so that the number of particles in the system is always constant.

The most time consuming part of a MD simulation is the calculation of the non-bonded energies and forces (electrostatic and van der Waals). In principle the non-bonded interactions should be computed between every pair of particles in the system including the images, but this would be too time consuming, inefficient and arguably unnecessary. For this reason the calculation can be approximated using a cutoff. It is assumed that the largest contribution to the potential and forces comes from the near neighbours and a truncation can be made: all the interactions between pairs of particles further than a spherical cutoff value are set to

zero. The choice of a correct cutoff radius is important. For van der Waals interactions, the potential becomes insignificant after a distance of about 2.5 σ in the Lennard-Jones potential. This means van der Waals interactions are short range interactions and in the case of atomistic simulations they can be truncated after, say, 10Å. In contrast, electrostatic interactions are long range interactions and their effects are felt at a considerably greater distance than the cutoff commonly used for simulations. To enforce the cutoff method without loosing accuracy, several techniques are available. First, two different cutoff radii can be used, a shorter one for short range van der Waals interactions and a longer one for long range electrostatic interactions.[12] Otherwise, more complicated methods such as the Ewald sum can be employed[21] for calculating the full electrostatic energy of a unit cell. According to the Ewald sum method, each point charge present in the system is surrounded by a charge distribution of equal magnitude and opposite sign; this distribution is conveniently taken to be Gaussian. This extra distribution acts like an ionic atmosphere to screen the interactions between neighbouring charges. The screened interactions are now short-ranged and the total screened potential is calculated by summing over all the particles in the central box and all the images in real space. Then, another charge distribution of opposite sign to the first added Gaussian distribution is also added and this cancelling distribution is summed in reciprocal space. A fast implementation of the Ewald summation is the Particle Mesh Ewald (PME) method.[22] In this case the reciprocal space term of the Ewald sum is approximated using Fast Fourier Transformations with convolutions on a grid where charges are interpolated to the grid points. This procedure greatly reduces the computational cost of the reciprocal space sum.

To reduce the time for the computation, a particle's neighbours list is created.[23] For a given particle, it contains a list of the particles within a distance slightly larger than the cutoff. To speed up the simulation the neighbour list is updated at given intervals and not at each simulation step. Between updates the program does not check through all the particles, but it calculates the distance between the particle of interest and only those particles appearing in the list, and consequently the time for the non-bonded calculation is significantly reduced.

Another problem of the cutoff method is the discontinuity in potential energy at the cutoff distance: since at this distance the potential suddenly drops from its real value to zero, at the cutoff distance interaction energies and forces fluctuate violently and this causes problems of energy and temperature conservation. A common solution consists in switching the potential between two cutoff distances:[12] the potential has its true value until the lower cutoff and then

it is multiplied by a switching function, whose values range between 1 at distance equal to the lower cutoff and 0 at distance equal to the upper cutoff. Force and potential are so smoothly reduced to zero.

2.8 Interfaces

When simulating liquid/liquid interfaces, only the component of the pressure tensor normal to the interface P_N gives a measure of the bulk pressure of the system, while the lateral components depend on the interfacial tension. Thus, in contrast to homogeneous fluid where only one pair of thermodynamic variables (pressure P and volume V) determine the size of the simulation cell, two sets of variables determine the size and shape of the interfacial system: P_N and V, together with the surface tension γ and the surface area A.[24,25] By extension of the Andersen constant pressure method and with the inclusion of a constant temperature algorithm, two ensembles can be generated: 1) constant normal pressure and surface area (NP_NAT), and 2) constant normal pressure and surface tension ($NP_N\gamma T$).

Let us fix the z axis of the simulation box along the normal to the interface, and the x and y axis define the interfacial plane. In this case, the box length along z is L_z and the normal pressure is the pressure tensor along z, i.e. $P_N = P_{zz}$. At each z position along the interface normal, the lateral pressure is the average between the pressure tensors along x and y, and the average lateral pressure along the z axis is:[24]

$$\overline{P_L} = \frac{1}{L_z}\int_0^{L_z} P_L(z)dz = \frac{1}{2L_z}\int_0^{L_z} [P_{xx}(z) + P_{yy}(z)]dz \tag{2.48}$$

The value of the surface tension, which is the observed γ in experiments, is obtained as:[24]

$$\gamma = \int_{-\infty}^{+\infty} [P_N - P_L(z)]dz = L_z(P_N - \overline{P_L}) \tag{2.49}$$

Only the interface contributes to the integral, since in bulk phase $P_N = P_L(z)$.

2.8.1 Constant Normal Pressure and Surface Area

In the constant normal pressure and constant surface area ensemble, L_x and L_y are fixed to maintain the constant surface area A=L_xL_y, while L_z fluctuates to adjust the internal normal pressure of the system P_N to the applied P_N^0. The extended system Hamiltonian for modelling in

this ensemble is in practice a one dimensional Andersen's pressure model.[24] Only the equation of motion along the z axis is modified:

$$\frac{d^2\boldsymbol{z}(t)}{dt^2} = \frac{\boldsymbol{f}_z(t)}{m_i} - \frac{dL_z(t)}{dt}\frac{1}{L_z(t)}\frac{d\boldsymbol{z}(t)}{dt} \tag{2.50}$$

with $L_z(t)$ controlled by:

$$\frac{d^2L_z(t)}{dt^2} = L_xL_y\frac{P_N(t) - P_N^0}{M} \tag{2.51}$$

2.8.2 Constant Normal Pressure and Surface Tension

In the constant normal pressure and constant surface tension ensemble L_x, L_y and L_z can all vary. In what follows, P_N^0 is the reference normal pressure, and γ^0 the reference surface tension. The equation of motion modifies to:[24]

$$\begin{aligned}
\frac{d^2\boldsymbol{x}(t)}{dt^2} &= \frac{\boldsymbol{f}_x(t)}{m_i} - \frac{dL_x(t)}{dt}\frac{1}{L_x(t)}\frac{d\boldsymbol{x}(t)}{dt} \\
\frac{d^2\boldsymbol{y}(t)}{dt^2} &= \frac{\boldsymbol{f}_y(t)}{m_i} - \frac{dL_y(t)}{dt}\frac{1}{L_y(t)}\frac{d\boldsymbol{y}(t)}{dt} \\
\frac{d^2\boldsymbol{z}(t)}{dt^2} &= \frac{\boldsymbol{f}_z(t)}{m_i} - \frac{dL_z(t)}{dt}\frac{1}{L_z(t)}\frac{d\boldsymbol{z}(t)}{dt}
\end{aligned} \tag{2.52}$$

Here:

$$\begin{aligned}
\frac{d^2L_x(t)}{dt^2} &= L_y\frac{\gamma^0 - L_z(P_N^0 - P_{xx}(t))}{M_x} \\
\frac{d^2L_y(t)}{dt^2} &= L_x\frac{\gamma^0 - L_z(P_N^0 - P_{yy}(t))}{M_y} \\
\frac{d^2L_z(t)}{dt^2} &= L_xL_y\frac{P_N(t) - P_N^0}{M_z}
\end{aligned} \tag{2.53}$$

CHAPTER 3

Biological Membranes

3.1 Cell Membranes

Biological membranes are complex molecular assemblies ensuring the cell integrity, acting as a selective barrier between the cell and the external environment. They also are the site of important metabolic processes.

The current interpretation of the membrane structure is based on the fluid mosaic model.[26,27] Membranes are made up of a bilayer of amphipathic lipid molecules whose hydrophobic region lies in the interior and whose hydrophilic region lies on the exterior. Under physiological conditions lipids are free to move, rotate and change their position as in the fluid phase subject to the constraint that bilayer integrity is maintained. Together with lipids, several types of proteins are located in the bilayer surface or interior, some of them simply floating on the membrane (peripheral proteins) and some other spanning it completely from one side to the other (integral proteins). In a given biological membrane, different regions can be found whose composition is richer in lipids than proteins, or vice versa, and for this reason the name mosaic is employed. Membrane proteins have various functions: some act as channels that allow specific molecules to cross the membrane; some act as markers which identify the cell as part of the same individual; some act as surface receptors which recognize and bind molecules produced by other cells; some are enzymes with important functions such as cell respiration and photosynthesis.[27]

Proteins can diffuse inside the lipid bilayer and in this way they can merge to form greater associations with specific functions. E.g., gramicidin is an integral protein that dimerises to form cation-selective channels.[28] Most of the membranes contain about 50% in weight of lipids and 50% in weight of proteins, but there are certain membranes with specific functions that contain a different lipid/protein ratio. For example the nervous system membranes, which are involved in electrical transmission, contain 80% lipids, while the mitocondria membranes, where the metabolic pathway involved in cell respiration is located, contain 80% proteins.

3.1.1 Phospholopid Molecules

The membrane main lipids are called phosphoglycerides.[27] They are built from a frame of glycerol. Two chains of fatty acids esterify two vicinal hydroxyl groups of the glycerol and they constitute the hydrophobic part of the lipid molecules, while a phosphate group esterifies the third hydroxyl on the glycerol, and also either a choline or ethanolamine or other polar groups. The fatty acids are called lipid tails and the polar groups bonded to the phosphate, together with the phosphate and the glycerol, are the lipid headgroup.

Among the most common phospholipids present in biological and model membranes is 1,2-dipalmitoyl-sn-glycerol-3-phosphatidylcholine (DPPC) represented in figure 3.1. It is a zwitterionic molecule with two hydrocarbon chains each containing 16 carbon atoms. In a pure DPPC bilayer the liquid-crystalline state is adopted at temperatures higher than 42 °C.[29,30] The hydrocarbon chain bonded with an ester linkage to the chiral carbon of the glycerol is given the name Sn2, while the other chain is given the name Sn1.

3.1.2 Lipid Bilayers

In water the phospholipids spontaneously form stable micelles and bilayers: the polar groups are arranged towards the water phase, while the fatty acids avoid contact with the water molecules and are aligned towards the interior of the bilayer. *In vivo*, the intra- and extracellular environments of a cell are aqueous, so the formation of the phospholipid bilayer creates a semipermeable barrier. Lipids can easily move laterally inside the layer, but rarely they also can change layer; this motion is called flip-flop.

The driving force for phospholipid self assembly is neither cohesion nor the formation of covalent bonds between the alkyl chains, but rather a hydrophobic effect,[30] which arises for entropic reasons. When a free lipid molecule is in water phase, only the headgroup is able to

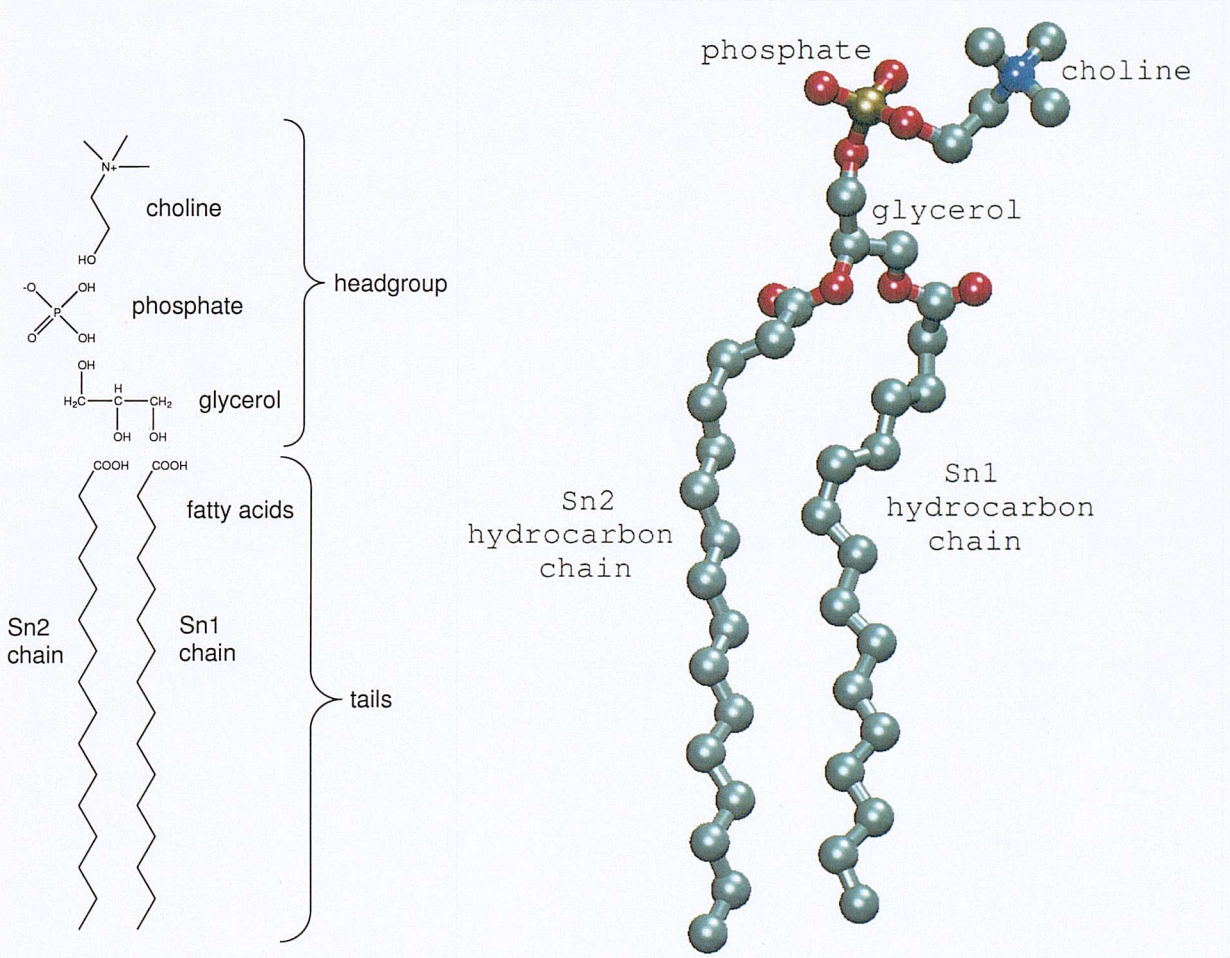


Figure 3.1: DPPC molecule (hydrogens are omitted). Blue atom is nitrogen, red atoms are oxygens, orange atom is phosphorus, cyan atoms are carbons.

form favourable interactions with the surrounding water molecules, namely hydrogen bonds and favourable electrostatic interactions. In contrast, the hydrogen bonds between water molecules have to be distorted when a hydrocarbon chain is dissolved in water. The formation of an ordered structure around the hydrophobic molecule to preserve hydrogen bonding causes a loss of configurational entropy. Enthalpic contributions from van der Waals interactions are comparatively small. Opposing lipid self association is the entropy of mixing of the lipid monomer with the solvent,[30] since there is an entropic advantage when the lipid monomers are free to mix with the solvent, whereas there are motional constraints and repulsive forces when the lipid molecules are packed in a bilayer structure.

The bilayer can undergo a phase transition according to the temperature and the level of hydration.[30] At high temperatures lipids are in a liquid-crystalline state, i.e. lipids can move, rotate around their longitudinal axis and diffuse in the bilayer, their chains are not ordered and much shorter than their maximum extension. This is the state in which cell membranes are functional and is called the L_α phase. At lower temperature membranes are in the gel state, i.e. the lipid molecules are tightly packed and well ordered, the volume per lipid molecule is lower, lipid movements are strictly limited, and the hydrocarbon chains are close to their maximum extension (all-trans). The phase transition temperature depends on the lipid composition:[30,31] usually the longer the hydrocarbon chains and the smaller the number of double bonds in these chains, the higher the phase transition temperature. The nature of the headgroup plays a minor role. The transition temperature decreases with increasing water content, reaching a limiting value when the water-binding capacity of the lipid has reached saturation.

X-ray diffraction is probably the most definitive method of determining bilayer structure.[30] X-ray scattering permits direct determination of the electron density profile along the bilayer normal:[30,32,33] the densest region is associated with the phosphate groups and the region with the lowest density is associated with the hydrocarbon tails. Therefore, the peak-to-peak distance in the electron density profile also gives a measure of the bilayer thickness.

Neutron diffraction experiments on selectively deuterated lipids yield the mean position of those deuterated segments along the bilayer normal.[30,32,34–36] For fully hydrated fluid phase lipid bilayers, it makes no sense to consider an atomic level structure because of the associated fluctuations, and the appropriate description for the positions of atoms in a lipid molecule is that of broad statistical distribution functions along the bilayer normal, which do not have to be Gaussian. Gaussians are however a convenient tool for use in analysis.[37] From the

mean positions of different segments, the mean conformation of the lipid chains can also be extrapolated: the chains are predominantly in the all-trans configuration, but gauche isomers are also present, and the lipid headgroup is nearly parallel to the bilayer plane.

^{2}H-NMR of specifically deuterated lipids allows for the calculation of order parameters,[4–6,30,38–40] trans/gauche populations[41] and trans/gauche isomerization correlation times.[41–43] Deuterium has spin = 1, which means that it has a quadrupole moment. In the case of anisotropic motions, the signal in the ^{2}H-NMR spectrum is a doublet. The difference $\Delta\nu$ between the two lines is related to the order parameter $S_{CD,n}$ of the carbon-deuterium bond at position n along the lipid tail and to the angle α between the bilayer normal and the magnetic field:[44]

$$\Delta\nu(\alpha) = \frac{3}{4}\frac{e^2qQ}{h}(3cos^2\alpha - 1)S_{CD,n} \tag{3.1}$$

The factor $\frac{e^2qQ}{h}$ is the ^{2}H quadrupole splitting constant and is known for paraffins. Therefore, the value of $S_{CD,n}$ can be obtained at $\alpha = 90$ °. It is related to the angle θ_n between the C-D bond at position n and the bilayer normal:

$$S_{CD,n} = \frac{1}{2}(3cos^2\theta_n - 1) \tag{3.2}$$

Experimentally, only the absolute value of $\Delta\nu$ is accessible and the sign of $S_{CD,n}$ can be deduced from geometrical considerations. In lipid bilayers the long molecular axis of a lipid molecule is essentially parallel to the bilayer normal and therefore the C-D bond is oriented more or less perpendicular to the bilayer normal. Such an orientation requires negative $S_{CD,n}$ values.

^{2}H- and ^{31}P-NMR measurements are used to study the orientation of selectively deuterated phosphorylcholine groups,[45,46] because both the ^{2}H quadrupole splitting and ^{31}P chemical shift depend on dihedral conformation and flexibility. Headgroups exhibit restricted flexibility characterized by rapid transitions between two conformations. The phosphorylcholine residue is found to be bent at the position of the phosphate group, so that the choline dipole is aligned parallel to the plane of the membrane. The ^{2}H-NMR technique is also employed to study lipid bilayer hydration.[47] The deuterium quadrupolar relaxation is a purely intramolecular mechanism, with intermolecular interactions having no effect. Therefore the spin lattice relaxation time T_1 can be interpreted in terms of an effective molecular correlation time. Studying the ^2H_2O T_1 as a function of water content in lipid bilayers shows that at low water content T_1 increases linearly with increasing number of water molecules per lipid from about 2 waters/lipid

up to 11-16 waters/lipid. At higher water content T_1 still increases linearly with the number of waters/lipid, but with a much smaller slope. The interpretation is that of a first hydration shell made of up to 11-16 water molecules bound to the lipid headgroups, with additional water filling the space between lipid lamellae or existing as bulk water.

From Raman spectra, the molecular structure (% of trans/gauche isomers) can be established, since conformational assignments of the Raman bands of the bond stretching vibrations can be established.[48]

The C=O stretching vibrational band of the carbonyl group in the ester linkage of the phospholipid molecule can be investigated by IR spectroscopy.[49] It has been seen that this group is hydrated, since its IR bands in water and organic solvents are different, proving that water enters the lipid bilayer to hydrate the carbonyl group. IR spectra of specifically deuterated lipids can yield the trans/gauche population at specific carbon positions along the hydrocarbon tails,[50,51] since distinct IR bands exist for the CD_2 group depending on whether the immediately adjacent C-C bond is in a trans or gauche conformation. IR spectra also contain bands due to CH_2 wagging of gauche conformers[52] and the mean number of gauche conformations per chain can therefore be obtained.

Fluorescence techniques can be used to measure the lateral diffusion coefficient of fluorescent lipid analogues in bilayers.[53,54] These methods are based on measurements of dynamic changes of fluorescence intensity due to changes in the number of fluorophores within a small illuminated area. The fluorescence of the lipid analogue is excited by a laser focused on the bilayer, then the excited fluorescence is collected and measured by photon counting. Such experiments sample the lipid translational diffusion involving exchange of lipids, which occurs on a time scale too long to be accessible in atomistic computer simulations. Neutron scattering experiments[55] can instead sample motions in the same time scale as computer simulations, that is they yield information about the "rattling in a cage" motions of lipid molecules and, to a lesser extent, about the rotation of the lipid as a whole molecule.

One of the greatest challenge in the study of lipid bilayer structure is the determination of the surface area per lipid, A. Different authors employing different techniques generally report different values and a wide range of results are found in the literature. A is mostly obtained indirectly by volumetric and X-ray measurements from the knowledge of the bilayer thickness and the volume of lipid and water molecules.[32,37,56] A also depends on the degree of hydration:[30] at low water concentrations A increases with increasing hydration, stabilising

at high water concentrations. The presence of headgroup-bound water increases the lateral pressure opposing the lipid assembly and consequently it increases the bilayer surface area up to a constant value when all lipid hydration sites are saturated (fully hydrated state). The surface area is related to the alkyl chain length and degree of unsaturation: the higher the number of double bonds and the higher the number of methylene groups along the lipid tails, the higher the surface area per lipid. A also increases when increasing the temperature. Usually, PhosphatidylCholines (PC) hydrate more than PhosphatidylEthanolamines (PE) and have a larger surface area per lipid.[57]

3.2 Previous Simulations of Biomembranes

Owing to the limited computer power, the early simulations of lipid assemblies employed relatively simple models. One of the most important is the mean field approach of Marcelja.[58] The energy of a lipid chain is given by the internal energy, calculated using standard force field methods described in chapter 2, and by the interaction energy with its neighbours. The latter has two contributions. One is the van der Waals interaction with the neighbouring molecules and depends on the angle formed by the molecular axis with the bilayer normal, the fraction of trans conformations, and the field strength which can be parameterized to reproduce experimental data like the order parameter. The second contribution is repulsive, it is due to the lateral pressure, which depends on the cross-sectional area of the chain, and to the distance of the terminal carbon from the bilayer surface. This model allows for the calculation of several thermodynamics properties and good qualitative agreement with experiments is obtained.

Karplus *et al.*[59] described the internal dynamics of DPPC with a Brownian dynamics simulation of a single molecule. The non-bonded potential energy was modelled with a mean field approach and the collision frequency for the random brownian force was adjusted to reproduce experimental NMR spin-lattice relaxation times. Non-collective reorientational motions (axial rotation and wobble) were also taken into account. A better agreement with experimental NMR and neutron diffraction data was then obtained, suggesting a high degree of chain disorder in biological membranes.

With increasing computer power, atomistic models are now used to simulate fully hydrated lipid bilayers of various size. The application of Molecular Dynamics (MD) methodology to the simulation of biological membranes was first performed in 1988 by Berendsen and Egberts.[60] The latter presented in his PhD thesis the simulation of a realistic phospholipid bilayer using

a hydrated DPPC membrane. Since then the studies have been extended to include small molecule diffusion and partition within the membrane environment[a] and even the simulation of membrane bound proteins[28,61–64] and lipid-DNA complexes.[65]

The most common lipid molecules used in MD simulations are phosphocholines (PC), which have a zero net charge, but also phosphatidylserines,[66–68] which posses a net negative charge, have been studied. In the latter case, the strong interactions between the lipid headgroups and the positive sodium counter-ions yield a reduction of the surface area; a reduction in tail order and water penetration are also revealed. Regarding the lipid tails, the most studied are dimyristoyl and dipalmitoyl residues, but recently branched[69] and unsaturated[70–76] chains have been simulated. A strong interaction is noted between the double bonds in the unsaturated chains and water. Polyunsaturated chains also reveal an unexpected high degree of flexibility. The gel phase of lipid bilayers has been studied,[77–79] but simulations are mostly dedicated to the biologically relevant liquid-crystalline phase.

Among the most common force fields developed for lipid simulations are the united-atom OPLS[73,80–85] and GROMOS/GROMACS[70,75,76,86–90] force fields, and the all-atom CHARMM[25,71,74,91–94] force field. They are usually parameterized by means of simulations of n-alkanes or n-alkenes and lecithin fragments. Some parameters were originally taken from similar functional groups in proteins and then refined (mainly charges and van der Waals parameters) to reproduce lipid bilayer experimental observations.

Lipid bilayers are usually simulated using Periodic Boundary Conditions (PBC), that is the simulation box is replicated in three dimensions to reproduce an infinite bilayer in a multilamellar geometry, which is common in experiments. The first ensembles employed were constant NVT[80,81,95] and NPT.[73,74,76,82,83,85,86,89,96] The great advantage of a fully flexible simulation cell (NPT) is that only an approximation of the initial size is needed, because the system will find its equilibrium density by itself, based on the force field. In the case of constant volume an accurate knowledge of properties such as repeat distance and area per headgroup is required, otherwise the system may be forced into dimensions it would not otherwise adopt. Unfortunately, an unreasonably wide range of experimental values are reported regarding, for example, the surface area per lipid and this makes difficult the choice of the simulation box size. More recently, NP_NAT and $NP_N\gamma T$ ensembles have been applied. Berendsen *et al.*[88] reported no significant differences in the bilayer physical behaviour when an isotropic pressure is applied to

[a]These simulations will be described in detail in chapter 5.

the simulation or when different values for the lateral and the perpendicular tensors are used to reproduce a non-zero surface tension. Feller *et al.*,[25,91,92,97] with a longer simulation and different force field and program, reached a different conclusion. A real biological flaccid vesicle membrane can freely undulate and consequently the projected area of a large patch fluctuates with no energy cost while the surface area per lipid is fixed. The size of this patch ($\sim$ 1 μm/side) is many orders of magnitude larger than is presently feasible in MD simulations ($\sim$ 50 Å/side) because of computer limitations. The use of PBC does not allow the simulation of fluctuations with wavelengths larger than the simulation box dimension. Although on a large scale it is true that the surface tension of a bilayer patch is zero, this being the main argument for those simulating lipid bilayers without any surface tension,[89] this is not necessarily true on the scale of a simulation system. It may be more appropriate to carry out simulations with a constant surface area or a non-zero surface tension to correct for the finite-size effect.

The simulation length and the number of lipids in the simulation box depend on computer limitations. The most frequent bilayer simulations contain 64[74,83,85,86,88,89,96] or 72[25,71–73,82,91,92,94] lipid molecules, but larger systems have been published.[70,75,76,84,87,90] The production run of the first atomistic membrane simulations were a few hundreds of ps long,[80,81,86–88,91] but now several ns are accessible.[72,76,84,85,93,94] This time scale allows for studying small conformational changes in the lipid molecules, such as dihedral trans/gauche interconversions, but the motions of the whole lipid is too slow to sample. Trans/gauche isomerizations along the lipid tails occur on the time scale of tens of ps closer to the tail terminal methyls and a few hundreds of ps towards the headgroup. A few ns are required for a lipid molecule to rotate around its long axis and tens of ns are needed for lateral diffusion or two lipids switching place within one bilayer leaflet. This is not necessarily a problem for understanding lipid bilayer physical properties, but the choice of the starting structure and the equilibration procedure are very important.

Starting structures may be generated from the lipid crystal structure and then the bilayer is "melted" to the fluid phase with a series of short high-temperature MD simulations.[80,82,83,86,88,89,96] Molecule arrangement and lipid internal conformation can be modified to make the passage to the fluid phase as trivial as possible. Water mass can be increased to allow for a longer time step in the MD algorithm. Alternatively, a set of lipids with conformations calculated from mean-field theory can be generated, so that they can be packed together at the desired density to form the bilayer.[25,71,91] A new smart way to achieve equilibration involves the use of a combined Molecular Dynamics - Monte Carlo method:[70,90,98] after a certain number of MD

steps are performed, one lipid molecule is picked at random and one of its alkyl chains or its headgroup, chosen at random, is regrown bond by bond to obtain a new conformation. This method allows the exploration of new regions in conformational space, and speeds up the time required to relax the system from its starting structure.

Electrostatic interactions play an important role and require an accurate treatment, particularly for the charged or zwitterionic molecules such as phospholipids.[99] Using a cutoff radius can introduce artifacts affecting a variety of structural and dynamical properties in simulations of biomolecules.[99] Since the early simulations, residue-based,[80–82,87,89] large[70,76,84,90] or cylindrical[75,86,88] cutoff have been used. Now, more computationally expensive techniques can be used, such as full Ewald summation[83,91,93,94,96] or Particle Mesh Ewald (PME).[25,71–74,85,92,100] Berendsen *et al.* find no difference in bilayer behaviour when using Ewald sum or truncation methods with large cutoff and they prefer the latter, which allows faster simulation runs.[88] Feller *et al.* find much better agreement with experiments when using the Ewald or PME methods, above all for lipid and water structural order.[25]

The development of lipid bilayer models has been astonishing over the last decade and still continues today. With the growth in computing power, the size and the length of the simulations has increased and the proper treatment of electrostatic interactions has become possible. From a technical point of view, improvements can still be made. Force fields are limited by not taking into account explicit polarizability and they are sometimes not consistent with parameters employed for proteins or diffusing solutes. New algorithms for treating long range interactions in a faster way, allowing longer time steps, or sampling long time-scale motions in a smarter way are a matter of research. Simulations now offer a detailed picture of the structure and the dynamics of lipid membranes and provide an atomistic detail that is unavailable from experiment.

CHAPTER 4

Computer Simulation Study of A Biological Membrane

In this chapter, the physical properties of a model biological membrane are studied by means of Molecular Dynamics (MD) computer simulations. The aim is to validate the computer model of such biological membrane prior to performing the permeability studies.

4.1 Simulation Protocol

The most interesting simulation protocol in the literature was thought to be that adopted by Feller *et al.*[91,92,97,101,102] and it was decided to use it for these studies. Those authors participated in creating and developing a force field specifically designed for lipid bilayer simulations, which they tested in different situations. The temperature and pressure coupling and the long range interactions are treated properly, and all the atoms are explicitly represented. The software package employed here correctly implements the force field, and also contains many functionalities specifically introduced for lipid bilayer simulations, for example the choice of NP_NAT and $NP_N\gamma T$ ensembles.

What follows is the description of the protocol adopted in this thesis. An all-atom representation (i.e. with explicit hydrogens) was employed, as previous literature suggests that an

all-atom representation may be necessary for the accurate estimation of the diffusion coefficients of permeating solutes.[103,104] The water model employed was TIPS3P,[105] which is a modified version of the original TIP3P model[106] and is included in the CHARMM force-field.[13] Version number 27 of the CHARMM force-field[100] and version number 27 of the CHARMM software package[13] were employed to run the simulations.

The starting structure was that of the last step of a recent Feller's simulation,[91] so the bilayer had already been minimized and equilibrated and its thermodynamic properties have been shown to reproduce experimental membrane behaviour with excellent agreement. The system was simulated under NP_NAT conditions, i.e. constant number of particles N, constant normal pressure $P_N = 1$ atm, constant surface area per lipid $A = 62.9$ Å^2, and constant temperature $T = 50$ °C (well above the DPPC phase transition temperature, which is around 42 °C[29]).

The membrane model included 72 DPPC molecules arranged in a square 36 × 2 bilayer. The simulation cell x, y and z starting dimensions were 47.5857 Å, 47.5857 Å and 66.8901 Å, respectively. The x-y plane was the bilayer plane, which was kept fixed to maintain a constant surface area at the value of 62.9 Å^2/lipid, recently determined experimentally[56] and confirmed with NP_NAT MD simulations by Feller *et al.*[91] The z direction was the bilayer normal, which was allowed to fluctuate in size under the influence of the constant pressure P_N applied along the bilayer normal. A total of 2094 TIPS3P water molecules were added on the two sides of the lipid bilayer. The number of water molecules per lipid was therefore 29.1, as determined in recent experiments[32,56] for a fully hydrated DPPC bilayer in the biologically active L_α phase.

The Langevin piston,[19] implemented in the Nosé-Klein version for non-cubic simulation cells, was used to maintain P_N at the chosen value, and T was maintained by means of the Hoover thermostat.[17] A mass of 500 amu and a collision frequency of 5 ps^{-1} were used for the Langevin piston, and a value of 1000 kcal ps^2 was used for the thermostat (fictitious) mass. The leap-frog algorithm was employed to solve the equation of motion, with a time step of 2 fs. For the Lennard-Jones potential a 12 Å cutoff was employed and it was switched smoothly to zero over the region between 10 and 12 Å. Electrostatic interactions were calculated via the PME method using a κ value of 0.23 and a fast-Fourier grid density of ca. 1 Å^{-1}. The real space part of the PME summation was truncated at 12 Å. A neighbour list, used for calculating the Lennard-Jones potential and the real space portion of the Ewald sum, was kept to 14 Å

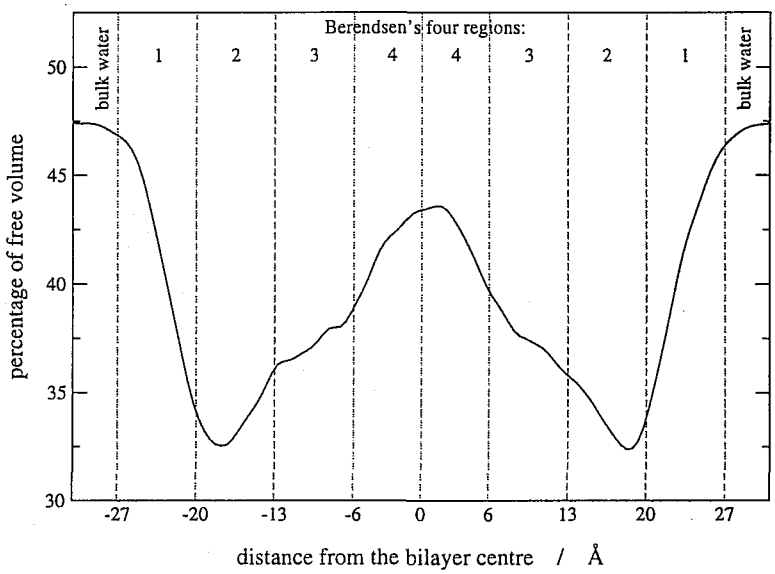


Figure 4.1: Free volume distribution (percentage) along the bilayer normal.

and updated every 50 fs. In all the simulations the SHAKE algorithm[20] was employed with a tolerance of 10^{-6} and all bonds involving hydrogens were constrained.

The simulation was 1 ns long and performed on a cluster of Linux PCs with 750 Mhz AMD Athlon processors, running in parallel using 4 processors.

4.2 Simulation Results

4.2.1 Four Region Model

Since the membrane has a very inhomogeneous character and a highly anisotropic behaviour, Berendsen *et al.*[61] proposed splitting the bilayer into four regions, each with its own special characteristics.

- **Region 1: low headgroup density**

  This region starts at the point where the presence of the membrane begins to result in a perturbation of the bulk water structure up to where the water density and the headgroup density are comparable. No sharp interface has been formed; water penetrates into the membrane and lipid headgroups protrude into the middle of the water layer. It is approximately 0.70 nm in width.

- **Region 2: high headgroup density**

  The water density drops to less than 1% with respect to the bulk phase. The membrane is hydrated up to the carbonyl groups at the beginning of the lipid tails and all the water

molecules are part of hydration shells of the phospholipid headgroups. This region has the highest density and the lowest free volume fraction. It is roughly 0.70 nm in width.

- **Region 3: high tail density**

  This region starts at the edges of the penetrating strands of water. It is characterized by a high tail density and low free volume and it consists of partially ordered chains. It is roughly 0.70 nm in width.

- **Region 4: low tail density**

  This is the middle part of the membrane. With respect to region 3, here the lipid chains are less ordered, the lipid density is lower and the free volume fraction is higher. It is roughly 0.60 nm in width.

Regions 1 and 2 represent the lipid/water interface and regions 3 and 4 are an almost completely hydrophobic environment. Figure 4.1 shows the free volume distribution along the bilayer normal resulting from the analysis on this system. The calculation of the free volume was done by searching for vacuum space points in bilayer slabs of 0.5 Å along the z-axis (corresponding to the bilayer normal), using a grid point spacing of 0.5 Å on the x-y plane. Vacuum points were defined as space points outside the van der Waals radii of the water and lipid molecules. The latter is considered to equal half the parameter σ in the Lennard-Jones (LJ) potential for each atom in the system, since σ is the distance at which the LJ energy is zero.

At a distance further than 27 Å from the bilayer centre, that is in the bulk water phase, the free volume percentage for this system is around 47%, in agreement with the percentage which can be calculated analytically from the density of bulk water. Inside the membrane, the lowest free volume is found in region 2 and the highest in region 4, in accordance with Berendsen's lipid bilayer model. However, Berendsen's simulations[107] were performed at temperature $\sim$ 30 °C higher (350 K) and employed a united-atom model, i.e. with no hydrogens along the lipid tails. Those simulations yielded a higher free volume percentage, above all in the membrane interior: the free volume in region 4 was even higher than in bulk water phase.[107] From this simulation, the lowest free volume percentage is about 32 % and the highest is about 43 %. Berendsen *et al.* obtained essentially the same value in region 2 ($\sim$ 34 %), but a higher percentage in region 4 ($\sim$ 55 %). They therefore concluded that the density of region 4 equals that in bulk hexane. This is not true for the membrane model simulated here. Moreover, in contradiction

with Berendsen's lipid bilayer, in the system simulated here there is no pocket of free space large enough to locate a molecule with the same size as water at any depth. As shown in later chapters, this has important consequences for the solute diffusion process. Unfortunately, to our knowledge, the measure of free volume distribution along a lipid bilayer normal is not accessible by experiments, and the publications using the CHARMM force field do not report free volume analysis. Stouch *et al.*[103,104,108,109] performed simulations at different temperatures and they found that the free volume percentage increases with temperature. This is due to the fact that at higher temperature the chains have a greater kinetic energy and consequently their packing density decreases. From those publications is not clear, however, whether the free volume inside the membrane is higher or lower than in bulk water, but all simulation studies show that the density decreases going from the carbonyl group towards the end of the lipid tails.

4.2.2 Atom Density Profile.

This gives the distribution of the position of a certain lipid or water atom along the axis perpendicular to the membrane. Figure 4.2 shows the results of the analysis on this system. The atom density η was derived by counting the number of atoms in slabs of width 0.5 Å along the z-axis and then dividing the number of occurrences N (number of atoms) by the slab volume V_S (nm^3) ($V_S = 47.5857 \times 47.5857 \times 0.5$ / 1000): $\eta = N/V_S$.

Results from these analyses are shown in figure 4.2 and are in accordance with previous simulations.[82,83,87–89,91,96] The simulation data reported in figure 4.2 are calculated and plotted for each half of the bilayer independently. The excellent agreement between the two data sets demonstrates that the simulation has been run for a sufficient length to ensure convergence.

It can be noted that water molecules penetrate up to the carbonyl groups of the lipids and never enter the regions 3 and 4. This is in excellent agreement with experimental observables.[49] The ammonium and phosphate groups reside in regions 1 and 2, while the methyl groups are located in region 4 and only rarely reach region 3. Moving from region 1, lipid density increases and it reaches the maximum in region 3, where there are about 110-120 atoms nm^{-3}, then density decreases and the minimum is in region 4, with about 100-110 atoms nm^{-3}. The water density has an inverse profile: the maximum is in the bulk phase, where there are 99 atoms nm^{-3} (corresponding to the experimental bulk water density of 33 molecules nm^{-3}), and then the density decreases, reaching zero at about 6 Å from the bilayer centre. The point where

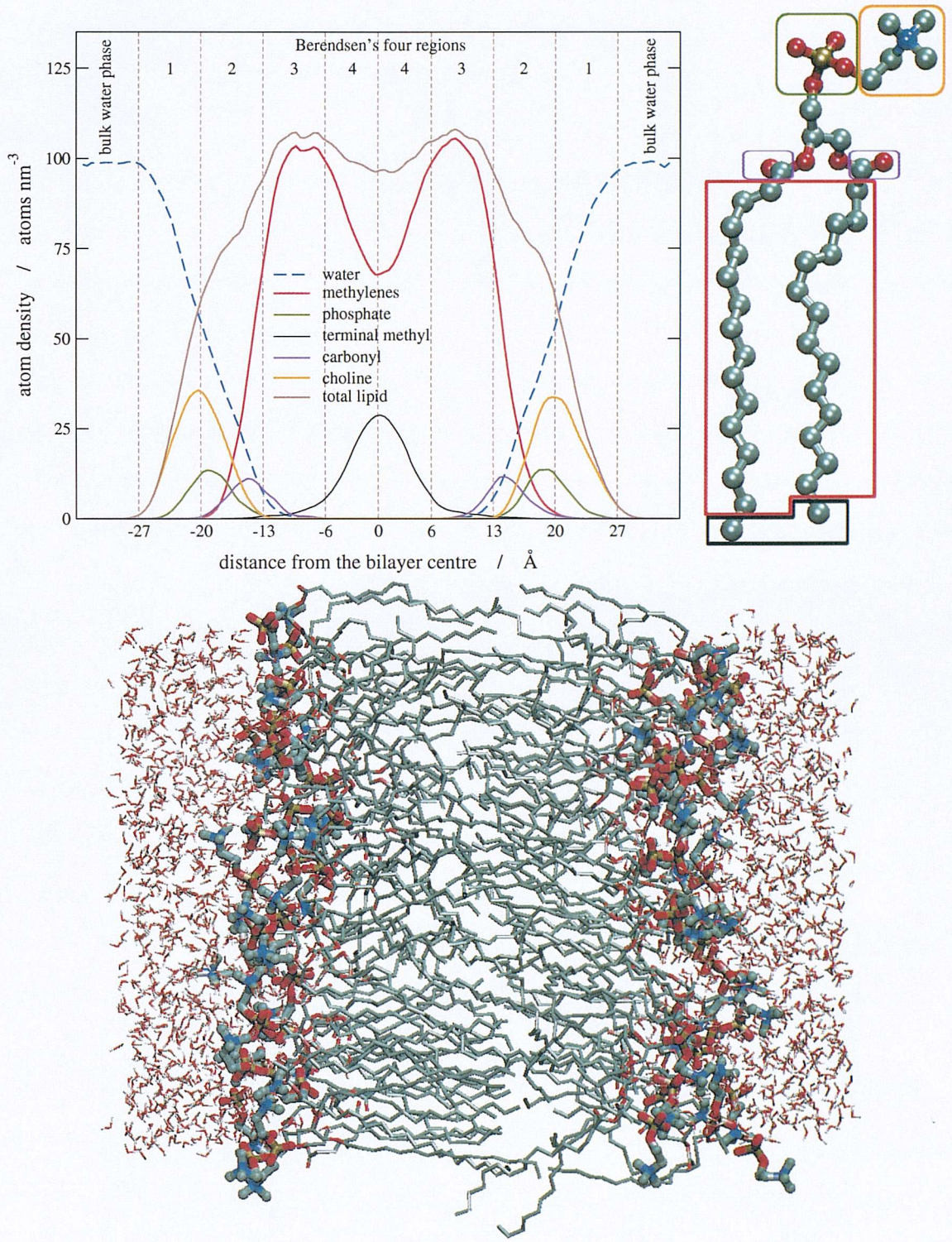


Figure 4.2: Top: atom density profile. Bottom: snapshot of bilayer; blue are nitrogen atoms, red oxygen atoms, orange phosphorus atoms, cyan carbon atoms; hydrogens are omitted.

lipid and water densities are equal (approximately 65 atoms nm^{-3}) is considered the boundary between regions 1 and 2 and the point where water density is lower than 1% is considered the boundary between regions 2 and 3.[61,88] Quantitative comparisons with other publications that do not distinguish the 4 regions in the bilayer or that use united-atom models (i.e. without explicit hydrogens) are difficult, but the typical atom density profiles show in outline the same results.[87,89,90,102]

The mean distance of selected lipid atoms from the bilayer centre was determined with neutron diffraction techniques.[34,35] Table 4.1 reports simulation and experiment results. For this simulation, errors are standard errors calculated dividing the total simulation into frames of 200 ps, as suggested by Feller[71]

The program, the force field and the protocol employed in this simulation are the same as in the reference[91] whose results are also reported in table 4.1. The same trend is found: excluding C15, the simulation yields slightly shorter distances than experiment. This was also obtained in other simulations employing different force fields.[83,96] However, experimental data do not provide a stringent test of the simulation accuracy, since the data have errors as large as ± 1.5 Å, and experiments were carried out at a lower level of hydration. The average difference between calculated and experimental values is anyway well within 1 Å. The observation that the choline C_α, C_β and C_γ atoms have almost the same value implies that the headgroups are aligned approximately parallel to the bilayer plane. The phosphorus-phosphorus distance in this simulation is found to be 38.1±0.1 Å, in agreement with the experimental value of 37±1 Å.[33] Values from previous simulations range between 35 Å[88] and 40 Å.[96]

Stouch *et al.*[103,104,108,109] found in their simulations that the presence of a small amount of penetrant particles in the membrane had no effect on the lipid atom distributions.

4.2.3 Electron Density Profile and Bilayer Thickness

To calculate the electron density profile from an MD simulation, a crude procedure of assuming atom-centred electron densities without electron reorganization can be used; the nitrogen atom can be assumed to have a net positive charge and the phosphorus a net negative charge. The electron density profile can be easily obtained from the atom density profile by multiplying the number of atoms in each slab by the number of electrons on each particular atom. The profiles resulting from these analyses are shown in figure 4.3 and agree with previous simula-

lipid atom	this simulation / Å	experimental data[34,35] / Å	reference simulation[91] / Å
C_γ	20.4 ± 0.1	21.8 ± 0.6	20.7 ± 0.2
C_β	19.7 ± 0.1	21.2 ± 1.0	20.0 ± 0.2
C_α	19.2 ± 0.1	21.0 ± 1.0	19.6 ± 0.2
CG-3	16.9 ± 0.1	17.4 ± 1.5	17.3 ± 0.2
C4	11.8 ± 0.1	12.2 ± 1.5	11.9 ± 0.2
C5	10.8 ± 0.1	10.5 ± 1.5	10.9 ± 0.2
C9	7.0 ± 0.1	8.1 ± 1.5	7.1 ± 0.2
C14	3.4 ± 0.1	3.6 ± 1.5	3.1 ± 0.2
C15	3.0 ± 0.1	1.9 ± 1.5	2.7 ± 0.2

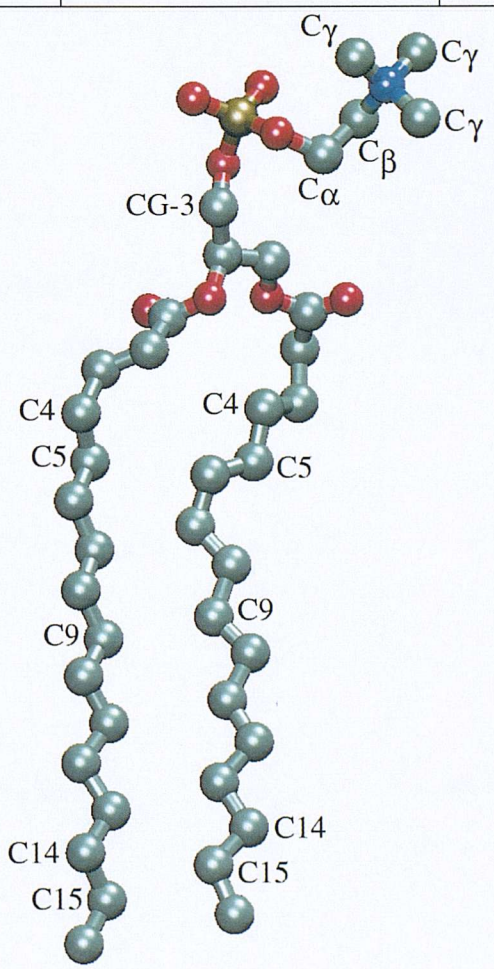


Table 4.1: Atom distances from the bilayer centre.

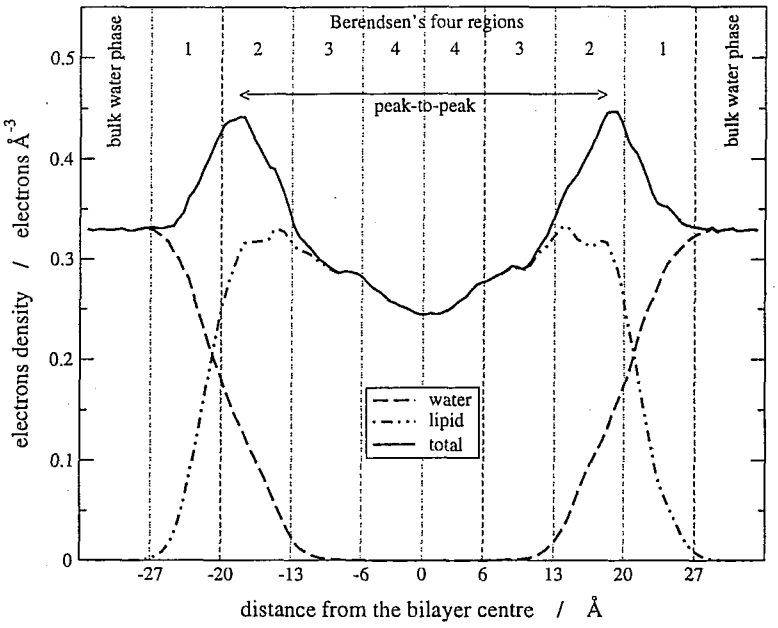


Figure 4.3: Electron density profile.

tions.[80,81,83,87,88,91,96,102] The same behaviour is reproduced: a minimum at the bilayer centre (about 0.25 e/Å^3), a maximum at the boundary between regions 1 and 2, corresponding to the headgroup region (about 0.44 e/Å^3), and a shoulder at the boundary between regions 2 and 3, that is at the interface between the hydrocarbon region and the headgroup region, which corresponds to the value of electron density in the water bulk phase (about 0.33 e/Å^3). Such profiles agree with experiments, which are obtained using X-ray spectroscopy.[32,56] The two peaks correspond to the electron rich phosphate fragments and the peak-to-peak distance is often used as a measure of the bilayer thickness. Recent experimental values range between 36.4 and 39.6 Å.[32,56] Previous simulations yielded 34[88] to 39[91] Å. The result of this simulation is 37.3±0.5 Å, well within the range of the experimental data.

Simulations also allow for the unphysical separation of the lipid and water contributions to the electron density (figure 4.3). Water electron density profile exactly reproduces the water atom density profile. Lipid electron density has a maximum in region 2, with about 0.33 e/Å^3, and coincides with the total profile in regions 3 and 4. The interface, regions 1 and 2, is the densest.

Feller showed[91] in his NP_NAT simulations that the larger the surface area, the smaller the peak-to-peak distance, consistent with a constant bilayer volume. This is explained in terms of decreased order leading to more instances of chain interdigitation. The best agreement of

the simulation results[91] with experiment was found for the value of surface area per lipid of 62.9 $Å^2$, that was employed in the simulation reported here. This value also agrees with that obtained by Nagle *et al.*[56] with a combination of X-ray experiments and theory.

As well as for the lipid atom distribution described in the previous section, Stouch *et al.*[103,104,108,109] found in their simulations that the presence of a small amount of penetrant particles had no effect on the bilayer thickness.

4.2.4 Volume of Lipid Molecules

In simulations a simple method is usually employed to calculate the volume per lipid V_L:[83,101]

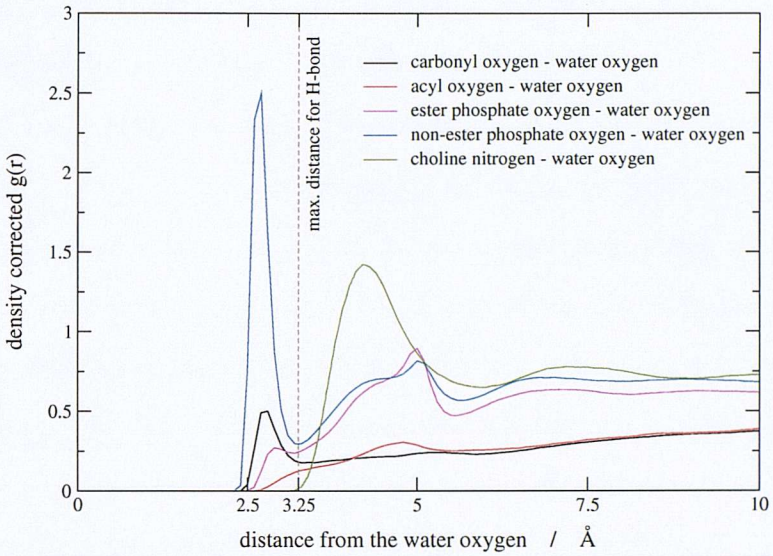
$$V_L = AD/2 - N_W V_W \tag{4.1}$$

Here A is the surface area per lipid (in our system it is fixed and its value is 62.9 $Å^2$), D is the height of the simulation box (in our system its average is 66.4 Å, with a standard error of 0.3 Å obtained dividing the simulation into frames of 200 ps), N_W is the number of water molecules per lipid (in our system it is 29.1), and V_W is the volume of one water molecule (30.3 $Å^3$, as derived from the bulk water density of $\sim$ 33 molecules nm^{-3} in this simulation). A recent experimental value of V_L is 1.232 $\pm$ 2 nm^3.[32,101] From these analyses a value of 1.207 $\pm$ 7 nm^3 is obtained, in quite good agreement with Feller's simulation[91] which yielded 1.219 nm^3. Both values are a bit lower than experiments. Other simulation results range between 1.217 and 1.230 nm^3.[83,101]

4.2.5 Hydration

The radial distribution function $g(r)$ of the water molecules around the lipid heteroatoms was calculated. A radial distribution function (RDF) gives the number of atoms that can be found at a distance r from a particular atom. This number is then normalized dividing by the number of atoms at the same distance in an ideal gas at the same density.

The RDF has been calculated for the ester and non-ester phosphate oxygens on the lipid headgroups, and for the acyl and carbonyl oxygens at the beginning of the hydrocarbon chains, with respect to the water oxygens. Results are plotted in figure 4.4. They agree with literature data.[3,80] The RDFs show that, among the lipid oxygens, the largest ordering of water is around the non-ester phosphate oxygens, while the acyl oxygens of the ester linkage between the fatty



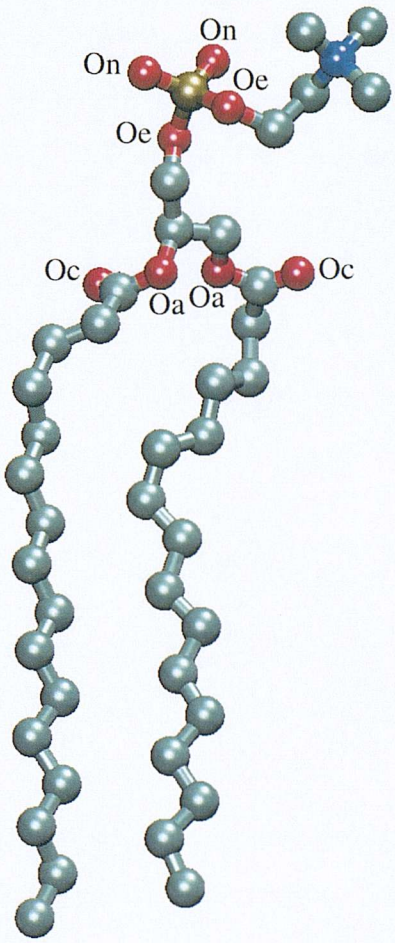


Figure 4.4: Top: lipid oxygen radial distribution functions. Bottom: identification of lipid oxygens: Oe = ester phosphate oxygens, On = non-ester phosphate oxygens, Oa = acyl oxygens, Oc = carbonyl oxygens.

acids and the glycerol practically have no hydration. This is in accord with the number of hydrogen-bonds involving the lipid oxygens, which are reported in table 4.2: the non-ester phosphate oxygens are involved in most of the H-bonds between water and lipids, while the acyl oxygens are involved in almost no H-bonds.

The criteria adopted for the identification of a H-bond are the following.[82] The distance between the water oxygen and the lipid oxygen is shorter or equal to 3.25 Å, and the angle between the vector linking the water and the lipid oxygens and one of the water hydrogen-oxygen bonds is less than or equal to 35°. The results of this simulation and a comparison with literature data are shown in table 4.2. Differences with literature are arguably insignificant. The RDF relative to the choline nitrogen (see figure 4.4) shows a peak at a distance larger than that used to define H-bonds. This result is consistent with the fact that the nitrogen atom is not able to make H-bonds with water,[82] although it is highly hydrated due to its net charge. Remarkably, the number of H-bonded water molecules per lipid from this simulation (5.2) is in excellent agreement with the experimental number of water molecules that are strongly bound to a phosphocholine and could not be removed by a common dehydrating agent, which is about 5.[82,110] The total number of bound water molecules per lipid from this simulation is 14.0, derived from the sum of H-bound and choline-bound water molecules. The latter are the choline nearest neighbours, defined as those water molecules located within the first minimum of the water oxygen - choline nitrogen RDF. The value from the simulation is perfectly within the range given in experiments, which is between 11 and 16.[47,82]

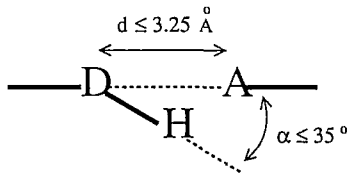
4.2.6 Trans/Gauche population

For the purpose of discussing the conformations adopted by the lipids, a conformation is assigned trans if it has a dihedral angle of $180 \pm 60°$, gauche$^+$ with an angle of $60 \pm 60°$, and gauche$^-$ with an angle of $-60 \pm 60°$.[80]

From experiments,[50,51] the percentage of gauche conformers at selectively deuterated methylene groups along the lipid tails determined with IR spectroscopy was about 20-30%. The mean number of gauche conformers per lipid chain as determined with IR techniques[50,52] and ^{2}H-NMR spectroscopy[6] is between 3.5 and 4.4. In previous simulations[80,81,83,87] the average gauche population along the hydrocarbon chains is about 20-30%, of which approximately the half is gauche$^+$ and the other half gauche$^-$. The rest of the conformers are trans, since the preferred

Property	Simulation values for a DMPC bilayer[82]	Values for this DPPC bilayer
H-bonds per lipid molecule	5.3 ± 0.14	5.5 ± 0.05
total number of H-bonded water molecules per lipid	4.5 ± 0.2	5.1 ± 0.06
number of doubly H-bonded water molecules per lipid	5.3-4.5=0.8	5.5-5.1=0.4
carbonyl oxygens:		
% of H-bonds	17%	18%
number of H-bonded water molecules	0.7	0.9
acyl oxygens:		
% of H-bonds	- %	1%
number of H-bonded water molecules	-	0.1
ester phosphate oxygens:		
% of H-bonds	8%	10%
number of H-bonded water molecules	0.3	0.6
non-ester phosphate oxygens:		
% of H-bonds	76%	70%
number of H-bonded water molecules	3.5	3.9
choline nitrogen:		
% of H-bonds	0%	0%
number of H-bonded water molecules	0	0
number of choline-bound water molecules(*)	7.6	8.8 ± 0.01

Criteria for identification of H-bonds:



A is H-bond acceptor (lipid oxygen) and D is H-bond donor (water oxygen).

Table 4.2: Hydrogen-bonds, comparison of literature and simulation data. (*)Choline-bound water molecules are waters not H-bonded to lipids and within the first peak in the water oxygen radial distribution function around the choline nitrogen, as shown in figure 4.4.

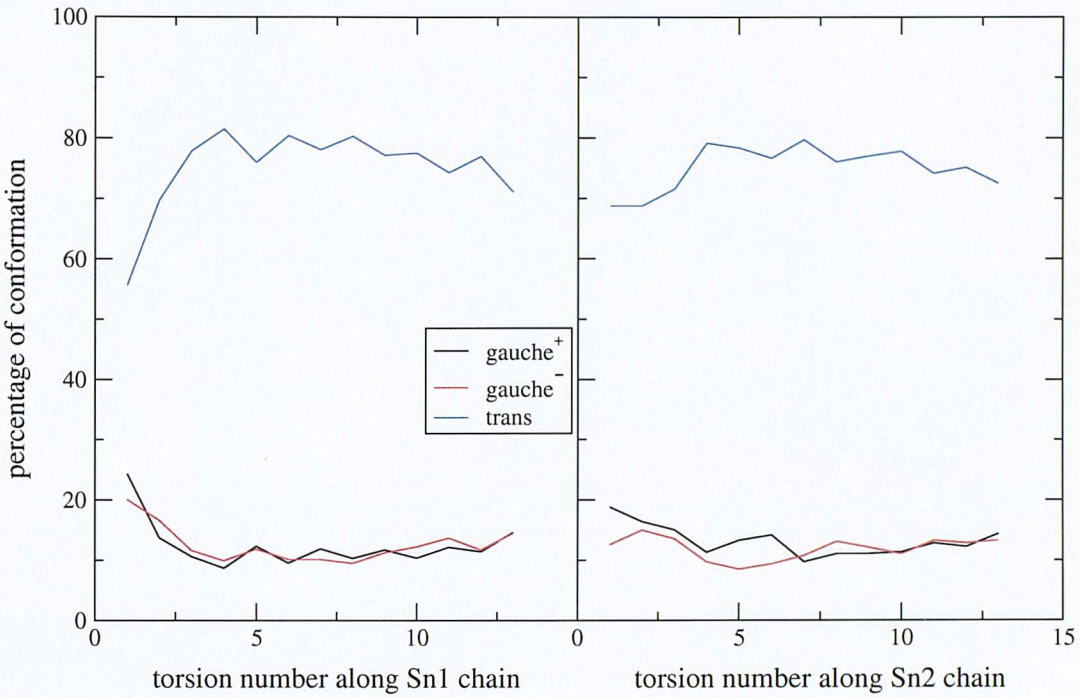


Figure 4.5: Conformational populations for the hydrocarbon chains as a function of torsion number.

alignment of the hydrocarbon chains is parallel to the bilayer normal. It has been noted that with decreasing temperature the fraction of gauche dihedrals decreases.[61] In figure 4.5 the trans/gauche populations are presented as a function of torsion number for the Sn1 and Sn2 hydrocarbon chains. Standard errors calculated dividing the whole simulation in frames of 200 ps are less than 5 degrees for all the torsions and between 1 and 3 degrees for most of them. Results from this simulation agree with the above mentioned published data and also reproduce the alternation in population observed by Essex *et al.*[80] An increase in gauche conformers in the lower part of the lipid chains, as expected from increasing disorder in the middle of the bilayer, is not revealed from experiments[50,51] and it is not produced by this simulation. The mean number of gauche conformers from this simulation is 3.3 ± 0.1, in agreement with experiments, although a little lower. With the same program, force field and simulation protocol, Feller *et al.*[100] also obtained a similar result.

It has been suggested[108] that the size and shape of the voids inside the bilayer are determined by the rates of rotation about the hydrocarbon single bond and that this is important to the process of diffusion. The residence time of a particular rotamer ranges from several tens of ps to hundreds of ps and these times vary with the position in the bilayer. Simulations[80,103,104,108,109] show that the fastest rates of torsional interconversion occur towards the terminal methyl end of the hydrocarbon chains, which is located in the middle of the bilayer, while the torsions near

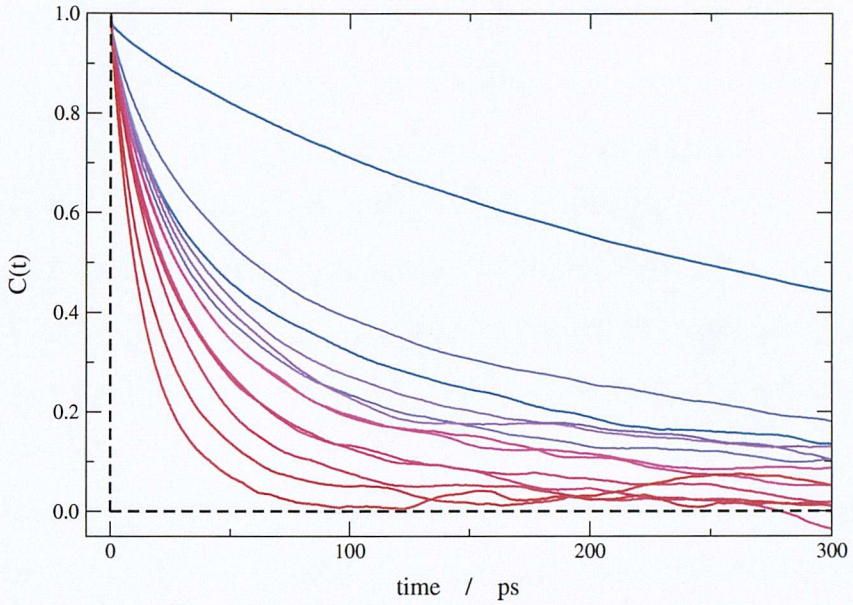


Figure 4.6: Time autocorrelation functions for torsional angles along the lipid tails. The blue-to-red shade corresponds to moving from the carbonyl position to the terminal methyl.

the ester linkage, that is near the interface, experience the longest residence times. The time autocorrelation function for the torsional angles ψ along the lipid tails has been calculated here as:[111]

$$C_\psi(t) = \frac{\langle \cos\psi(t)\cos\psi(0)\rangle - \langle\cos\psi(0)\rangle^2}{\langle\cos^2\psi(0)\rangle - \langle\cos\psi(0)\rangle^2} \tag{4.2}$$

The angular brackets denote averages over all the time origins and all the lipids. The results are plotted in figure 4.6. The curves had to be fitted with double exponentials:

$$C_\psi(t) = C_1\exp(-t/\tau_{short}) + C_2\exp(-t/\tau_{long}) \tag{4.3}$$

With the exception of the $C(t)$ for the two or three torsions closest to the terminal methyl, the curves do not decay to zero in a time which is short compared with the complete simulation time. Therefore statistics are poor and detailed numerical comparisons of the correlation times are not appropriate. The short decay time τ_{short} can be identified as the motion within the trans or gauche potential well, and is quite similar for all the torsions, with a value between 20 and 30 ps. The long decay time τ_{long} is associated to the transition between trans and gauche states and varies along the lipid chain. For the first three torsions, which are those closest to the carbonyl group, its value is longer than 1 ns, that is the total length of this simulation. For torsions 4 to 8 (assuming torsion 1 is the closest to the carbonyl group and torsion 13 the closest to the terminal methyl group) its value is between 200 and 300 ps. For torsions 9 to 11, it is between 60 and 90 ps, and for the last two torsions it is about 20 ps.

Experimentally, direct measurement of the above correlation times is not possible. Lipid tail motions have been investigated with the measurement of ^{2}H and ^{13}C spin-lattice relaxation times.[38,42,112,113] These are thought to be related to lipid motions as a whole, such as the axial rotation and the tilting of the lipid molecule with respect to the bilayer normal, and to lipid internal motions, such as trans/gauche interconversions. This interpretation leads to the conclusion that, in agreement with simulation results, the mean life of trans and gauche conformers is of the order of hundreds of ps in the upper part of the lipid chains, and tens of ps in the lower part.

The trans/gauche population on the headgroup is responsible for the orientation of the lipid dipole moment, which is usually identified by the phosphorus-nitrogen (P-N) vector. Results from these simulations are shown in table 4.3. The table also contains the headgroup torsions in the crystal.[114] The names to which the torsions refer are indicated in figure 4.7. On the phosphate and choline groups the torsion distribution during liquid crystalline phase simulations is relatively broad, but gauche conformations are found to be dominant,[80,81,83,87] bringing the P-N vector into a conformation almost parallel to the bilayer surface. From ^{31}P-NMR,[45] it is found that the conformation in the crystal structure is maintained in the L_α liquid crystalline phase. Above all, torsions $\alpha 2$ and $\alpha 3$, which are located at the phosphodiester linkage, are in the gauche conformation and lead to a bending of the phosphorylcholine group, so that it becomes aligned parallel to the membrane plane. This simulation is then in agreement with experiments. Raman spectroscopy[48] offers the evidence that a gauche conformation for the $\alpha 5$ torsion is preferred in both gel and liquid crystalline phases. This is reproduced by this simulation. Torsion angles $\alpha 1$ and $\alpha 4$ require a separate consideration. While some simulation[83] yields a predominance of gauche, in this and some other liquid phase simulations[80,81] they are predominantly trans in accordance with ^{31}P-NMR experiments. Regarding the glycerol group, in both the liquid crystal phase and gel phase, the dihedrals $\theta 3$ and $\theta 4$ are predominantly trans and gauche, respectively. This is shown in this and previous simulations[83] and in NMR experiments.[115]

4.2.7 Order parameter

^{2}H-NMR experiments enable the evaluation of the hydrocarbon chain order parameter $S_{CD,n}$ according to equations 3.1 and 3.2. From computer simulations $S_{CD,n}$ can be evaluated directly

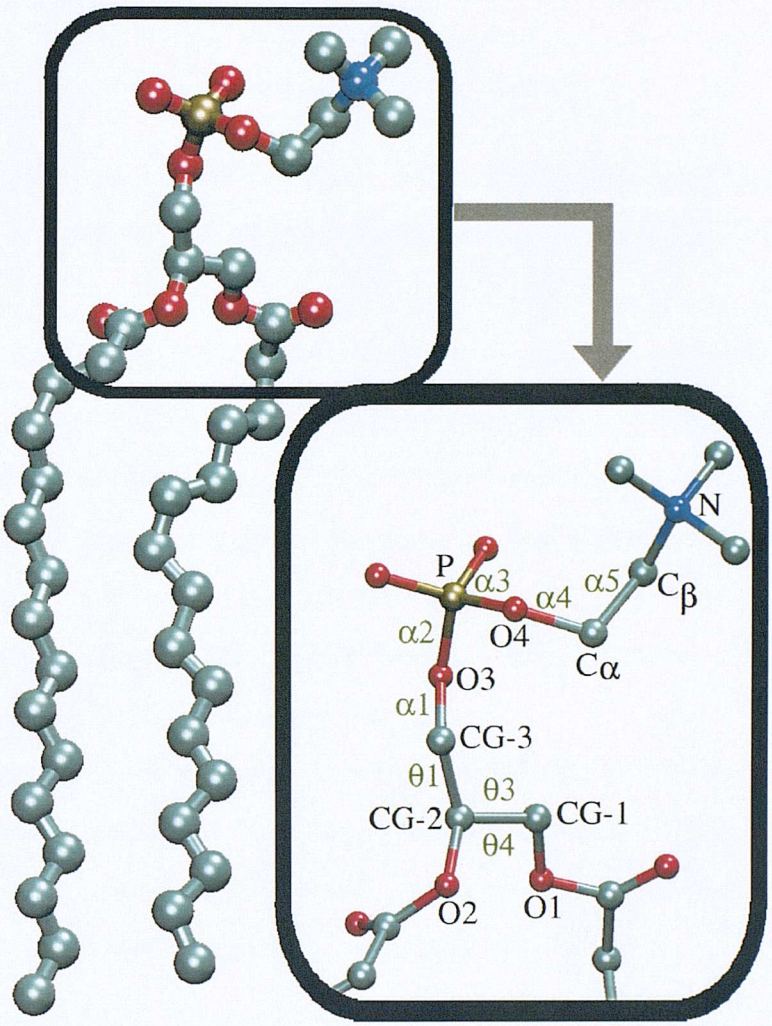


Figure 4.7: Head group dihedral angles used in the analysis. Definitions: θs are glycerol torsions and αs are phosphatidylcholine torsions; $\theta 1$ = O3 - CG-3 - CG-2 - CG-1; $\theta 3$ = CG-3 - CG-2 - CG-1 - O1; $\theta 4$ = O2 - CG-2 - CG-1 - O1; $\alpha 1$ = CG-2 - CG-3 - O3 - P; $\alpha 2$ = CG-3 - O3 - P - O4; $\alpha 3$ = O3 - P - O4 - Cα; $\alpha 4$ = P - O4 - Cα - Cβ; $\alpha 5$ = O4 - Cα - Cβ - N.

dihedral	% gauche$^+$	% gauche$^-$	% trans	crystal[114] mol.A	crystal[114] mol.B
$\theta 1$	79.92	7.22	12.86	trans	gauche$^+$
$\theta 3$	21.57	11.55	66.87	trans	trans
$\theta 4$	66.89	21.57	11.53	gauche$^+$	gauche$^+$
$\alpha 1$	10.99	5.74	83.27	trans	trans
$\alpha 2$	42.78	39.86	17.36	gauche$^+$	gauche$^-$
$\alpha 3$	39.20	38.73	22.07	gauche$^+$	gauche$^-$
$\alpha 4$	18.60	23.59	57.81	trans	trans
$\alpha 5$	36.55	45.71	17.74	gauche$^-$	gauche$^+$

Table 4.3: Headgroup trans/gauche population in percentage.

if an all-atom model is employed, after rewriting equation 3.2 into:[81,87,89–92,116]

$$S_{CD,n} = \frac{1}{2}\langle 3\cos^2\theta_n - 1\rangle \tag{4.4}$$

Here θ_n is the angle between the C-H bond vectors of the methylene group at chain position n and the bilayer normal, and the brackets denote average over simulation steps and number of lipids.

An order parameter is a measure of the degree of order (or, equivalently, disorder) in the system. It can range from -0.5 (indicating full order along the bilayer normal) to 1.0 (full order perpendicular to the bilayer normal). A value of zero is considered to indicate that full isotropic motions are occurring and that the atoms are distributed randomly. Complete disorder is expected in the liquid state at equilibrium, but if molecules which form a liquid crystalline phase are simulated, such as phospholipid bilayers, then it is expected that, on average, the molecules tend to line up in a common direction.

In lipid bilayers, the hydrocarbon chains tend to be aligned parallel to the bilayer normal because of interfacial constraints and consequently S_{CD} values are negative. However, in the literature their absolute values $\mid S_{CD} \mid$ are indicated, because the absolute value only is accessible from experiments. If we number the methylene groups along the hydrocarbon chain moving from the carbonyl group towards the tail end, typical $\mid S_{CD} \mid$ from ^{2}H-NMR experiments[4–6,38,39] and from simulations[61,81,87–92,96,116] have the value of approximately 0.18 - 0.22 for the methylenes 2 to 9 (a plateau at the beginning of the chains), then they decrease and reach the approximate value of 0.08 - 0.10 for the last methylene before the final methyl group. Values from this simulation agree with literature data and are shown in figure 4.8 together with experimental values. The standard error for the simulation results is at most 0.02.

Contributions to the order parameter profile include the trans/gauche isomerization and the reorientation of the lipid molecule as a whole.[4] The plateau for the upper part of the chain near the glycerol region arises because the motion of the methylenes is restricted. The degree of disorder towards the end of the chains near the middle of the bilayer is due to the greater torsional and translational freedom of these methylenes.

No significant differences can be seen between the two lipid tails apart for the first methylene, which is carbon position number 2 on the chain. ^{2}H-NMR experiments[5] show that S_{CD} at the carbon position number 2 of the Sn2 chain is lower than that of the Sn1 chain. This is thought to be indicative of different orientations of the two chains with respect to the bilayer normal.[5]

The CHARMM force field[100] does not completely reproduce this behaviour. The first S_{CD} value of the Sn1 chain is a little higher than that of the Sn2 chain, but still much lower than the experimental. This limitation is admitted in the original publication of the force field.[100] From this simulation, results for the Sn1 chain are further from the experimental than those for the Sn2 chain, but still sufficiently close.

Stouch *et al.*[103,104,108,109] found that towards the carbonyl groups there is a small decrease in order in the presence of benzene molecules, while at the end of the chains the simulations with and without solute are almost identical. This is explained in the following way: the centre of the bilayer has more free space than the headgroup region, so the solute molecules cause little effect on the ordering of the chains, while towards the carbonyl groups there is less free space and the presence of the solute is felt more strongly.

4.2.8 Dipole Moment

Phosphatidylcholines are zwitterionic molecules, because the phosphate group has a negative charge and the ammonium group has a positive charge. This results in the presence of an electric dipole located on the headgroup and in MD simulations it is usually identified by the P-N vector. The orientation of the dipole parallel to the bilayer surface is favourable in terms of free energy of electrostatic interactions, but there are steric requirements which oppose this parallel orientation.[61] This results in a wide distribution of tilt angles from 0° to 90°. Different experiments confirm that the lipid headgroup tend to lie on the membrane planes. ^{2}H- and ^{31}P-NMR measurements[45] show that the phosphodiester torsions lead to a bending of the phosphorylcholine fragment; neutron diffraction on selectively deuterated lipids[34,35] yields similar distances from the bilayer centre for all the choline carbon atoms; Raman spectroscopy[48] reveals the choline group is aligned parallel to the bilayer surface.

The average P-N angle with respect to the membrane surface is found to be quite different in different simulations, ranging from 5°[81] to 30°,[86] and this may depend on the relatively short length of the simulations. However, these angles are small and reproduce the experimental observable that the dipole moment tends to lie parallel to the bilayer surface.

The angle between the bilayer normal and the average P-N vector of this simulation has a mean value of 72.5±1.2° pointing towards the water phase, that is the angle between the P-N vector and the membrane plane is 27.5°. This is in excellent agreement with previous studies.[74,96] Figure 4.9 plots the instantaneous angle between the P-N vector and the bilayer

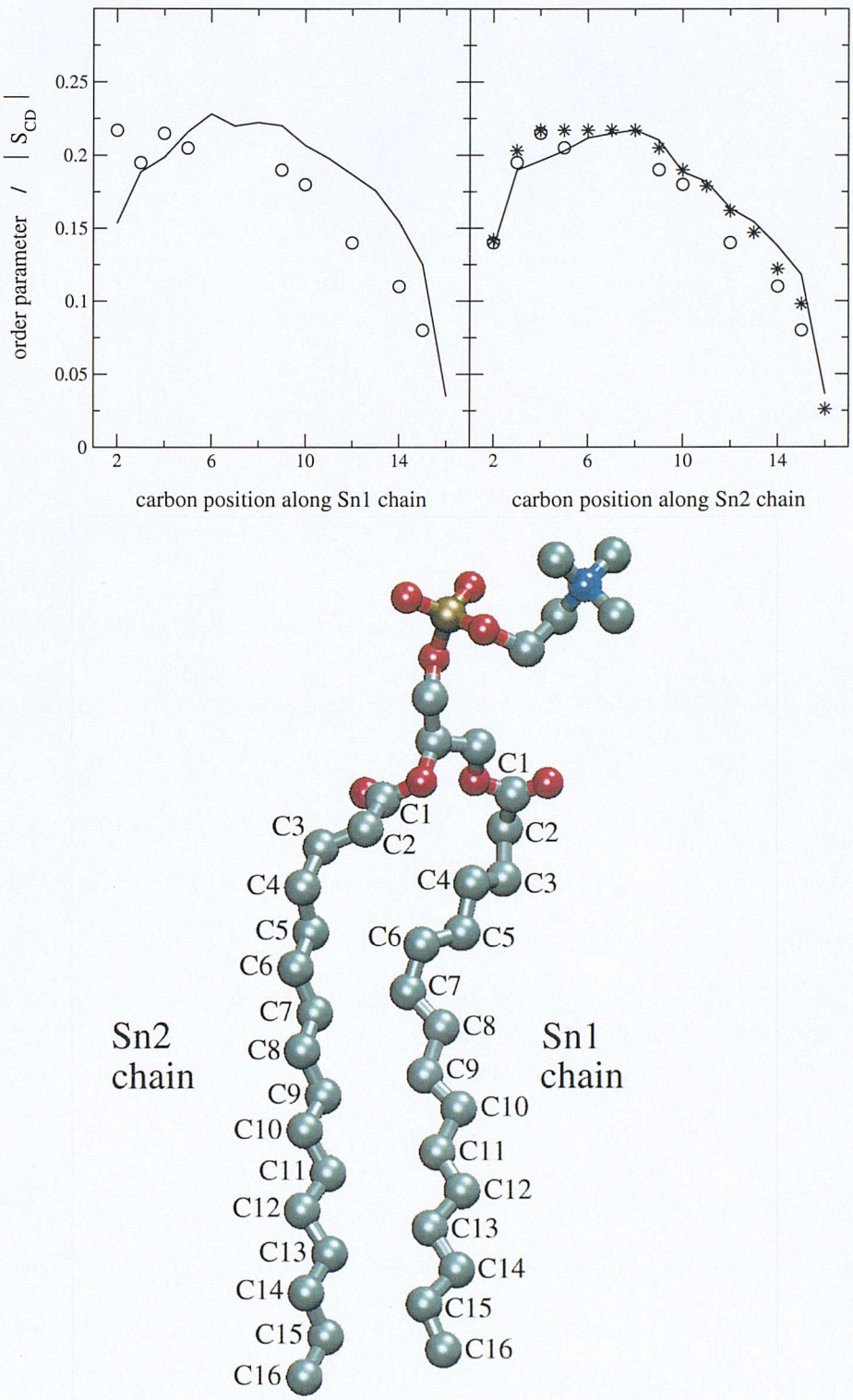


Figure 4.8: Top: order parameter along the lipid tails; open circles are experimental values from Seelig *et al.*,[4,5] stars are experimental values from Douliez *et al.*[6] Bottom: carbon positions are defined as shown.

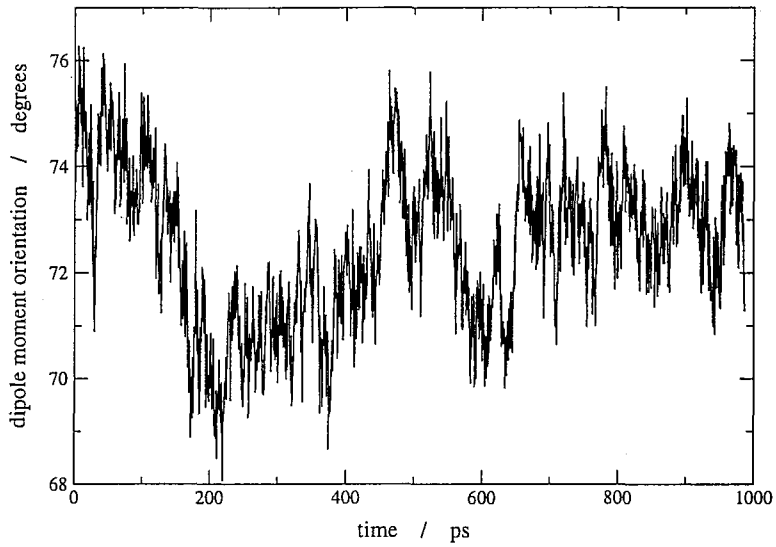


Figure 4.9: Angle between the P-N vector and the bilayer normal as a function of simulation time.

normal as a function of simulation time averaged over all lipids.

4.2.9 Electrostatic Potential

The electrostatic potential $\Psi(z)$ across the membrane is plotted in figure 4.10. The electrostatic potential $\Psi(\boldsymbol{r})$ at a point $\boldsymbol{r}$ is defined as the work done to bring a unit positive charge from infinity to that point.[12] Within a medium of uniform dielectric constant ϵ, the Poisson's equation relates $\Psi(\boldsymbol{r})$ to the charge density $\rho(\boldsymbol{r})$:[12] $\nabla^2\Psi(\boldsymbol{r}) = -\rho(\boldsymbol{r})/\epsilon$. For a lipid bilayer, the electrostatic potential $\Psi(z)$ along the bilayer normal, z axis, was obtained calculating the charge density $\rho(z)$ as a function of depth and then integrating twice along the bilayer normal:[74]

$$\Psi(z) = -\frac{1}{\epsilon_0}\int_0^z dz' \int_0^{z'} \rho(z'')dz'' \tag{4.5}$$

Here ϵ_0 is the vacuum permittivity. The charge density $\rho(z)$ was obtained dividing the simulation cell into slabs of 0.5 Å and counting the individual atomic charges in each slab averaged over the simulation time. The value of the potential is taken to be zero in the centre of the bilayer because of symmetry and the potential $\Psi(z)$ in equation 4.5 is actually a difference between the potential at position z and that at position 0.

MD simulations allow for the unphysical separation of the lipid and water contributions. These profiles are in qualitative agreement with all previous publications,[61,70,74,83,87–89] and absolute values are in agreement with most of them. The contributions to the total potential

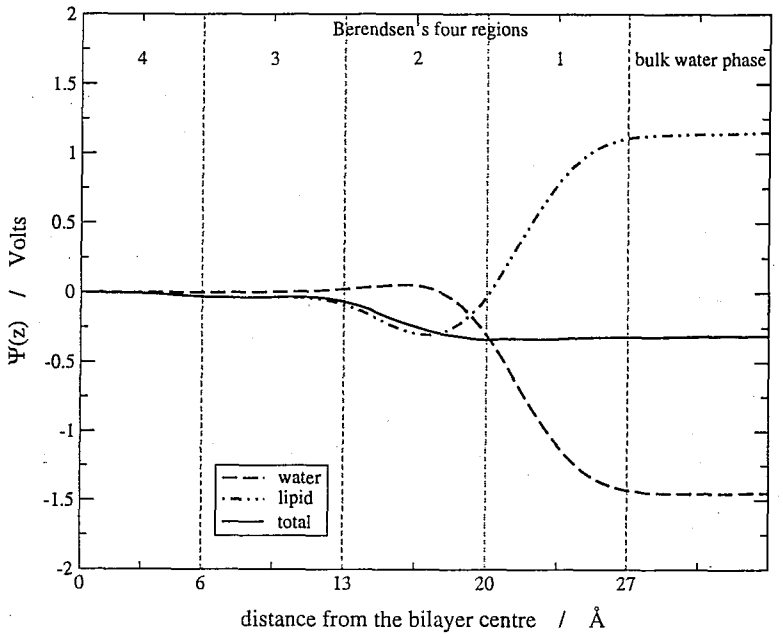


Figure 4.10: Electrostatic potential along the bilayer normal.

from the water and the lipid are opposite to each other and the net potential (which is the only experimentally measurable quantity) is a relatively small difference between two large quantities.

Since in the lipid molecule the negative phosphate group is on average located closer to the middle of the bilayer than the positive ammonium group, the lipid contribution is negative approaching the interface and then it is positive. Owing to the orientation of the water molecules, with their negatively charged oxygen atoms pointing towards the bulk phase and their positively charged hydrogens towards the membrane, the water contribution is negative in bulk phase and increases on entering the membrane. The orientation of the water molecules is such to compensate the charge density arising from the lipid headgroups, but also simulations of uncharged lipids find an intrinsic tendency of the water to slightly orient its dipole outwards the water phase.[61] The overall electrostatic potential is negative in water bulk phase with respect to the bilayer interior, indicating that the contribution of water prevails over that of lipids. The difference between the membrane interior and the water bulk phase from these simulations is about 280 - 300 mV. Previous publications report values ranging between 250 mV[88] and 600 mV.[89] Experimental values for phosphocholine/water interfaces are not conclusive, but range between 200 mV and 575 mV,[88] suggesting that results from these simulations are of the right order of magnitude.

4.2.10 Lipid Diffusion Coefficient

The diffusion coefficient D of a particle is related to its Mean Squared Displacement (MSD):[11,12]

$$D = \frac{1}{2n} \lim_{t\to\infty} \frac{d}{dt}\langle[r(t) - r(0)]^2\rangle \approx \frac{1}{2nt}\langle[r(t) - r(0)]^2\rangle = \frac{1}{2nt}\text{MSD} \tag{4.6}$$

Here $\langle[r(t) - r(0)]^2\rangle$ is the MSD of the diffusion, which is directly accessible from computer simulations, and n is the dimensionality of the diffusion. The approximate equality in the above formula becomes exact at long times, when the simulation is significantly longer than the time scale of the diffusion process under study. The displacement is averaged in a simulation over all possible time origins and over all possible time intervals to increase the statistical precision. However, MSD values closer to zero have higher precision, because for them there is a larger number of data points, while values at high t have lower precision. Thus MSD values for t approaching the total simulation time are very noisy and not reliable. In these analyses, MSDs are calculated for time intervals between 0 and 500 ps, since the simulation lasted 1 ns. Setting n=2 in equation 4.6 and considering the displacement of the lipid centre of mass on the x-y plane, i.e. $\text{MSD}_{||} = \langle[x(t) - x(0)]^2 + [y(t) - y(0)]^2\rangle$, the lipid in-plane (lateral) diffusion coefficient $D_{||}$ can be obtained. Instead, setting n=1 and considering the lipid centre of mass displacement along the z-axis, i.e. $\text{MSD}_{\perp} = \langle[z(t) - z(0)]^2\rangle$, the lipid out-of-plane diffusion along the bilayer normal (perpendicular) $D_{\perp}$ is obtained.

$\text{MSD}_{||}$ and $\text{MSD}_{\perp}$ from these simulations are plotted in figure 4.11. Linear fitting of these curves between 100 and 500 ps yields a $D_{||}$ (i.e. on the x-y plane) of 1.1 $\times$ 10^{-6} cm^2 s^{-1} and a $D_{\perp}$ (i.e. along the z axis) of 2.4 $\times$ 10^{-7} cm^2 s^{-1}. Previous simulations yielded values a little larger in the order of 10^{-6}-10^{-5} cm^2 s^{-1}.[61,80]

From experiments such as fluorescence techniques[53,54] lipid lateral diffusion coefficients in L_{α} phase are in the order of magnitude of 10^{-8}-10^{-7} cm^2 s^{-1}. The discrepancy between calculated and experimental data is due to the short time scale of current computer simulations. The latter can only yield information about the "rattling in a cage" motions of lipid molecules and, to a lesser extent, about the rotation of the lipid as a whole molecule, but not translational diffusion involving exchange of lipids.[92] Neutron scattering experiments[55] can instead sample motions in the same time scale as computer simulations and yield lipid lateral and perpendicular diffusion coefficients in the order of 10^{-7}-10^{-6} cm^2 s^{-1}, with which this and previous simulations are in agreement.

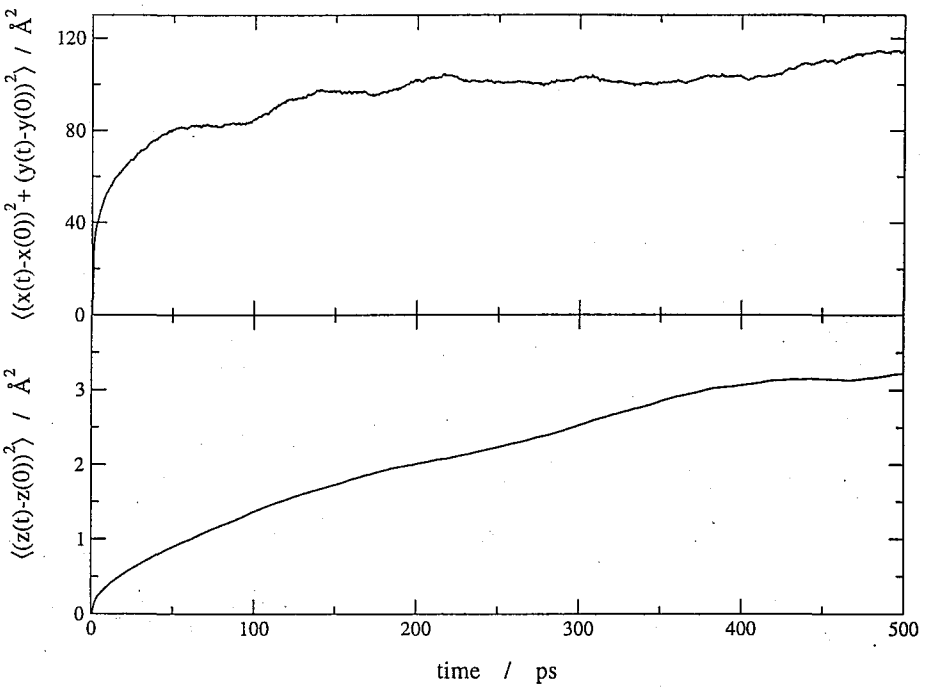


Figure 4.11: Top: $MSD_{\parallel}$; bottom: $MSD_{\perp}$.

4.2.11 Energy and Pressure Components

The energy of the system and the dimension of the simulation box along the bilayer normal (z-axis) are plotted in figure 4.12 as a function of time to monitor the integrity of the simulation.

The total energy of the system was not expected to be constant, because the chosen ensemble was NP_NAT and energy does not appear among the constrained degrees of freedom. However, since the system was already equilibrated and its physical properties oscillate around their equilibrium values based on the force field, the total energy of the system does not vary systematically during the simulation. The same behaviour is shown by its components, namely the kinetic and the potential energies. Instead of the kinetic energy, its physically related property, i.e. the temperature, is plotted. In contrast to the potential and total energy, the temperature was constrained at 323 °C and its value was expected to be quite stable, since the CHARMM program has a proven thermostat. In this simulation the temperature is indeed stable, its mean value being 323.0 ± 2.5 °C. The values of the surface area per lipid, A, during the simulation are not shown because CHARMM keeps the x-y plane exactly constrained. The value of A employed in this simulation is that of a recent experimental determination[56] confirmed by Feller *et al.* in recent simulations,[91,100] i.e. 62.9 Å^2. Only the z-dimension is allowed to change and to adjust its value according to the pressure applied along the bilayer normal. However, the same considerations made for the total energy are valid here too, that is the dimension

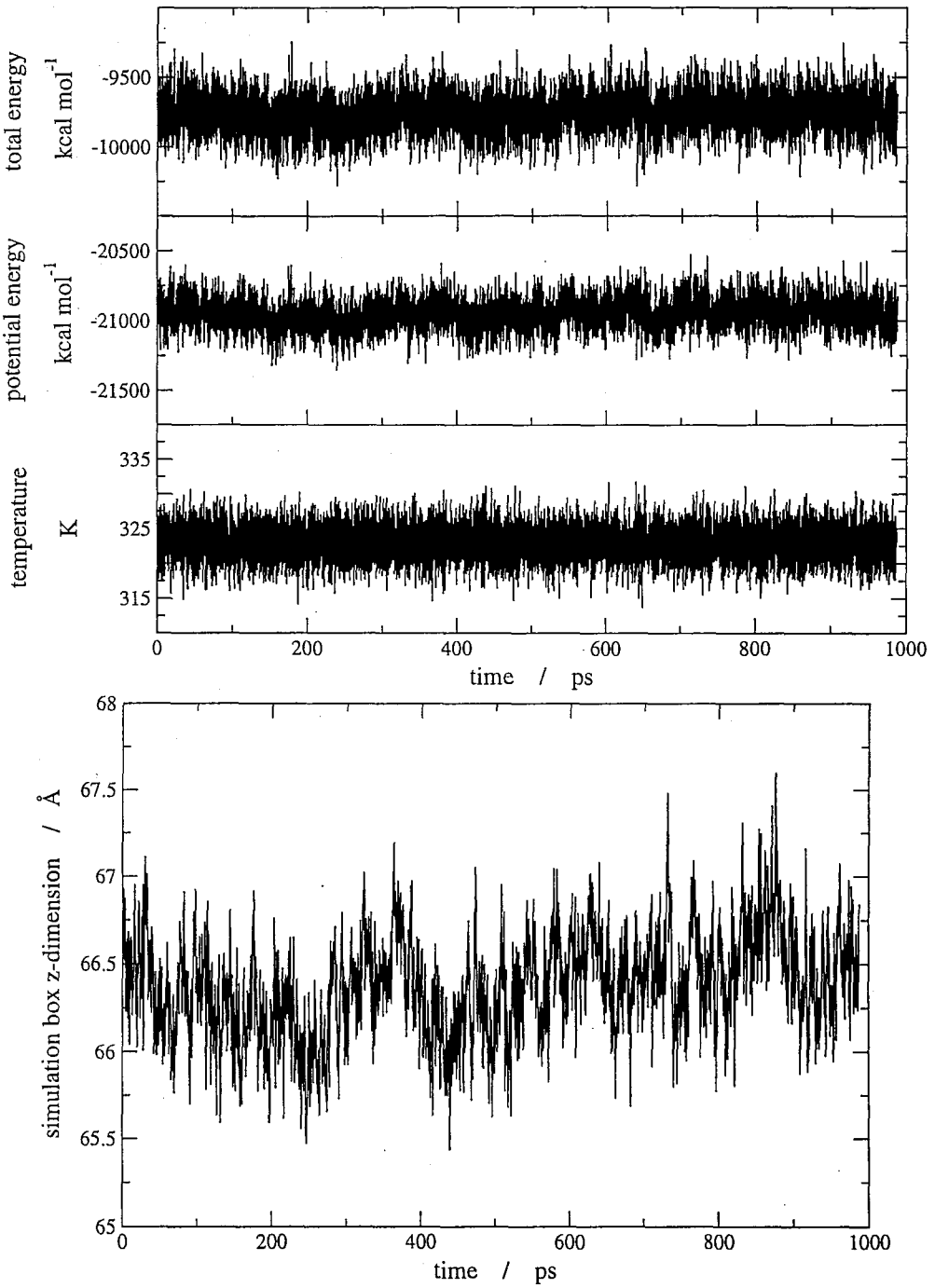


Figure 4.12: Top: total energy (potential + kinetic), potential energy and temperature; bottom: box z-dimension.

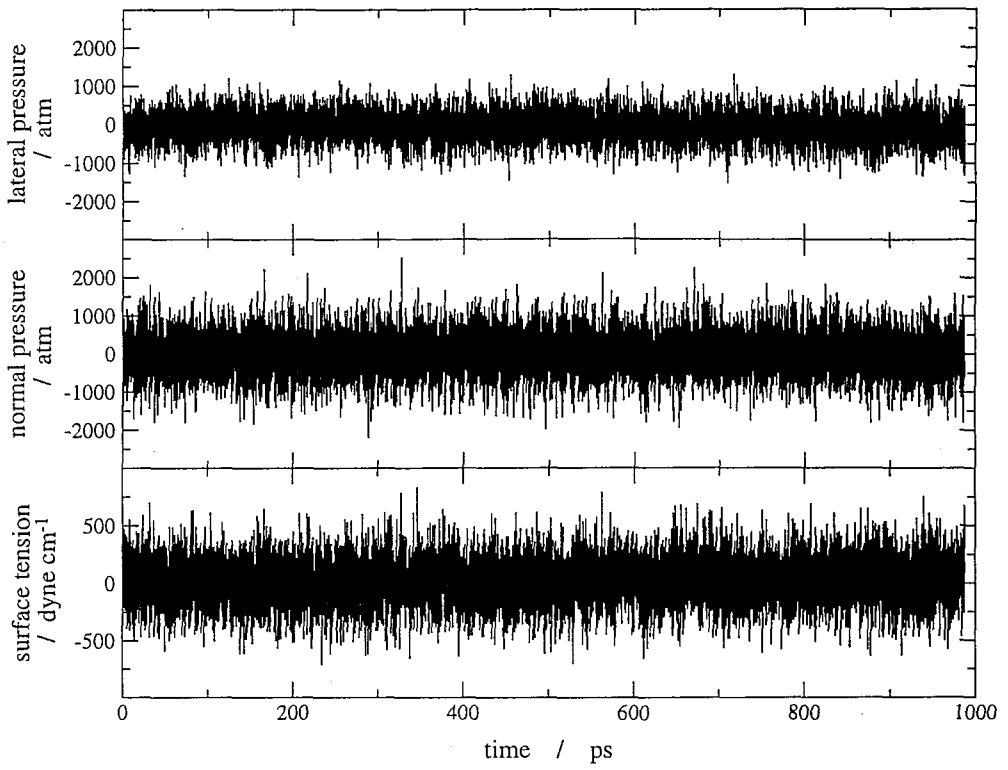


Figure 4.13: Lateral and normal components of the pressure tensor, and surface tension.

of the simulation box along the z-axis is expected to fluctuate around its equilibrium value without systematic drifts. This can be seen in figure 4.12, where the simulation z-dimension is plotted against time. In this simulation the mean value of the box z-dimension is 66.38 $\pm$ 0.29 Å. Since periodic boundary conditions are employed, the lipid bilayer is repeated in the three dimensions of space and what is actually simulated is a multilamellar geometry along the z-axis. The dimension of the simulation box along this axis has then to be compared with the so called repeat spacing of multilamellar lipid vesicles which is experimentally measurable. The most accurate value for this property seems to be around 67 Å.[32] Feller *et al.*[101] using the same force field, the same program and the same simulation protocol obtained a value of 66.91Å. In both this and Feller's simulation, the simulated value is then in good agreement with the experiments, although a little lower.

The pressure of the simulation box was calculated with equation 2.38. Figure 4.13 plots the components of the pressure tensor. Since in the simulation box the z-axis is along the bilayer normal and the x-y plane defines the bilayer surface, the normal pressure is $P_N = P_{zz}$ and the average lateral pressure $\overline{P_L}$ is calculated according to equation 2.48 with L_z the box length along z. As is common in MD simulations, both P_N and $\overline{P_L}$ oscillate heavily. This is due to the

fact that the product of the positions with the derivative of the potential energy (i.e. the virial) changes quickly with the position.[12] The mean values averaged over the simulation are -64.1 ± 36.6 atm for $\overline{P_L}$ and 1.3 ± 2.4 atm for P_N. The latter is in agreement with the reference value at which P_N was constrained, suggesting that the Langevin piston method is working properly. The presence of a negative lateral pressure gives rise to a non-zero surface tension γ. From MD simulations, γ is calculated according to equation 2.49. In these simulations γ is 44.0 ± 12.6 dyne cm^{-1}, in agreement with the value of 39 ± 8 dyne cm^{-1} reported in the original NP_NAT simulation by Feller.[91]

4.3 Conclusions

In this chapter the MD simulation of a fully hydrated DPPC phospholipid bilayer has been presented. The treatment of electrostatic interactions, the temperature and pressure coupling, and the simulation protocol were rigorous and chosen from among the best approaches to simulate such biological membrane models. This MD simulation was quite computationally demanding. Although simpler procedures have been applied to such systems and shown to reduce the expense of such calculations, they have been shown to produce artifacts and are not recommended.

The work by Feller *et al.* has been repeated and correctly reproduced. The simulation also compares favourably with other publications. The system is stable and the results are in good agreement with available experimental data. Several physical properties of the DPPC bilayer have been analysed. Atomistic simulations allow for structural and dynamic properties to be described with molecular detail. The atom distribution and the electron density agree with a four region model of the membrane. The choline fragment and the phosphate non-ester oxygens on the lipid headgroups are the main responsible for the lipid hydration. In proximity of the bilayer centre, the hydrocarbon chain order decreases and its mobility increases.

The conclusion is therefore that this membrane computer model can be used to study the mechanism of solute permeation and, eventually, can help to understand the mechanism of drug absorption after administration.

CHAPTER 5

Permeability Coefficients

This chapter describes how experimental permeability coefficients can be measured for small organic solutes across lipid bilayers and for real drug molecules across cell monolayers. Previous simulations dealing with solute partition and diffusion inside biological membranes are also reviewed. A technique which allows for the permeability coefficients to be obtained from molecular dynamics simulations is presented prior to its application in the following chapters.

5.1 Experimental Determination of Permeability Coefficients for Small Organics

Permeability coefficients for small organic molecules can be measured across planar lipid bilayers formed on a $\approx$1 mm^2 hole in a polyethylene or Teflon partition separating two magnetically stirred chambers[2,117–123] or, alternatively, across lipid bilayers forming large unilamellar vesicles[122,124–134] by means of radioactive tracers, ^{1}H-NMR measurements or fluorescence techniques, by simply withdrawing samples of the receiving solution or collecting the vesicles at intervals of time.

Since the barrier properties of lipid bilayers are shown to be pH independent, the observed pH dependence of the apparent permeability coefficient for weak acids and weak bases may therefore be attributed to the different permeability coefficients of the neutral and ionic forms,

and to the relative fractions of permeant in neutral and ionized form with solution pH, as governed by the solute pK_a. Thus for electrolytes:[2,111,118,127–130,133–139]

$$P_{obs} = P_n f_n + P_i f_i \tag{5.1}$$

where n stands for neutral and i for ionized species respectively and f indicates the fraction of each species. Assuming $P_i \approx 0$, that is only the neutral species crosses the membrane at a significant rate, then the permeability coefficient reduces to the contribution of the neutral form only, i.e. $P_{obs} = P_n f_n$.[2,118,135] Some publications report P_{obs} values, but most of them report P_n data.

In the absence of carriers or channels, permeation of molecules across membranes is thought to occur via the lipid bilayer by a three-step process: the partitioning of the substance from the aqueous phase on one side into the bilayer, its diffusion across the bilayer, and its partitioning from the bilayer into the aqueous phase on the other side. This model is termed solubility-diffusion model of permeation.[119–121,126,128,129,133,138,140–145] The overall membrane resistance R to solute permeation, defined as the inverse of the permeability coefficient P, can be expressed as the integral over the local resistances across the membrane:[133,136,138]

$$R = \frac{1}{P} = \int_0^d \frac{dz}{K(z)D(z)} \tag{5.2}$$

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. However, if transport is governed primarily by a distinct region within the lipid bilayer, the barrier region, it may be possible to describe P by the following simplified solubility-diffusion model:[2,118,128,133,136,138]

$$P = \frac{K_{barrier} D_{barrier}}{d_{barrier}} \tag{5.3}$$

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 hydrocarbon-like 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 conclusion is that the zone of maximum partition accounts for only a small fraction of the resistance to permeation and that 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.[133,138] 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. Owing to the hydrocarbon nature of the membrane interior, permeability coefficients for the most common small organics are often highly and linearly correlated with partition coefficients in long chain alkanes like hexadecane, hexadecene or decadiene.[2,118,128,133,136,138] However, this approach fails to consider interfacial constraints which lead to chain-order gradients in the bilayer interior and which are not present in bulk solvents.[127–129] The chain packing may reduce the permeability coefficients by increasing viscosity (.i.e. reducing diffusion coefficients), by decreasing solute penetration and accommodation (i.e. reducing the partition coefficient), or both.[128]

Moreover, water and other small hydrophilic molecules penetrate cells more rapidly than would be expected on the basis of their solubility properties in a lipid solvent. Furthermore, the introduction of an alkyl group into the first member of a homologous series of nonelectrolytes decreases the permeation rate in many cells instead of increasing it, as might be expected from a consideration of their increased partition into a bulk lipid solvent. Many authors[2,118,119,136,140,141,143,146,147] argued that this "sieve-like" property can be explained by treating the membranes as a homologous polymer network and describing diffusion as regulated by the amount of free volume or "holes" between the hydrocarbon chains.

5.2 Experimental Determination of Permeability Coefficients for Drugs

Permeability coefficients for drug molecules are not usually measured across isolated lipid bilayers, but rather more complex model systems are employed that more closely correlate with the extent of drug absorption *in vivo*.

The most common *in vitro* model for selection of new chemical entities in drug discovery are Caco-2 cell monolayers.[148–178] Caco-2 cells originate from a human colorectal carcinoma and, when seeded on a polycarbonate filter and allowed to grow on it for up to about 3-4 weeks, spontaneously differentiate into polarized monolayers with morphological and functional similarities to the small intestinal epithelium. The basolateral side is towards the filter and the

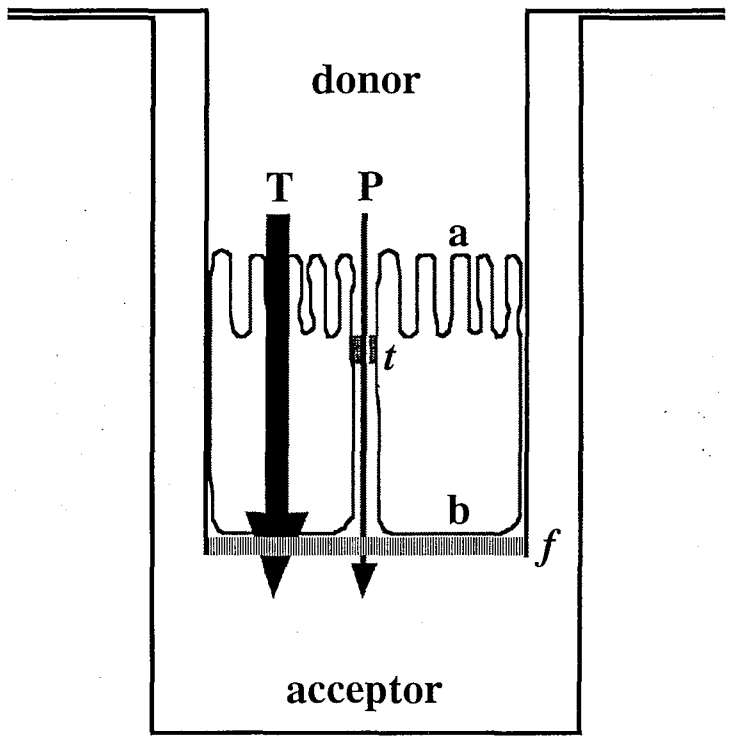


Figure 5.1: Measurement of permeability across Caco-2 cell monolayers. **T** = transcellular route; **P** = paracellular route; ***t*** = tight junction; ***f*** = filter support; **a** = apical surface; **b** = basolateral surface; **donor** = donor compartment; **acceptor** = acceptor compartment.

apical side towards 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 radio-labelled (^{14}C and ^{3}H) compounds, with ultraviolet spectroscopy, with liquid chromatography, or with combinations of liquid chromatography and mass spectroscopy.

Caco-2 cell monolayer experiments are schematically represented in figure 5.1. Passive solute permeation can occur through hydrophilic aqueous pores located at the tight junctions between cells (paracellular route) or through the lipid bilayer of the cell membrane (transcellular route). Assuming that only the neutral species can cross cell membranes, then:

$$P_{cell} = f(P^n_{trans} + P^n_{para}) + (1 - f)P^i_{para} \tag{5.4}$$

where P^n is the intrinsic permeability of the neutral species, P^i the intrinsic permeability of the ionised species, and f the fraction of the neutral species depending on solution pH and solute pK_a.

Compounds with MW < 200 are able to pass the intestinal mucosa by the paracellular route, while compounds with MW of about 350 ± 150 can readily diffuse through the cell membrane without any restriction. Increasing the MW further, approximately MW > 500, leads to a decrease in membrane diffusion. Compounds with $P_{Caco-2} < 10^{-7}$ cm s^{-1} exhibit very poor

oral absorption, compounds with P_{Caco-2} between 1 and 10 $\times$ 10^{-6} cm s^{-1} are moderately absorbed, and compounds with $P_{Caco-2} > 10^{-5}$ cm s^{-1} have excellent oral absorption.

5.3 Previous Simulations

The use of MD simulations to study aspects of solute diffusion through lipid bilayer membranes was applied first by Stouch *et al.*[103,104,108,109] They calculated the diffusion coefficient D from the solute Mean Squared Displacement (MSD), according to equation 4.6. In that equation, $r(t) - r(0)$ is the solute displacement. Since biological membranes have a highly inhomogeneous character along the bilayer normal, in crossing the membrane from one side to the other the permeating particles experience different environments and consequently show different diffusion behaviour (that is different D) at different depths in the membrane. A consequence of this is that the above formula for calculating D is no longer correct when the particles change membrane region, i.e. the above formula fails at $t \rightarrow \infty$ (bilayers cannot be treated as a bulk alkane phase). Stouch *et al.* calculated different D in different regions of the hydrocarbon core of the membrane. The simulation time did not allow for a complete membrane crossing by the permeating particle, but rather the particle motions occured in a limited region for which the local D could be calculated.

Results of simulations with benzene molecules in a all-atom DMPC bilayer[103,104,108] show that the diffusion coefficient calculated from the MSD varies with position in the bilayer. Those benzene molecules that spend a large fraction of their time in the centre of the hydrocarbon region diffuse 2-4 times faster than the benzene molecules that spend more time closer to the headgroup region. This gradient parallels the distribution of free volume inside the lipid bilayer and the time of trans/gauche interconversion along the alkyl chains. Benzene molecules rattle around in a particular void for a relatively long period of time, and infrequently take jumps to another void. Near the centre of the bilayer voids are larger than closer to the headgroup/water interface, so the rate of diffusion is greater; towards the methyl terminal end of the hydrocarbon chains the rates of torsional interconversions are faster and therefore jumps between voids occur more frequently and diffusion is greater.

Stouch also performed the first atomistic simulation of a drug-like molecule in a lipid bilayer.[109] In contrast to what was found for benzene, the rate of diffusion for the drug-like molecule does not vary with location in the bilayer. This is explained on the basis that, owing to its larger size, voids are never available for the drug in any region of the bilayer, so that its

diffusion coefficient is not related to the free volume distribution. An important contribution to the drug behaviour is also found to be the optimization of hydrogen bonding, especially with water.

The same method has subsequently been applied by other authors in the simulation of benzylalcohol[179] and halothane at low[180] and high[181] concentrations. The latter studies were more focused on understanding the modifications induced by high concentrations of the anaesthetic on the lipid bilayer structure.

The calculation of diffusion coefficients from the MSD has been criticised on the basis that it is only valid at long time, but during the simulation the solute can change its location (depth) and move into a new region of the lipid bilayer with different chemical and physical properties, and therefore different diffusive behaviour. The alternative of choosing only those solute molecules that remain in a given location (depth) for a sufficiently long time, introduces an unacceptable bias. These argument have led to the application of new simulation techniques to study diffusion, for example the use of non-equilibrium MD simulations,[139] in which an external force is applied on the solute to induce a larger displacement and avoid problems associated with long-time tailing of correlation functions.

Marrink and Berendsen[3,182] studied the permeation process of water, oxygen and ammonia across a united-atom lipid bilayer applying the solubility-diffusion model shown in equation 5.2. For each permeant, partition coefficients (rewritten in terms of free energy of transfer $\Delta G(z)$ from water into the membrane) and diffusion coefficients $D(z)$ were calculated by constraining the solute at chosen depths z inside the lipid bilayer. The overall resistance profile could therefore be obtained[a].

The free energy profiles of water and ammonia are very similar. In both cases $\Delta G(z)$ increases moving from the water phase towards the interior of the bilayer, since both molecules are polar and leaving the water environment results in them losing the possibility of forming favourable hydrogen bonds. However, the profile shows a dip in the centre of the bilayer and the highest free energy value is located in the upper part of the lipid tails, where the lipid density is highest and the free volume lowest. The free energy profile for oxygen is completely different. Being a hydrophobic penetrant, it prefers to dissolve in the hydrocarbon region than in water, so $\Delta G(z)$ decreases moving towards the centre of the bilayer, and the lowest value is reached exactly halfway between the two bilayer leaflets.

[a]The mathematical treatment for the calculation of the above quantities will be explained in section 5.4.

In contrast, the diffusion coefficient profile for all three compounds is the same. The value of $D(z)$ decreases moving from the water phase into the headgroup region and then increases moving towards the middle of the bilayer, where it reaches its highest value, even higher than in water and very close to the value in alkane solvents.

Combining $\Delta G(z)$ and $D(z)$, the resistance profile can be obtained. For water and ammonia the resistance to permeation is greater in the interior of the membrane than in the bulk water phase, since they are polar molecules, while for oxygen the resistance has an inverse profile, oxygen being more hydrophobic. Moreover, the resistance to permeation is smaller in the centre of the bilayer than in the region of the upper part of the lipid tails, since in the middle of the membrane $\Delta G(z)$ is lower and $D(z)$ is higher than closer to the carbonyl group.

Molecular dynamics simulations were also employed to study solute partitioning between water and simple membrane models by Pohorille *et al.*[183,184] The free energy $\Delta G(z)$ of solute transfer from the water phase into the membrane at depth z is calculated from the solute excess chemical potential $\Delta\mu^{exc}$, which represents the work to bring a solute molecule from the gas to a fixed position in the solvent.The value of $\Delta\mu^{exc}$ is calculated with the Widom particle insertion method,[185,186] which can be applied only to small solutes that are able to find in the solvent voids large enough to accommodate them.

The solutes studied by Pohorille *et al.* were methane and fluoromethanes,[183] which show anaesthetic action. As expected, for all of them the excess chemical potential is lower in the interior of the bilayer than in water phase. However, while the $\Delta\mu^{exc}$ *vs.* z profile decreases monotonically for CH_4 and CF_4, the profiles for mono-, di- and trifluoro methanes exhibit interfacial minima and these compounds reside predominantly in the interfacial region. This distinct behaviour is explained by the presence on the latter molecules of a net dipole moment.

The authors also employed the Umbrella sampling method.[187] Since the free energy difference between different depths in the bilayer may be very large, the simulated permeant particle would soon leave its starting position and rapidly move towards the most favourable region inside the membrane, such that all the bilayer regions would never be sampled sufficiently. Pohorille *et al.* restrained their permeants around chosen positions by adding a biasing potential $U_b(z)$ to the potential energy already defined by the force field employed. The free energy $\Delta G(z)$ of solute transfer from water to depth z in the bilayer is then related to the probability $P(z)$ of finding the solute at that depth.

This method is first applied to study the partition of several anaesthetic compounds, such

as desflurane, isoflurane, nitrous oxide, fluoroalkanes and alcohols.[184] The polar compounds are observed to concentrate at the water/bilayer interface and the free energy profiles show interfacial minima previously observed with fluoromethanes. In contrast, non-polar compouds are predominantly located in the hydrocarbon core of the bilayer. This strongly influences the anaesthetic power of these compounds, with the most polar that concentrate at the interface being the most powerful.

5.4 Z-Constraint Algorithm

In this thesis, the method employed by Berendsen *et al.*[3,182] has been chosen among those reported in the literature to study the permeation process of, first, small organics and, second, real drugs across a DPPC bilayer. This method requires the centre of mass of the permeant particle to be constrained at chosen depths in the membrane. The bilayer normal of the simulation system presented in this thesis is along the z axis, as described in chapter 4, so the centre of mass of the chosen particle has to be constrained at chosen z positions for all the time of the computer simulation. This is not possible within the CHARMM software package and a modification inside the code of the program was required. This section deals with the modifications introduced in the CHARMM program and with the statistical mechanics justifications of why the particle under study has to be constrained.

Free energy cannot be accurately determined from a "standard" MD simulation because such simulations sample the lower energy regions but never adaquetely sample the high energy regions that make important contributions to the free energy calculation. In the case of particle permeation the free energy change does not correspond to chemical mutations, but instead it is a function of the particle position inside the membrane; that is, a function of intermolecular coordinates. The free energy surface along the chosen coordinates is known as a *Potential of Mean Force* (PMF). It is possible to determine directly the derivative of the PMF by measuring the average force exerted on the particle as a function of the chosen intermolecular coordinates. In the case of membrane permeation, one can measure the average force exerted on the permeant molecule at different depths z in the membrane. To sample adequately all the energy regions that contribute to the free energy profile at each depth z, the permeant molecules are "forced" to stay at that z coordinate during the simulation to experience both low (= favourable) and high (= unfavourable) energy states. This is done by *constraining* the permeant particle at the desired z coordinate. The method to calculate the PMF is here referred to as the z-constraint

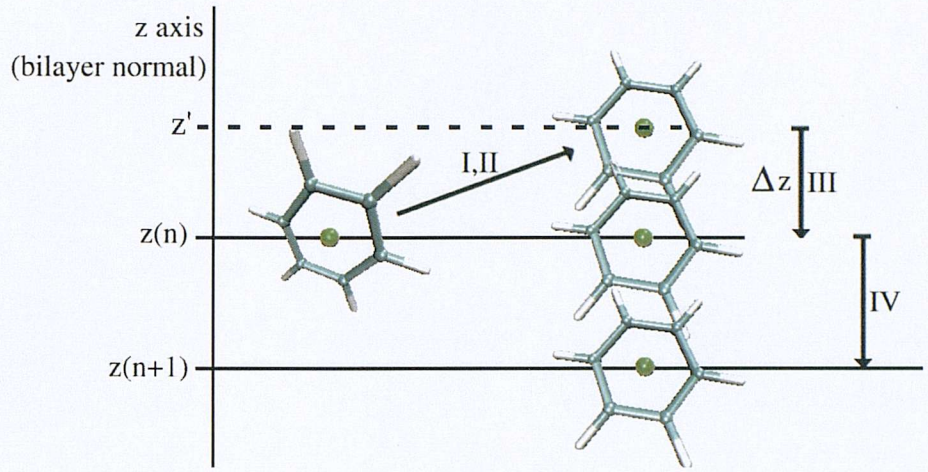


Figure 5.2: The z-constraint algorithm applied, as an example, to a benzene molecule, whose centre of mass is represented with a green sphere. **I:** leap-frog, **II:** SHAKE, **III:** z-constraint, **IV:** other constraints (barostat). Position $z(n)$ is the constrained depth for simulation step n, position $z(n+1)$ is the constrained depth for simulation step n+1. The $\boldsymbol{F}(z)$ value for step n is obtained from $\Delta \boldsymbol{z}$.

method.

5.4.1 Overview

The z-constraint algorithm is represented schematically in figure 5.2.

Step **I**:

At the beginning of the n^{th} simulation step, the solute centre of mass is located at position $z(n)$. The leap-frog algorithm solves the equation of motion without any constraint.

Step **II**:

The SHAKE algorithm is applied to constrain selected covalent bonds. Since satisfying a constraint may violate another one, several iterations are required. At the end, the new position z' of the centre of mass of the permeating particle along the z-axis is calculated.

Step **III**:

Here, the z-constraint is applied. Just one interaction is required: if the distance of the solute from the centre of the lipid bilayer is not identical to that at the beginning of the step, i.e. if $z' \neq z(n)$, the atomic coordinates of the solute are uniformly translated by $\Delta \boldsymbol{z} = z(n) - z'$ to re-set the centre of mass at the desired depth $z(n)$. The permeant is free to rotate and to move on the xy plane, but its centre of mass is constrained at the desired z coordinate.

Step **IV**:

From the SHAKE'd and z-constrained coordinates, new velocities and virial are calcu-

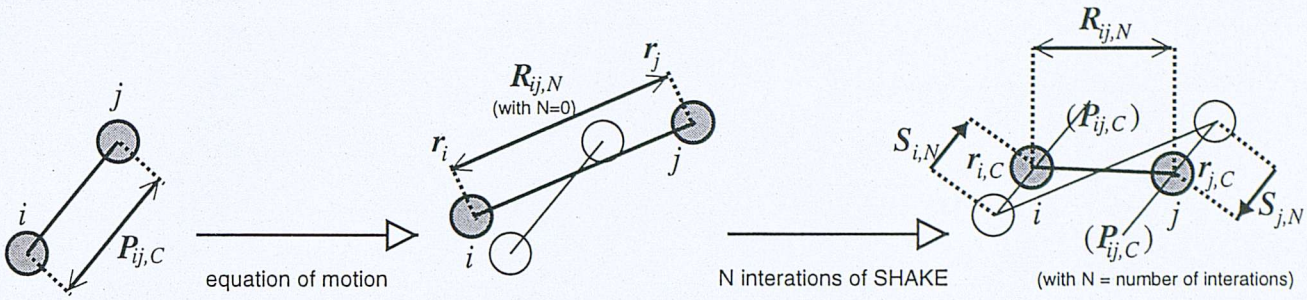


Figure 5.3: The SHAKE algorithm applied, as an example, to a system of two bonded particles i and j. Filled circles are the particles in the current positions and open circles are the previous positions as a reference. The cartesian coordinates $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are those obtained right after solving the equation of motion without any constraint, $\boldsymbol{r}_{i,C}$ and $\boldsymbol{r}_{j,C}$ are those obtained after the application of SHAKE, i.e. $\boldsymbol{r}_{i,C} = \boldsymbol{S}_{i,N} + \boldsymbol{r}_i$

lated. Owing to the presence of the barostat, the coordinates are subsequently moved again along z to maintain a constant P_N with the solute reaching the final position $z(n+1)$. To fully apply the z-constraint on the permeating molecule, after the application of the barostat the centre of mass of the solute should be calculated again and, in case it moved, translated again along the z axis to re-set the z coordinate to $z(n)$. However, since the barostat scales uniformly the coordinates of both solute and lipids, it is thought that it is more appropriate to let the permeant move together with the surrounding lipid phase and not to re-set the z position again. Therefore, the relative rather than the absolute distance between the solute and the bilayer centre of mass was constrained. The coordinate obtained at the end of the MD step after the application of the barostat, i.e. $z(n+1)$, was the value to be satisfied during the application of the z-constraint in the subsequent step, which is n+1.

Some concern may arise from the choice of letting the z-constraint fluctuate because of the pressure coupling. A physical picture can help to understand the reason for this choice. Allowing the barostat to be applied on the constrained permeant maintains the local environment around the soute. If instead the z-constraint is applied in full, by shifting the local environment, the barostat would possibly give an extra contribution to the force acting on the solute and eventually bias the results. In the original publication,[3] Berendsen adopted an identical procedure.

5.4.2 Force Calculation

In dynamics with distance constraints, the total force $\boldsymbol{F}_{i,tot}$ acting on particle i equals the sum of the force derived from the potential energy $-d\mathcal{V}_i/d\boldsymbol{r}_i$ plus the constraint force $\boldsymbol{F}_i$,

i.e. $\boldsymbol{F}_{i,tot} = -d\mathcal{V}_i/d\boldsymbol{r}_i + \boldsymbol{F}_i$.[12] This can be incorporated in an integration algorithm like the leap-frog:[12]

$$\begin{aligned} \boldsymbol{r}_{i,C}(t+\Delta t) &= \boldsymbol{r}_{i,C}(t) + \boldsymbol{v}_i\,(t+\Delta t/2)\,\Delta t = \\ &= \boldsymbol{r}_{i,C}(t) + \Delta t\,[\boldsymbol{v}_i\,(t-\Delta t/2) + \boldsymbol{a}_i(t)\Delta t] = \\ &= \boldsymbol{r}_{i,C}(t) + \boldsymbol{v}_i\,(t-\Delta t/2)\,\Delta t + \boldsymbol{a}_i(t)(\Delta t)^2 = \\ &= \boldsymbol{r}_{i,C}(t) + \boldsymbol{v}_i\,(t-\Delta t/2)\,\Delta t + (\Delta t)^2\boldsymbol{F}_{i,tot}(t)/m_i = \\ &= \boldsymbol{r}_{i,C}(t) + \boldsymbol{v}_i\,(t-\Delta t/2)\,\Delta t + (\Delta t)^2(-d\mathcal{V}_i(t)/d\boldsymbol{r}_i)/m_i + (\Delta t)^2\boldsymbol{F}_i(t)/m_i = \\ &= \boldsymbol{r}_i(t+\Delta t) + (\Delta t)^2\boldsymbol{F}_i(t)/m_i \end{aligned} \tag{5.5}$$

with $(t+\Delta t)$ and (t) the new and the old simulation time steps, respectively. Here $\boldsymbol{r}_i(t+\Delta t)$ is the position that particle i would have assumed when solving the equation of motion without the application of the constraint, and $\boldsymbol{r}_{i,C}(t+\Delta t)$ and $\boldsymbol{r}_{i,C}(t)$ the positions reached after the application of the constraint in the new and old step, respectively. Let us define $\boldsymbol{S}_i(t+\Delta t) = \boldsymbol{r}_{i,C}(t+\Delta t) - \boldsymbol{r}_i(t+\Delta t)$, the shift caused by the constraint on the position $\boldsymbol{r}_i(t+\Delta t)$ otherwise assumed. The constraint force $\boldsymbol{F}_i(t)$ can be derived from the above equation 5.5:

$$\boldsymbol{F}_i(t) = m_i\frac{\boldsymbol{S}_i(t+\Delta t)}{(\Delta t)^2} \tag{5.6}$$

An identical result is obtained using different algorithms to solve the equation of motion.[12] When the constraint is the cartesian distance between particles i and j, as in SHAKE, $\boldsymbol{S}_j(t+\Delta t)$ on particle j is such that $\boldsymbol{F}_j(t) = -\boldsymbol{F}_i(t)$.

In constraint dynamics, the constraint force $\boldsymbol{F}_i(t)$ acting on i can be written as the derivative of the constraint with respect to the coordinates of i, multiplied by a multiplier λ which is to be determined.[11,12,20] When applying SHAKE,[11,12,20,188] the constraint is the cartesian distance between particle i and particle j, i.e. $[\boldsymbol{r}_{i,C}(t) - \boldsymbol{r}_{j,C}(t)]^2 - D_{ij}^2 = 0$, with D_{ij} the desired constraint distance. Therefore:

$$\begin{aligned} \boldsymbol{F}_i(t) &= \lambda\frac{\partial\left\{[\boldsymbol{r}_{i,C}(t) - \boldsymbol{r}_{j,C}(t)]^2 - D_{ij}^2\right\}}{\partial\boldsymbol{r}_{i,C}} = \\ &= 2\lambda\,[\boldsymbol{r}_{i,C}(t) - \boldsymbol{r}_{j,C}(t)] = \overset{*}{\lambda}\,[\boldsymbol{r}_{i,C}(t) - \boldsymbol{r}_{j,C}(t)] = \overset{*}{\lambda}\boldsymbol{P}_{ij,C} \end{aligned} \tag{5.7}$$

Actually, $\overset{*}{\lambda}$ can be used, which also includes the factor 2 arising from the derivative. Here $\boldsymbol{P}_{ij,C} = \boldsymbol{r}_{i,C}(t) - \boldsymbol{r}_{j,C}(t)$ is the i-j vector obtained at the end of time step t after the application

of the constraint. In SHAKE[11,12,20,188] it is more convenient to introduce a different multiplier $g = \overset{*}{\lambda}\Delta t^2$. Consequently:

$$\boldsymbol{F}_i(t) = \frac{g}{\Delta t^2}\boldsymbol{P}_{ij,C} \qquad \text{and} \qquad \boldsymbol{F}_j(t) = -\frac{g}{\Delta t^2}\boldsymbol{P}_{ij,C} \tag{5.8}$$

and including equation 5.6:

$$\boldsymbol{S}_i(t+\Delta t) = \frac{g\boldsymbol{P}_{ij,C}}{m_i} \qquad \text{and} \qquad \boldsymbol{S}_j(t+\Delta t) = -\frac{g\boldsymbol{P}_{ij,C}}{m_j} \tag{5.9}$$

We can now define:

$$\boldsymbol{S}_{tot}(t+\Delta t) = \boldsymbol{S}_i(t+\Delta t) - \boldsymbol{S}_j(t+\Delta t) = g\boldsymbol{P}_{ij,C}\left(\frac{1}{m_i} + \frac{1}{m_j}\right) \tag{5.10}$$

In SHAKE,[11,12,20,188] $\boldsymbol{S}_i(t+\Delta t)$ and $\boldsymbol{S}_j(t+\Delta t)$ are calculated in an iterative fashion in order to satisfy the distance constraint:

$$D_{ij}^2 - \boldsymbol{R}_{ij,N}^2 = 0 \pm D_{ij}^2 T^2 \tag{5.11}$$

with $\boldsymbol{R}_{ij,N} = \boldsymbol{R}_{ij,N-1} + \boldsymbol{S}_{ij,N}$. Here $\boldsymbol{S}_{ij,N}$ is $\boldsymbol{S}_{tot}(t+\Delta t)$ at SHAKE iteration number N, $\boldsymbol{R}_{ij,N-1}$ is the i-j vector obtained at the end of the previous SHAKE iteration (iteration N-1), and $\boldsymbol{R}_{ij,N}$ is the i-j vector at the current SHAKE iteration (the *Nth* iteration). Let us also indicate $\boldsymbol{S}_i(t+\Delta t)$ at the *Nth* SHAKE iteration with $\boldsymbol{S}_{i,N}$. The equality in equation 5.11 has to be satisfied within a certain tolerance T chosen by the user. Right after solving the equation of motion (e.g. the leap-frog algorithm), $\boldsymbol{r}_i(t+\Delta t) - \boldsymbol{r}_j(t+\Delta t) = \boldsymbol{R}_{ij,N}$ is obtained, with $N = 0$, and $\boldsymbol{S}_{i,N}$ is not calculated. Equation 5.11 is tested with this value of $\boldsymbol{R}_{ij,N}$. If not satisfied,

then $N = 1$ and $\boldsymbol{S}_{i,N}$ is calculated from equation 5.11 (setting $T = 0$) as follows[b]:

$$D_{ij}^2 - (\boldsymbol{R}_{ij,N-1} + \boldsymbol{S}_{ij,N})^2 = 0$$

$$D_{ij}^2 - \boldsymbol{R}_{ij,N-1}^2 - 2\boldsymbol{R}_{ij,N-1}\boldsymbol{S}_{ij,N} = 0$$

$$\boldsymbol{S}_{ij,N} = \frac{D_{ij}^2 - \boldsymbol{R}_{ij,N-1}^2}{2\boldsymbol{R}_{ij,N-1}}$$

$$g\boldsymbol{P}_{ij,C}\left(\frac{1}{m_i} + \frac{1}{m_j}\right) = \frac{D_{ij}^2 - \boldsymbol{R}_{ij,N-1}^2}{2\boldsymbol{R}_{ij,N-1}} \tag{5.12}$$

$$g = \frac{D_{ij}^2 - \boldsymbol{R}_{ij,N-1}^2}{2\left(\boldsymbol{R}_{ij,N-1} \cdot \boldsymbol{P}_{ij,C}\right)\left(\frac{1}{m_i} + \frac{1}{m_j}\right)}$$

$$\boldsymbol{S}_{i,N} = \frac{D_{ij}^2 - \boldsymbol{R}_{ij,N-1}^2}{2m_i L_{ij}\left(\frac{1}{m_i} + \frac{1}{m_j}\right)}\boldsymbol{P}_{ij,C}$$

with L_{ij} the scalar product between $\boldsymbol{R}_{ij,N-1}$ and $\boldsymbol{P}_{ij,C}$. The vector $\boldsymbol{S}_{i,N}$ is thus along the direction of $\boldsymbol{P}_{ij,C}$, because $\boldsymbol{F}_i(t)$ acts on that vector. At this point, the equality to zero in equation 5.11 is checked again and, if not satisfied, N is incremented by one and $\boldsymbol{S}_{i,N}$ is calculated again. This procedure is repeated until equation 5.11 is satisfied. At this point one has $\boldsymbol{R}_{ij,N} = \boldsymbol{r}_{i,C}(t + \Delta t) - \boldsymbol{r}_{j,C}(t + \Delta t)$ with $N =$ total number of iterations. The SHAKE procedure is schematically shown in figure 5.3.

In contrast to SHAKE, which is applied to three-dimensional vectors, in the z-constraint D_{ij} is the constraint distance along the z axis only from the centre of mass of the permeant and that of the lipid bilayer, and $\boldsymbol{S}_{i,N}$ is the amount by which the z coordinate of the centre of mass of permeant i has to be re-set. The centre of mass of the lipid bilayer is not moved, $\boldsymbol{F}_i(t)$ is only applied on the solute i. This is like the bilayer had infinite mass. This is true if one remembers that, because of periodic boundary conditions, what is actually simulated is an infinite bilayer. After the equation of motion, if the distance between solute and bilayer centres of mass is not identical to D_{ij}, the solute is translated along z by Δz (see again figure 5.2). After the translation (one iteration) , D_{ij} and $\boldsymbol{R}_{ij,N}$ (with $N = 1$) coincide. Therefore, $S_i(t + \Delta t) = \boldsymbol{S}_{i,N}$ with $N = 1$, and also $S_i(t + \Delta t) = \Delta z$ with $\Delta z = z' - z(n)$ in figure 5.2.

[b]The term $\boldsymbol{S}_{ij,N}^2$ in equation 5.12 is neglected for computational efficiency. The iterative nature of the procedure assures that the equation is finally solved within the specified tolerance.

Consequently, the z constraint force is:

$$\boldsymbol{F}_i(t) = m_i \frac{\Delta \boldsymbol{z}}{(\Delta t)^2} \tag{5.13}$$

5.4.3 Positions and Velocities

Having applied SHAKE or the z constraint, the velocities must also be modified. In what follows, $\boldsymbol{r}_i(t + \Delta t)$ is the position of the particle after solving the equation of motion and $\boldsymbol{r}_{i,C}(t + \Delta t)$ is the position obtained after applying the constraint. In analogy, $\boldsymbol{v}_i(t + \Delta t)$ and $\boldsymbol{v}_{i,C}(t + \Delta t)$ are the particle's velocities before and after the application of the constraint.

When using SHAKE, as described with equation 5.5:

$$\boldsymbol{r}_{i,C}(t + \Delta t) = \boldsymbol{r}_i(t + \Delta t) + \boldsymbol{S}_i(t + \Delta t) \tag{5.14}$$

In the z-constraint, the z coordinate only must be modified:

$$\boldsymbol{r}_{z,i,C}(t + \Delta t) = \boldsymbol{r}_{z,i}(t + \Delta t) + \Delta \boldsymbol{z} \tag{5.15}$$

When using a constraint in MD simulations, not only must the distance constraint be satisfied, that is $r_{ij}^2 =$ constant, but also its first derivative must be zero, that is:[11]

$$\frac{d(\boldsymbol{r}_{ij}^2)}{dt} = 2\boldsymbol{r}_{ij}\boldsymbol{v}_{ij} = 0 \tag{5.16}$$

Thus the relative velocities of the two bonded atoms along the bond vector must be zero. By analogy with equation 5.5, when the leap-frog algorithm is employed, as in these simulations, the velocities to be corrected are:

$$\begin{aligned}
\boldsymbol{v}_{i,C}\,(t + \Delta t/2) &= \boldsymbol{v}_{i,C}\,(t - \Delta t/2) + \Delta t \boldsymbol{a}(t) \\
&= \boldsymbol{v}_{i,C}\,(t - \Delta t/2) + \Delta t \boldsymbol{F}_{i,tot}(t)/m_i \\
&= \boldsymbol{v}_{i,C}\,(t - \Delta t/2) + \Delta t(-d\mathscr{V}_i(t)/d\boldsymbol{r}_i)/m_i + \Delta t \boldsymbol{F}_i(t)/m_i \\
&= \boldsymbol{v}_i\,(t + \Delta t/2) + \Delta t \boldsymbol{F}_i(t)/m_i \\
&= \boldsymbol{v}_i\,(t + \Delta t/2) + \boldsymbol{S}_i(t + \Delta t)/\Delta t
\end{aligned} \tag{5.17}$$

with $\boldsymbol{F}_i(t)$ the constraint force and $\boldsymbol{F}_i(t)/m_i$ the acceleration induced by $\boldsymbol{F}_i(t)$. In the z-constraint, the z-component only must be corrected:

$$\boldsymbol{v}_{z,i,C}\,(t + \Delta t/2) = \boldsymbol{v}_{z,i}\,(t + \Delta t/2) + \Delta \boldsymbol{z}/\Delta t \tag{5.18}$$

This correction ensures that the velocity of the solute centre of mass along z is zero, relatively to the bilayer centre of mass. From the corrected positions and velocities, the virial and kinetic energy are eventually calculated.

5.5 Summary

Before performing permeability studies, it is important to know with what type of experiments the simulation results can be compared. Experimental solute partition coefficients in water/membrane systems do not probe the membrane region offering the main contribution to permeation resistance, and solute diffusion coefficients inside the membrane are not directly accessible by experiment. On the other hand, permeability coefficients have been recently measured carefully. Some concern may arise when comparing drug permeability calculated across a DPPC bilayer computer model with that measured across cell monolayers.

From this chapter it is clear that the simulations can simply focus on the neutral form of the permeant molecule, because ions do not cross biological membranes at a significant rate. Available experimental data are also corrected to consider the intrinsic permeability coefficient of the neutral form.

Among the various approaches found in the literature, the application of the z constraint method to this problem seems the most complete. In contrast to the MSD calculation, it avoids the problems associated with the heterogeneous nature of the lipid bilayer, and in contrast with the particle insertion method, it can be applied to large solutes (e.g. drug molecules) and dense solvents (e.g. lipid bilayers). As explained in the next chapter, one only computation is required, namely the value of Δz at each simulation step, to allow the calculation of both free energies and diffusion coefficients at different depths in the lipid bilayer.

CHAPTER 6

Computer Simulation Study of Small Molecule Permeation

The aim of this chapter is to report the use of computer simulation techniques to study small molecule permeation through a phospholipid bilayer at an atomic detail which is not possible with experiments. The *macroscopic* property of selected organics, namely the permeability coefficient, as extrapolated from the *microscopic* size of a computer simulation system is compared with the available experimental data. Where the calculated and the measured values agree, the atomistic simulation can help in elucidating the individual contributions of different membrane regions to the overall resistance to permeation towards different chemical entities.

6.1 Solute Candidates

For the aim of these simulations, suitable solutes were chosen to represent the most common functional groups. Although they differed in size, they were reasonably small, and although fully flexible all-atom models were used, they did not have large internal degrees of freedom. This choice simplified the simulation technical issues. All of the solutes had tested parameters within the CHARMM force field, consistent with those employed for the lipid and water molecules. Experimental permeability coefficients across lipid bilayers are available for many simple organic acids, a very few amines, some amides, polyalcohols and complex aromatic compounds. The

final list of molecules chosen for these simulations was: acetamide, acetic acid, benzene, ethane, methanol, methylacetate, methylamine, water.

6.2 Simulation Protocol

In the simulations described here, the method for the computation of permeability coefficients employed by Marrink and Berendsen[3,182] to study the permeation process of water, ammonia and oxygen has been used for the calculation of permeability coefficients of the small organic molecules listed above. However, the force field, the program, the simulation protocol and the number of molecules in these simulations differ from those used in the original publication. The simulation protocol is identical to that employed for the pure lipid bilayer simulations described in chapter 4.

Briefly, the simulation box contained 72 dipalmitoylphosphatidylcholine (DPPC) molecules arranged in a 2×36 bilayer, together with 2094 water molecules (full hydration). Lipids and water were modelled using version 27 of the CHARMM force field for lipids.[100] An equilibrated starting structure of the lipid bilayer was kindly obtained from A. D. MacKerell and S. E. Feller, who participated in developing the force field. The simulation protocol was the same as that used in some of the latest Feller's simulations.[91,100] The LJ potential was switched smoothly to zero over the region from 10 and 12 Å. Electrostatic interactions were calculated via the Particle Mesh Ewald (PME) method using a κ value of 0.23 and a fast-Fourier grid density of ca. 1 $Å^{-1}$. The real space part of the PME summation was truncated at 12 Å. The SHAKE algorithm[20] was used to constrain all covalent bonds involving hydrogens. The leap-frog algorithm[14] was employed to solve the equation of motion with a time step of 2 fs. A neighbour list, used for calculating the LJ potential and the real space portion of the PME, was kept to 14 Å and updated every 50 fs. Coordinates were saved every ps for subsequent analysis. Three-dimensional periodic boundary conditions were applied. Only the cell length normal to the membrane (L_z) was allowed to vary during the simulation to maintain a constant normal pressure (P_N) of 1 atm. The other cell dimensions (L_x and L_y) were kept fixed to maintain a constant surface area per lipid (A) of 62.9 $Å^2$. The pressure was maintained by the Langevin Piston algorithm[19] with a mass of 500 amu and a collision frequency of 5 ps. The temperature (T) was maintained at 50 °C, well above the phase transition temperature of DPPC bilayers, by means of the Hoover thermostat.[17] A value of 1000 kcal ps^2 was used for the thermostat

(fictitious) mass. The ensemble was therefore NP_NAT.

Topologies and parameters for the small organic solutes were available from the latest CHARMM force field for lipids and proteins. The molecules were manually inserted in the lipid bilayer at the desired z depth and a crude short minimization was performed to eliminate bad contacts. The steepest descent algorithm was employed for the minimizations.

All the simulations were run in parallel with 4 processors, using version 27 of the CHARMM software package.[13] A few simulations were run on a cluster of PCs with 750 Mhz AMD Athlon processors, but most were run on a cluster of PCs with 1000 Mhz Intel Pentium IIIs and 1500 Mhz Intel Pentium IVs .

6.3 Calculation of Permeability

The z-constraint, as introduced by Marrink and Berendsen[3,182] and implemented in CHARMM as described in chapter 5, was employed in these simulations to constrain different solutes at different depths (i.e. z positions) in the membrane, and values of the constraint force $\boldsymbol{F}_i$ (see equation 5.13) were collected during the simulation for each z. With one single computation ($\boldsymbol{F}_i$), the z-constraint allows for the calculation of both equilibrium and dynamic properties.

The equilibrium property is the solute partition behaviour between water and different regions in the membrane and is expressed by the free energy of solute transfer from water into various depths z of the membrane, i.e. $\Delta G(z)$. The force acting on the centre of mass of the solute at a chosen z depth is obtained at each time step as the negative of the force required to maintain the z-constraint and is indicated as $\boldsymbol{F}(z,t)$. Eventually, the free energy difference $\Delta G(z)$ between depths z_1 and z_2 is accessible as the potential of the mean force [a]:

$$\Delta G(z) = -\int_{z_1}^{z_2} \langle \boldsymbol{F}(z) \rangle_t dz \tag{6.1}$$

Here $\langle ... \rangle_t$ indicates average over the simulation time. When z_1 and z_2 are taken as the two sides of the lipid bilayer, the free energy profile for the entire membrane thickness is obtained.

[a] Rigorously speaking, it has been shown that in general cases the derivative of the free energy is *not* equal to the (negative of the) mean constraint force, but rather the correct formula, even though it includes the mean constraint force, is much more complicated.[189–195] A correction is needed to remove the bias generated by the constrained molecular dynamics trajectory, due to the fact that the time derivative of the constrained reaction coordinate is set to zero. The application of the constraint affects the phase space distribution of the system and makes the momentum distribution coordinate dependent. The formula simplifies considerably in the special case where the reaction coordinate to be constrained is the magnitude of the internuclear separation between two particles[189–195] or the distance between a particle and a plane. No correction terms are needed and the mean constraint force can be used directly. The formula reported above in equation 6.1 is based on this consideration. Despite its appeal, this simple approach is an exception and it is not valid when reaction coordinates other than geometrical distances are to be constrained.

The solute $\Delta G(z)$ is related to its partition coefficient $K(z)$:

$$K(z) = \exp(-\Delta G(z)/RT) \tag{6.2}$$

The dynamic property is the solute diffusion coefficient at different z depths in the membrane. Studying solute diffusion in a medium,[196,197] it can be demonstrated that the local time-dependent friction coefficient of the diffusant molecule $\xi(t)$ is related to the time autocorrelation function of the fluctuations of the instantaneous $\boldsymbol{F}(z,t)$ from the mean $\langle \boldsymbol{F}(z) \rangle_t$ with the following equation:

$$\xi(t) = \frac{\langle \Delta \boldsymbol{F}(z,t) \cdot \Delta \boldsymbol{F}(z,0) \rangle}{RT} \tag{6.3}$$

Here R is the gas constant, T the absolute temperature, and:

$$\Delta \boldsymbol{F}(z,t) = \boldsymbol{F}(z,t) - \langle \boldsymbol{F}(z) \rangle_t \tag{6.4}$$

Assuming that $\xi(t)$ is large and decays rapidly compared to other time scales in the system, a satisfactory description of the full dynamics is provided by the static friction coefficient $\tilde{\xi}$:

$$\tilde{\xi} = \int_0^\infty \xi(t) dt = \int_0^\infty \frac{\langle \Delta \boldsymbol{F}(z,t) \cdot \Delta \boldsymbol{F}(z,0) \rangle}{RT} dt \tag{6.5}$$

When studying diffusion across a free energy barrier,[196] the above condition is met if the slope of the free energy barrier over a distance covered by the particle during the decay time of its friction coefficient is lower than the thermal fluctuation, i.e. RT per mole. That is, the condition is met if the particle remains in a region of constant free energy. In this case, $\tilde{\xi}$ is related to the local diffusion coefficient $D(z)$ of the permeating solute at depth z:

$$D(z) = \frac{RT}{\tilde{\xi}} = \frac{(RT)^2}{\int_0^\infty \langle \Delta \boldsymbol{F}(z,t) \cdot \Delta \boldsymbol{F}(z,0) \rangle dt} \tag{6.6}$$

Eventually, from $\Delta G(z)$ and $D(z)$ the overall resistance R to permeation is obtained according to the solubility-diffusion model integrating over the local resistances $\mathfrak{R}(z)$ at different depths in the membrane:

$$R = \int_{z_1}^{z_2} \mathfrak{R}(z) dz = \int_{z_1}^{z_2} \frac{\exp(\Delta G(z)/RT)}{D(z)} dz = \frac{1}{P} \tag{6.7}$$

and the permeability coefficient P of the solute is defined as the inverse of R.

In this study, $\Delta G(z)$, $D(z)$ and P were calculated for water and other seven small solutes representing the most common chemical functional groups: acetic acid, acetamide, methylamine, methanol, ethane, methylacetate, benzene. For each of the eight molecules simulated, ten depths (z coordinates) were sampled from 0 to 30.5 Å from the bilayer centre, constraining the solutes at those z-depths. Results from one leaflet were considered valid for the other layer too, for reasons of symmetry. For each z depth, 5 xy positions on the bilayer plane were sampled, this giving a total of 10 x 5 = 50 different positions sampled for each molecule. In a single simulation only 5 of the 10 z depths were studied, this leading to 50 / 5 = 10 simulations for each molecule. In other words, each single simulation contained 5 solutes whose centre of mass was constrained at 5 different depths. Solutes were at least 6 Å apart from each other along the bilayer normal, but in most cases they were more than 12 Å apart. Moreover, they were not in the same xy position, but rather at five different positions on the bilayer plane. That is, only one solute was present at each z, and different xy positions for the same z were sampled in separate simulations. This care in choosing starting positions was expected to avoid strong solute-solute interactions and clustering. In each simulation the 5 solutes were identical. They were inserted in the membrane with their centre of mass at the desired z position and the system was minimized to eliminate bad contacts. Finally, 2.1 ns MD simulations were performed and the first 100 ps were discarded as equilibration. Therefore, for each molecule, P was calculated from 20 ns of simulations and for the eight molecules a total of 160 ns of data collection was performed.

6.4 Four Region Model

Since the membrane has a very inhomogeneous character when moving from one side to the other, each individual layer has been split into four regions as described in section 4.2.1:

Region 1: low headgroup density, 20 to 27 Å from the bilayer centre.

Region 2: high headgroup density, 13 to 20 Å from the bilayer centre.

Region 3: high tail density, 6 to 13 Å from the bilayer centre.

Region 4: low tail density, 0 to 6 Å from the bilayer centre.

The same distinction between membrane regions has been adopted here in the interpretation of the permeation results. In what follows, unless indicated otherwise, the x-axis of all plots are

along the bilayer normal, i.e. they report the z depths at which the molecules were constrained. The four regions into which each leaflet can be divided are also separated by vertical lines. Each lipid monolayer thickness is about 27 Å and further from the bilayer centre there is bulk water.

6.5 Free Energies

Free energy profiles $\Delta G(z)$ for the permeants studied, calculated with equation 6.1, are plotted in figure 6.1. Errors in $\Delta G(z)$ are calculated propagating the errors in the force $\boldsymbol{F}(z)$. The latter are standard errors calculated from the difference of the mean force $\langle \boldsymbol{F}(z) \rangle_t$ in each of the five individual simulations at each depth from their average.

6.5.1 Water

To validate the methodology, water permeation was studied first and results from these simulations were compared with the original publication by Marrink and Berendsen.[3] The free energy profiles have an identical shape: $\Delta G(z)$ increases smoothly in region 1 and then very rapidly in regions 2 and 3, whereas region 4 present a small dip of 1.5 kJ mol^{-1} in the literature, and 1.8 kJ mol^{-1} in these simulations. The increase in $\Delta G(z)$ is due to the loss of favourable electrostatic interactions and hydrogen bonds when leaving the bulk water phase for the hydrophobic core of the bilayer. In contrast, in the middle of the bilayer $\Delta G(z)$ decreases because of the lower local density.

However, $\Delta G(z)$ values in these simulations are always about 3 kJ mol^{-1} lower than those reported by Marrink and Berendsen, with the exception of region 1. The highest value of $\Delta G(z)$ is reached in both simulations in region 4, but it is 26 $\pm$ 2 kJ mol^{-1} in the literature[3] and 22.9 $\pm$ 1.9 kJ mol^{-1} in these simulations. Since the bilayer core has the nature of long chain alkanes, Marrink and Berendsen compare this value with the experimental ΔG^0(water$\rightarrow$hexadecane) value, which is 24.9 kJ mol^{-1}.[1] Values from these simulations are in fair agreement, although a little lower.

6.5.2 General Trend

Regarding the other solutes studied here, a general trend can be highlighted from figure 6.1. For all the compounds $\Delta G(z)$ increases moving from the water phase into the membrane, crossing the headgroup regions 1 and 2. This is thought to be due to the increase in density, which makes

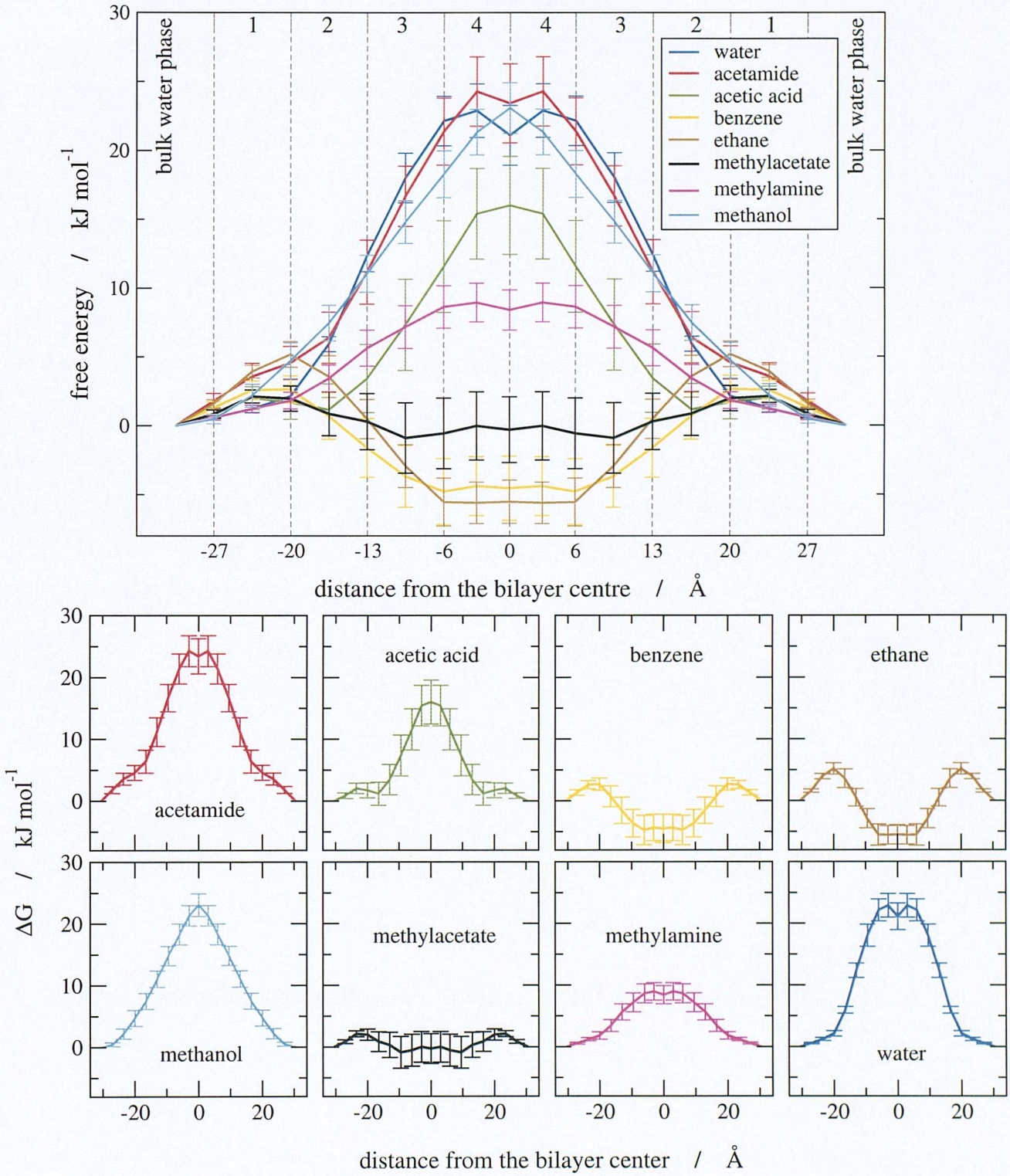


Figure 6.1: Free energy profiles. Top: bilayer is divided into the four regions. Bottom: for clarity, each profile is plotted alone. Error bars are standard errors calculated from the difference of the force in the five individual simulations from their average.

the solubilization of a solute more difficult because of the lower free volume available to locate permeant molecules. Another reason is that water molecules are strongly hydrogen-bonded to the lipid headgroups and they must be displaced to allow solute permeation. The free energy then continues to increase in regions 3 and 4 for hydrophilic compounds, while it decreases for hydrophobic compounds. For the former, the hydrophobic core of the lipid tails represents the main barrier to permeation. For the latter, partition is more favoured in the middle of the bilayer than at the lipid/water interface. Among the hydrophilic compounds, acetamide and methylamine show a small dip in $\Delta G(z)$ in the bilayer centre as well as water, whereas acetic acid and methanol do not. Even though methylacetate is a hydrophilic and polar compound (its ΔG(water→hexadecane) is positive), for the purpose of this analysis it is included among the hydrophobic compounds, since its $\Delta G(z)$ in the middle of the bilayer is lower than at the water/lipid interface. The explanation for this effect may reside in the fact that, in contrast to the other hydrophilic solutes, methylacetate does not posses any hydrogen bond donors, or in other words it has no polar hydrogens. The lipid headgroups do not posses any either. Therefore, methylacetate is unique among the hydrophilic solutes in the sense that it can form no H-bonds with the surrounding lipids.

It is difficult to correlate the free energy curves with the size or with the hydrophobicity/hydrophilicity of the compounds when the two descriptors are both involved. Useful comparisons can be made between two or more compounds when the size of the permeants is the same and only the chemical groups differ. In this case, it can be assumed that the volume displaced, and hence the work required to locate the solutes under study into the membrane at a given depth, is the same, and the difference in free energy is only due to the different chemical affinity between the solute molecule and the surrounding phase.

6.5.3 Acetamide and Acetic Acid

A comparison can be made between acetic acid and acetamide, which differ in the presence or the absence of a hydrogen atom, while the van der Waals radii of the other atoms are identical or very similar. $\Delta G(z)$ for acetamide is always higher than that for acetic acid. Considering that in equation 6.7 $\Delta G(z)$ contributes to increase the resistance and decrease the permeability, this plot predicts that acetamide's permeability is lower than that of acetic acid. Experimentally, this is indeed true. Acetamide's permeability is reported to be 1 to 2 orders of magnitude lower than that of acetic acid.[127,128,135,136] The difference between the two curves resides

primarily in region 1, where acetic acid partition seems more favored than that of acetamide. A possible explanation may be due to the fact that the carboxylic group can form more favourable interactions with the highly charged phospholipid headgroups than can the amide group. In fact, as will be shown in section 6.9.5, acetic acid forms longer-lived hydrogen bonds with the surrounding lipids. Deeper in the membrane, the two curves have the same gradient and both reach their maximum value in region 4: 24.3 $\pm$ 2.5 kJ mol^{-1} for acetamide and 16.0 $\pm$ 3.6 kJ mol^{-1} for acetic acid.

Owing to the hydrocarbon nature of the middle of the membrane, permeabilty is often correlated with the solute partition coefficient K in long chain alkanes. Considering the case of acetic acid and acetamide, it is reported[2,118,120,121] that permeabilities P_n (see equation 5.1) measured through egg lecithins and K in hexadecane/water systems are highly (correlation coefficient > 0.95) and linearly (slope ≈ 1) correlated. Hexadecane is chosen since its chain length is comparable to that of the fatty acid residues in most lipids and it is conveniently available. Reported K(hexadecane/water) values are 5.3×10^{-4} for acetic acid and 2.1×10^{-5} for acetamide. These in turn mean ΔG(water$\rightarrow$hexadecane) = 18.6 kJ mol^{-1} for acetic acid and ΔG(water$\rightarrow$hexadecane) = 26.6 kJ mol^{-1} for acetamide. Other authors[135,136] more recently have found that, for egg lecithins, hexadecene and 1,9-decadiene yield better correlations, since the presence of a double bond in the reference organic solvent more closely reproduces the broad distribution of unsaturation in egg lecithins. Reported K(hexadecene/water) values are 7.8×10^{-4} for acetic acid and 3.8×10^{-5} for acetamide, i.e. ΔG(water$\rightarrow$hexadecene) = 17.7 kJ mol^{-1} for acetic acid and ΔG(water$\rightarrow$hexadecene) = 25.2 kJ mol^{-1} for acetamide. Values from the two reference solvents are not very different and indeed in fair agreement with the highest values of $\Delta G(z)$ calculated from these simulations for these two solutes. Moreover, the simulations also reproduce the qualitative trend whereby acetic acid partition is higher than that of acetamide. However, $\Delta G(z)$s from the simulations are slightly lower than the experimental ΔG(water$\rightarrow$reference solvent) and this would predict a slightly higher permeability coefficient from simulations than from experiments. This is the case, as will be shown later in the discussion.

Experimental acetic acid partition coefficients in water/lipid bilayer systems K(membrane$\rightarrow$water) exist. It must be remembered that the zone of maximum partition accounts for only a very small fraction of the resistance to permeation and therefore K(membrane$\rightarrow$water) is not the

partition coefficient in the barrier region of the membrane. Acetic acid partition coefficients between cholesterol-free DMPC vesicles and water are reported to be around 1.5 for a wide range of temperatures in the liquid-crystalline phase.[127] This yields in turn ΔG(water→membrane) $\approx$ -0.45 kJ mol^{-1}. This is thought to correspond to acetic acid partition into the polar and highly charged headgroup region. In these simulations, acetic acid $\Delta G(z)$ is never negative. However, in regions 1 and 2 $\Delta G(z)$ is between 1 and 2 kJ mol^{-1} and, considering the error bars, $\Delta G(z)$ reaches the value of -0.7 kJ mol^{-1} in region 2. This result suggests that the simulations can reproduce the experimental observation and also confirm that acetic acid tends to partition into the headgroup region of the lipid bilayer as previously argued.

6.5.4 Ethane, Methylamine and Methanol

Comparisons can also be made between ethane, methanol and methylamine, whose molecular sizes are very similar.

As expected for alkanes,[198,199] ethane partition is favoured in the hydrophobic core of the bilayer and the free energy decreases moving from the headgroup regions 1 and 2 towards the hydrocarbon chain regions 3 and 4. The partitioning of ethane into the membrane is expected to be higher than in water, due to the hydrocarbon nature of the lipid chains. The headgroup region of the bilayer, however, is highly charged, dense and rich in water-lipid H-bond networks. Ethane partitioning in this region is then expected to be lower than in water and this is the reason why the free energy increases entering the water/lipid interface. In the literature, ethane ΔG^0(water→hexadecane) is -4.6 kJ mol^{-1},[1] while $\Delta G(z)$ in region 4 from these simulations is -5.6 $\pm$ 1.5 kJ. The calculated value is a bit lower than the experimental ΔG^0(water→hexadecane), as in the case of acetamide and acetic acid. Pohorille *et al.*[184] calculated $\Delta G(z)$ for methane along the normal of a bilayer made up of glycerol 1-monooleate molecules, which, with respect to a phospholipid, miss a second hydrocarbon chain and, above all, the highly charged phosphocholine group. Their profile was therefore negative at all depths and did not show the positive value across regions 1 and 2 as in these simulations. The lowest value of $\Delta G(z)$ in that publication was about -6.3 kJ mol^{-1}. Considering the differences in solute and membrane model, the agreement is satisfactory.

Profiles for methanol and methylamine are very close in regions 1 and 2. Owing to the increase in solvent density moving from the water phase towards the lipid bilayer, $\Delta G(z)$

for the two polar solutes and also for ethane increases in the same direction. Methanol and methylamine $\Delta G(z)$ are however lower than ethane $\Delta G(z)$ in region 1 and at the beginning of region 2, because of the polar and hydrophilic nature of the former solutes. Methanol $\Delta G(z)$ is always higher than methylamine $\Delta G(z)$, but this difference is only significant in regions 3 and 4. From these simulations methanol permeability is predicted to be lower than that of methylamine, owing to the lower partitioning of the alcohol into the hydrophobic core of the membrane. The experimental permeability for methanol is not available, and similarly for many simple alcohols such as ethanol, propanol and butanol, since these compounds can dissolve lipid membranes. However, experimental ΔG^0(water→hexadecane) is reported to be 15.9 kJ mol^{-1} for methanol and 14.0 kJ mol^{-1} for methylamine.[1] From these simulations, the maximum value of ΔG(water→membrane) is 23.1 ± 1.8 kJ mol^{-1} for methanol and 9.0 ± 1.4 kJ mol^{-1} for methylamine, both reached in region 4. These results are not in quantitative agreement with experiment, and for methylamine ΔG(water→membrane) is again lower than ΔG^0(water→hexadecane) as for ethane, acetamide and acetic acid. The results do however correctly rank the compounds.

The experimental free energy of methanol transfer from water into DMPC liposomes in the liquid-crystalline phase is reported to be 0.76 kcal mol^{-1}, i.e. 3.18 kJ mol^{-1}.[200] This value is in agreement with those obtained in these simulations for region 1.

6.5.5 Methylacetate and Benzene

The largest compounds studied here are methylacetate and benzene.

Despite the fact that methylacetate is much less hydrophilic than its analogue acetic acid, $\Delta G(z)$ for both compounds in region 1 and at the beginning of region 2 are almost the same. The two profiles largely differ in regions 3 and 4, where the more hydrophobic ester has a more favorable partition, despite its larger size requiring a larger cavity. Experimental methylacetate ΔG^0(water→hexadecane) is reported to be 2.7 kJ mol^{-1}.[1] Calculated $\Delta G(z)$ in the middle of the bilayer for this compound ranges between 0.3 ± 2.0 and -0.9 ± 2.6 kJ mol^{-1} in regions 3 and 4. Simulation results are again lower than the experimental difference in solvation free energy in hexadecane/water systems.

In the literature, the free energy of transfer of ethylacetate, which has an additional -CH_2- group with respect to methylacetate, from water into DMPC liposomes in the liquid-crystalline

phase is found to be -0.66 kcal mol^{-1}, i.e. -2.76 kJ mol^{-1}.[200] As for methylacetate, this is thought to reflect ethylacetate partitioning into the hydrocarbon region of the lipid bilayer. This solute has a ΔG^0(water→hexadecane) of -0.63 kJ mol^{-1}. Therefore, ethylacetate experimental ΔG(water→membrane) is lower than its experimental ΔG^0(water→hexadecane). The same behaviour is found here for methylacetate: its calculated $\Delta G(z)$ from these simulations is lower than its experimental ΔG^0(water→hexadecane). This suggests that the simulation trends are correct.

Benzene $\Delta G(z)$ has a similar profile to that of ethane: $\Delta G(z)$ increases on entering the highly charged headgroup region and decreases moving deeper along the hydrocarbon tails. The experimental ΔG^0(water→hexadecane) for benzene is -12.3 kJ mol^{-1}.[1] The lowest $\Delta G(z)$ value from these simulations is -4.8 ± 2.4 kJ mol^{-1} in region 4. The calculated ΔG(water→membrane) is therefore higher than the experimental ΔG^0(water→hexadecane). This may be due to the fact that, because of lateral packing in lipid bilayers due to interfacial constraints which are not present in bulk organic solvents, accommodating a relatively large solute such as benzene requires more work than accommodating the other smaller compounds studied in these simulations. Therefore, ΔG(water→membrane) for benzene deviates more from ΔG^0(water→hexadecane) than for the other compounds. This proves again that biomembranes do not behave like bulk solvents. Unfortunately, experimental values of benzene partition coefficients between water and lipid bilayers are not reliable for comparisons with these simulations, because of the benzene propensity to adsorb at the lipid/water interface.[201]

6.5.6 Bilayer / Bulk Solvents Comparisons

6.5.6.1 Selectivity Coefficients

The selectivity coefficient s between solute partition K into two different solvents X and Y is defined as:[138,202]

$$\log K(X) = s \log K(Y) + r \tag{6.8}$$

When the two solvents yield $s = 1$, they show identical chemical selectivity towards the solutes under study. If $K(X)$ is taken to be the partition coefficient in the lipid bilayer, correlations between the latter and reference organic solvents Y can be studied. The correlation between the experimental partition coefficients in water/hexadecane systems, K(water→hexadecane), and the calculated partition coefficients at different depths in the lipid bilayer, $K(z) = \exp[\Delta G(z)/RT]$,

	z-depth/Å	water/hexadecane		water/octanol	
		slope	correlation coefficient	slope	correlation coefficient
bilayer regions 2	16.5	0.11	0.64	0.21	0.64
	13.0	0.32	0.88	0.61	0.88
3	9.5	0.54	0.93	1.01	0.92
	6.0	0.71	0.95	1.31	0.93
4	3.0	0.77	0.95	1.41	0.92
	0.0	0.76	0.93	1.38	0.90

Table 6.1: Correlation between calculated partition coefficients at different z-depths in the bilayer and experimental partition coefficients in water/hexadecane and water/octanol systems. Experimental partition coefficients are from Tianhai *et al.*[1] and Walter *et al.*[2]

has been studied plotting $\log K(z)$ *vs.* $\log K$(water→hexadecane) (graphs not shown). Since in drug design water/1-octanol partition coefficients are widely used to estimate drug ability to permeate cell membranes, $\log K(z)$ *vs.* $\log K$(water→octanol) plots have also been examined. This solvent is commonly considered a good model for the biomembrane chemical environment because of the presence of a polar head and a hydrocarbon chain. Slopes of such log-log plots correspond to the selectivity coefficients s between the reference organic solvents and the lipid bilayer. The relevant results are reported in table 6.1.

Considering hexadecane, the best (highest correlation coefficient) and most linear (slope closest to unity) correlation is at $z = 3$ Å and the second most linear correlation is at $z = 0$ Å. This was expected, because hexadecane has a similar chemical nature to that of region 4 of the lipid bilayer. Considering 1-octanol, the most linear correlation is shifted towards the interface and is located at $z = 9.5$ Å; at this depth the correlation coefficient is also the second highest. Because of the presence of a hydrophilic and H-bond donor/acceptor group on octanol which is not present on hexadecane, this solvent mimics a chemical environment in between that of the middle of the bilayer and that of the headgroup region.

Even though the chemical selectivity of biomembranes is similar to that of hexadecane and 1-octanol ($s \approx 1$), solute partitioning into lipid bilayers is somewhat lower than in these bulk solvents. This was clear for benzene, as previously mentioned, but the log-log plots reveal that this is on average a general trend for all the compounds studied here. The intercepts of those plots for the z-depths reported in table 6.1 are all negative. In particular, for hexadecane at $z = 3$ Å the intercept is -0.23 and for 1-octanol at $z = 9.5$ Å the intercept is -1.13 (these z positions are those with s closest to 1). Such findings are also reported in the literature.[203]

It is also noted that for $\log K(z)$ *vs.* $\log K$(water→hexadecane) the slope is systematically lower than 1, whereas for $\log K(z)$ *vs.* $\log K$(water→octanol) the slope is systematically higher than 1. A slope < 1 indicates that the lipid bilayer is more polar than hexadecane, and a slope > 1 indicates that the lipid bilayer is less polar than 1-octanol. Identical findings are reported in experiments,[138] suggesting the simulations are correct. This also explains why the value of s increases for both reference solvents when moving towards the bilayer centre. The conclusion is that on average the polarity of the lipid bilayer is somewhere in between that of octanol and that of hexadecane.

6.5.6.2 Barclay-Butler constants

Both experiments[200,201,204–208] and theoretical models[202,203,209] support the hypothesis that solute partition into bulk organic solvents is not a good model of solute partition into lipid bilayers and biomembranes, primarily from an entropic point of view. The Barclay-Butler constants a and b relate the entropy ΔS and the enthalpy ΔH of solution of different solutes in a given solvent:[202]

$$\Delta S = a + b\Delta H \tag{6.9}$$

A large b constant characterizes a solvent in which entropy changes are relatively more important than enthalpic effects. The fact that in experimental water/membrane partitioning studies the b constant has a larger value than in bulk organic liquids shows that solubilization into membranes is affected to a larger extent by entropic factors. In bulk liquids, molecules are free to move and rearrange. In contrast, inside membranes, because of interfacial constraints, lipids align solutes in a preferred direction and freeze some of their internal degrees of freedom. Even for noble gases b has a higher value than in organic solvents. This shows that to host a permeant molecule the lipid chains are forced to rearrange themselves and restrict even more their mobility. It can therefore be suggested that the similar selectivities of lipid bilayers and some organic solvents only occurs because of a fortunate choice:[202] the enthalpic effect in a solvent is compensated by the entropic factors in the lipid bilayer.

6.5.7 Data Convergence

The accuracy of the simulations was monitored looking at the convergence of the value of $\langle \boldsymbol{F}(z) \rangle_t$. It was expected that $\langle \boldsymbol{F}(z) \rangle_t$s from the five separate simulations would have converged

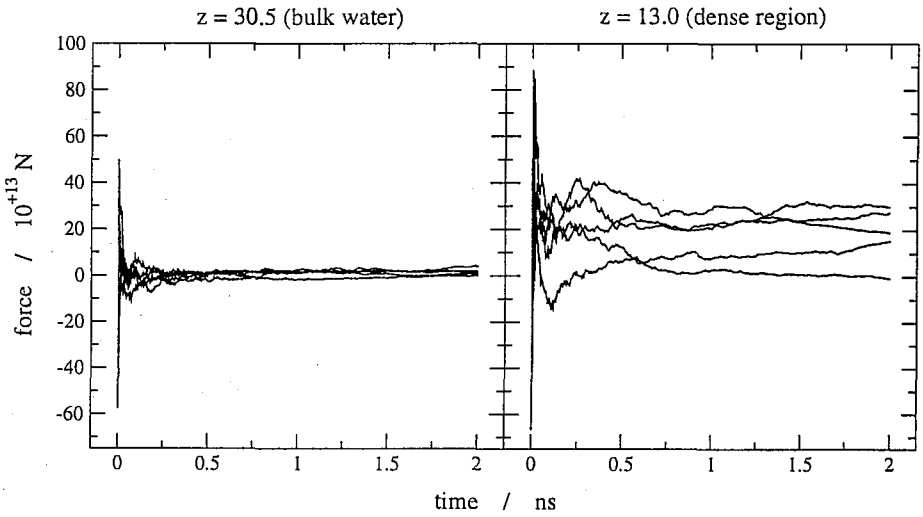


Figure 6.2: Value of $\langle \boldsymbol{F}(z) \rangle_t$ as a function of the simulation time from the five separate simulations.

to the same value, because the bilayer physical properties which can be studied within the time scale of current MD simulations are shown to fairly converge. Since the solute behaviour inside the membrane is thought to depend on the behaviour of the surrounding lipid molecules, the value of $\langle \boldsymbol{F}(z) \rangle_t$ was expected to converge too. $\langle \boldsymbol{F}(z) \rangle_t$s are plotted for water as an example in figure 6.2 at $z = 30.5$ Å, when the solutes are still in bulk water, and $z = 13.0$ Å, which is in the densest part of the membrane and corresponds to the edge between the headgroup region and the hydrocarbon core of the bilayer. For the other depths, convergence profiles between these two extremes are found.

It can be seen that even a 2 ns long simulation is not sufficient to ensure convergence, and this causes the significant errors in the $\Delta G(z)$ profiles. The permeant behaviour is highly affected by those slow lipid motions which possess a long time scale and are not sampled completely in a single 2 ns MD simulation. This observation enforced the necessity of sampling the same z depth in several separate simulations at different x-y positions, to obtain a reasonable value of $\Delta G(z)$ which is not too biased by the choice of the starting structure. A favourable aspect is represented by the fact that at each depth the curves from the five simulations cross each other during 2 ns. This indicates that the independent simulations sample common regions of phase space and suggests that in the end the phase space is reasonably sampled.

In the literature[3,182] only 120 ps were simulated when studying water, oxygen and ammonia permeation with the z-constraint method and only one single xy position was sampled for each depth. The simulations reported here are 16.7 times longer and sample five different positions at each depth in the membrane, but complete convergence of data is still not achieved. The

precision of the literature results is therefore called into question.

6.6 Diffusion Coefficients

Diffusion coefficient $D(z)$ profiles for the permeants studied, calculated with equation 6.6, are plotted in figure 6.3. Errors in $D(z)$ are standard errors calculated from the difference of the diffusion coefficients in each of the five individual simulation at each depth from their average.

6.6.1 General Trend

For all the compounds, $D(z)$ decreases entering the membrane in regions 1 and 2, then it is quite uniform in region 3 and increases slightly in region 4. Diffusion coefficients inside the membrane are lower than in bulk water phase because of the higher density of the lipid environment. Water, the smallest solute, has the highest $D(z)$ at all depths. However, the differences between the other permeants are very small and arguably not significant, as $D(z)$ of benzene, the largest solute, is not always the lowest.

6.6.2 Water

Marrink and Berendsen,[3,182] who introduced the z-constraint method, reported a water diffusion coefficient in the middle of the bilayer twice as large as in the bulk water phase, since in their simulations $D(z)$ increased very rapidly in regions 3 and 4. They reported the same behaviour for ammonia and oxygen, and, since ammonia has a smaller size than water and oxygen has a larger one, the increase of $D(z)$ in regions 3 and 4 was even higher for NH_3 and lower for O_2. Values of $D(z)$ in the hydrophobic core of the membrane were found to be close to those in bulk alkanes such as hexadecane and cyclohexane. This result clearly indicated that size was the dominant parameter regulating diffusion in phospholipid membranes. In contrast, in these simulations $D(z)$ is very similar for all the compounds, whose sizes range from that of water to that of benzene, and $D(z)$ is lower than that in the bulk water phase at all depths. Some considerations may help to understand these results. Marrink and Berendsen's simulations were run at 350 K (77 °C), this temperature being 27 K higher than that employed here and higher than that commonly employed in experiments. The authors assumed an Arrhenius behaviour and used published activation energies to compare their values with experimental diffusion coefficients in bulk organic solvents. Their water model overestimated the diffusion

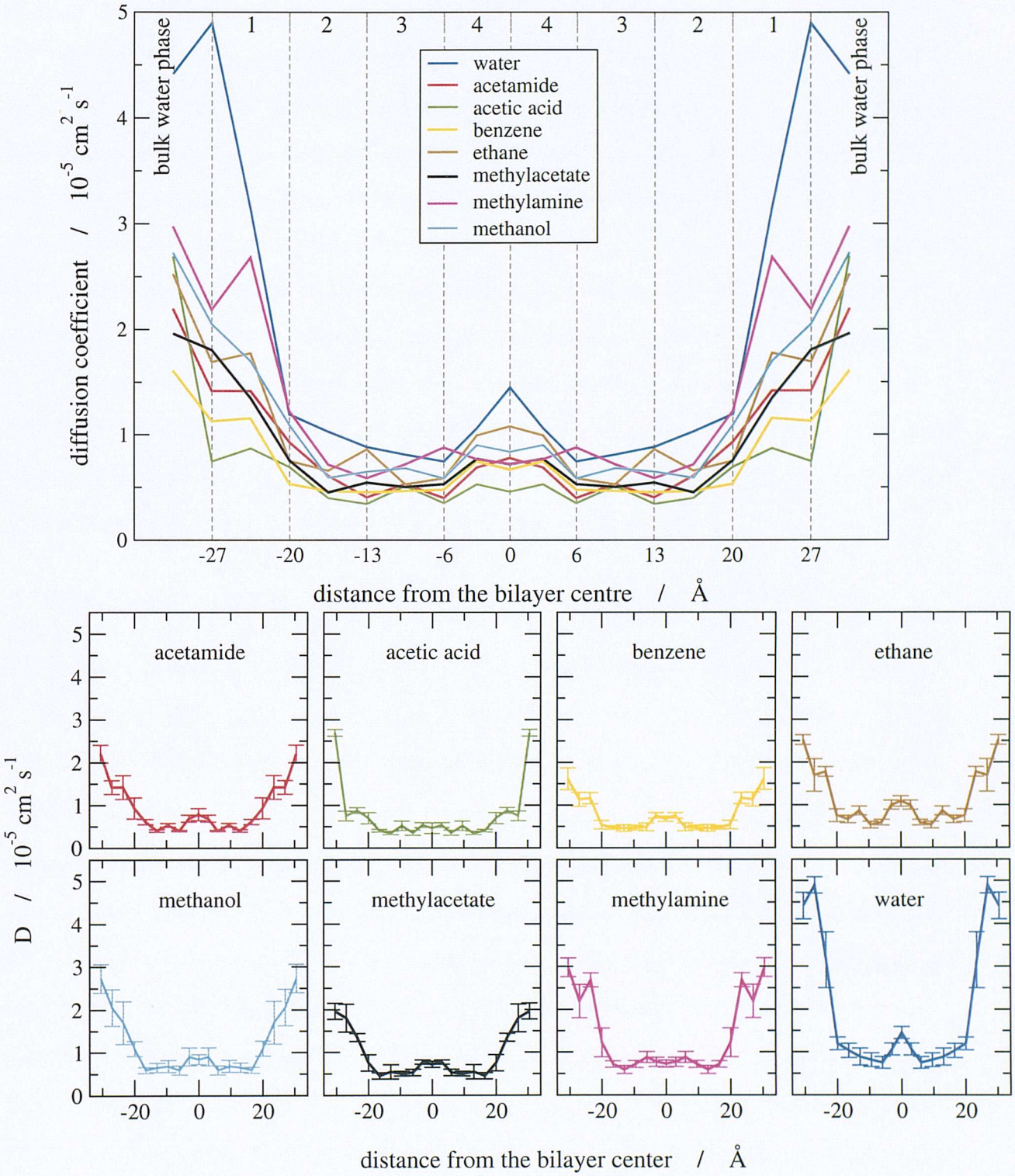


Figure 6.3: Diffusion coefficient profiles. Top: bilayer is divided into the four regions; for clarity error bars are not shown. Bottom: each profile is plotted alone with error bars. Error bars are standard errors calculated from the difference of the diffusion coefficients in the five individual simulations from their average.

coefficient in the bulk water phase, so it is also expected to yield higher values inside the membrane. The values of $D(z)$ for water, ammonia and oxygen were much higher than the experimental diffusion coefficient of oxygen in pure DPPC bilayers ($\approx 1 \times 10^{-5}$ cm^2 s^{-1} at 318 K), even after temperature correction. The lipid model used by Marrink and Berendsen was united-atom, whereas in these simulations an all-atom model is employed. As suggested by Stouch *et al.*,[104] united-atom models result in artificially high diffusion rates. Moreover, free volume analysis of the lipid bilayer employed here revealed that there is no free space large enough to locate a solute molecule with the same size as water at any depth, and that the free volume inside the membrane is always lower than that in bulk water. This is in contradiction with Marrink and Berendsen's lipid system[3] for which in region 4 the free volume is larger than in bulk water and pockets of free space large enough to host water molecules do exist. This important difference between the two lipid systems (CHARMM versus GROMACS parameters) is reflected in different solute diffusive behaviour. In support of this work, we can say that the all-atom force field is more realistic and therefore more likely to be right.

6.6.3 Benzene

Our diffusion results are in closer agreement with other publications. Stouch *et al.*[103,104] simulated benzene diffusion at different depths in the hydrocarbon region of a DMPC bilayer and calculated its diffusion coefficient from the mean squared displacement (see section 5.3). They reported $D(z)$ values of 0.46×10^{-5} cm^2 s^{-1} at the terminal methyl group of the lipid tails and 0.13×10^{-5} cm^2 s^{-1} at the carbonyl group between the fatty acid chains and the glycerol fragment, in both cases at 320 K. The simulations presented here yield values of 0.75×10^{-5} cm^2 s^{-1} in region 4 and 0.45×10^{-5} cm^2 s^{-1} in region 2. No experimental data are available about benzene $D(z)$ in lipid bilayers. Stouch *et al.* compared their values with those of di-*tert*-butyl nitroxide in dipalmitoyl lecithins,[132] which range between 0.05×10^{-5} cm^2 s^{-1} at the carbonyl position and 0.19×10^{-5} cm^2 s^{-1} at the chain terminal methyl. Therefore, Stouch's and these simulations yield values which are at most one order of magnitude higher than experiment, but both agree with experiment in reproducing faster solute diffusion in the lower than in the upper part of the lipid tails. However, experimental diffusion coefficients for di-*tert*-butyl nitroxide were not determined directly, but rather estimated assuming a preferential solute partition in the middle of the bilayer rather than at the lipid/water interface. Considering the uncertainties

in the force field employed here, and in the interpretation of the experiments, and considering the different chemical structure and properties of di-*tert*-butyl nitroxide, the agreement of the simulations is fairly good. The different value of $D(z)$ between these and Stouch's simulations may arise from the different way of calculating $D(z)$, and from the differences in the force field applied. These simulations also allow for the calculation of $D(z)$ in the headgroup region, not only in the hydrocarbon core of the bilayer. Stouch's unconstrained simulations allow for the mechanism of diffusion to be better understood: benzene molecules clearly rattle inside a particular void for a relatively long period of time and infrequently take jumps to a neighbouring void, the overall diffusion coefficient depending on both the restricted motions in a pocket of free space and on the size and frequency of the jumps. Another important difference between this and Stouch's lipid systems is the free volume distribution. As in Marrink and Berendsen's simulations, in Stouch's bilayer centre the free volume fraction is higher than in bulk water, whereas in the simulations reported here it is not, and voids large enough to host a benzene molecule were found.[103] It must be said that Stouch used a surface area per lipid $A \approx 66$ Å^2, whereas in these simulations $A = 62.9$ Å^2. These discrepancies produce different diffusion coefficients. Unfortunately, the experimental free volume fraction in lipid bilayers is not available and previous publications using the CHARMM force field do not report their free volume distribution.

6.6.4 Diffusion in Bulk Water

Experimental diffusion coefficients are available in the bulk water phase for some of the solutes studied here and can be compared with the value of $D(z)$ at $z = 30.5$ Å. These are shown in table 6.2. In all cases, diffusion coefficients from these simulations are higher than those determined experimentally. The higher temperature employed in the simulations (50 °C) compared with the experiments (25 °C) may be the reason for this difference. The order of magnitude is the same in all cases anyway and the agreement is fairly good. Regarding methylamine, table 6.2 reports the value for ethylamine. Considering that the diffusion coefficients in water for di- and tri-ethylamine in the same publication[214] are 1.59×10^{-5} cm^2 s^{-1} and 1.32×10^{-5} cm^2 s^{-1} respectively, one could estimate that the value for methylamine should be higher than that for ethylamine and hence closer to that calculated from these simulations.

For some of the solutes, diffusion coefficients at different temperatures have been experimentally measured, namely acetamide,[210] acetic acid (two temperatures only),[216] benzene,[217]

solute	$D(30.5)$ in these MD sim. (50 °C)	experimental D(water)	(T)	[ref.]
acetamide	2.197 ± 0.205	1.32 ± 0.03	(25)	[210]
acetic acid	2.692 ± 0.079	1.201	(25)	[211]
benzene	1.609 ± 0.253	≈ 1.1	(25)	[212]
ethane	2.524 ± 0.113	1.52 ± 0.03	(25)	[213]
methanol	2.728 ± 0.334	1.70 ± 0.03	(25)	[210]
methylacetate	1.961 ± 0.185	-		
methylamine	2.976 ± 0.219	1.94‡	(25)	[214]
water	4.412 ± 0.312	2.299	(25)	[215]

Table 6.2: Calculated and measured diffusion coefficients. The associated references are given in [ ]. D values are in 10^{-5} cm^2 s^{-1}, T in ( ) are temperatures in °C. ‡ Value for ethylamine.

ethane,[213] methanol[210] and water.[215] Using a simple linear regression experimental diffusion coefficients have been extrapolated at 50 °C. Results are reported in figure 6.4. The solid line is the linear regression between calculated and "experimental" diffusion coefficients, while the dashed line is the ideal 1:1 correlation. Although the simulations do not yield the expected values, however, considering the crude approach employed to obtain experimental data at 50 °C and that, for example, common water models tend to overestimate their self diffusion coefficient, the simulation results are reasonably valid.

6.6.5 Force Fluctuation Autocorrelation Function

According to equation 6.6, diffusion coefficients can be obtained from the time autocorrelation function of the instantaneous force fluctuation from the mean. A general time autocorrelation function is defined as:[12]

$$C(t) = x(t)x(0) \tag{6.10}$$

where t is time and x is the variable for which the time autocorrelation function is to be calculated. For these simulations this variable is the instantaneous force fluctuation with respect to the mean. Owing to limitations in computer power, simulations are finite in time, so a time autocorrelation function is calculated from computer simulations averaging over all possible

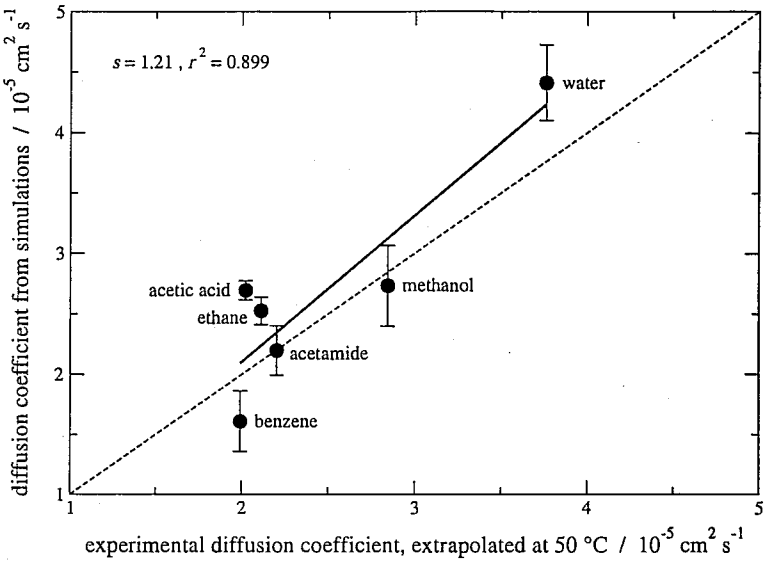


Figure 6.4: Comparison between the diffusion coefficients from the simulations and those extrapolated at 50 °C from experimental data. The solid line is the linear regression, whose slope s and correlation coefficient r^2 are reported in the left top corner of the graph. The dashed line is the straight line y = x.

time origins:

$$C(t) = \langle x(t)x(0) \rangle \tag{6.11}$$

As an example, force fluctuation autocorrelation functions for water and benzene are reported in figure 6.5. Standard time autocorrelation functions can be fitted to an exponential curve, since they have a characteristic exponential decay with relaxation time τ:

$$C(t) = C(0) \exp(-t/\tau) \tag{6.12}$$

Here $C(0)$ is the value of the time autocorrelation function at $t = 0$. For these simulations, the force fluctuation time autocorrelation function had to be fitted with a double exponential:

$$C(t) = C_1 \exp(-t/\tau_1) + C_2 \exp(-t/\tau_2) \tag{6.13}$$

In this case, $C(0) = C_1 + C_2$. In figure 6.5 an example for each of the four regions of the lipid bilayer is plotted. Moving from region 1 to region 4, $C(0)$ tend to decrease but the decay time increases. The latter effect prevails, so that the integral of the autocorrelation function increases, moving from region 1 to region 3. According to equation 6.6, the diffusion coefficient decreases when the integral of the autocorrelation function increases. In region 4, D values are slightly higher because the increase in decay time is not long enough to compensate for the decrease in $C(0)$, so that the integral is lower than in region 3.

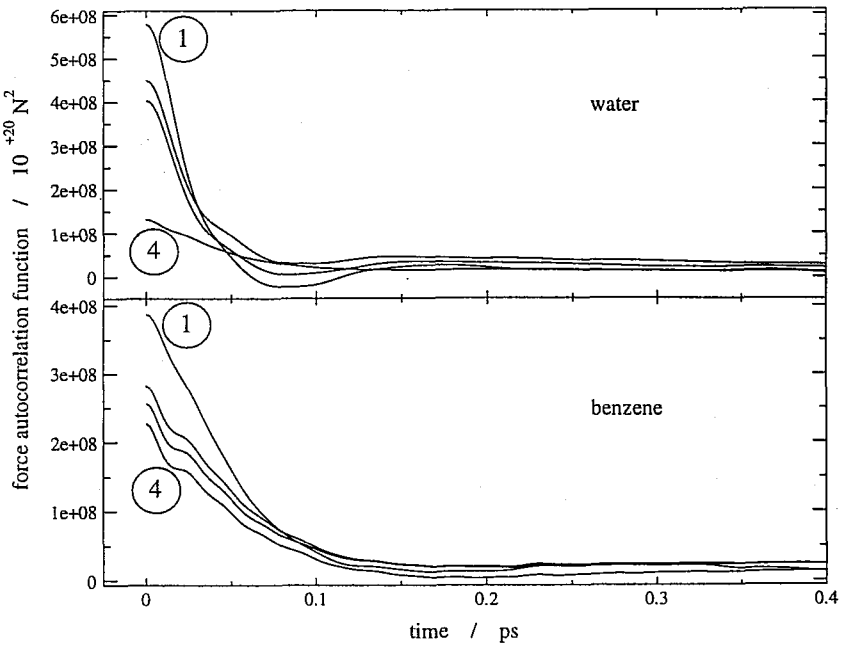


Figure 6.5: Force fluctuation autocorrelation functions of water and benzene. An example for each of the four regions of the lipid bilayer: moving from region 1 to region 4, $C(0)$ tend to decrease and the decay time to increase.

6.6.6 Friction Relaxation Times

That a double exponential is required for a good fit to the autocorrelation function shows that solute motions inside the bilayer have two distinct time scales (τ_1 and τ_2 in equation 6.13). The motions with a short decay time have typical τ_{short} of 0.02 to 0.03 ps for water, 0.04 to 0.05 ps for benzene, and 0.03 to 0.04 ps for all the other compounds. This corresponds to the immediate response of the solute to the local environment, i.e. the friction due to the surrounding lipids that keeps the permeants in their local cages. The value of τ_{short} tends to increase going deeper into the membrane and the value is a little higher for larger solutes. Although the change is very small and may be not related to the permeant location, however, considering that closer to the middle of the bilayer the pockets of free space accommodating permeants are larger, the increase in τ_{short} seems reasonable. The motions with a long decay time have a typical τ_{long} of 1 to 12 ps. This value appears not to be systematically related to solute size. It is instead related to permeant location, as it increases moving from the lipid/water interface to region 4. This longer time-scale motion is related to the overall mechanism of diffusion through the membrane. Assuming a hopping type of diffusion, τ_{long} is related to the residence time of the penetrants in their cage of free volume between subsequent hops. Even though values for τ_{long} in regions 3 and 4 are higher, these findings are in some agreement with those of Marrink and Berendsen[182] regarding water, ammonia and oxygen permeation, where their τ_{short}s were < 0.1

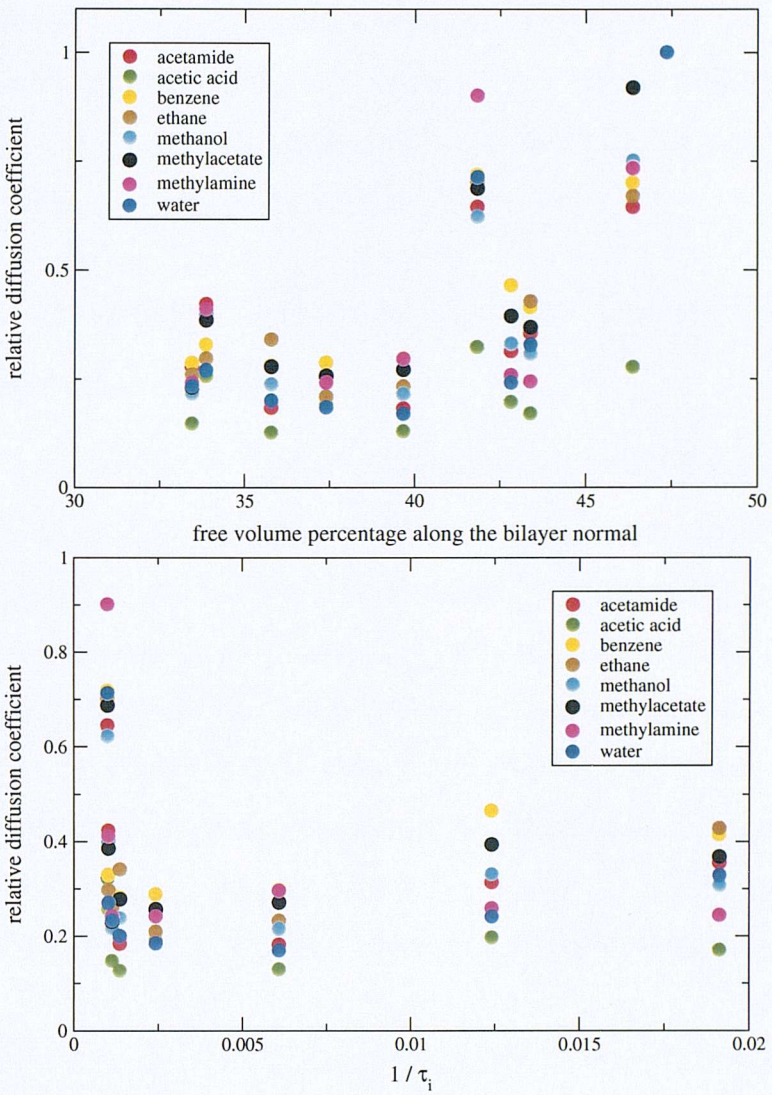


Figure 6.6: Top: relative diffusion coefficients versus free volume percentage; the former are D values divided by D at 30.5 Å from the bilayer centre (that is in bulk water). Bottom: relative diffusion coefficients versus the inverse of lipid trans/gauche interconversion times, as calculated in chapter 4.

ps and τ_{long}s $\approx$ 2 ps.

6.6.7 Correlations Between Solute Diffusion and Bilayer Free Volume

Owing to their atomistic detail, MD simulations allow for correlations of solute diffusion coefficients with static and dynamic properties of the solute and lipid molecules to be investigated.

From these simulations there is no strong correlation between solute diffusion coefficients and free volume distribution along the bilayer normal. This is shown at the top of figure 6.6, where the relative diffusion coefficients [i.e. $D(z)/D(30.5)$] are plotted versus free volume percentage along the bilayer normal. The only correlation is that D values decrease on entering the membrane and are only slightly higher in the bilayer centre. The lack of any correlation can

be explained on the basis that free volume is never sufficient to ensure the presence of holes large enough to accommodate solutes of the size of water. The absence of a relationship between solute diffusion coefficient and free volume fraction along the bilayer normal was previously reported in the simulation of a drug analogue in a lipid bilayer.[109]

6.6.8 Correlations Between Solute Diffusion and Lipid Internal Motions

It is suggested[111] that diffusion in chain-like solvents such as lipid bilayers can be related to the characteristic times τ_i for solvent internal motions such as trans/gauche interconversion rates along the lipid hydrocarbon chains. These characteristic times are different at different positions along the lipid chains, being hundreds of ps close to the carbonyl group and tens of ps close the terminal methyl group. Lipid chain trans/gauche transitions are thought to be the main mechanism responsible for free volume redistribution inside the membrane, since lipid translation and rotation as a whole require much longer times. This would predict lower solute diffusion in the upper part of the lipid tails, where τ_i is longer, and a higher value at the end of the tails, where τ_i is shorter. Although from these simulations diffusion coefficients in the centre of the membrane are slightly higher, there is no evident correlation between lipid τ_i and solute D. This is clear from the bottom of figure 6.6, where relative diffusion coefficients are plotted against the inverse of the lipid trans/gauche interconversion times $(1/\tau_i)$, calculated as indicated in chapter 4. Since there is no correlation between solute D and the free volume distribution, an absence of correlation between solute D and lipid τ_i would also be expected, since τ_i mainly acts to redistribute the free volume. Moreover, the longest decay times of solute friction coefficients from these simulations are always much shorter than lipid τ_i at all the depths, and this suggests that lipid internal flexibility may be only one of the various contributions to solute diffusion.

6.6.9 Correlations Between Solute Diffusion and Solute Molecular Volume

The extent of the correlation between solute volume V_d and diffusion coefficient D was also studied. The value of V_d was calculated using the same method applied to calculate the free volume along the bilayer normal. The solute was placed in a reasonably large box, for which points outside the van der Waals radii of the solute atoms were counted. This gave the free

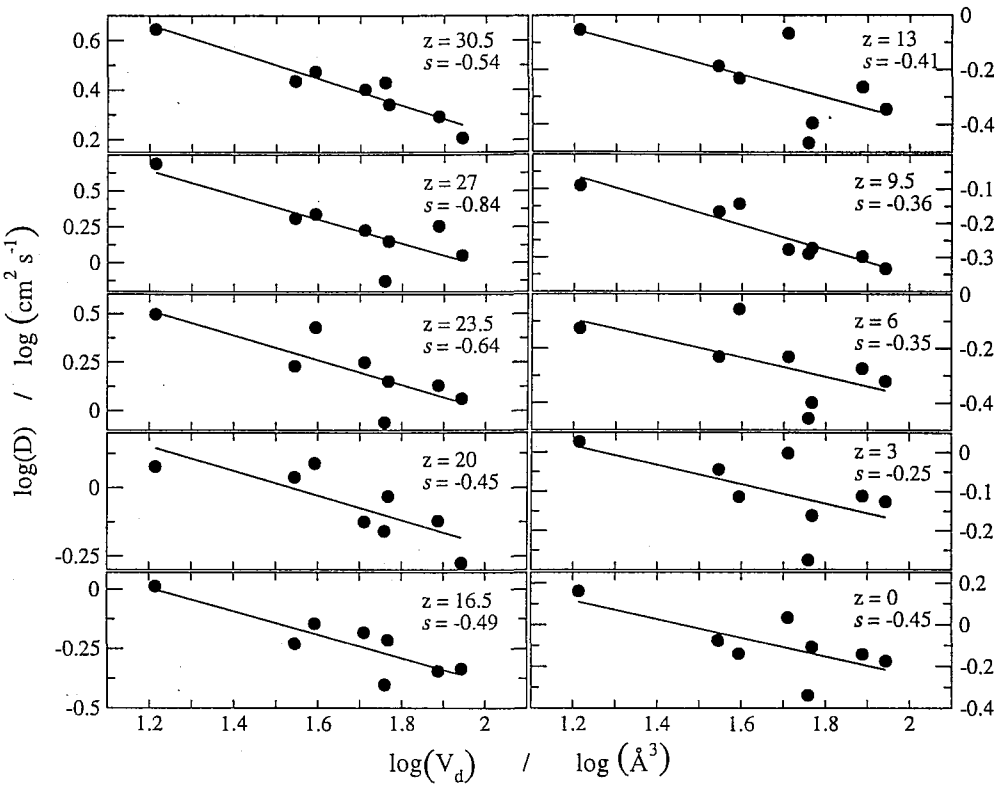


Figure 6.7: Log-log plots of diffusion coefficients as a function of solute volume at different z-depths (in Å) in the membrane. Solid lines are linear regression. Slopes s are also reported.

volume of said box. V_d was then equal to the difference between the total volume of the box and the free volume.

Figure 6.7 contains log-log plots of D *vs* V_d. Investigating the nature of the $D \propto (V_d)^s$ relationship, the slopes of those plots yield the value of the exponent s. In figure 6.7 the distances z from the bilayer centre and the slopes s are also reported. The slope at $z = 30.5$ Å, where the solutes are still in bulk water, is in agreement with the experimental observation that $\log(D)$ is approximately $\propto -0.6 \log(V_d)$[2,136,139] for small solutes in sphere-like solvents. With the exception of the D values 30.5 Å distant from the bilayer centre, i.e. in bulk water, for which the correlation coefficient is 0.96, unfortunately the data are quite scattered and correlation coefficients are between 0.59 and 0.86. The interesting feature is that the size dependence of D is only higher than that in water at the interface ($z = 27$ and 23.5 Å), and lower in the membrane interior. It must be said that, although the membrane system is very different, this is in contrast to simulations of a bilayer of fatty acids,[139] where it was found that solute diffusion along the bilayer normal had a greater dependence on solute volume than in bulk water, above all in the upper part of the chains, which are expected to mimic the upper part of the lipid tails in phospholipid bilayers. That result was confirmed by the experimental

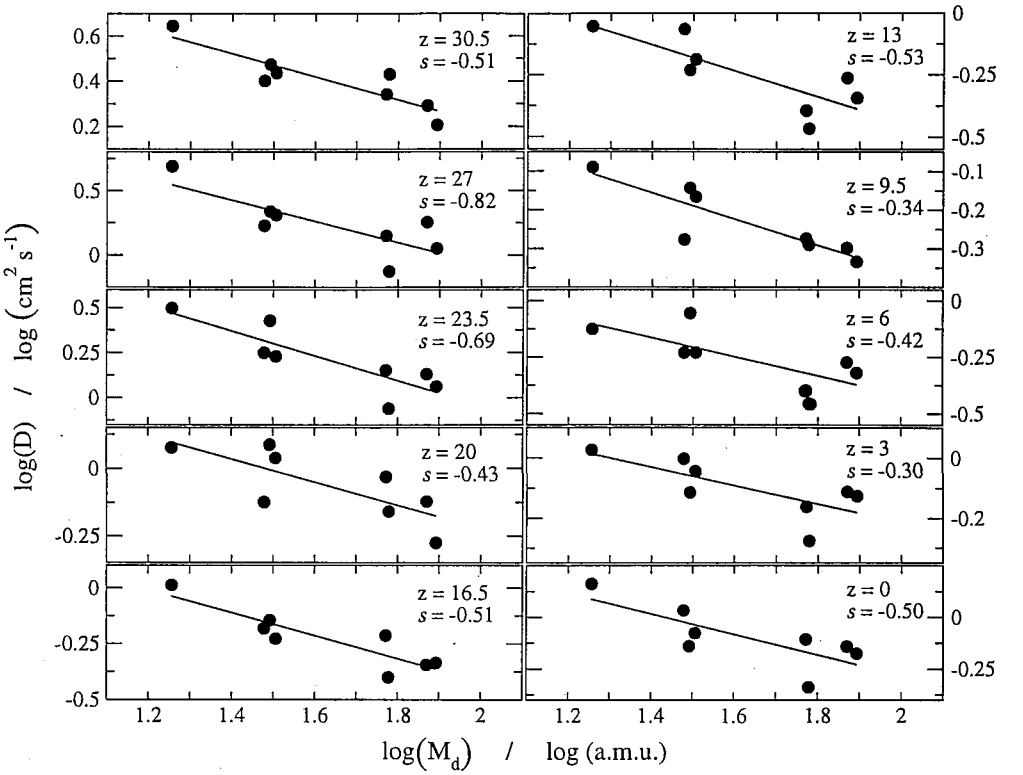


Figure 6.8: Log-log plots of diffusion coefficients as a function of solute mass at different z-depths (in Å) in the membrane. Solid lines are linear regression. Slopes s are also reported.

observation that diffusion in long chain alkanes and polymers shows a greater dependence on solute size than in simple liquids.[139] However, in those simulations the solutes studied were simple spherical particles representing noble gases, whereas in the simulations reported here much more complex shapes are involved.

6.6.10 Correlations Between Solute Diffusion and Solute Mass

The dependence of D on solute mass was also studied. Log-log plots are reported in figure 6.8 with the same conventions as in figure 6.7. It is comforting to see that $D(30.5)$, that is the diffusion coefficient in bulk water, has $s \approx$ -1/2 in excellent agreement with experimental observations for small molecules diffusing in water.[2,136,218,219] The trend found for the log$D$-logV_d relationship is also found for the log$D$-log$M$ relationship: higher s values at the interface and then lower values inside the membrane, with respect to the water phase. This is in contrast with the experimental observation that diffusion coefficients in polymers and cell membranes have a higher mass dependence than in simple liquids.[2,136,218–221] These studies led to the conclusion that diffusion in biological membranes resembles that in soft polymers and depends

on the formation frequency and the size distribution of pockets of free space among which solutes can jump. The simulations performed here do not support these results derived from polymers and cell membranes. This aspect is investigated in more detail later in this chapter.

6.6.11 Experiments vs. Simulations

The assumptions associated with the experiments that determine the solute diffusion dependence on both solute mass M and volume V_d must be assessed. The main limitation of these experiments resides in the fact that diffusion coefficients across the membrane are not accessible directly, but rather they must be extrapolated from the measurement of the permeability coefficients. According to the solubility-diffusion model, assuming the main resistance to permeation comes from a uniform distinct barrier region in the membrane, the solute permeability coefficient P can be expressed as a function of the solute partition coefficient K and diffusion coefficient D inside the barrier region and the thickness d of that region:

$$P = \frac{KD}{d} \tag{6.14}$$

To study the dependence of D on solute M or V_d, the logarithm of the ratio between the experimentally measured permeability coefficient P across membranes and the partition coefficient K in organic solvent / water systems, $\log(P/K)$, is plotted versus $\log(M)$ or $\log(V_d)$.[2,136,218] K in organic solvent / water systems is used as an estimation of K in the membrane barrier region. This ascribes the entire molecular size dependence in P to the diffusion coefficient and neglects the possibility of size dependent solute partitioning. It was found[136] that changing reference organic solvent or using models which ascribe the size dependence of P to both partitioning and diffusive behaviours decreases the slopes s to values much closer to that in simple liquids (-0.5), although still higher. This suggests that the size dependence of diffusion in lipid bilayers may be less than previously thought, and thus closer to the results of these simulations. Finally, it must not be forgotten that these simulations are limited to solutes with similar size and exclude large molecules like nucleosides or steroids, which are included in the interpretation of the experimental results.

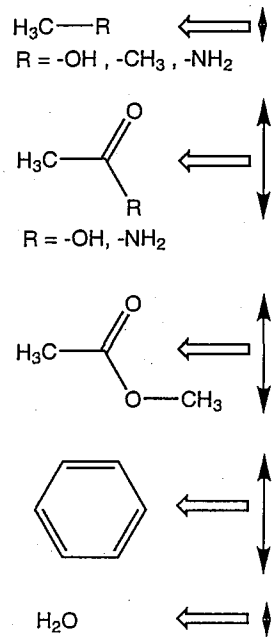


Figure 6.9: Calculation of solute cross-sectional area: the single headed arrows indicate the side whose area is considered to be S_d, and the double headed arrows represent the extension of that area.

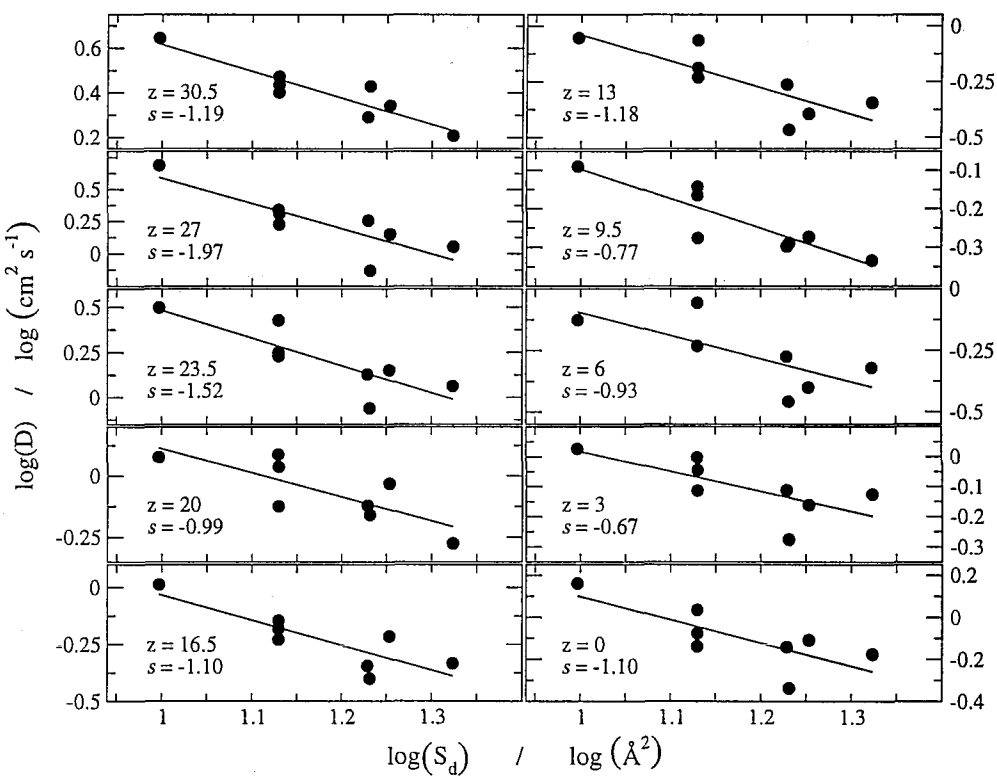


Figure 6.10: Log-log plot of diffusion coefficients as a function of solute cross-sectional areas S_d at different z-depths (in Å) in the membrane. s are slopes of linear regressions.

6.6.12 Correlations Between Solute Diffusion and Solute Cross-Sectional Area

Finally, it has been suggested[111] that diffusion in chain-like solvents is related to the solute cross-sectional area, S_d. In contrast to classical free volume diffusion theories, solute displacement can occur not only when a void of volume comparable to that of the diffusant is created, but also when the cross-sectional area of the newly created void is equal to or greater than that of the diffusant. This would predict $D \propto (S_d)^s$ with s negative. Since in this lipid system voids large enough to locate these solutes have never been found, the dependence of D on S_d appears particularly interesting. In fact, these simulations do show this relation. Solute S_d was calculated here by taking the solutes in their equilibrium conformation, that is the one corresponding to the bonds and angles set to the values used as equilibrium values in the force field parameters. The size of each atom was then taken from its van der Waals radius in the force field. The molecular side whose area was considered to be the S_d is shown in figure 6.9 for each solute.

Figure 6.10 plots solute $\log(D)$ as a function of solute $\log(S_d)$. The solid lines are linear regressions. With the exception of $z = 30.5$ Å, with a value of 0.93, correlation coefficients for these regressions are not high (between 0.71 and 0.87), but this may depend on the low number of data points (8 solutes) and the uncertainties intrinsic in the MD simulation technique. The main feature is that at all depths solute diffusion does depend on solute cross-sectional area, in agreement with previous theories.[111] The diffusion dependence on S_d also appears to be even larger than the dependence on V_d, about twice as large. The correlation coefficients are also a little higher. As for the relationship with solute volume and mass, the dependence on S_d inside the membrane is lower than in the water phase, with the exception of the interface.

6.6.13 Relative Diffusion Coefficients

The above observations that the D dependence on either V_d, M or S_d inside the membrane is lower than in water, leads to the conclusion that, since inside the membrane D values are lower than in water, the decrease of diffusion coefficient when moving from the water phase towards the membrane interior is larger for the smallest solutes and lower for the largest solutes. This is confirmed by plotting relative diffusion coefficients as a function of z-depth, in figure 6.11. Relative diffusion coefficients are calculated as the ratio between the $D(z)$ and $D(30.5)$. The

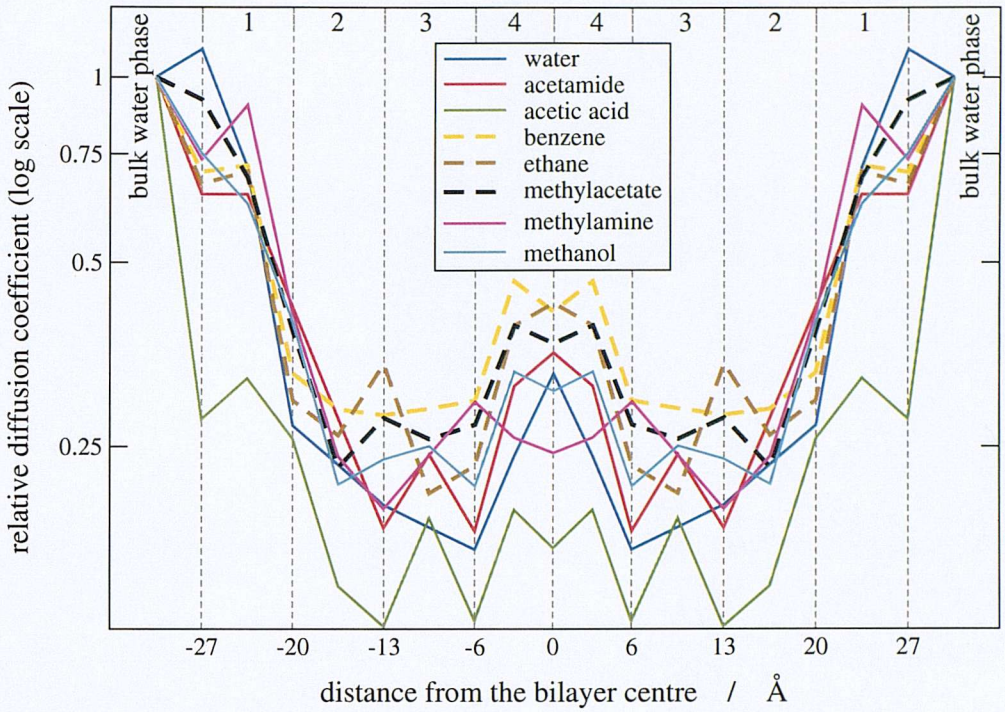


Figure 6.11: Relative diffusion coefficients. Hydrophobic compounds are plotted with dashed lines for clarity.

ordinate is scaled in a logarithmic fashion for clarity. A second feature is that, when the solute size is similar, the decrease in D tends to be less for the hydrophobic compounds than for the hydrophilic ones. To our knowledge, no similar findings have been reported in the literature and are difficult to explain with common interpretations. The chemical nature of the diffusant in biological membranes may be more important than usually thought.

6.6.14 Approximations

In the calculation of diffusion coefficients it is assumed that during the decay time of the time-dependent friction coefficient the particle remains in a region of constant free energy, so that the effective friction could be fully described by considering only the static friction coefficient $\tilde{\xi}$ (see equations 6.3, 6.5 and 6.6). A variation in free energy of the order of RT per mole is however allowed, where R is the gas constant and T the absolute temperature, since this is the extent of thermal fluctuations. To determine whether this assumption is valid, the following calculations may be performed. Since the diffusion coefficient is related to the particle mean squared displacement (equation 4.6), the calculated $D(z)$ can be used to obtain the mean displacement covered by the particle in a time identical to the decay time of its friction coefficient. The difference in free energy between two positions along the bilayer normal whose distance is equal to the displacement calculated from $D(z)$, can then be compared with RT.

solute	ΔG / kJ mol^{-1}
acetamide	1.33
acetic acid	1.11
benzene	0.46
ethane	0.69
methanol	0.93
methylacetate	0.23
methylamine	0.45
water	1.41

Table 6.3: Calculated maximum free energy difference over the solute displacement within the decay time of its friction coefficient.

Table 6.3 reports for each solute the largest free energy difference along such a distance. The reference value of RT is 2.69 kJ mol^{-1}. It is clear that for all the compounds this approximation is valid: the change in free energy over the typical distance derived from the decay time of the friction coefficient is less than that expected for normal thermal fluctuations. This implies that the motion is within the limit of Markovian dynamics and is diffusive.[3] In the physical picture, random motions involving the molecules of the medium give rise to impacts on the solute particle, and the response to the force acting on the particle is almost instantaneous.[197] This behaviour is typical of small and rigid permeants. If, in contrast, the particle is large and flexible, the dynamics is more likely to be non-Markovian, since successive dynamic events become correlated and memory effects are no longer negligible.[196,222–227]

A second source of approximation is the fit performed on the force fluctuation autocorrelation functions shown in section 6.6.5. These fits are very sensitive to the parameters chosen for the exponential function (see equation 6.13): small changes of C_1, C_2, τ_1 and τ_2 do not yield significant changes in the correlation coefficient for the fit (so they are statistically equally acceptable), but yield quite different values of diffusion coefficients. Furthermore, the fitting operation is performed by minimizing the sum of the squares of the differences between the points of the fitting curve and those of the original autocorrelation function. Owing to the complex profile of the latter, the minimization algorithm can be easily trapped in a local minimum and the best fit must be found by combining automatic algorithms with manual adjustment, guided by intuition and a visual comparison of the two curves. The fitting therefore remains partly subjective, above all for the positions in the dense part of the membrane, and the qualitative

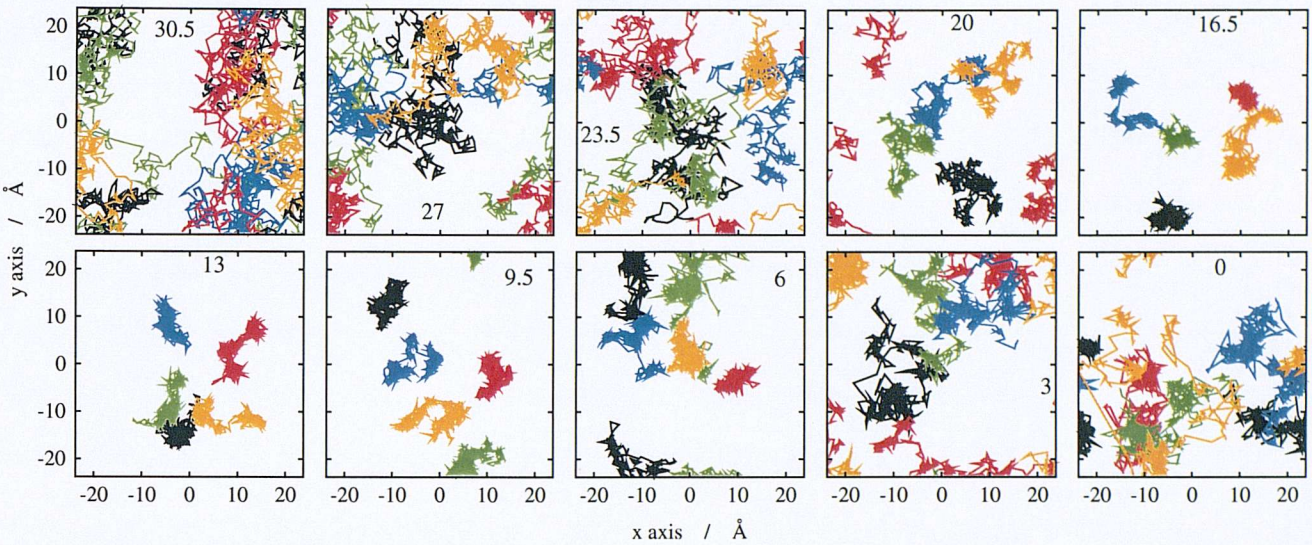


Figure 6.12: Water motions on the x-y plane at different depths. The five different colours refer to the five water molecules constrained at the same z-depth but at different xy-positions in the starting structure.

trend should therefore be considered rather than the exact numerical values.

6.6.15 Motions on the x-y Plane

The application of the z constraint may bias the diffusion observed in the x-y plane, the diffusion on the bilayer plane of a z-constrained particle may differ from that of a free particle.[182] However, it is instructive to examine the solute displacement on the x-y plane. This is plotted for water as an example in figure 6.12. To avoid making the picture unreadable, only 0.5 ns are shown. Periodic boundary conditions were applied during the simulations, so that, when a water molecule exits the simulation box from one side, it enters the box from the opposite side. In the figure, the five different colours refer to the five water molecules constrained at the same z-depth but at different xy-positions in the starting structure.

It is interesting to note how the magnitude of the x-y displacement changes with depth and parallels the values of $D(z)$. At z position equal to 30.5 (i.e., 30.5 Å from the bilayer centre) the water molecules are completely free to move in the bulk phase. Then, going deeper into the membrane to z position 9.5 Å, water motions are more and more restricted, since, because of the interfacial constraints, lipids are tightly packed together. Between z positions 16.5 and 9.5 waters are trapped in cages delimited by the surrounding lipids: waters rattle inside these pockets and only rarely abandon one to fill a neighbour. Closer to the bilayer centre, the available pockets of free space are larger and the lipid tails' trans/gauche interconversions occur more often. Consequently, water motions in the middle of the bilayer are larger than in the

upper part of the lipid chains, although they are still smaller than in the bulk water phase.

This picture is closely related to the diffusion coefficients calculated from the force autocorrelation function according to equation 6.6. Using this equation the diffusion coefficient along the bilayer normal (z-axis) is calculated, whereas figure 6.12 is a picture of water diffusion on the bilayer plane (x and y-axis). Owing to the heterogeneous nature of lipid bilayers, diffusion parallel or perpendicular to the bilayer normal is different. However, on a local scale (i.e. at a given z-depth), the environment around a permeant is expected to be identical in all directions (isotropic) and diffusion along the x, y or z axis should be identical.

6.7 Local Resistances and Permeability Coefficients

Local resistances $\Re(z)$ for the permeants, calculated with equation 6.7, are plotted in figure 6.13. For clarity, error bars are not plotted. These errors are large, since, owing to the exponential form of the mathematical relationship between $\Delta G(z)$ and $\Re(z)$, even a small error in $\Delta G(z)$ is amplified in $\Re(z)$. Other publications[3,182] reporting $\Re(z)$ also omit to give error bars. However, this is not a significant problem for the calculation of the overall permeability coefficient P, since this is derived from the integration of $\Re(z)$ along the z axis from one side of the bilayer to the other. Calculated P values from these MD simulations are reported in table 6.4. Where available, experimental values are also reported for comparison.

The simulation results are very promising. The water permeability coefficient is of the right order of magnitude with respect to the experimental data. For the other compounds with available experimental data, the discrepancy between simulation and experiment is only one order of magnitude, with the simulation value being higher than experiment.

Despite the fact that no experimental data are available for methanol, some extrapolations are possible. It may be surprising that the permeability for methanol, an alcohol, is lower than that of acetic acid. If one considers the observed apparent permeability coefficient P_{obs} (see equation 5.1), then methanol P_{obs} is indeed higher than that of acetic acid, because at common pH values the acid is mainly present as an ion, while the neutral form only can cross biomembranes at a significant rate. When instead considering the intrinsic permeability P_n of the neutral form of the solute (see equation 5.1), methanol permeability is lower than that of acetic acid. In the literature, P for ethanediol across egg PC bilayers at 25 °C is 8.8×10^{-5}

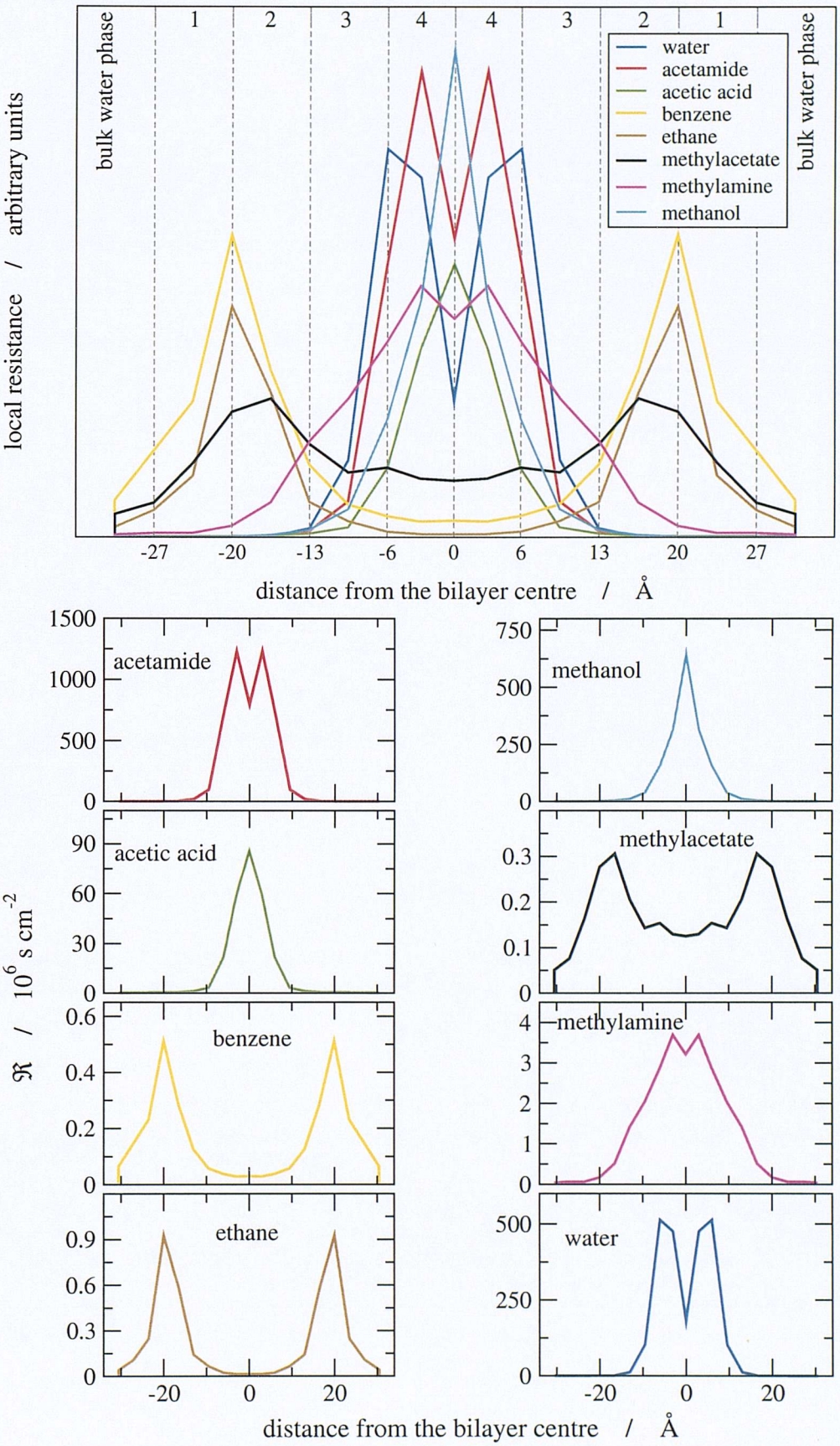


Figure 6.13: Local resistance profiles. For clarity errors bars are not plotted. Top: individual profiles are arbitrarily scaled to aid comparisons. Bottom: each profile is plotted alone.

solute	these MD simulations, 50 °C	experimental egg lecithins, T °C [ref.]	experimental pure DPPC, T °C [ref.]
acetamide	$6.57\ (\pm\ 1.92) \times 10^{-3}$	1.7×10^{-4}, 25 [2] 1.7×10^{-4}, 25 [121] $2.9\ (\pm\ 0.3) \times 10^{-4}$, 25 [135] $2.9\ (\pm\ 0.3) \times 10^{-4}$, 25 [136]	$\approx 4.5 \times 10^{-4}$, $\approx$48 [145] 1.65×10^{-4}, 25 [120]
acetic acid	$1.27\ (\pm\ 0.55) \times 10^{-1}$	6.9×10^{-3}, 25 [2] 6.6×10^{-3}, 22 [118] $5.0\ (\pm\ 0.2) \times 10^{-3}$, 25 [136]	$\approx 1.1 \times 10^{-2}$, $\approx$50 [129] $3.3\ (\pm\ 0.4) \times 10^{-2}$, 50 [128] $2.0\ (\pm\ 0.4) \times 10^{-2}$, 36 [127] †
benzene	9.91 (± 1.11)	-	-
ethane	6.75 (± 0.82)	-	-
methanol	$1.94\ (\pm\ 0.44) \times 10^{-2}$	-	-
methylacetate	9.50 (± 1.11)	-	-
methylamine	1.25 (± 0.16)	8.0×10^{-2}, 25 [2] $\approx 9.0 \times 10^{-1}$, 50 [124] ‡	-
water	$1.33\ (\pm\ 0.28) \times 10^{-2}$	3.4×10^{-3}, 25 [2] 2.2×10^{-3}, 25 [121] $1.9\ (\pm\ 0.9) \times 10^{-3}$, 25 [136]	2.5×10^{-2}, 30 [144] 2.2×10^{-2}, 25 [120] $31.5\ (\pm\ 1.1) \times 10^{-3}$, 37 [228] $7\ (\pm\ 3) \times 10^{-2}$, 77 [3] ††

Table 6.4: Calculated and measured permeability coefficients. In square brackets are references. P values are in cm s^{-1}, T are temperatures in °C.
†: measured in pure DMPC.
‡: The publication simply reports $P^{app} = 1.0 \times 10^{-6}$ at pH 6.7 and temperature 25 °C, and an activation energy of about 18 kcal mol^{-1}; with this information, considering methylamine $K_b = 3.38$ and assuming that only the neutral form crosses membranes at a significant rate, a value of 8.3×10^{-2} cm s^{-1} at 25 °C and the value reported in the table at 50 °C for the intrinsic permeability coefficient of the neutral form P_n^m were obtained using equation: $P^{app} = P_n^m f_n$, with f_n the fraction of neutral form.
††: value obtained in MD simulation by Marrink and Berendsen[3]

cm s^{-1} [2,121] and P for glycerol in the same system is 5.4 $\times$ 10^{-6} cm s^{-1}.[2,121] In the same system, P for 1,2-propanediol is reported to be 2.8 $\times$ 10^{-4} cm s^{-1}, almost identical to that of 1,4-butanediol which is 2.7 $\times$ 10^{-4} cm s^{-1}.[2,121] With a crude extrapolation, P for methanol could be thought to be around 10^{-4}-10^{-3}. The value from these simulations is about one order of magnitude larger, as for the other compounds.

For methylacetate, the most relevant experimental data are those regarding the relative permeability of mono- and di-acetate derivatives of glycerol:[141] the partial esterification of the polyol makes P about 8 and 16 times higher, respectively. Here the esterification of methanol with acetic acid makes P of methylacetate two orders of magnitude higher than that of methanol. Considering that the comparison is made with molecules whose size and number of hydroxyl groups are very different, that the esterification of glycerol in the experiments was incomplete, that the lipid composition and the temperature of the experiments and the simulations differ, the value calculated here for methyalcetate appears of the right order of magnitude or may indeed be one order of magnitude higher, as for the other compounds.

The experimental benzene permeability was not found in the literature. These simulations yield quite a high value. However, benzoic acid has P two orders of magnitude higher than that of acetic acid.[2,118,136]

The discrepancy between theory and experiment, with calculated Ps being one order of magnitude higher than experiment, may be due to the fact that the permeation free energy surface was not sampled completely, as Klein *et al.* suggested in a all-atom simulation study of halothane in a DPPC bilayer.[180] However, their simulation was much shorter and no constraints were applied to the solutes.

6.7.1 Water

The resistance profile for water has the same shape as that published by Marrink and Berendsen.[3] However, the amplitude is about 5 times larger because of the different values of $\Delta G(z)$ and $D(z)$ obtained here. This is reflected in the calculated permeability coefficient being about 5 times smaller. It is important to note that the resistance profile has its highest value at the border between region 3 and 4 and presents a dip in the middle of the bilayer. Regions 1 and 2 offer no significant contributions. The resistance is largely determined by the free energy profile and the influence of diffusion coefficients seems almost negligible. The typical shape of the water resistance profile can be explained by noting that the largest contribution to R comes

from the upper part of the lipid chains, where the density is the highest, whereas the lower part, which is more fluid and less dense, causes a dip in the middle of the curve.

6.7.2 Hydrophilic Compounds

Similar behaviour is found for the other hydrophilic solutes. As expected, the main barrier to permeation is represented by the hydrocarbon core of the membrane, where both solute partition and diffusion are less favorable than in the water phase. Regions 1 and 2 offer no resistance. However, compared with water, the highest value of R is shifted towards the interior of the membrane and is located in region 4. Moreover, while for acetamide and methylamine a dip in the curve is present in the middle of the bilayer, for acetic acid and methanol the peak is reached exactly in the bilayer centre. This behaviour is also shown by the latter molecules in the $\Delta G(z)$ profile as plotted in figure 6.1. The nitrogen-containing molecules have then a higher sensitivity to the physical state of the surrounding phase, while for the hydroxyl-containing molecules the main contribution comes from the chemical affinity. As for water, R is primarily determined by $\Delta G(z)$ and the contribution of $D(z)$ seems almost negligible.

6.7.3 Hydrophobic Compounds

For the hydrophobic compounds, namely benzene and ethane, the resistance profile is completely opposite. The main contribution to R is offered by the headgroup region, which is highly polar and charged. For methylacetate most of the resistance is due to the presence of the lipid headgroups, but regions 3 and 4 also offer a significant contribution to R. As for the hydrophilic compounds, R is primarily determined by $\Delta G(z)$.

As explained in the free energy section, although methylacetate is a hydrophilic and polar compound (its ΔG(water→hexadecane) is positive), for the purpose of this analysis it is included among the hydrophobic compounds, since its $\Delta G(z)$ in the middle of the bilayer is lower than at the water/lipid interface. The explanation for this effect may reside in the fact that, in contrast to the other hydrophilic solutes, methylacetate does not posses any hydrogen bond donors, or in other words it has no polar hydrogens. The lipid headgroups do not posses any either. Therefore, methylacetate is unique among the hydrophilic solutes in the sense that it can form no H-bonds with the surrounding lipids. This is also reflected in the local resistance profile: the main resistance to permeation is closer to the interface than in the bilayer centre.

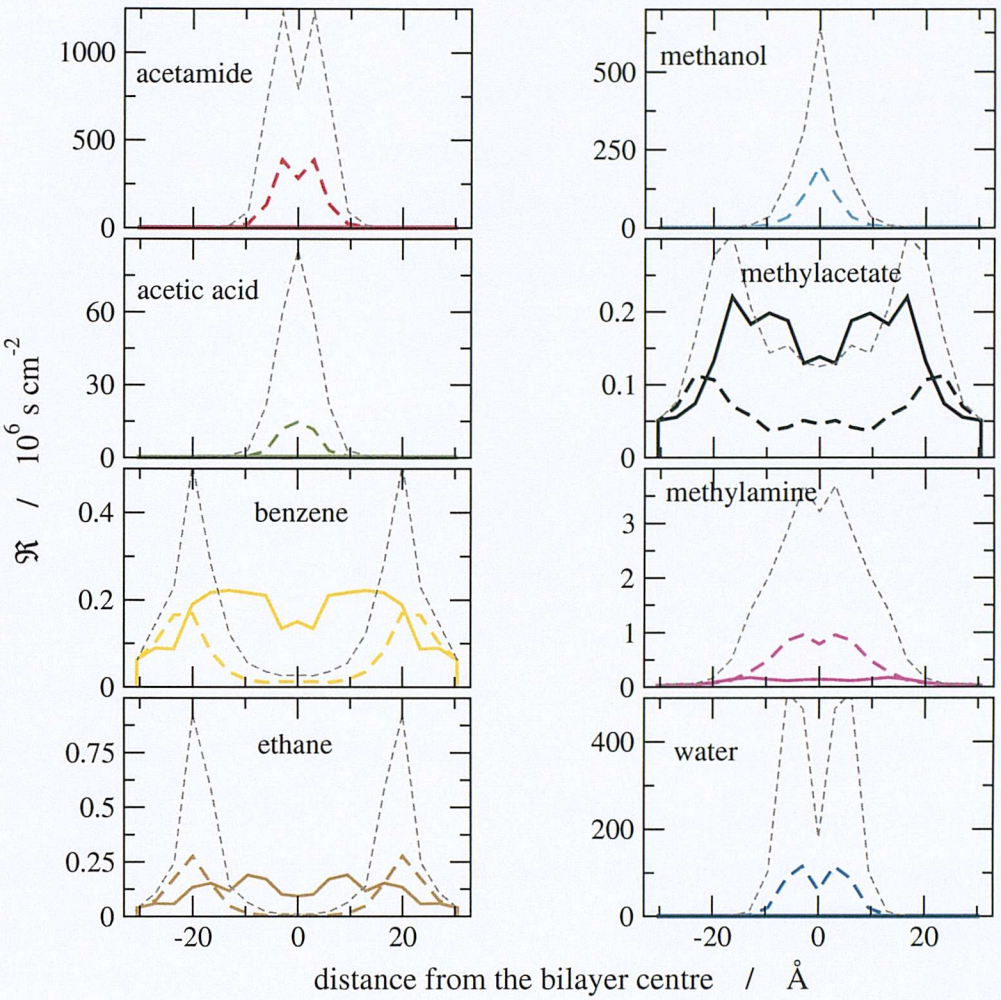


Figure 6.14: Contributions from free energies and diffusion coefficients to the resistance. Light gray short dashed lines plot the actual resistance profile $\Re(z)$, coloured long dashed lines are $\Re[noD](z)$, coloured solid lines are $\Re[no\Delta G](z)$ (see text for definitions).

6.7.4 Contributions to Resistance from Partitioning and Diffusive Behaviours

To understand the separate contributions of $D(z)$ and $\Delta G(z)$ to $\Re(z)$, the local resistances were also calculated by either considering the diffusion coefficient into the membrane to be identical to that in bulk water, or by considering there to be no free energy barrier in crossing the membrane. In the first case, $\Re(z)$ was calculated with $D(z)$ equal to $D(30.5)$ at all depths and it is referred to as $\Re[noD](z)$. In the second case, $\Re(z)$ was calculated with $\Delta G(z) = 0$ at all depths and it is referred to as $\Re[no\Delta G](z)$. These plots are reported in figure 6.14. Light gray short dashed lines plot the actual resistance profile $\Re(z)$, coloured long dashed lines are $\Re[noD](z)$, coloured solid lines are $\Re[no\Delta G](z)$.

It is clear that for the hydrophilic compounds, for which the main barrier to permeation is in the middle of the membrane, $\Re[noD](z)$ is far closer to the actual $\Re(z)$ than $\Re[no\Delta G](z)$,

suggesting that the contribution from $\Delta G(z)$ is far more significant, i.e. the partition properties of the solutes prevail over their diffusion behaviour in controlling the overall permeation process. However, although $\Re[no\Delta G](z)$ almost disappears onto the x-axis of those plots, considering only $\Delta G(z)$ for studying permeation is not enough, since the contribution from $D(z)$ has the important role of scaling the profile obtained by considering only $\Delta G(z)$ up to the correct value, which is approximately 3 to 5 times larger. That is, the diffusive behaviour inside the membrane is not negligible.

Results for the hydrophobic compounds are different. For the purpose of these analysis, methylacetate is included among the hydrophobic compounds, since the main barrier against its permeation is located at the interface instead of the middle of the membrane. For these solutes, the contribution from $D(z)$ is more important, as $\Re[no\Delta G](z)$ is as large or even larger than $\Re[noD](z)$. The two curves are very close in the part of the membrane offering the main resistance to permeation, i.e. at the interface. In contrast, in the middle of the membrane the curve of the actual $\Re(z)$ is much closer to $\Re[noD](z)$. Methylacetate is a special case. Its $\Delta G(z)$ profile is very close to zero, so that its $\Re[no\Delta G](z)$ is practically the same as $\Re(z)$. In this case, the diffusion behaviour is even more important than the partition properties of the solute.

6.7.5 Correlation Between Permeability and Partition Coefficient

By analogy with equation 6.8 (which relates the partition coefficient of a solute in two different solvents), and assuming that the main contribution to the permeation resistance comes from a distinct and uniform barrier region inside the lipid bilayer, a linear free energy relationship correlating the permeability coefficient P with the partition coefficient $K_{org/w}$ in a reference organic solvent can be derived:[118,133,138]

$$\log P = s \log K_{org/w} + i \tag{6.15}$$

The slope s measures the relative chemical affinities of the solutes for the barrier domain in the bilayer versus the organic solvent chosen for the correlation. The solvent yielding $s = 1$ in equation 6.15 exactly matches the chemical selectivity of the bilayer barrier domain.[133,138]

From experiments, many solvents yield fairly good results. Depending on the set of solutes, the type of membrane, and the method employed to measure P and $K_{org/w}$, some authors have

found that the solvents yielding s closest to 1 are saturated long chain hydrocarbons like hexadecane,[2,118,120,121] confirming that the main barrier to permeation is the hydrocarbon core of the lipid bilayer, whereas some others have had better results with more polar/polarizable solvents like unsaturated hydrocarbons,[127,128,136,138] suggesting that the barrier domain is probably located closer to the interface. These empirical observations are in accordance with the so-called Overton's rule,[229] which states that solute permeability is correlated with solute partitioning in oil/water systems.

These simulations allow for the direct calculation of K at every depth of the lipid bilayer, so no reference solvent is needed. Considering the K values obtained with equation 6.2 from the z-depths which yield the highest $\Delta G(z)$ or $\mathfrak{R}(z)$ values, should be equivalent to considering the K values in the solvent which best mimics the barrier region of the lipid bilayer. However, plotting the logarithm of P values obtained from these simulations *vs.* the logarithm of the above defined K values, yields $s \approx 0.88$, with a correlation coefficient higher than 0.99. If K in hexadecane are employed instead, $s \approx 0.50$, with a correlation coefficient ≈ 0.90.

The failure to reproduce the experimental observation ($s \approx 1$) by these simulations may be explained by the low number of solutes studied here and by the uncertainties intrinsic in the MD simulations. But also the shape of the resistance profiles plotted in figure 6.13 can give an important indication. In experiments, the resistance profile is approximated to a rectangular shape: the resistance is zero outside the barrier domain and at a constant value higher than 0 inside the barrier. These simulations instead reveal that the shape is far more complicated and smooth, and the width of the barrier region is quite different for different solutes. Therefore, considering the partition at the z-depth with the highest $\Delta G(z)$ or $\mathfrak{R}(z)$ is an oversimplification which does not hold for the permeability coefficients calculated from these simulations. This result does not imply, however, that in the experiments the choice of a good reference organic solvent yielding $s = 1$ is not a reasonable model for the barrier opposing solute permeation.

6.7.6 Permeability Dependence on Molecular Size

As mentioned in section 6.6.11, direct measurement of diffusion coefficients inside the membrane from experiment is not possible. To study the dependence of D on solute size, the logarithm of the ratio between the experimentally measured permeability coefficient P across membranes and the partition coefficient K in an organic solvent / water systems, $\log(P/K)$, is plotted versus $\log(M)$ or $\log(V_d)$.[2,136,218] K in organic solvent / water systems is used as an estimation of K in

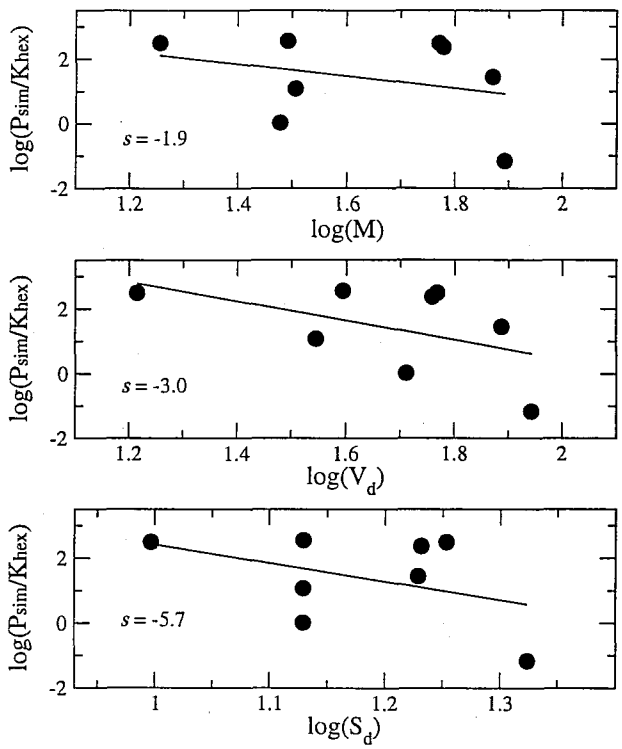


Figure 6.15: Log-log plots of permeability-partition ratio versus solute mass M, volume V_d and cross-sectional area S_d. s values are slopes of linear regressions.

the membrane barrier region. According to equation 6.14, the ratio between permeability and partition coefficients, the so-called "hydrophobicity-corrected" permeability, yields the value of the solute diffusion coefficient D in the membrane barrier region, divided by the barrier region thickness d. The same approach to study permeability size-dependence is used here. Figure 6.15 contains the log-log plots of the ratios between permeability coefficients as calculated from these simulations and experimental partition coefficients in hexadecane versus solute mass M, volume V_d and cross-sectional area S_d. The s values reported in these graphs are the slopes of linear regressions. That is the exponent in the following relationship:

$$\frac{P}{K} = \frac{D}{d} \propto A^s \tag{6.16}$$

where A is a solute property related to the size, i.e. either mass M, volume V_d or cross-sectional area S_d.

The correlation coefficients for these plots are low, ≈ 0.4. Correlation coefficients from some experiments are very low too, varying between 0.12 and 0.70[2,136] depending on the choice of solutes and reference solvent employed for K. The reason for the poor correlation from the simulations may reside in the low number of data points (eight solutes, whereas in experiments more than 20 solutes are examined[2,136,218]), in the narrow range of sizes of the solutes studied

(some experiment also includes large molecules such as nucleosides and steroids[136]), and in the uncertainties intrinsic to the simulations. However, in figure 6.15 there is the indication of a correlation between solute size and P-K ratios. Thus, although the correlation is too poor here to conclude anything significant, the experimental trend is broadly reproduced. This observation has led experimental scientists to think that the diffusive behaviour inside the membrane is strongly dependent on solute size, more than it is in water or simple liquids and rather similar to that in synthetic polymers.

These simulations allow for the direct calculation of partition coefficients K at every depth of the lipid bilayer. Instead of K in hexadecane, if one uses in the above equation 6.16 the K values obtained with equation 6.2 from the z-depths which yield the highest $\Delta G(z)$ or $\Re(z)$, correlation coefficients for the $\log(P/K)$ *vs.* $\log(M)$ or $\log(V_d)$ plots are only slightly better ($\approx$ 0.5), but slopes are $\approx$ -0.6.

The following conclusion can be drawn here. Although correlation coefficients are low, the simulations reproduce the experimental findings that $\log(P/K)$ *vs.* $\log(M)$ or $\log(V_d)$ yields negative slopes with magnitude higher than 0.6. However, when using "real" K values between the water and the lipid bilayer, instead of the K values in a bulk organic solvent like hexadecane, a much smaller size dependency is obtained. This agrees with the $D(z)$ *vs.* $\log(M)$ or $\log(V_d)$ plots (see sections 6.6.9 and 6.6.10), which yield low slopes. Therefore, the experimental approach using K in bulk solvents overestimates the size dependence of diffusion coefficients inside the membrane.

6.8 Bilayer Structure

It has been reported in previous publications that the presence of small solutes at low concentration does not disrupt significantly the structure of lipid bilayers.[104,179,180] This is the reason why in these simulations the surface area was chosen to be kept constant at the value which was shown to yield the best agreement with experimental bilayer physical properties.[91] To validate the assumption that five small organic molecules could be simulated in such a model bilayer, bilayer thickness and lipid order parameters were calculated from trajectories saved in these simulations and compared with those from the simulation of the pure DPPC bilayer described in chapter 4. Comparisons showed that with and without permeants the lipid bilayer structure is essentially the same.

The electron density along the bilayer normal can be determined with X-ray spectroscopy

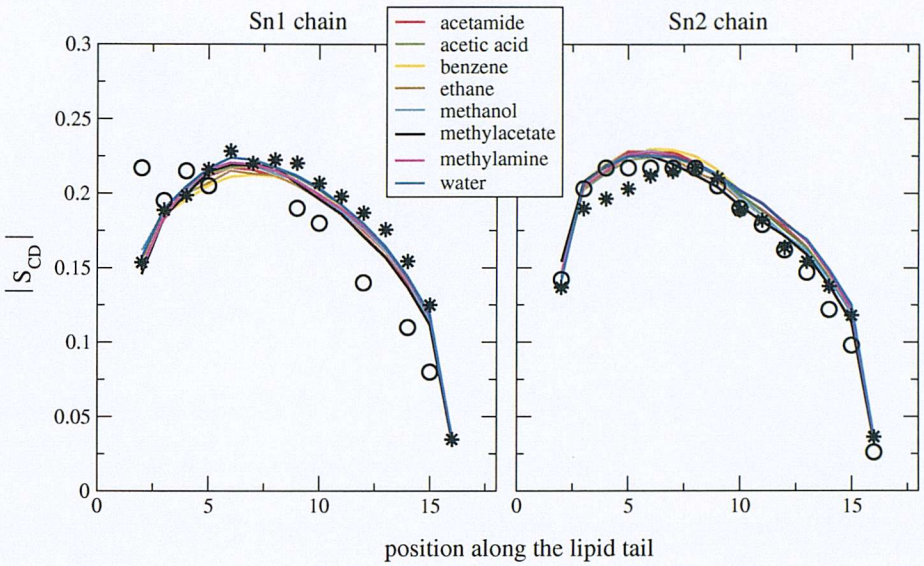


Figure 6.16: Lipid tail order parameters. Values from the pure DPPC simulation are given as stars and open circles are experimental values.[4–6]

and the peak-to-peak distance is used as a measure of the bilayer thickness. Experimental values range between 36.4 and 39.6 Å[91] and a value of 37.3 ± 0.5 Å was obtained in the pure DPPC simulation. When inserting the solutes, simulations yielded the following results: acetamide 37.3 ± 0.1, acetic acid 37.8 ± 0.1, benzene 37.4 ± 0.1, ethane 37.8 ± 0.1, methanol 37.4 ± 0.1, methylacetate 38.1 ± 0.1, methylamine 38.1 ± 0.1, water 38.1 ± 0.1. It must be said that some of these values are a little higher than that obtained in the simulation of the pure DPPC bilayer. However, they are well within the experimental range. Moreover, the pure DPPC simulation lasted 1 ns, whereas permeation studies consisted of many 2 ns simulations. It must be noted that the largest deviation from the pure lipid simulation is obtained with water, the smallest permeant, and the smallest deviation with benzene, the largest solute. It thus appears that these fluctuations are not due to the presence of the solute. In the original paper using the NP_NAT ensemble and the CHARMM force field, a peak-to-peak distance of 38.9 Å was obtained.[91]

The order parameter of the C-H bonds along the lipid chains can be determined by means of ^{2}H-NMR spectroscopy and it is used as a measure of the lipid order. In figure 6.16 the order parameters S_{CD} for the pure DPPC bilayer and for the permeant/lipid systems are plotted as absolute values. Values from the pure DPPC simulation are given as stars and open circles correspond to experimental values.[4–6] The coloured lines are values from these permeation studies. It is clear that the presence of the solutes does not significantly affect the lipid structure. Furthermore, no differences in S_{CD} were found in a simulation of 4 halothane

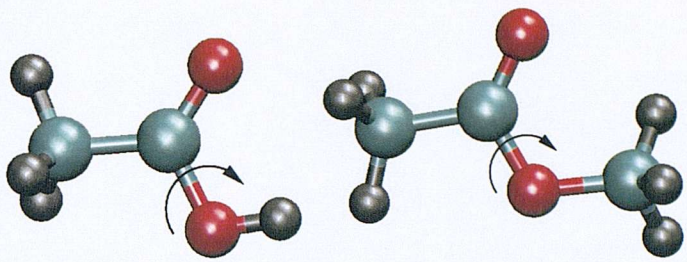


Figure 6.17: Significant dihedral torsion in acetic acid (left) and methylacetate (right). Hydrogens are in gray, carbons in cyan, and oxygens in red.

molecules in a bilayer of 64 DPPC molecules.[180] However, a small decrease in the order of the upper part of the chains was observed in a simulation of 4 benzene molecules in a bilayer of 36 DMPC lipids.[104] Experiments reveal a decrease in order parameter only at high solute concentration.[104]

6.9 Atomic Detail

6.9.1 Solute Flexibility

The compounds studied here are quite rigid. Among the internal degrees of freedom only bonds involving hydrogen atoms were constrained. However, significant conformational changes do not in general occur. The main variations in molecular structure are related to the dihedral angles in the carboxylic group of acetic acid and the ester group in methylacetate. These are shown in figure 6.17. The dihedral angle ϕ is defined as the torsion O=C-O-H in acetic acid and the torsion O=C-O-C in methylacetate.

In acetic acid ϕ is quite stable, being at $0 \pm 30°$ in most of the simulations and for most of the time. This conformation can be referred to as "closed" conformation, since the carboxylic hydrogen is oriented towards the oxygen with which it is not bonded. Among the 50 (= 10 depths $\times$ 5 x-y positions) acetic acid molecules simulated, all had ϕ at 0° in the starting structure and only six undertook a rotation to 180° in the equilibration period. Of these six, only four undertook a second conformational change back to 0°. For two of them this change occurred in the first 200 ps of the production run, and for the other two it occurred after 1.1 and 1.5 ns, respectively. When $\phi = 180°$, the conformation can be referred to as "open", since the carboxylic hydrogen is oriented towards the outside of the molecule. These analyses revealed that there is no relationship between the value of the dihedral angle and the solute position along the bilayer normal.

All of the 50 methylacetates had $\phi = 0 \pm 30°$. One only showed a rotation to 180° after

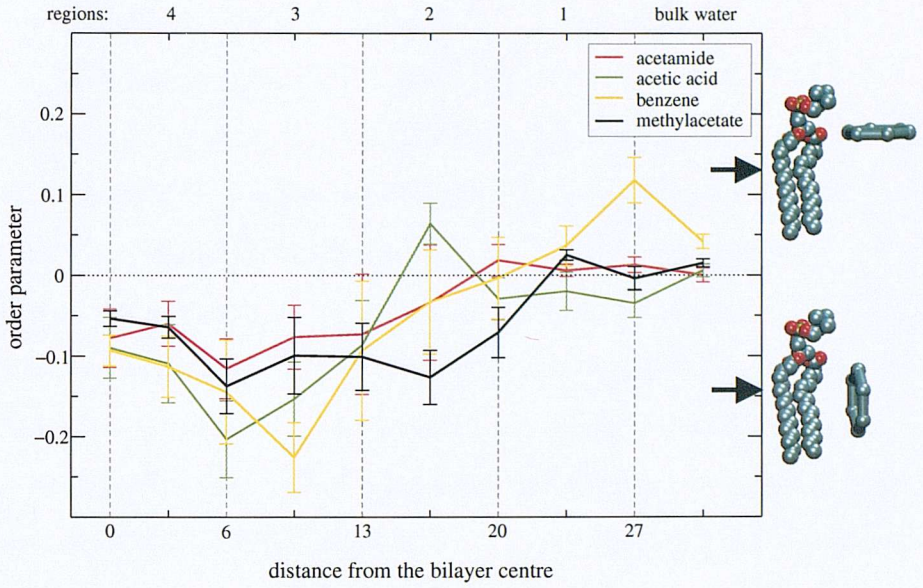


Figure 6.18: Solute order parameter as a function of depth. For clarity, a picture indicates the physical meaning of positive and negative values of S. In that picture, the size ratio between lipid and solute (benzene) molecules is not respected.

700 ps and ϕ remained at this value for the rest of the simulation. As for acetic acid, the value of the torsion is uncorrelated with the solute position, i.e. the solute internal conformation is not affected by the location along the bilayer normal.

6.9.2 Solute Orientation

To monitor the solute behaviour inside the membrane, the orientation of its principal axis was calculated. For the largest compounds, namely benzene, methylacetate, acetic acid and acetamide, the angle θ between the normal to the molecule's principal plane and the bilayer normal was obtained from trajectory analysis. A measure of the degree of order in a particular system is given by the so-called order parameter S:

$$S = \frac{1}{2}\langle 3\cos^2\theta - 1\rangle \tag{6.17}$$

where the brackets denote an ensemble average. S varies between 1.0 (indicating full order perpendicular to the bilayer normal) and -0.5 (indicating full order along the bilayer normal). A value of zero is considered to indicate that full isotropic motion is occurring. Figure 6.18 plots the solute order parameter as a function of depth in the bilayer. Standard errors are calculated from the difference of the mean $\langle S\rangle$ in each of the five simulations from their overall average value for each depth.

It can be seen that moving from the water phase towards the membrane interior, the perme-

ants tend to be aligned parallel to the lipid tails. This was expected, since the lipid molecules are tightly packed because of interfacial constraints, and the free space for solute permeation is therefore parallel to the bilayer normal. For all the compounds, the greatest disorder, i.e. S closest to 0, occurs in bulk water and in region 1, as expected. The highest order parallel to the lipid tails, i.e. S closest to -0.5, occurs in region 3. In region 4, in the lower part of the lipid chains, the solutes are preferentially ordered along the bilayer normal but to a lesser extent than in the upper part of the chains, primarily because of the lower density and the higher lipid mobility in region 4. The value of S is closest to -0.5 for benzene, which is the largest compound among those studied. However, the difference in order parameter among the solutes is very small and arguably the size is not the most important factor determining the solute orientation.

Interestingly, at 27 Å far from the bilayer centre, i.e. at the interface between bilayer and bulk water phase, the benzene molecule tends to lie on the lipid heads parallel to the bilayer surface, and this may be related to the behaviour of the phosphorus-nitrogen vector in the choline fragment, which tends to lie on the bilayer plane. Remarkably, the same behaviour was found in simulation studies of indole molecules inside a POPC bilayer:[230] the order parameter of these aromatic molecules is reported to be positive in the headgroup region and negative in the membrane interior, with values in region 4 closer to 0 than those in region 3.

Results from these simulations therefore suggest that all the solutes have a preferred orientation, to a greater or lesser extent. This is in contrast to that reported in previous simulations of benzene molecules inside a DMPC bilayers.[103,104,108] However, the surface area per lipid in that system was about 66 $Å^2$, while here it is 62.9 $Å^2$. When the benzene molecule is more tightly packed in the lipid environment, its orientation is preferentially parallel to the lipid chains. Moreover, the simulation temperature was the same as that employed here, but DPPC bilayers enter the biologically relevant liquid-crystalline phase at 42 °C, while DMPC enters at 22 °C.[30] When the lipid bilayer is more fluid, permeants can move and rotate to a larger extent. No direct experimental studies have been found in the literature regarding benzene or the other small solutes simulated here. However, the preferred orientation parallel to the bilayer normal was reported for long chain alkanes,[198,204] fluorescent probes[231] and drug molecules.[232,233]

The findings from these simulations regarding the preferred solute orientation inside the lipid bilayer agree with the experimental observation that water/membrane partitioning is highly affected by entropic effects, as mentioned in section 6.5.6.2. In those analyses, the theoretical

model, which uses the Barclay-Butler relationship between entropy and enthalpy of solvation (equation 6.9), had no experimental confirmation, since the experiments do not yield atomic details. In contrast, these simulations offer such details and support the previous hypothesis: when moving from water into the membrane, solutes lose degrees of freedom, because they are forced to stay oriented parallel to the lipids. For molecules larger than those studied here, this effect is expected to be even stronger, and so it can be argued that partitioning into lipid bilayers, and in turn permeabilities, are also highly size-dependent properties.

6.9.3 Solute Orientational Correlation Times

To evaluate the solute ability to rotate around its centre of mass at different depths in the membrane, the following time autocorrelation functions were studied:

$$C(t) = \langle P_2[\vec{\mu}(t) \cdot \vec{\mu}(0)] \rangle \tag{6.18}$$

P_2 is the Legendre polynomial of rank 2 and $\vec{\mu}$ is the unit vector along the normal to the molecular plane as defined previously to calculate the angle θ with the bilayer normal (see section 6.9.2). The equation turns into:

$$C(t) = \frac{1}{2}\langle 3\cos^2\alpha - 1 \rangle \tag{6.19}$$

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

$$C(t) = C(0)\exp(-t/\tau) \tag{6.20}$$

From the fitting, the relaxation times, τ, for solute reorientation were obtained and plotted in figure 6.19. A general trend is observed: as expected, τ increases entering the membrane and reaches its highest values in regions 2 and 3, the densest parts of the bilayer. Benzene has the highest values of τ, since it is the largest permeant and its rotations in the packed lipid environment are the most difficult. The standard errors in figure 6.19 were calculated from the differences of the five independent simulations from their average. This shows that for the longest τs, the variability between the five simulations was the highest.

Benzene diffusion in the hydrocarbon region has been previously simulated.[104] Reorientational correlation times in that publication were found to be $\approx$ 25 ps. These simulations yield values of the same magnitude in region 4. However, in region 3 these simulations yield values

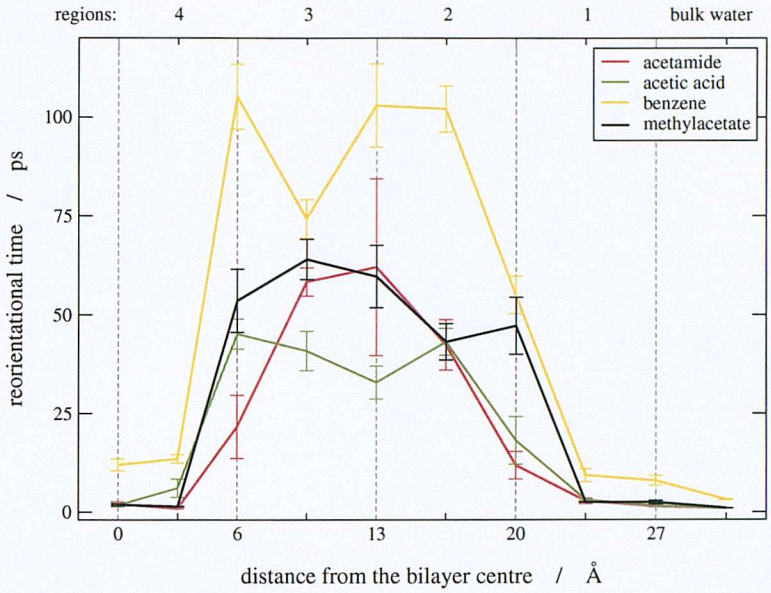


Figure 6.19: Solute reorientational correlation times as a function of depth in the membrane.

3 to 4 times higher. This is associated with the more restricted benzene sampling observed in these studies.

6.9.4 Polar Group Orientation

Many of these solutes are polar compounds, and it was interesting to investigate the orientation of the polar group with respect to the apolar part of the molecule. A polar vector was identified as the vector linking the carbon atom and its heteroatom(s): the C-O bond in methanol, the C-N bond in methylamine, the vector connecting the carboxyl carbon and the centre of geometry of the two carboxyl oxygens in acetic acid, the vector connecting the amide carbon and the centre of geometry between amide oxygen and nitrogen in acetamide. The direction of the vector is from the carbon towards the heteroatoms. From the angle between the polar vector and the bilayer normal an order parameter was calculated and plotted in figure 6.20. In this case a value of 1 indicates full order parallel to bilayer normal and a value of -0.5 full order perpendicular to the bilayer normal. Results agree with those plotted in figure 6.18: the polar vector is mostly oriented parallel to the lipid chains, above all in the densest regions 2 and 3. To understand whether the polar group is oriented towards the water phase or the hydrocarbon core of the membrane, the ratio between the z component of the polar vector and its total length was calculated and plotted in figure 6.21. This yields the cosine of the angle between the polar vector and the bilayer normal, in other words the Legendre polynomial of first order. A value of 1 indicates that the polar group is oriented towards the water phase and a value of -1

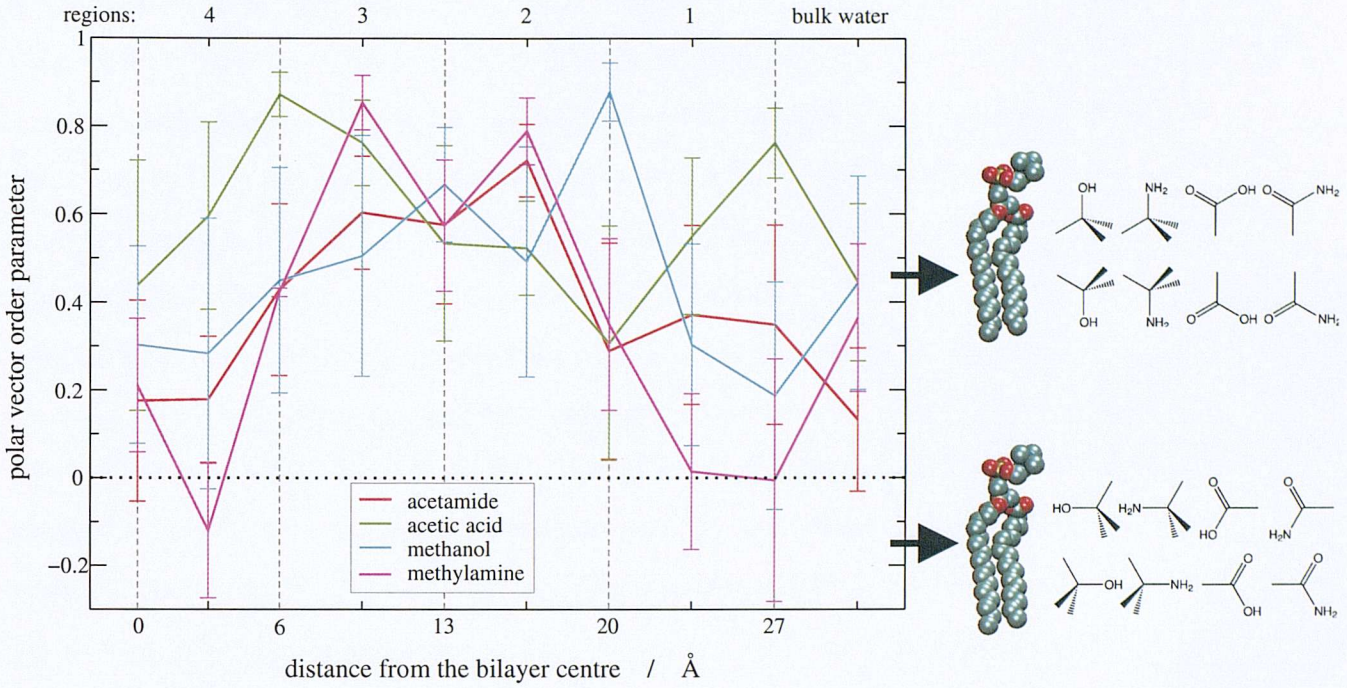


Figure 6.20: Polar vector orientation with respect to the bilayer normal.

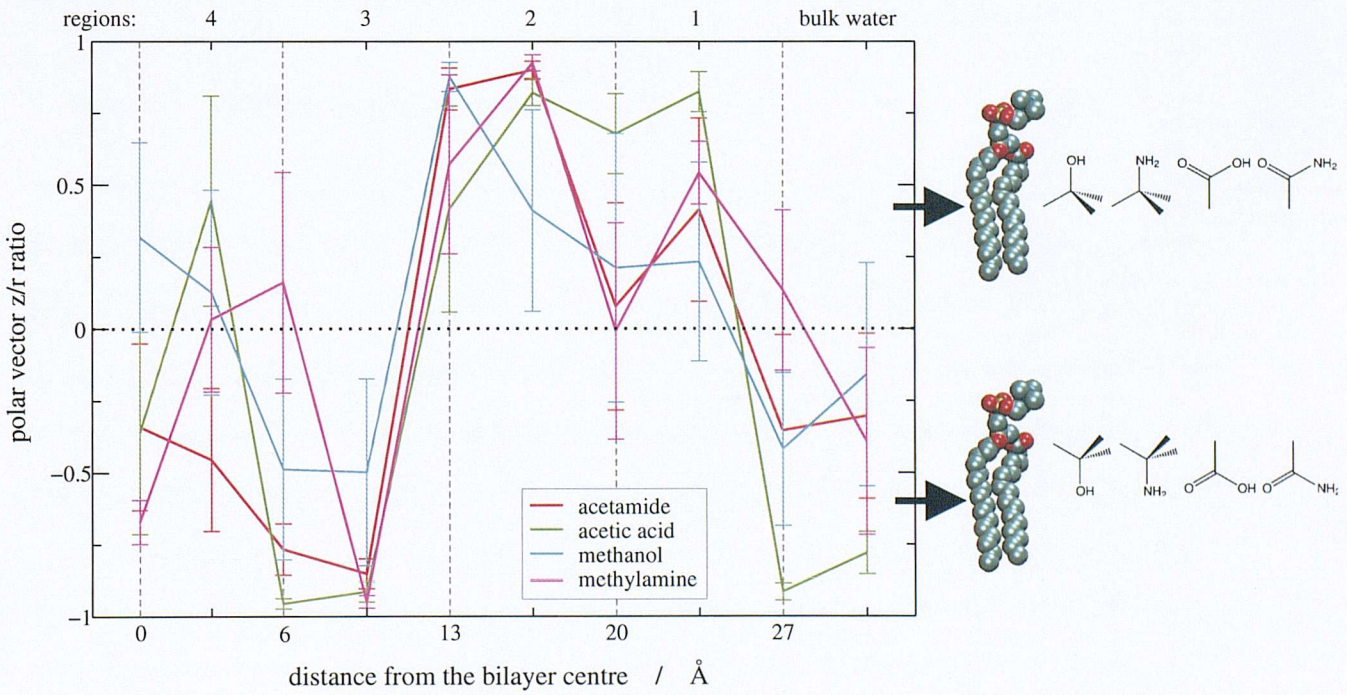


Figure 6.21: Ratio between the z component of the polar vector and its total length.

indicates that the polar group is oriented towards the middle of the bilayer. A value of zero does not necessarily mean that the polar vector is oriented perpendicular to the bilayer normal, but rather that on average there are as many vectors oriented towards the water phase as towards the membrane interior. This is the reason why neither the polar vector order parameter nor the z/length ratio are sufficient, when considered alone, to give a complete picture of the solute orientational behaviour. Interestingly, polar groups tend to be oriented towards the lipid headgroups when the solutes are closer to the interface, while they tend to be oriented towards the centre of the bilayer when the solutes are in the middle of the membrane. Lipid headgroups are highly charged and have oxygens available for H-bonds, so it is thought that short range electrostatic interactions and local hydrogen-bonding networks may be responsible for the behaviour of the permeants at the interface. Moreover, when the polar group is oriented towards bulk water, the alkyl part can penetrate the hydrocarbon core of the membrane. This kind of solute orientation is also reported for drug molecules by both experiments[232–235] and simulations.[109] On the other hand, it is hard to understand the behaviour observed at the end of region 3 where the polar groups tend to be oriented towards the centre of the membrane. It is expected that, when the solutes are in the middle of the bilayer, the interactions with the headgroups and waters are negligible and they do not affect the solute orientation. This is indeed what happens in region 4.

6.9.5 Hydrogen Bonds

The number of hydrogen-bonds between the permeants and the surrounding waters and lipids was computed and plotted in figure 6.23 as a function of depth. The following criteria were adopted for the identification of a H-bond:[82] the distance between the donor and the acceptor atoms (oxygen or nitrogen) is less than or equal to 3.25 Å and the angle between the vector linking acceptor and donor atoms and the donor-hydrogen bond is less than or equal to 35°. This geometry is show in figure 6.22. The number of H-bonds involving a molecule is counted during the simulation and then divided by the number of coordinate sets analysed. The number plotted in figure 6.23 is thus the mean number of H-bonds per simulation frame, or in other words the mean instantaneous number of H-bonds. This is also averaged over the five solutes constrained at the same z depth and, from the differences between them, the standard errors are obtained.

Despite the strong polarization of the O-H bond, which is modelled in atomistic simulation

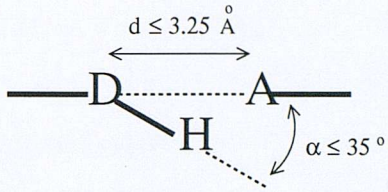


Figure 6.22: Geometrical criteria for hydrogen bonds. D is donor, A acceptor, H hydrogen.

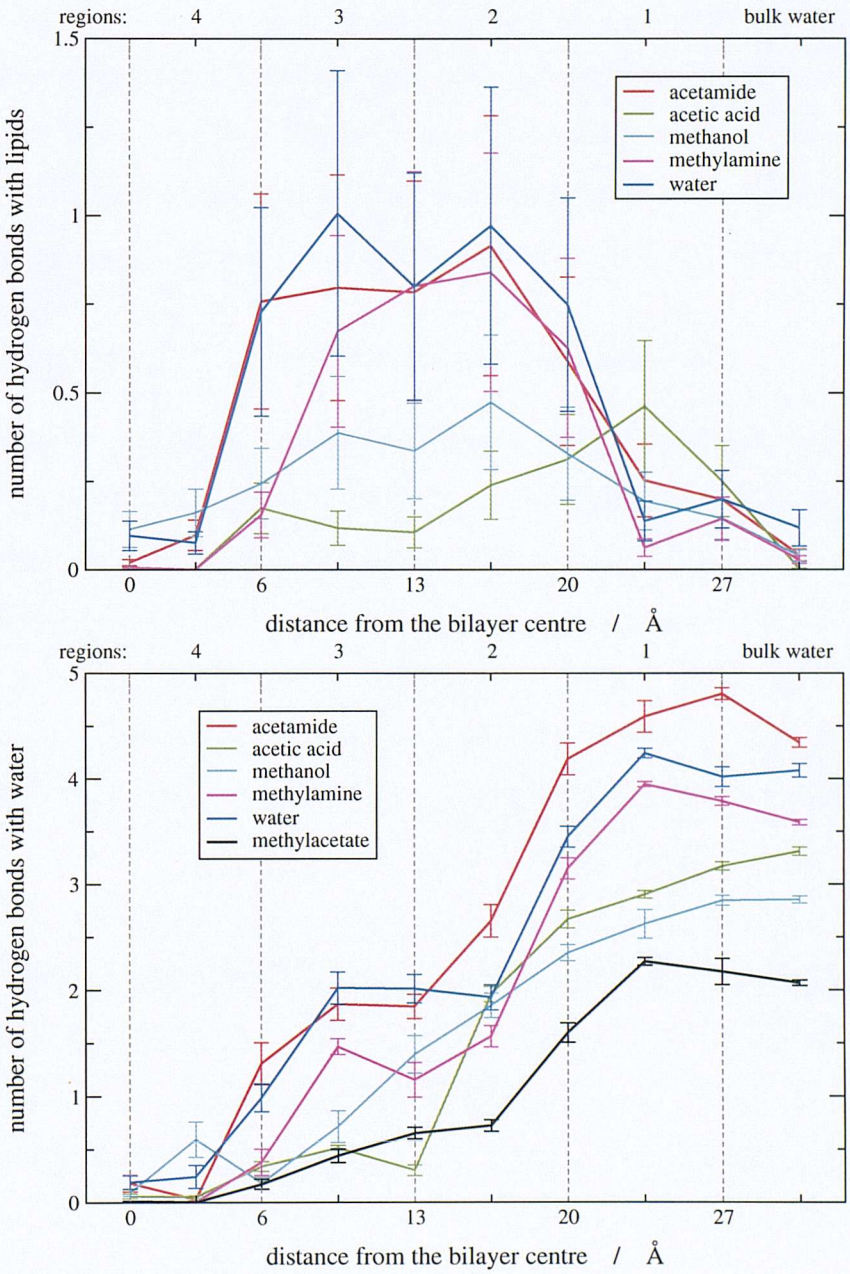


Figure 6.23: Hydrogen bonds.

with a charge of -0.834 e on the oxygen and +0.417 e on the hydrogen, water is not always the solute involved in the largest number of H-bonds. Nitrogen-containing solutes, namely acetamide and methylamine, generally form more H-bonds than oxygen-containing solutes, namely acetic-acid and methanol, although the differences are very small. The mean H-bond lifetime was also calculated and plotted in figure 6.24. This is defined as the sum over all the H-bonding events of the duration of H-bonds between two atoms, divided by the number of different H-bonds formed by those atoms. Although the H-bond lifetime is very similar for all the solutes, for acetic acid and methanol they generally last longer. This observation may be related to the fact that, even though their size is almost the same, methanol has lower diffusion coefficients than methylamine, and acetic acid has lower diffusion coefficients than acetamide: since the solutes containing the hydroxyl group are involved in longer-lived H-bonds, their mobility may be reduced. It can also be noted that the mean H-bond lifetime involving solutes located in regions 2 and 3 is usually longer than for solutes located in regions 1 and 4.

6.9.6 Acetic Acid Anomalous Behaviour

It should be noted that H-bonds between polar permeants and water were non-zero even in the hydrocarbon core of the membrane. The hydration in the middle of the bilayer was analyzed in detail for one of the simulations with acetic acid, as an example, since the mean lifetime of H-bonds at $z = 0$ Å for this solute was surprisingly high. At the end of the simulation, that is after 2 ns, it was revealed that a column of water molecules entered the membrane and reached the acetic acid molecule which is located in the middle of the hydrocarbon core of the bilayer. This is shown in figure 6.25. The acetic acid orients its polar fragment towards the water/lipid interface and towards the water column. A hydrogen bond network is clear: oxygen and hydrogen atoms alternate in the acetic acid and water column. It must be noted that the value of $\langle \boldsymbol{F}(z) \rangle_t$ for the hydrated acetic acid in the middle of the bilayer is less than that for the non-hydrated molecule at the same depth, i.e. hydrated acetic acid crosses a lower free energy barrier during membrane permeation. After studying the "history" of the water column, it was found that a single water molecule entered the hydrocarbon core, reaching the acetic acid in the middle of the membrane and H-bonding to it. This occurred 0.55 ns after the beginning of the simulation. This "complex" was very stable and lasted for about 1 ns. After another 0.6 ns, in which the complex freely rotated around the solute centre of mass and moved on the x-y plane, other water molecules began moving towards the water/acetic acid complex, which at

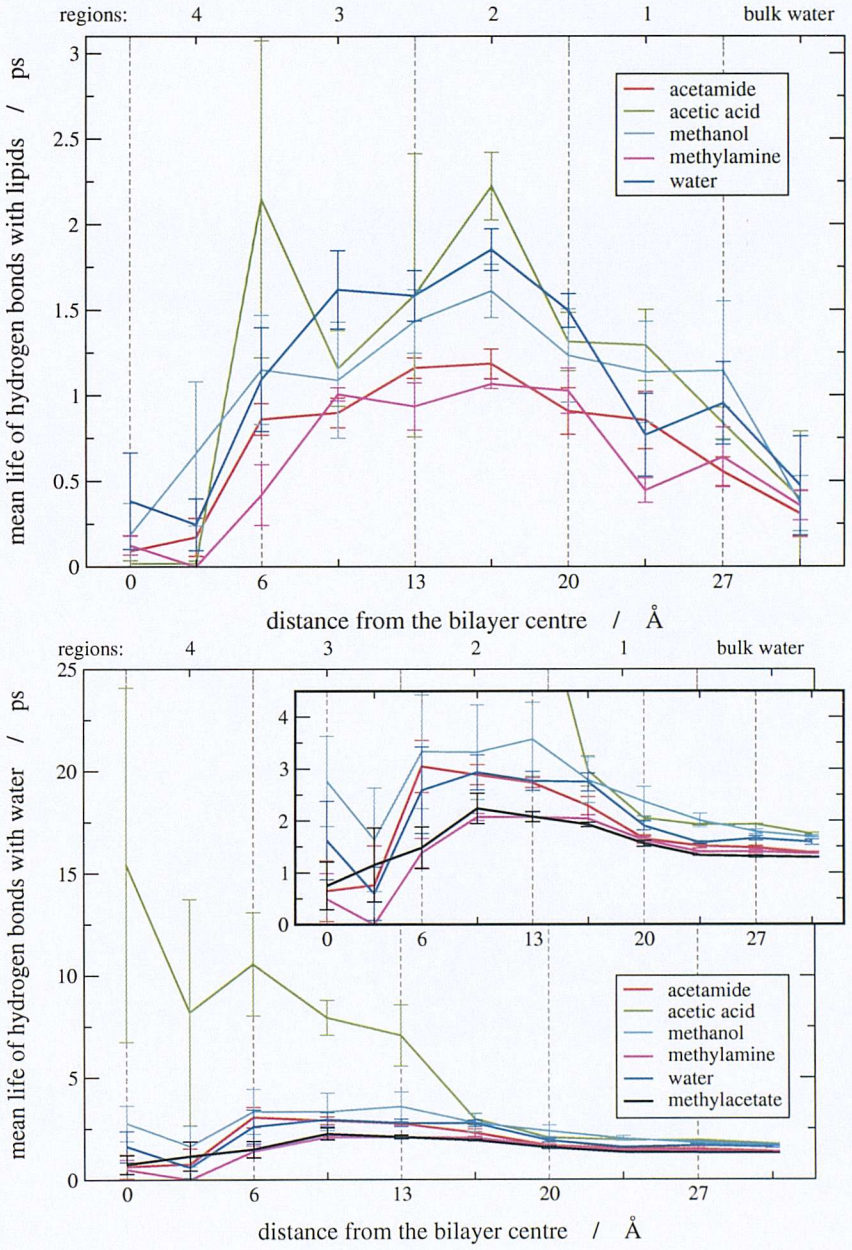


Figure 6.24: Hydrogen bond mean lifetime.

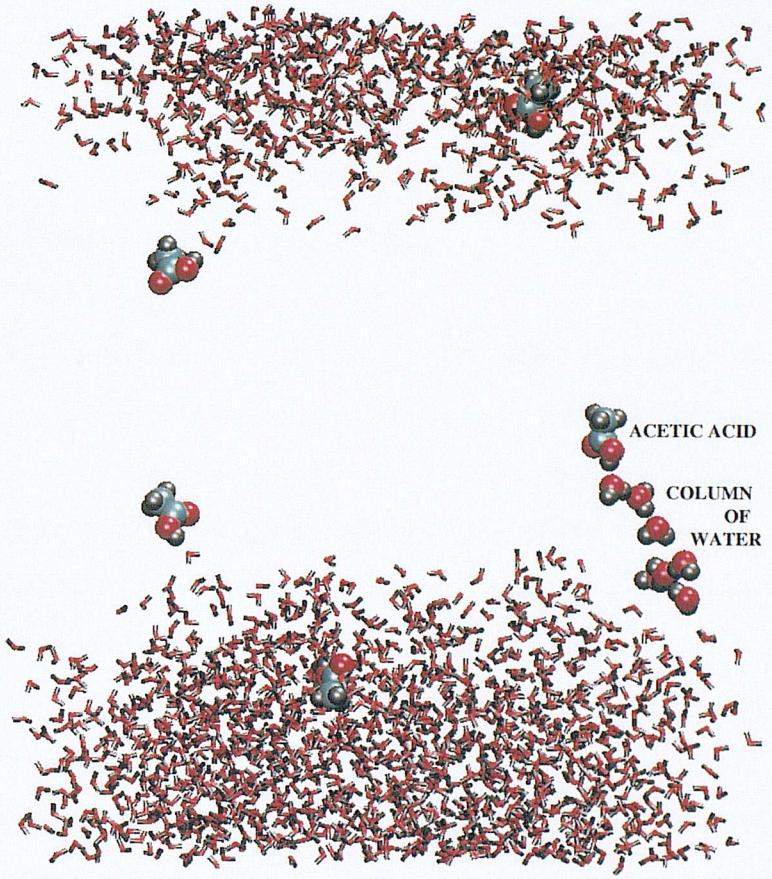


Figure 6.25: Hydration of acetic acid. The water layers on the two sides of the membrane are shown. For clarity, lipids are omitted. Acetic acid and water molecules forming a column hydrating the solute are highlighted.

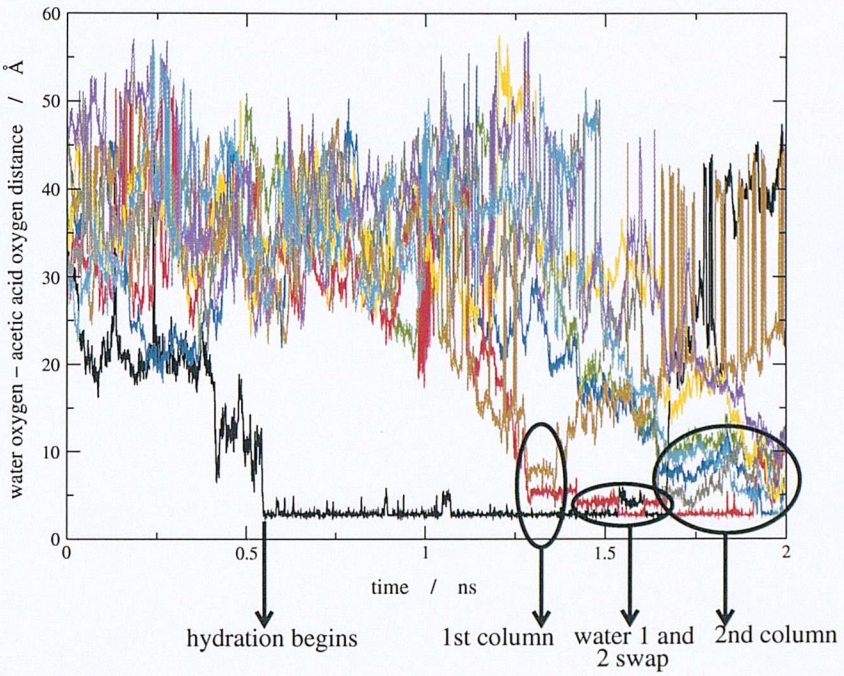


Figure 6.26: History of the water molecules hydrating the acetic acid constrained in the middle of the bilayer.

that moment was oriented with the water towards the headgroup region. After 1.3 ns from the beggining of the simulation a first column of water was formed, similar to that of figure 6.25. After this, the complex stopped rotating and reduced its motion in the x-y plane. After another 0.2 ns this first column was disrupted and only two water molecules remained hydrogen-bonded to the acetic acid in the middle of the bilayer, including the first which reached the solute 0.9 ns earlier: one molecule hydrated the carbonyl oxygen of the acetic acid and the other hydrated the hydroxyl group. This new three-membered di-hydrated complex was again free to rotate around the centre of mass of the acetic acid and to move in the x-y plane. After another 0.2 ns, a second column of water was formed, but this time with water molecules belonging to the opposite side of the bilayer. This column of water was still present at the end of the simulation. However, the individual water molecules in the column exchanged their relative positions, some left and new water molecules arrived. The closer the water to the acetic acid, the longer it stayed in the column. The first water reaching the solute remained complexed for about 1 ns, until 1.55 ns after the simulation started. Eventually, it swapped its location with the second water in the column twice in the space of 0.1 ns, and then finally left. Remarkably, it moved towards the opposite side of the bilayer to that from which it entered. Therefore, in less than 2 ns this water molecule crossed the entire thickness of the lipid membrane. The history of the column of water molecules is shown in figure 6.26.

These observations suggest that water may readily facilitate the passage of polar molecules

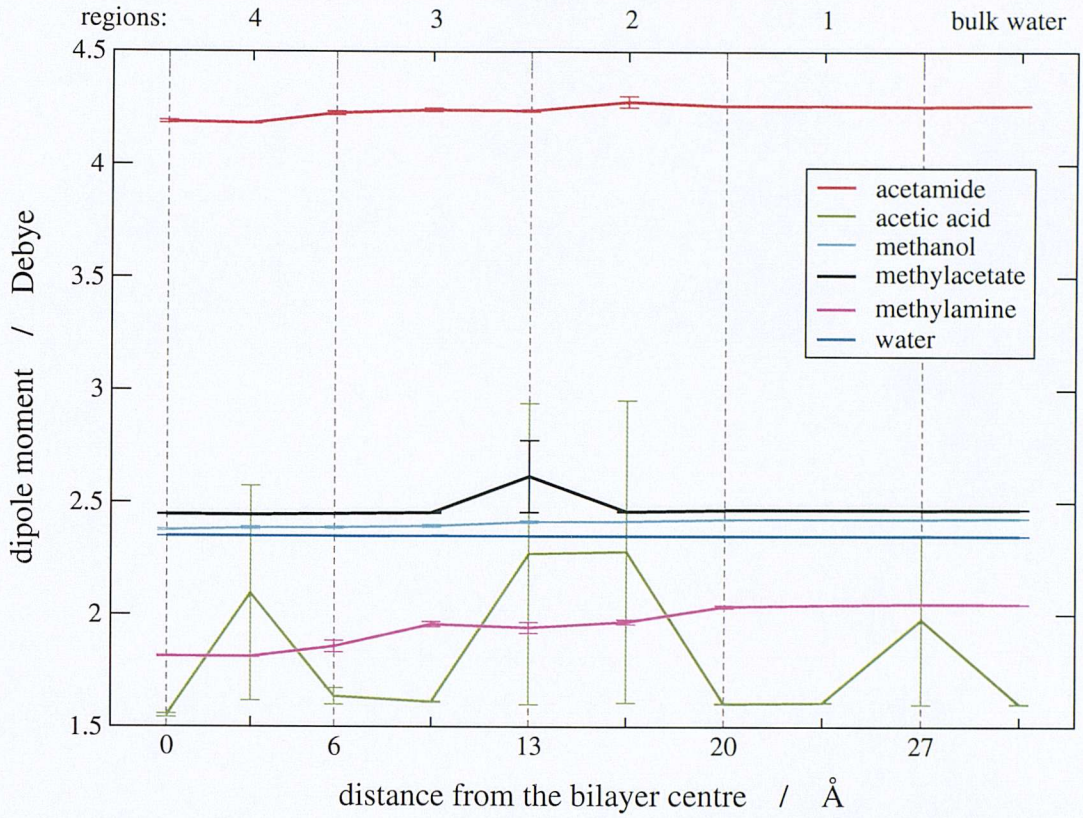


Figure 6.27: Solute dipole moments as a function of depth inside the membrane.

through lipid bilayers.

6.9.7 Dipole moment

The dipole moment for the polar solutes was calculated at different depths and plotted in figure 6.27. Errors were calculated from the differences of the five simulations from their average. Some observations are notable.

Although a (very small) decrease in the value of the dipole moment may be observed for some of the solutes when moving deep into the membrane, in accordance with its hydrophobic nature, the compounds studied here are quite rigid and their dipole moment does not vary significantly during the simulation or between different x, y or z positions. An exception is acetic acid. As described in section 6.9.1, concerning solute flexibility, the carboxylic group has been found to be predominantly in the so called "closed" conformation with the dihedral angle $\phi \approx 0°$. In this conformation the acetic acid dipole moment from these simulations is approximately 1.6 D. In a few cases $\phi \approx 180°$ and the solute dipole moment was about 4.9 D. Since figure 6.27 reports an average over the total simulation time and the five solutes at each z-depth, this explains the variations in acetic acid dipole moment. As argued in the section

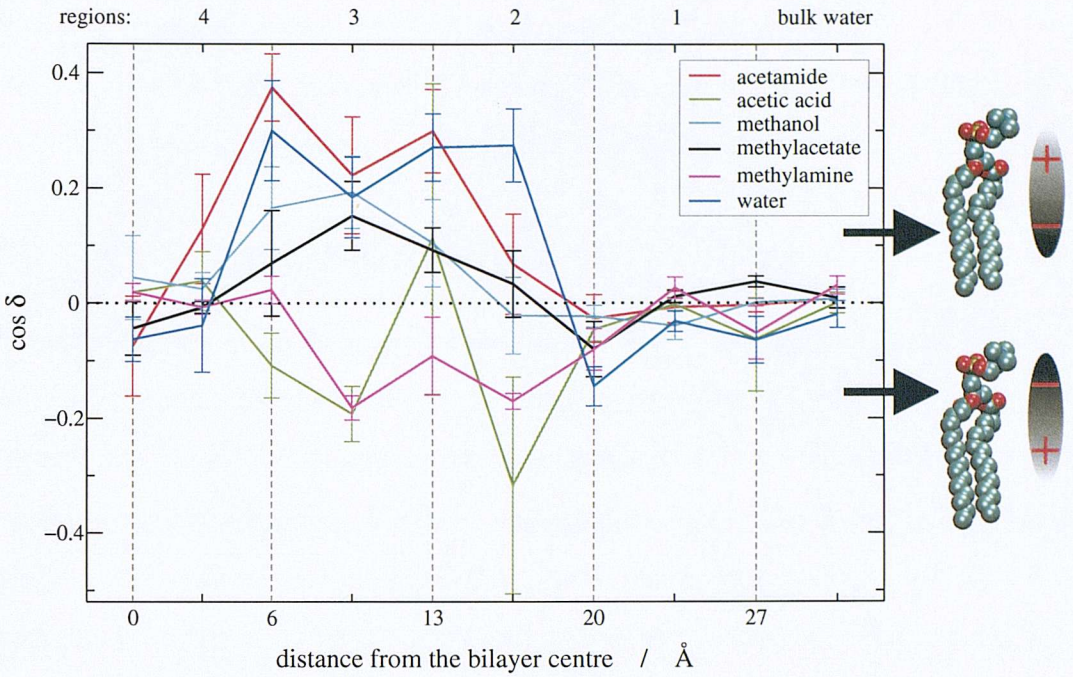


Figure 6.28: Cosine of the angle between solute dipole moment and bilayer normal. This angle is named δ.

dealing with solute flexibility, such variations do not appear to be related with solute location.

The cosine of the angle between the dipole moment and the bilayer normal is plotted in figure 6.28. In this analysis, this angle is called δ. As shown on the right hand side of the plot, when $\cos\delta$ is negative, the solute dipole moment is oriented towards the membrane interior, and when $\cos\delta$ is positive, the dipole moment is oriented towards the membrane exterior. If the cosine is zero, this does not automatically mean that the dipole moment is oriented perpendicular to the lipid molecules, but also that an identical number of dipoles are oriented upwards and downwards.

A wide range of orientations is found in the membrane. The dipole moment has an unpredictable behaviour for all the solutes at all depths. Much more information is obtained regarding the solute behaviour from the steric and polar interactions studied in the previous sections. It is clear that generally van der Waals interactions or local strong hydrogen-bonding networks prevail over the long range electrostatic interactions in determining the overall solute orientation among the lipids.

In region 1 the dipole moment of most solutes is oriented towards the membrane interior. This is true above all for water, in agreement with previous lipid bilayer simulations[236] and with the results of the pure DPPC simulation presented in chapter 4 (see section 4.2.9). Since this behaviour is common to all waters residing at the interface, this gives the main contribution to the electrostatic potential profile along the bilayer normal. The electrostatic potential across

the membrane was calculated for the solute/bilayer systems using equation 4.5 and profiles identical to that in figure 4.10 were obtained. This confirms that the presence of a small amount of organic molecules does not disrupt the bilayer structure.

6.10 Limitations and Errors

In the computational methods applied here, there are inevitably sources of error.

The force field employed may not be the most appropriate for these calculations. However, simulations of pure DPPC bilayers have been shown to yield a realistic physical picture of such systems and to reproduce available experimental data.[91,100] Therefore, this membrane computer model is expected to be a good choice for permeation studies. On the other hand, the parameters for the small organics may not be adequate. These parameters are available from the distributed version of the CHARMM force field for lipids and proteins, and so are expected to be consistent with the parameters employed to model the lipid molecules. However, they are optimized for simulations in the water phase and their performance in a lipid bilayer may not be as good. This may be the reason for the low free energies from these simulations with respect to the experimental free energies in hexadecane/water systems. It would however be necessary to optimize solute parameters at each of the membrane locations, but this is not reasonable.

Although bilayer anisotropy is present along the bilayer normal, and hence the force acting on a permeant varies with depth, these studies reveal that the mean forces at the same depth but different x-y positions indeed differ, and complete convergence is not reached even after 2 ns. It was found that the value of $\langle \boldsymbol{F}(z) \rangle_t$ is more sensitive to the position on the bilayer plane than to the force field employed (data not shown) and therefore any bias in the starting configuration would have affected the results. In this study it was considered appropriate to use five independent, separate different positions per depth. Since 2 ns were simulated for each of these, a total of 10 ns were sampled for each z coordinate. The solutes were small, rigid organic molecules for which the limit of Markovian dynamics holds. These molecules can easily move and rotate in the bilayer, their flexibility is restricted and few internal conformational degrees of freedom need to be sampled. Consequently, it is thought that the results reported here are suitable for drawing fairly reliable conclusions about solute partitioning and diffusion.

Concerns may arise about the way of calculating solute diffusion coefficients from the force fluctuation autocorrelation function (ACF) and the choice of abandoning the well established and much more "reassuring" mean square displacement (MSD). The first proof that the force

fluctuation ACF works reasonably well for these studies is in the agreement between the calculated diffusion coefficients at 30.5 Å from the bilayer centre and the experimental diffusion coefficients in water. The second proof is in the relationships $\log(D) \approx -0.6\log(V_d)$ and $\log(D) \approx -0.5\log(M)$ in the calculated diffusion coefficients at $z = 30.5$ Å, in excellent agreement with empirical observations of small solute diffusion in water and sphere-like solvents. As a third proof, the diffusion coefficient of water and benzene was calculated in a box of 512 TIPS3P waters by means of the classical MSD at 323 K and compared with that obtained with the z-constraint method. Water and benzene were chosen as the two "extreme" solutes among those studied here: the former very polar and small, the latter hydrophobic and large. The MSD calculated diffusion coefficients were in excellent agreement with those obtained from the force fluctuation ACF at $z = 30.5$ Å and were 4.7×10^{-5} cm^2 s^{-1} for water and 1.4×10^{-5} cm^2 s^{-1} for benzene. As explained, the MSD cannot be used in the lipid bilayer environment because of its inhomogeneous character. Therefore the z-constraint *vs.* MSD test was done in bulk water only. The "bulk water test" should offer the definitive support for the methodology chosen in the permeation studies.

A particular note must be made regarding the parameters employed to simulate methylamine. This residue, like the other compounds studied here, is present in the latest CHARMM force field for proteins,[237] which is said to be consistent with the latest CHARMM force field for lipids.[100] However, just after the simulations described here were completed and analyzed, a publication by MacKerell *et al.*[238] appeared with updated and refined parameters for simple amines. However, the results obtained using the old parameters are consistent with the overall trends observed.

When the pressure is kept constant in lipid bilayer simulations, the most obvious choice is using an isotropic pressure ensemble, where the normal and lateral components with respect to the water/lipid interface are identical, because experiments clearly show that such a system is tension free. If also the temperature is constrained, in MD simulations such an ensemble is the NPT. However, it has been argued that a non-zero surface tension is appropriate to correct for the finite size of the simulation system compared with macroscopic real membranes whose experimental properties are to be reproduced.[97] It has been shown that with the CHARMM force field the $NP_N\gamma T$ and NP_NAT ensembles yield the best agreement with experiments[91,92] and are equally reliable. For this reason a "simpler" NPT ensemble seemed inappropriate for these simulations. The $NP_N\gamma T$ could therefore be chosen for these permeation studies. A

fully flexible simulation cell could adjust its dimensions because of the presence of the solutes. However, the appropriate value of γ is highly sensitive to the size and the force field of the simulated system, and there is no guiding procedure for its optimization. Previous works[92] clearly demonstrate the difficulty in determining the value of γ to a high precision. For a CHARMM pure lipid bilayer a large range between 35 and 45 dyn cm^{-1} was given, but the presence of permeants is expected to significantly alter the bilayer surface tension. That range was obtained with a 1 ns long simulation, but it was noted that convergence and sufficient equilibrium sampling of area fluctuations would require a few tens of ns and that additional inaccuracies in the potential energy parameters (like those for the small organics in these studies) could significantly alter the value. The conclusion is that also the CHARMM authors prefer using a constant surface area than constant surface tension ensemble in their work on lipid bilayers. The final choice of ensemble for these simulations was therefore NP_NAT. When a small amount of a small solute enters the membrane, the lipid structure is not altered, as shown by these and previous simulations, so that the area per lipid should also arguably remain constant. Moreover, the same packing constraints existing in pure lipid membranes are expected to still act when solutes try to permeate: these studies were not intended to understand modifications to lipid structure by solutes, but rather the forces operating on solutes when permeating the membrane.

The use of a non-flexible simulation cell for simulating solute diffusion within lipid bilayers is not new. Stouch *et al.* performed such simulations using a constant volume / constant temperature ensemble and they employed the same value of surface area per lipid for both pure DMPC bilayers[95,239] and solute/membrane systems.[103,104,109]

6.11 Conclusions

In this chapter, permeability coefficients for eight small organic molecules representing the most common chemical functional groups have been calculated by means of MD simulations and compared with available experimental data. Calculated values are generally one order of magnitude higher.

The distinguishing feature of the solubility-diffusion model is that it incorporates information about both the equilibrium and the dynamic behaviour of the solute in the bilayer. Equilibrium properties are described by the free energy, which depends on the magnitude and the direction of the mean force acting on the permeant molecule and yields the solute ability

to partition into the membrane from water. Dynamic properties are described by the diffusion coefficient, which depends on the amplitude of the fluctuations of the instantaneous force from the mean and on its relaxation time, and yields the ability of the solute to move inside the membrane from one side to the other. The advantage of MD simulations is that the different contributions from the different regions of the lipid bilayer, that is free energy, diffusion and local resistance as a function of z, can be studied at a molecular level, whereas experiments can only approximate the membrane as a uniform barrier slab.

For hydrophilic compounds the main barrier to permeation is represented by the hydrocarbon core of the lipid bilayer, whereas for hydrophobic compounds the main contribution to the permeation resistance is offered by the lipid headgroups. For these small solutes, the resistance is mainly determined by solute partitioning in the different bilayer regions. However, the diffusive behaviour is not negligible in determining the overall permeation process, since it scales the contribution of the free energy of solute transfer. The diffusive behaviour is above all important in the region which offers the main resistance and for hydrophobic compounds.

For the small and rigid organics studied here, diffusion occurs within the limit of Markovian dynamics at all depths and diffusion coefficients have a higher dependence on solute cross-sectional area than volume. Despite their differences in size, all the solutes, to a greater or lesser extent, tend to permeate the lipid bilayer with a preferred orientation parallel to the bilayer normal, above all in the densest part of the membrane. Polar groups are preferentially oriented towards the lipid headgroups, where they can be involved in hydrogen-bonds. Hydration has also been observed in the middle of the membrane, which is considered an almost completely hydrophobic environment. van der Waals and short range electrostatic interactions prevail over the long range forces in determining the solute behaviour. These studies clearly reveal that the bilayer structure is not disrupted by the presence of a small amount of permeating molecules.

The unique conclusion of this work is that calculated diffusion coefficients inside the bilayer are dependent on solute size to a lesser extent than in water or simple liquids. This is not in contrast with the experimental observation that the ratio of permeability coefficients across biological membranes and partition coefficients in reference organic solvents has a strong dependence on solute size, because this trend is reproduced by the simulations. However, it is in contrast with the interpretation of that experimental observation, which ascribes the size dependence to the solute diffusion only. From these simulations, the size dependence shown by permeability is instead to be ascribed to the partitioning behaviour of the solutes into the lipid bilayer. This is

confirmed but the fact that the size dependence of the permeability-partition ratio significantly decreases when using partition coefficients inside the bilayer instead of those in bulk solvent. These simulations clearly show with an atomic detail that the membrane core does not behave like bulk solvent. The fact that $\log K$(membr)-$\log K$(water/hexadecane,water/octanol) plots have negative intercepts show that solute partitioning into lipid bilayers is lower than in bulk solvents. Solutes show a preferred orientation parallel to the bilayer normal and have longer reorientational times than in bulk water, so there is a loss of degrees of freedom due to lipid packing and ordering when solutes move from a liquid into a biomembrane.

The simulation method employed here yielded results in fair agreement with experiments and seems suitable for studying interactions between membranes and real drug molecules. These calculations would be computationally expensive and time consuming, so they are not currently used in drug discovery. However, computer technology evolves quickly and these calculations are becoming tractable.

CHAPTER 7

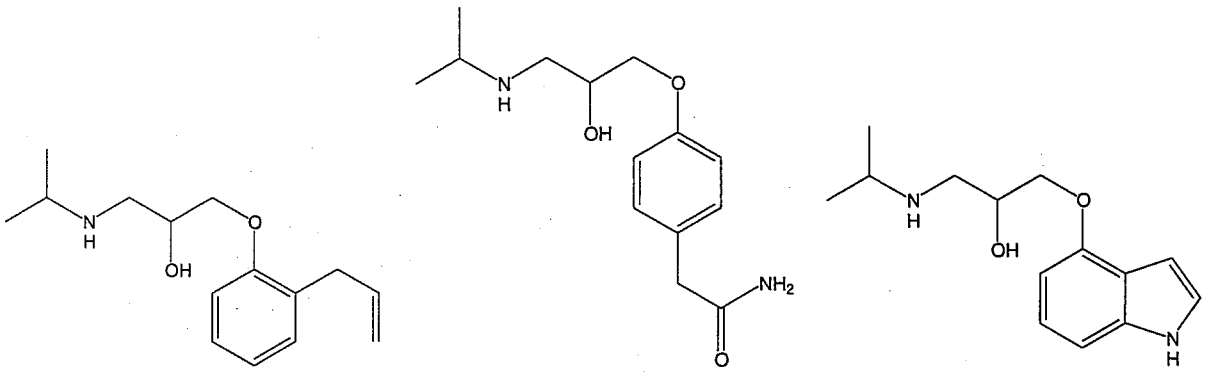
Drug Candidates

The ultimate application of the simulation technique previously employed to study small molecule permeation through a phospholipid bilayer is the study of drug permeation. This chapter describes the choice and the parameterisation procedure for interesting drug molecules.

7.1 Drug Selection

Experimental permeability coefficients across pure DPPC or other lipid bilayers are not available for real drug molecules. However, as described in chapter 5, many data are published regarding drug behaviour across Caco-2 cell monolayers. Most of the experiments are carried out with β-blockers. This class of drugs is also suitable for simulation studies for several reasons:

- they do not posses "exotic" functional groups for which force field parameters are very difficult to obtain;
- they are not very big and their internal flexibility is (hopefully) reasonably sampled in the simulation time scale;
- they differ from each other only in one side chain, so most of the parameters are transferable, simplifying the parameterization procedure; therefore, errors in the parameters are expected to cancel out in the relative permeabilities between the drugs and trends are likely to be accessible.



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

Figure 7.1: Simulated drugs.

- many experimental observations are available regarding their absorption mechanism across Caco-2 cells;
- their absorption after oral administration and their lipophilicity display a wide variability, although the differences in chemical structure are very small; it will therefore be interesting to further investigate this behaviour at an atomic level;
- they have similar molecular weight and pK_a values, which minimizes the influence of these parameters on the results;
- experimental physical properties are available which in turn are useful for parameterizing the empirical force field employed in the simulations.

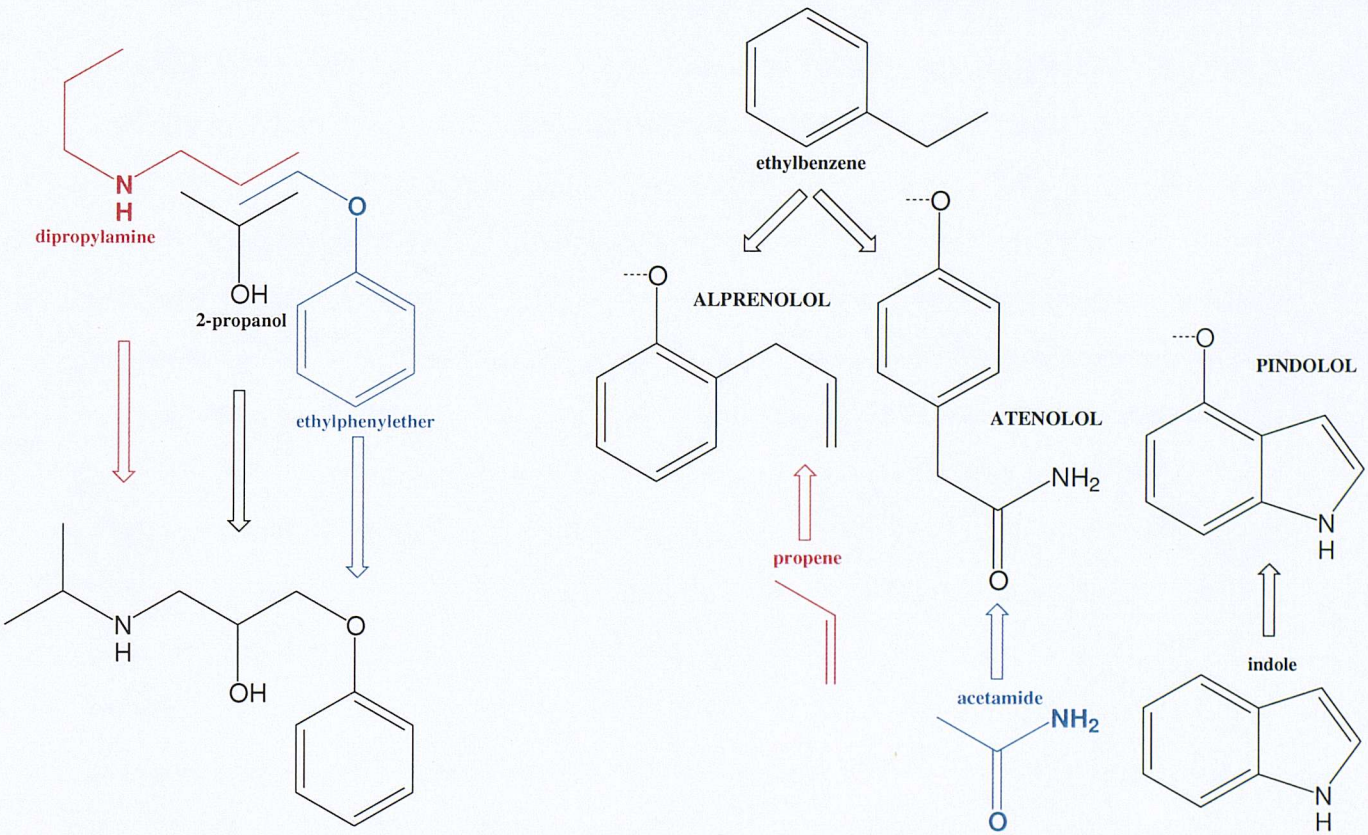
The final list of simulated drugs is given in figure 7.1. They have a common (3-(N-isopropyl)amino-2-idroxy)propyl-phenyl ether frame and a different substituent on the aromatic ring. They act as antagonists of β-adrenoreceptors and are used in various cardiovascular diseases.[240]

7.2 Drug Molecule Parameterization

Parameters for the simulation of the above drugs are not directly available in the CHARMM force field. To obtain new parameters consistent with the existing CHARMM force field, a long procedure is required.[237,238] Briefly, after taking initial parameters from similar atoms already present in the force field, the non-bonded parameters (partial atomic charges and Lennard-Jones parameters) are optimized first, by reproduction of minimum interaction energies and geometries with water and rare gases (helium and neon) obtained from quantum-mechanical

calculations. Internal parameters (i.e. those associated with bonds, bond angles and dihedral angles) are optimized secondly, with equilibrium lengths and angles taken from experimental gas phase or crystal structures and force constants parameterized to reproduce vibrational spectra. Condensed phase simulations are performed to test the complete set of parameters and, if the difference between calculated and experimental data (such as heats of vaporization or sublimation) is larger than a certain tolerance, the procedure is repeated in an iterative fashion, until convergence is reached. In the case of large molecules, smaller model compounds for which experimental data are available are chosen in such a way that their chemical groups can be combined to obtain the target large molecule. Parameters for the large molecule are not changed when transferred from the model compounds, but simulations involving the target molecule can be performed and parameters can be refined to reproduce experimental data.

In these studies, a shorter but similar and still rigorous method was chosen. As suggested in the CHARMM protocol reported above, the large drug molecules were divided into small model compounds on the basis of chemically obvious fragments and depending on the availability of adequate experimental data. A certain degree of overlap between the small compounds ensured that they could be connected to recreate the target drug molecules. For the β-blockers studied here, the chosen model compounds are shown in figure 7.2. Note that ethylbenzene is used to study the side chain connection with the aromatic ring in all drugs. Acetamide, propene, 2-propanol and indole were already present in the CHARMM force field, so they did not require further calculations. In contrast, parameters for ethylbenzene, ethylphenylether and dipropylamine were needed. Non-bonded and internal parameters for the model compounds were not optimized as described above in the full procedure, but rather condensed phase simulations were performed, to test the quality of those parameters directly borrowed from similar atoms present in the available CHARMM force field. Such simulations consisted of calculations of free energies of solvation, whose reproduction is considered a stringent test for evaluating the reliability of a force field. Eventually, after reconnecting the small compounds into the real drug molecules, the parameters were not furtherly refined by means of condensed phase simulations on the drugs. Test simulations have been performed, but they showed very poor convergence and large fiuctuations, so the results were not considered reliable.



(a) Common frame

(b) Side chains

Figure 7.2: Selection of model compounds.

7.2.1 Free Energy Perturbations

One of the most common techniques to compute free energy differences is the so called Free Energy Perturbation (FEP) method.[241] The free energy difference between two different states A and B is given by:

$$\Delta G_{\mathrm{AB}} = -k_{\mathrm{B}}T \ln \left\langle \exp \left[\frac{-\Delta \mathscr{H}_{AB}}{k_B T} \right] \right\rangle_A \tag{7.1}$$

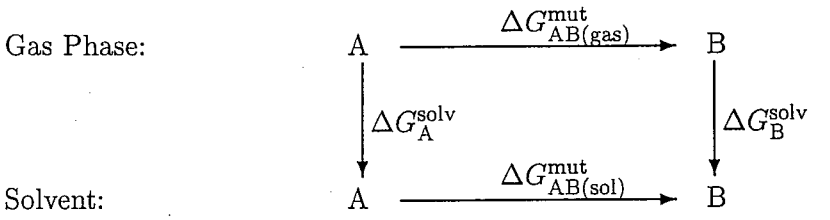
Here k_B is the Boltzmann constant, T the temperature, $\mathscr{H}$ the Hamiltonian, and $\Delta \mathscr{H}_{AB} = \mathscr{H}_B - \mathscr{H}_A$. The averaging is performed over configurations for state A. Such an expression is exact, but it only converges if the two states A and B are similar to each other to keep $\Delta \mathscr{H}_{\mathrm{AB}}$ small. This is because the end points commonly sample quite different regions of configuration space. What may be the low energy states for one Hamiltonian may become very high energy states for another. Therefore it is frequently necessary to define several arbitrary non-physical intermediate states between A and B to increase the similarity between successive states and thus overlap to a greater extent the configurations that they sample. This arbitrary partitioning is allowed because free energy is a state function independent of the path taken between two points. Conventionally each state is defined by the variable λ which ranges from 0 to 1.

The conventional way to perturb molecules between states A and B is the so-called single topology method. State A gradually changes into state B. Dummy atoms, atoms with zero charge and Lennard-Jones parameters, are used to grow or remove atoms. The parameters in the Hamiltonian that differ between A and B are altered according to the value of λ. Typically this dependence is made to be linear. For example, if A and B differ by a certain bond length, then the intermediate bond length l_λ is given by

$$l_\lambda = \lambda l_{\mathrm{A}} + (1 - \lambda) l_{\mathrm{B}} \tag{7.2}$$

Alternatively, the whole Hamiltonian itself can be scaled in this way. The total free energy is then the sum of the component free energies for each section obtained using an analogous formula to equation 7.1 except with different end points. One point to note is that FEP does not require configurations to be sampled from state B. If this were done, then the reverse free energy may be also calculated. The value of this free energy should be the negative of the former.

The calculation of the solvation free energy requires two simulations, one in solvent and one in gas phase and is given by equation 7.3 using the following thermodynamic cycle:



$$\Delta\Delta G_{AB} = \Delta G^{solv}_{B} - \Delta G^{solv}_{A} = \Delta G^{mut}_{AB(sol)} - \Delta G^{mut}_{AB(gas)} \tag{7.3}$$

Free energies have to be calculated only for the smaller A to B mutations. If the molecule is assumed to be rigid, then even $\Delta G^{mut}_{AB(gas)}$ does not need to be calculated. The reason for this is that the free energy change calculated in solvent is now not strictly $\Delta G^{mut}_{AB(sol)}$ but $\Delta G^{mut}_{AB(sol)}$ with the intramolecular free energy change going from A to B removed. This intramolecular free energy change is the same in gas phase and solvent since the molecule is rigid. In other words, it is equal to $\Delta G^{mut}_{AB(gas)}$. Therefore $\Delta\Delta G_{AB}$ simply equals the free energy change in solvent.

7.2.2 Free Energy of Solvation

Free energies of solvation for the chosen model compounds were calculated in water. For consistency, this solvent was represented with the TIPS3P model which is also employed with the lipid bilayer. Simulations were performed with the program BOSS,[242] which is widely proven to yield accurate free energies using Monte Carlo (MC) simulations and the FEP method. It is also used by the authors of the CHARMM force field. The energy function employed in this program is very different from the one used in CHARMM, so BOSS had to be heavily modified. The differences are due to the Urey-Bradley terms and the potential associated with dihedrals and van der Waals interactions.

The molecule was mutated from itself to a non-interacting particle in aqueous solution. The mutation to nothing was divided up into a number of smaller stages consisting in deleting a heavy atom with its bonded hydrogens (e.g. a methyl group, a hydroxyl, an amine) and replacing it with a hydrogen atom. Where possible, perturbation pathways were chosen to minimise changes in the force field, such as charges and Lennard-Jones parameters. By calculating the free energy changes for all the stages in the mutation tree, the absolute free energy for any molecule was then calculated by summing the components. Each of these mutations between molecules was subdivided further into a number of windows defined by the coupling parameter

λ which varies from 0 to 1.

The simulations were performed in the NPT ensemble, with T = 25°C and P = 1 atm. MC move size was adjusted to yield approximately 40% acceptance ratios, with application of Metropolis sampling criteria. Internal geometries were not fixed except for bond lengths. Also aromatic rings were kept rigid, i.e. bond lengths, bond angles and dihedral angles in aromatic rings were kept fixed at their equilibrium values according to the force field parameters employed. The justification for this is that such molecular fragments are quite rigid anyway and interact with surrounding atoms mainly by van der Waals interactions. This approach also makes the MC simulations faster, because less calculations are required. MC calculations include 2×10^6 configurations of equilibration and 5×10^6 configurations of averaging. The cutoff used was 11 Å and, if any of the solute-solvent atom distances was less than the cutoff, the interaction between the entire solute and solvent molecule was included. All non-bonded interactions were quadratically feathered to zero over the last 0.5 Å. A coupling parameter λ was used to gradually perturb solute A in solute B and a $\Delta\lambda$ of ± 0.05 was used, such that the total ΔG_{AB} was the sum of results from 10 simulation windows. For each perturbation, a single solute was placed in the centre of a periodic box containing 512 TIPS3P water molecules and a number of water molecules identical to the number of solute heavy atoms was removed. The water molecules discarded were those with the highest interaction energy with the solute. This water box was cubic with edges $\approx$ 24.5 Å long. Atoms that disappeared were mutated to dummy atoms with their bond lengths reduced to 0.5 Å.

7.2.3 Results and Discussion

The mutation tree performed with the FEP method is shown in figure 7.3. In that figure, vdWNH_3 and vdWbenzene are ammonia and benzene molecules modelled without charges, so that they interact with the surrounding solvent by means of van der Waals interactions only. UAbenzene is united-atom benzene. The mutation from vdWbenzene to UAbenzene consists of deleting the hydrogen atoms and increasing the Lennard-Jones (LJ) parameters on the carbons. LJparticle is a single site particle with no charge and with LJ parameters identical to those of united-atom methane. Finally, the symbol Ph in figure 7.3 stands for phenyl group. Mutations in gas phase from the benzene ring to nothing and from ammonia to nothing have $\Delta G = 0$ because such molecules are rigid. Also the mutation of the non-interacting particle (nothing) from the gas to the water phase has $\Delta G = 0$ by definition. Parameters for the intermediate

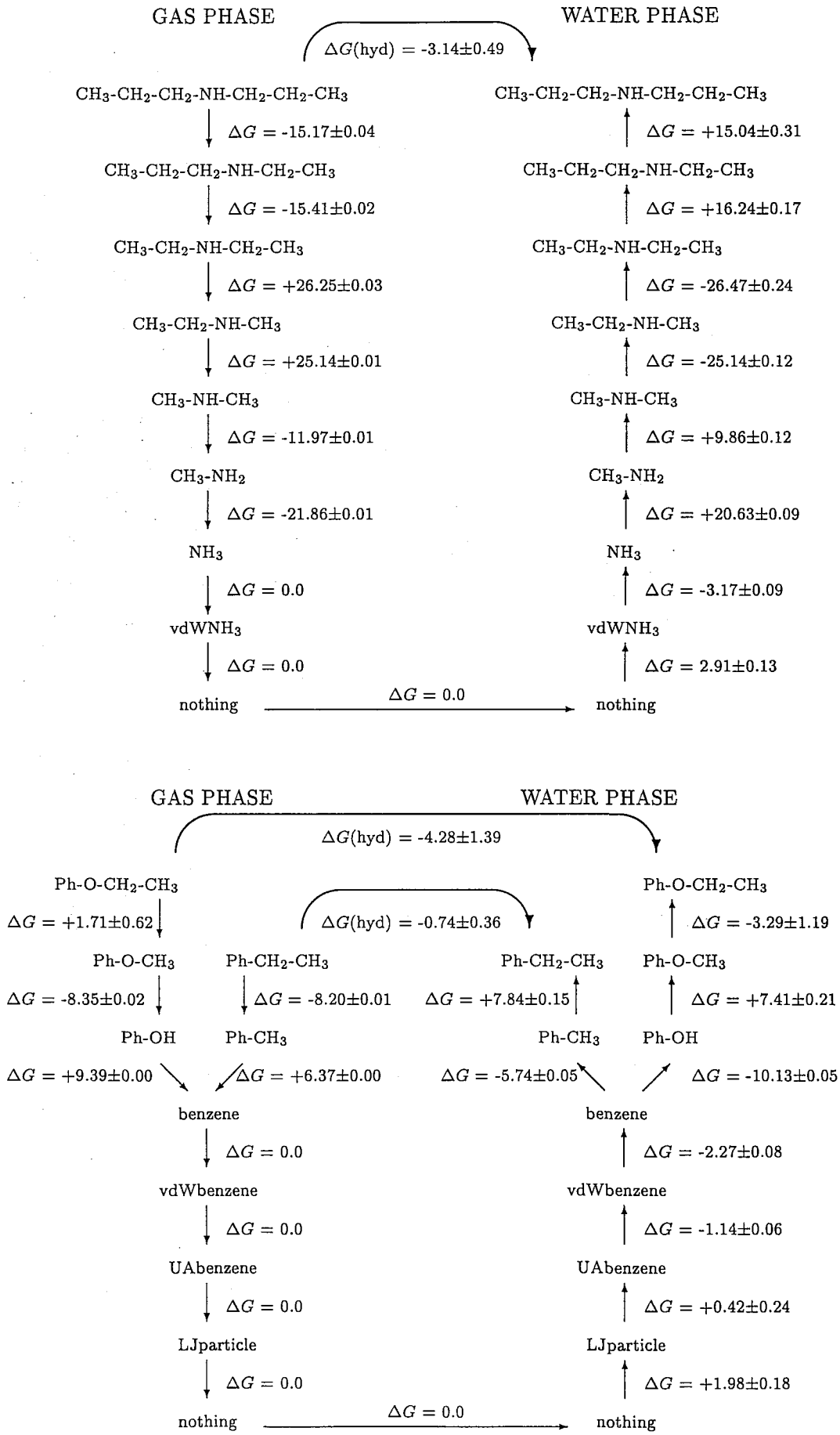


Figure 7.3: Free energy tree showing the mutations performed to calculate the free energies of hydration. ΔG values are in kcal mol^{-1}.

solute	ΔG_{sim} / kcal mol^{-1}	ΔG_{exp} / kcal mol^{-1}
dipropylamine	-3.140 ± 0.493	-3.66
ethylphenylether	-4.277 ± 1.392	-4.28
ethylbenzene	-0.737 ± 0.355	-0.80

Table 7.1: Calculated free energies of solvation (ΔG_{sim}) versus experiments (ΔG_{exp}).[1]

compounds (i.e. those between the molecule under study and nothing) were not optimised, because the mutations were chosen to minimise the changes in the force field to obtain a better convergence of the overall ΔG value. Therefore, ΔG values reported in figure 7.3 cannot be used to obtain free energy of solvation for the intermediate compounds.

All the mutations were performed in the direction from the largest to the smallest molecule, that is deleting a group of atoms. Therefore, the state A for the average in equation 7.1 is the state of the largest molecule. However, in figure 7.3 mutations in the gas phase only (i.e. those on the left hand side) are drawn in this direction. Those in water phase (i.e. on the right hand side) are written in the opposite direction (i.e. from the smallest to the largest molecule) to aid the visualization of the free energy cycle. The value of ΔG in this case has the same magnitude and opposite sign than that for the mutation in the opposite direction. In figure 7.3, the overall free energy of hydration ΔG(hyd) is also reported at the top of the two cycles. The values of ΔG(hyd) are also highlighted in table 7.1, where they are compared with experimental values.[1] Errors are calculated in BOSS as standard errors from the fluctuations in the averages of five runs of 10^6 configurations. Calculated free energies are in good agreement with experiments, the errors are fairly low and of the same order of magnitude as previous publications from this laboratory.[243] The (few) most noisy perturbations were re-run doubling the number of configurations, but the error did not decrease.

Since parameters applied to small model compounds yielded reasonably good results, they were used without any further modification for the entire drug molecules. A more rigorous approach should have been their further validation by means of free energy calculations for those drug molecules. Unfortunately, this was not possible because of the large size and the large number of internal degrees of freedom of the drugs. Once placed in a box of water, MC moves were accepted with a ratio of 40% only if the ranges for dihedral moves and solute

translation and rotation were ridiculously small. Large energy barriers separated different drug internal conformations and, once the starting structure was chosen, the other conformers were in practice not accessible in the time scale of common MC simulations. Values of ΔG for individual perturbations were very noisy and rather inaccurate, irrespective of the number of configurations. Since the accurate calculation of drug free energies and behaviour in water or other solvents was not the objective of this thesis, it was thought that the simple direct transfer of parameters from model compounds to drugs was acceptable.

7.3 Drug Molecule Conformations

Given the difficulty of sampling the drug internal flexibility in a box of explicit water molecules and in turn the high sensitivity of the free energy results to the starting structure, this was also expected to be a problem once the drugs were place inside the lipid bilayer, which contains regions with density higher than that of bulk water. It appeared to be very difficult to apply special techniques such as parallel tempering[244] for accurately sampling all the drug conformations, since the lipid bilayer system is large and the simulations computationally demanding. It seemed more feasible to carefully choose representative drug conformations as starting structures for the bilayer simulations.

7.3.1 Simulation Protocol

The most populated drug conformations were generated in implicit solvent. The latter was modelled using a dielectric constant other than 1 to scale the electrostatic interactions; from all other points of view the simulations were the same as in the gas phase. This allowed large conformational changes of the drug molecules to be accessible.

The dielectric constants used were 78.3 to mimic the water phase at 25 °C, and 2.0 to mimic the hydrophobic core of the lipid bilayer, as reported in the literature.[245,246] MC simulations were carried out with the program BOSS. The simulations were performed with T = 25 °C. MC move size was adjusted to yield approximately 40% acceptance ratios, with application of Metropolis sampling criteria. Internal geometries were not fixed except for bond lengths and aromatic rings, as explained for the small model compounds in section 7.2.2. MC calculations include 1×10^6 configurations of equilibration and 50×10^6 configurations of averaging.

Among the drug internal degrees of freedom, six dihedral angles were considered as the most important and the most difficult to sample. They are shown in figure 7.4: five are in the main

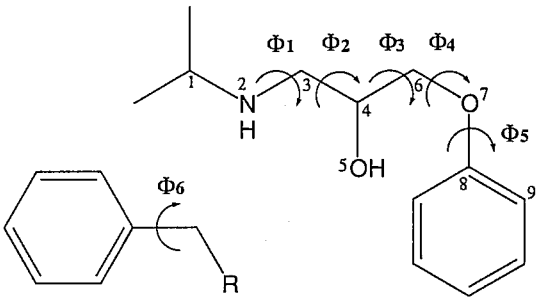


Figure 7.4: Critical dihedrals in drug molecules. Φ1 = 1-2-3-4, Φ2 = 2-3-4-5, Φ3 = 5-4-6-7, Φ4 = 4-6-7-8, Φ5 = 6-7-8-9.

chain and one in the side chain. To allow for complete conformational sampling, the BOSS program was further modified. Every time a dihedral angle variation was chosen to be the MC move, a change of 120° (corresponding to a trans/gauche interconversion) was performed on one of those dihedrals chosen at random. A random number between 0 and 1 was then generated and the 120° variation was accepted only if the random number was equal or lower than 0.4. This ensured that the large angle change had a probability of 40% to be accepted. The history of these dihedrals was plotted in the output file from BOSS for a total of 6×10^4 data points over the 50×10^6 configurations. To ensure that phase space was accurately sampled, the same simulation was performed from six different starting structure. Each of those six dihedrals had a different value in each of the six starting conformations. The six values for each dihedral were generated randomly with the constraint that they had to be at least 25° different, in order to cover significant points between 0 and 360°, and that bad structures with atom overlapping had to be avoided.

7.3.2 Results

The populations from the six different starting structures for each dihedral angle were remarkably similar between the simulations. This suggests that phase space was adequately sampled. Some dihedrals clearly have one preferred conformation only. In contrast, some others can significantly interconvert among two to four different conformations. However, correlations reveal that only a few combinations of configurations are actually significantly populated. Such combinations are shown in table 7.2.

For alprenolol one conformer was clearly preferred in both implicit solvents. The percentage of saved structures which possessed this configuration was $\sim$ 30%, whereas other configurations had populations lower than $\sim$ 10%. For atenolol, several dihedrals had a non-zero population for different conformations. Correlation analysis among the conformers yielded eight different

drug	Φ1	Φ2	Φ3	Φ4	Φ5	Φ6
alprenolol	292.5	82.5	180.0	300.0	127.0	202.5
atenolol	292.5	82.5	180.0	300.0	127.0	202.5
	292.5	82.5	180.0	300.0	127.0	23.5
	292.5	82.5	180.0	300.0	307.5	202.5
	292.5	82.5	180.0	300.0	307.5	23.5
	292.5	300.0	180.0	300.0	127.0	202.5
	292.5	300.0	180.0	300.0	127.0	23.5
	292.5	300.0	180.0	300.0	307.5	202.5
	292.5	300.0	180.0	300.0	307.5	23.5
pindolol(*)	300.0	292.5	180.0	285.0	292.5	
	82.5	292.5	180.0	285.0	292.5	
	82.5	292.5	180.0	180.0	240.0	

Table 7.2: The most populated combinations of dihedral angles. (*) For pindolol, the side chain is a rigid ring and there is no Φ6.

structures preferred over the others. The four conformers with $\Phi 2 = 82.5°$ were preferred in implicit water and the population of saved structures which possessed one of these four configurations was $\sim$ 10%. This implies that the four configurations together accounted for $\sim$ 40% of all the saved structures generated in the water phase. Other single conformers had populations lower than 4%. The four with $\Phi 2 = 300°$ were preferred in the hydrophobic environment. Similarly to the results in the water phase, each of these configurations had a population of $\sim$ 10% and all together accounted for $\sim$ 40% of the structures generated in the hydrophobic environment. Other conformers had populations lower than 4%. For pindolol, the side chain is a rigid ring and the drug thus contains one less torsion. Three conformers were found to be the most populated. By analogy with atenolol, the coformers with $\Phi 2 = 82.5°$ were the most populated in implicit water. The percentage of each of them was $\sim$ 25%, so together they accounted for half of the structures generated in the water phase. Other conformers had percentage lower than 5%. The configuration with $\Phi 2 = 300°$ was the most populated in the hydrophobic environment, representing $\sim$ 25% of the structures generated in this implicit solvent. The other conformers appeared with percentages lower than $\sim$ 5%.

One could argue that it would be reasonable to use drug conformers obtained in implicit water for simulating drugs at the lipid/water interface only, and to use conformers obtained

in implicit hydrophobic environment for simulating drugs in the membrane interior. However, owing to the complex and inhomogeneous character of lipid bilayers, it was preferred to take into account all the conformers obtained in both environments for simulations in all the membrane regions.

7.4 Conclusions

In this chapter, interesting drug molecules have been selected for studying the mechanism of drug permeation through cell membranes, and a quick and rigorous procedure has been developed to obtain force field parameters for the drug/membrane simulations which are consistent with the CHARMM force field.

Free energies of solvations have been calculated for small organic compounds by heavily modifying the program BOSS and results are in excellent agreement with experiments. Such calculations have allowed the validation of the parameters for the small compounds, and these in turn have yielded reasonable parameters for the desired drug molecules.

Finally, conformational studies for the drug molecules have been performed prior to their simulation inside the lipid bilayer environment, to select their most significant structures. Such investigations have been done in implicit solvents mimicking the water phase and the hydrophobic core of the lipid bilayer. The selected drug conformers will be used in the next chapter for studying their mechanism of permeation across the bilayer.

CHAPTER 8

Computer Simulation Study of Drug Permeation

The simulation technique previously employed to study small molecule permeation through a phospholipid bilayer is here further applied to real drugs.

8.1 Simulation Protocol

The same simulation protocol employed for the simulation of small molecule permeation was also used here with real drugs (see section 6.2).

The differences between the simulations with small organics and those with real drugs reside in the number of molecules placed in the lipid bilayer at the same time, and in the number of membrane positions sampled. The β-blockers studied are half the size of a lipid molecule, so only two of them were simulated in the membrane at the same time. Since this in turn requires a larger number of simulations, only one z depth per bilayer region was sampled, where the regions are defined as per the four-region membrane model presented in the previous chapters. In contrast to small organics that can freely rotate around their centre of mass in the membrane environment, the drugs are elongated molecules and it was therefore not expected that full rotation would be observed. However, it was considered extremely unlikely for the

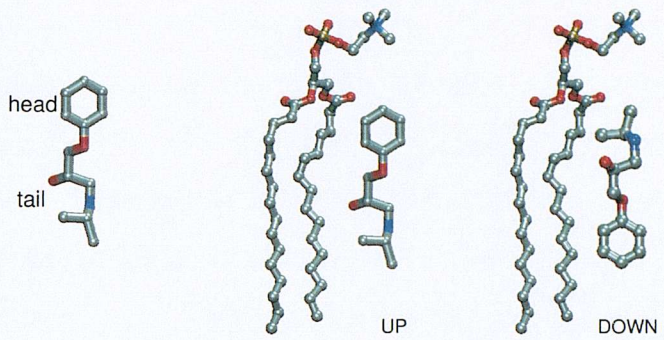


Figure 8.1: Up and down orientations of the drugs with respect to the surrounding lipids. Drug side chains are not shown for clarity.

drugs to cross the membrane with an orientation perpendicular to the lipid molecules, since the previous studies with small organics revealed that permeants tend to be aligned parallel to the lipids. Therefore, each z-depth was sampled twice: once with the drug having the aromatic ring oriented towards the middle of the bilayer and the isopropylamine fragment towards the water phase, and once with the aromatic ring oriented towards the membrane exterior and the isopropylamine fragment towards the interior. If we name the aromatic ring as the *drug head* and the main chain as the *drug tail*, it is convenient to refer as *up* the orientation where the drug head is towards the membrane exterior and *down* the orientation where the drug head is towards the centre of the bilayer. These orientations are represented in figure 8.1.

A comment is required regarding the method of drug insertion inside the lipid bilayer. Whereas for the small organics a crude minimization was sufficient to eliminate bad contacts, for some of the drugs a more complex procedure was necessary. The bond lengths were scaled by 0.1 to reduce the size of the permeant. After insertion at the desired position in the membrane, a few steps of minimization were performed. The bond lengths were then progressively increased: after each increase the drug was inserted in the lipid bilayer at the same position as before, and using the lipid coordinates obtained from the previous insertion with the smaller molecule. A few steps of minimization then followed. This procedure was repeated until the bond lengths corresponded to the equilibrium values in the bond stretching parameters. Each minimization consisted of 50 steps of steepest descent.

For each drug conformer selected (see table 7.2 in the previous chapter) six simulations were performed. Six depths were sampled along the z axis, once with up orientation at one x-y position and once with down orientation at another x-y position. This would yield 12 simulations, but two drug molecules were inserted in the bilayer at the same time, reducing the total number of simulations to six. The very first 100 ps of every simulation was discarded as equilibration and all the timings given below refer to the subsequent production runs. The

alprenolol simulations were performed first and showed that, even though force convergence was not completely achieved, as was the case for the small organics, the free energy profile calculated by averaging the force over 4 ns was almost identical to that calculated using the first 3 ns only. Therefore, for atenolol and pindolol each simulation was 3 ns long. A total of 72 simulations, corresponding to 222 ns of data collection, was performed. The simulations were run on different Linux PC clusters with either 1 GHz Pentium III, 1.5 GHz Pentium IV, or 1.4 GHz AMD Athlon.

Free energies $\Delta G(z)$, diffusion coefficients $D(z)$, local resistances $\mathfrak{R}(z)$ and permeability coefficients P along the bilayer normal were calculated for alprenolol, atenolol and pindolol using the same approach employed for the small organics (see section 6.3). The four region model is employed to describe the membrane environment:

Region 1: low headgroup density, 20 to 27 Å from the bilayer centre.

Region 2: high headgroup density, 13 to 20 Å from the bilayer centre.

Region 3: low tail density, 6 to 13 Å from the bilayer centre.

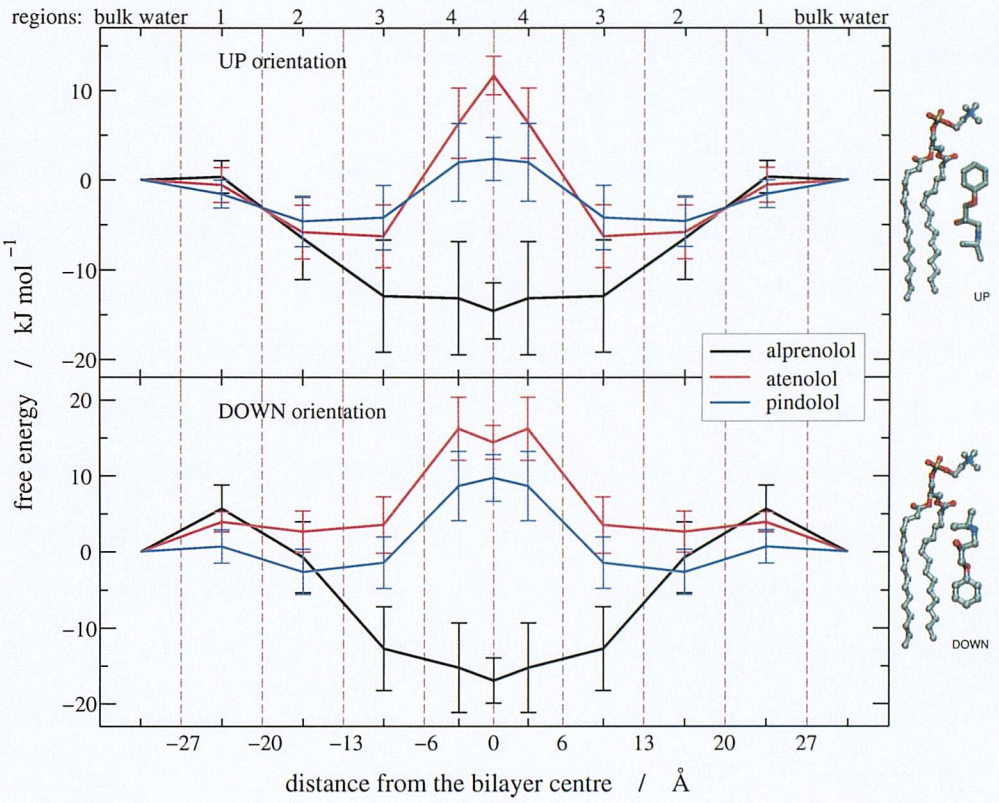
Region 4: high tail density, 0 to 6 Å from the bilayer centre.

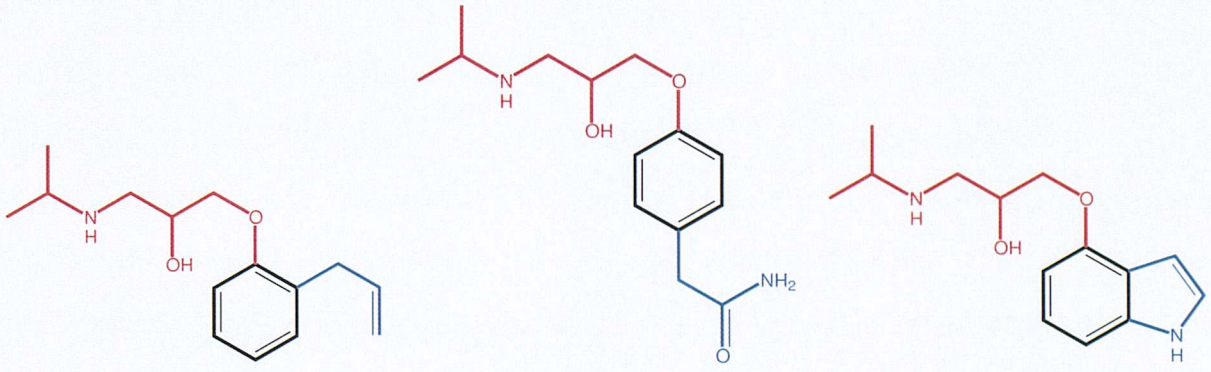
Six z-depths were sampled: at 30.5 Å from the bilayer centre (where the drug centre of mass is still in bulk water, even though the lipid headgroups can be in contact with the drug extremities), the bilayer centre itself ($z = 0.0$ Å), and one position in each bilayer region (23.5, 16.5, 9.5 and 3 Å from the bilayer centre).

8.2 Free Energies

The free energy profiles $\Delta G(z)$ for the up and down orientations of the three drugs are plotted in figure 8.2. For atenolol and pindolol, several conformers were simulated and at each z-depth the error on $\Delta G(z)$ is calculated from the difference in mean force $\langle \boldsymbol{F}(z) \rangle_t$ among the independent simulations. For alprenolol, since one conformer only was simulated, the error is calculated by dividing the simulation into frames of 200 ps, as for the pure DPPC bilayer simulation described in chapter 4. The drug chemical formulae are also given to allow a better understanding of the $\Delta G(z)$ profile.

It is clear that alprenolol has an inverse partition behaviour with respect to the other drugs, as its $\Delta G(z)$ profile presents a deep well in the membrane interior, whereas for atenolol and





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

Figure 8.2: Top: free energy profiles for the up and down orientations of the three drugs. Bottom: drug chemical formulae: in red the drug tail, in black the drug head, and in blue the drug side chain.

pindolol $\Delta G(z)$ in the middle of the membrane is higher than closer to the lipid/water interface. That is: alprenolol prefers to partition into the hydrophobic core of the membrane rather than in water, while the converse applies to a greater extent for atenolol and to a lesser extent for pindolol. This solute ranking is in agreement with experimental partition coefficients of these drugs into organic solvent/water systems. The reference solvent most commonly used in drug design is 1-octanol, and the logarithm of the drug octanol/water partition coefficient is usually indicated with $\log P$.

In the literature, alprenolol $\log P$ has values of around 3,[151,170,247–252] atenolol $\log P$ is about 0.2[151,170,248–251,253] and pindolol $\log P$ is around 1.6.[170,248,250–252,252] These values are all positive, meaning that the three drugs all prefer to partition into the organic phase (octanol) than in water. From these simulations, only alprenolol prefers the membrane interior to water. This different partitioning behaviour may be due to the hydrophobic nature of the lipid tails in contrast to the presence of a hydroxyl group on the octanol molecule, and to the fact that in experiments the octanol phase is saturated with water.[254] However, for at least one of the two orientations (up/down), atenolol and pindolol also present a zone of negative $\Delta G(z)$ between the middle of region 1 and the middle of region 3. Considering the $\log P$ values given above yields ΔG(water$\rightarrow$octanol) $\approx$ -18.6 kJ mol^{-1} for alprenolol, -1.2 kJ mol^{-1} for atenolol, -9.9 kJ mol^{-1} for pindolol [a]. From these simulations, alprenolol $\Delta G(z)$ is between -15 and -17 kJ mol^{-1} and between -13 and -15 kJ mol^{-1} in the down and up orientations respectively when located in region 4. Although these values are close to those derived from experimental $\log P$, they are less negative. Atenolol $\Delta G(z)$ is between 0 and -6 and between 0 and 4 kJ mol^{-1} in the down and up orientation respectively when located in regions 1 to 3. The sum of these effects would agree with the value around -1.2 kJ mol^{-1} from $\log P$s. Pindolol $\Delta G(z)$ is between 1 and -3 kJ mol^{-1} and between -2 and -5 kJ mol^{-1} in the down and up orientations respectively when located in regions 1 to 3. These values are in agreement but still less negative than those obtained from $\log P$s.

The higher drug partitioning into octanol than into the lipid bilayer can be explained on the basis of a different chemical affinity (enthalpic grounds). Another reason may be the tight lipid packing (entropic grounds): in contrast to bulk solvents, in lipid bilayers interfacial constraints restrict the accessibility and the mobility of both lipids and permeant molecules. More detailed

[a] If $K(x/y)$ is the x/y partition coefficient between two solvents, then the free energy of transfer is: $\Delta G(y \rightarrow x) = -2.303RT \log K(x/y)$.

comments have been made in section 6.5.6 (Bilayer / Bulk Solvents Comparisons): the same conclusions drawn for the small organics are even more valid for larger drug molecules.

Unfortunately, given the wide use of octanol in drug design studies, very few experimental data have been found in the literature regarding the partition behaviour of these drugs between water and different organic solvents. One publication has been found reporting the atenolol log partition coefficient in heptane = -1.28,[253] which yields ΔG(water→heptane) ≈ 7.9 kJ mol^{-1}. This value is lower than those obtained from these simulations in bilayer region 4 when the drug is in the down orientation (between 14 and 16 kJ mol^{-1}), but agrees with the values for the up orientation (between 6 and 12 kJ mol^{-1}), suggesting the simulation is qualitatively correct. Partition coefficients for the three drugs have been measured in 1,2-dichloroethane/water systems:[249] the log values are 3.26 for alprenolol, -1.24 for atenolol, 1.31 for pindolol. In this case, atenolol, but not pindolol, prefers the water phase to the organic solvent. This solvent is chosen because is found to incorporate the same contribution from hydrogen-bonding as an alkane, but has far better dissolution properties.

The overall conclusion here is that the solute ranking from these simulations regarding the drug partitioning behaviour reproduces the experimental observables. Moreover, the simulations also show that using octanol log Ps to mimic drug partitioning into biological membranes is not completely correct, as this solvent overestimates the ability of the drug to dissolve in the membrane.

Detailed analyses regarding the drug behaviour inside the membrane, for example the ability of such drugs to change their orientation up→down or down→up, are reported later in this chapter.

8.3 Diffusion Coefficients

Figure 8.3 reports the diffusion coefficients $D(z)$ for the three drugs in the up and down orientations obtained from these simulations.

8.3.1 General Trend

As for the small organics, the $D(z)$ values in the membrane interior are lower than in water at all depths. The absolute values are of the order of 10^{-6} cm^2 s^{-1}, i.e. one order of magnitude smaller than for the small organics. The diffusion coefficients at 30.5 Å from the bilayer centre

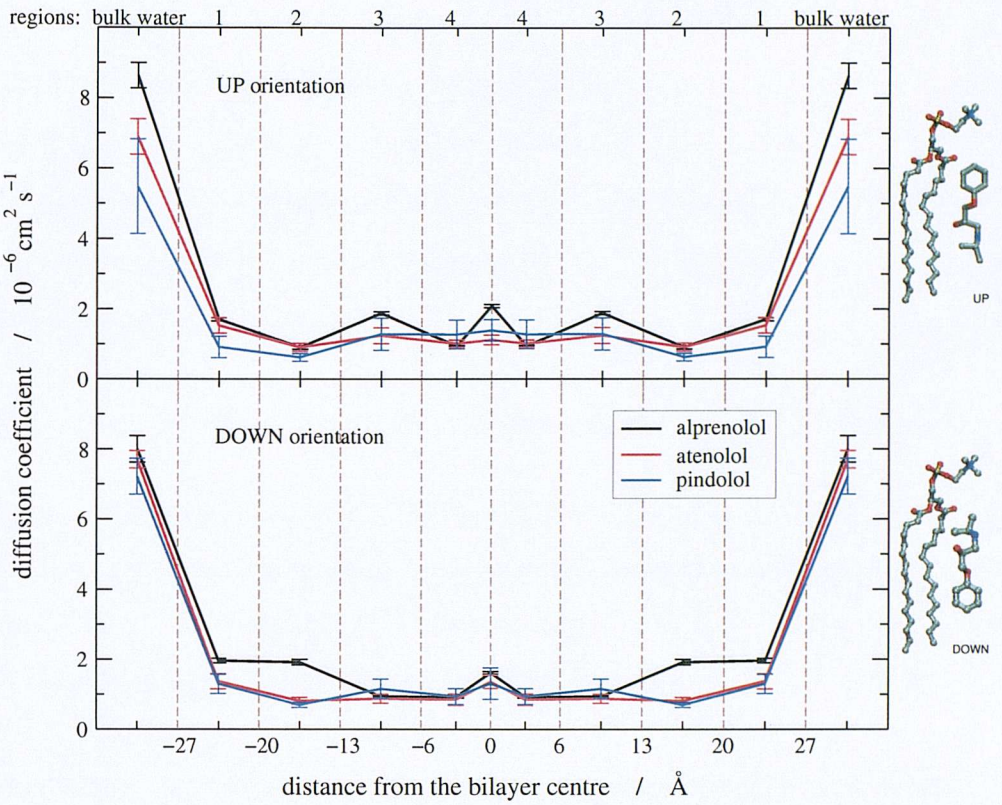


Figure 8.3: Diffusion coefficients for the up and down orientations of the three drugs.

calculated in these simulations agree favourably, although they are a little higher, with those calculated in bulk water at 37 °C, assuming the drug molecule to be spherical and using the Stokes-Einstein equation[170,b]. Diffusion coefficients from these simulations are essentially the same as those obtained by Stouch *et al.* in a simulation of a nifedipine analogue in a bilayer of 36 DMPC molecules.[109] Although a little higher, D values from these and Stouch's simulations agree favorably with experimental data for a molecule of similar size obtained with magnetic resonance spectroscopy.[132,255]

Figure 8.3 also shows that drug diffusion in the bilayer centre is not higher than in the upper part of the lipid tails, while D values for the small organics in region 4 were slightly higher. This behaviour, also observed by Stouch *et al* in a simulation of a drug analogue in a DMPC bilayer,[109] can be explained assuming that larger drug molecules are less sensitive to the mobility of the lipid tails and to the presence of small pockets of free space among them. Smaller solutes can jump more easily in the space arising within the bilayer, whereas larger drugs require a common and simultaneous displacement of a large amount of bonded atoms and therefore jumps are much more difficult. As for the small solutes, no strong direct correlation was found

[b]Diffusion coefficients D calculated from the Stokes-Einstein equation are: $D = k_B T/6\pi\eta R$, with R the molecular radius and η the solvent viscosity.

between D values and lipid internal motions (such as trans/gauche interconversion rates along the hydrocarbon tails) or the bilayer free volume distribution.

Owing to the close similarity in molecular weight and size among the drugs and to the very small number of drugs simulated, it is not worthwhile looking for statistical correlations between the diffusion coefficients and the drug molecular properties, as was performed with the small organics presented in chapter 6.

8.3.2 Force Fluctuation Autocorrelation Function

According to equations 6.3 to 6.6, the diffusion coefficients are calculated from the time autocorrelation function of the fluctuations of the instantaneous force from the mean. As for the small organics described in chapter 6, the autocorrelation functions for the three drugs had to be fitted with double exponentials and have a profile basically identical to that shown in figure 6.5.

The short decay time τ_{short} is the same as for the small solutes, that is between 0.03 and 0.05 ps, with the lower values at the interface and the higher values in the bilayer centre. The physical interpretation of τ_{short} is that it corresponds to the solute response to the friction of the surrounding lipids keeping the diffusant in its local cage. The long decay time τ_{long} is a few ps at $z = 30.5$ Å and between 10 and 50 ps in the membrane interior. These times are much longer than those for the small organic molecules, and are not correlated to permeant location in the membrane interior. Physically, τ_{long} is related to the overall mechanism of diffusion through the membrane. Assuming a hopping type of diffusion, τ_{long} is related to the residence time of the penetrants in their cage of free volume between subsequent hops. Owing to their larger size, it is reasonable that τ_{long} values for larger drugs are longer than for smaller organic molecules.

8.3.3 Approximations

As described for the small organics in section 6.6.14, in the calculation of the diffusion coefficients it was assumed that during the decay time of the time-dependent friction coefficient the particle remains in a region of constant free energy, so that the effective friction could be well described by considering only the static friction coefficient $\tilde{\xi}$ (see equations 6.3, 6.5 and 6.6). A free energy variation in the order of RT per mole is however allowed, where R is the gas constant and T the absolute temperature, since this is a thermal fluctuation. To check that this

drug	ΔG_{MAX} / kJ mol^{-1}
alprenolol	1.35
atenolol	3.57
pindolol	3.07

Table 8.1: Calculated maximum free energy difference over the solute displacement within the decay time of its friction coefficient.

assumption is valid, one can estimate the maximum free energy difference between two points along the bilayer normal whose distance equals the displacement covered by the particle in a time identical to the decay time of its friction coefficient, assuming the calculated D is related to the particle mean squared displacement (equation 4.6). Table 8.1 reports for each drug the maximum free energy difference ΔG_{MAX}. The reference value of RT is 2.69 kJ mol^{-1}.

For alprenolol, ΔG_{MAX} is similar to the values obtained for some of the small organics and the approximation is valid. For atenolol and pindolol ΔG_{MAX} is higher than RT and thus the assumptions for equation 6.6 do not hold. This means that the time-dependent friction coefficient $\xi(t)$ is still correct, but the concept of local diffusion coefficient $D(z)$ becomes ill defined, since it cannnot be derived from $\xi(t)$ with equation 6.6. However, these ΔG_{MAX} values for atenolol and pindolol are obtained in the bilayer centre, while the free energy differences calculated in regions 1 to 3 are all lower than 0.7 kJ mol^{-1}. Therefore, only D values in region 4 are not valid. This is not expected to change significantly the solute ranking of the overall resistances to permeation, since the free energy profiles $\Delta G(z)$ are clearly different and, as for the small organics, the main contribution to R comes from $\Delta G(z)$, particularly when $\Delta G(z)$ is positive. Moreover, the drugs studied here are all very similar in size, shape and flexibility, so that their diffusion behaviour is expected to be very similar. The difference in resistance to permeation should therefore reside in the different partitioning behaviour [that is $\Delta G(z)$] and $D(z)$ should have little effect on the overall drug ranking.

8.4 Local Resistances

Local drug resistances $\Re(z)$ are plotted in figure 8.4. For clarity, error bars are not plotted for the same reasons explained in section 6.7 for the small organics.

The up and down orientations are plotted separately, since for each drug they yield pro-

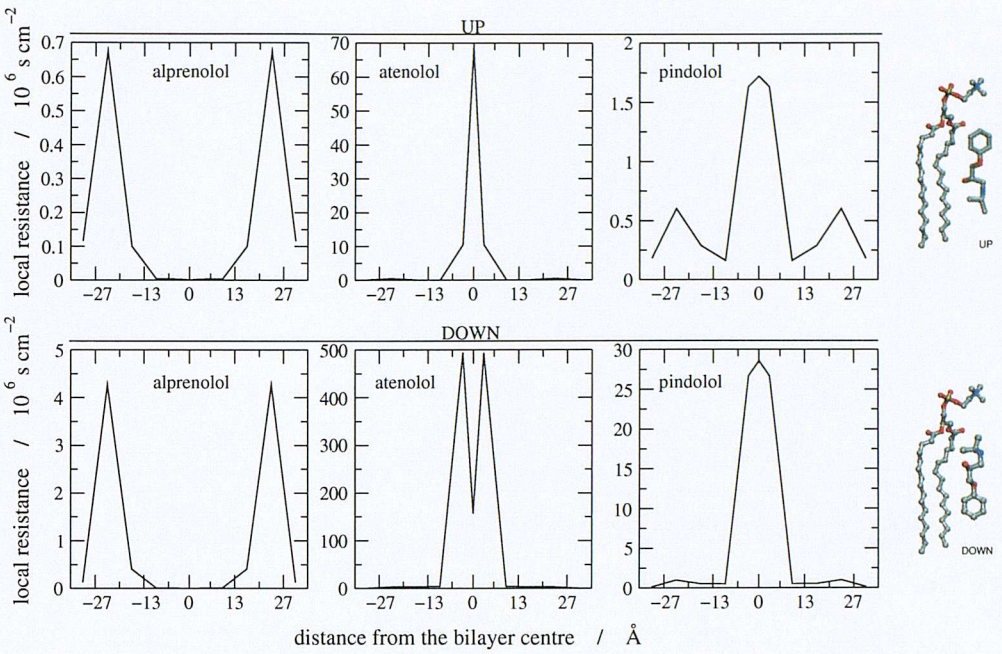


Figure 8.4: Local resistances for the up and down orientations of the three drugs.

files which are about one order of magnitude different, with the down orientation yielding the highest resistance. The small differences in $\Delta G(z)$ profiles between up and down are enhanced because of the exponential form of the resistance equation. This suggests that sampling each z depth with both orientations was required to obtain a complete description of the resistance to permeation and to have an insight into the possible path adopted by these drugs in crossing such membrane systems.

Owing to limitations in computational resources, it was thought not to be possible to simulate a complete drug inversion (up→down or down→up). Unfortunately, simulating both orientations (up and down) separately does not say anything about the equilibrium distribution of these orientations. However, if we admit that in real cell membranes drug molecules are able to freely invert their orientation during the long time scale of the permeation process, a permeation path can be suggested by taking at each z depth the lowest value of $\mathfrak{R}(z)$ between the up and down orientations. This would give the path with the minimum overall resistance. This would imply that at each z depth both simulated orientations, and all the others in between, are completely accessible to the drug molecules, and that the energy barriers between them can freely be crossed. This would allow the permeation process to proceed along the lowest resistance path. This path is plotted in figure 8.5.

The profile of the minimum resistance is almost identical to that obtained with the up orientation for all drugs. However, in the zones of lowest resistance, it is the down orientation

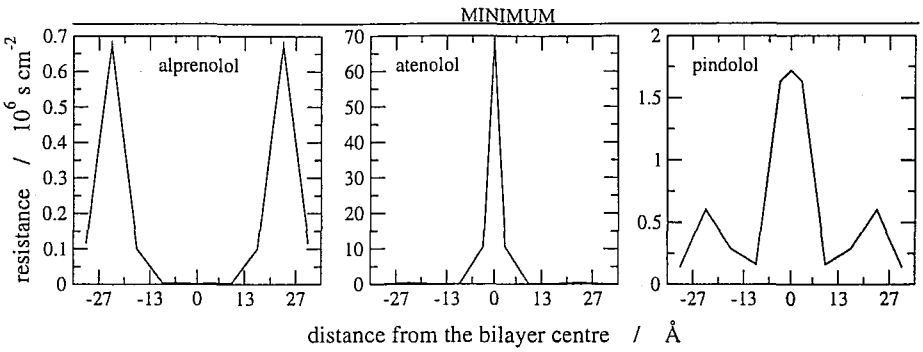


Figure 8.5: Minimum resistance profile for the three drugs.

drug	these simulations		P_{APP}(Caco-2)	P_C(Caco-2)	P^n_{trans}(Caco-2)	P_o(hxd)
alprenolol	min.	17.00 ± 6.94	31×10^{-6}	24.2×10^{-5}	47.60×10^{-4}	46.52
	UP	16.93 ± 6.94				
	DOWN	3.00 ± 2.22				
atenolol	min.	0.63 ± 0.39	0.26×10^{-6}	0.102×10^{-5}	0.37×10^{-4}	-
	UP	0.63 ± 0.39				
	DOWN	0.04 ± 0.04				
pindolol	min.	5.50 ± 3.12	9.9×10^{-6}	5.35×10^{-5}	2.97×10^{-4}	-
	UP	5.46 ± 3.06				
	DOWN	0.54 ± 0.50				

Table 8.2: Permeabilities from the up and down orientations, and from the minimum resistance path, together with experimental permeabilities across Caco-2 cell monolayers and hexadecane. All permeabilities are in cm s^{-1}. Experimental Caco-2 and hexadecane permeabilities were taken from the references cited in the text.

that gives the minimum resistance. These zones are the bilayer centre for alprenolol and the interface for atenolol and pindolol. Even though numerical values differ, for each drug the shape of the resistance profiles in the two orientations is identical. For alprenolol, the main resistance comes from the highly charged and hydrated regions 1 and 2. For atenolol, the main resistance is offered by the hydrophobic bilayer centre. Pindolol has an intermediate character, but the overall profile is more similar to that of atenolol.

8.5 Permeability Coefficients

Table 8.2 reports the permeability coefficients obtained by integrating the local resistances, together with various experimental permeabilities.

Separate permeability coefficients have been calculated for the up and down orientations and for the path with minimum resistance. In all cases, the permeabilities span a range of two

orders of magnitude, with alprenolol having the highest P, atenolol the lowest P, and pindolol an intermediate value. Unfortunately, these results are quite noisy and the errors are larger than they were for the small organic solutes. This does not however prevent the correct solute ranking with respect to experiments.

Caco-2 cell monolayer experiments have been described in section 5.2. Permeability coefficients from different laboratories can vary significantly.[148,152,159,160,172,256] A source of variability in the Caco-2 measurements is due to the cell line itself. Caco-2 are a heterogeneous cell population which is exposed to different selection pressures in different laboratories. Therefore, the properties of Caco-2 cells in one laboratory may easily differ from those in another. Moreover, Caco-2 monolayers also vary within a laboratory with time because of subtle differences in culture medium, extracellular support, temperature, pH gradient, and hydrodynamic properties of the system. In table 8.2 the following experimental data are reported:

- P_{APP}(Caco-2):[148,149,151,160,168,176,257,258] This is the apparent permeability coefficient directly derived from the measured flux J across the cell monolayer: $P_{APP} = J/A\Delta C$ Here A is the area of the filter support and ΔC the difference in solute concentration between the donor and the acceptor chambers.

- P_C(Caco-2):[154,156,165,259] This is the cellular permeability coefficient, obtained after correcting P_{APP} for the permeability of the filter support P_f and for the presence of an unstirred water layer (UWL) on both sides of the cell monolayer. The UWL can become the main barrier to permeation for drugs with a high permeability across Caco-2 cells. The permeability P_{UWL} across the UWL is related to the stirring rate v of the medium by a constant K, so that: $1/P_{APP} = 1/Kv + 1/P_C + 1/P_f$.

- P^n_{trans}(Caco-2):[170] This is the permeability of the drug neutral form through the transcellular route (see section 5.2). It can be obtained from the fraction f of neutral form (which depends on medium pH and drug pK_a) and from P_C (described above): $P_C = f(P^n_{trans} + P^n_{para}) + (1-f)(P^i_{trans} + P^i_{para})$ Here, the subscript *para* stands for the paracellular route and the superscript i stands for the ionized form of the drug.

- P_o(hxd):[260] This is the intrinsic permeability of the drug neutral form across a layer of ≈ 10 μm of hexadecane. The value reported in the table is extrapolated for a layer of 54 Å, which is the thickness of the lipid bilayer employed in these simulations.

For each type of Caco-2 permeability, when several values are found in the literature, an average is used here. The experimental P_{APP}(Caco-2) can differ by as much as 6 times, whereas P_C(Caco-2) values are quite similar.

8.5.1 Comparison with Caco-2 Permeabilities

It is clear that although the order of magnitude is different, all the Caco-2 data yield the same solute ranking. This is due to the strong similarity of the drugs under study, so that they have almost identical diffusion coefficients in the unstirred water layer and across the filter support, and they have similar pK_a values. From this point of view the simulations reproduce the experimental trend remarkably well. The main difference is the order of magnitude of the numerical permeability coefficients. The reason why the calculated values are 10^4 to 10^6 times higher than experiments may reside in the particular composition of the Caco-2 cells compared to a pure DPPC bilayer. The increase of unsaturated phospholipids in the Caco-2 membrane, for example, has been shown to reduce Caco-2 permeability[261] and the presence of cholesterol is known to reduce small solute permeabilities across lipid bilayers.[117,127,128,135,140,201,262] Caco-2 are entire cells; they have a cytoplasm with a gel-like consistency, a network of protein fibres (cytoskeleton), separated compartments (organelles) and enzymes. Moreover, the permeabilities of small organic molecules across egg lecithins is at least one order of magnitude lower than across pure DPPC bilayers (see table 6.4 in chapter 6 and references cited). However, this is not a problem in drug design as long as the solute ranking is correct.

Caco-2 permeabilities are employed in correlation studies with the fraction (FA) of drug absorption in humans after oral administration as a replacement for simpler physico-chemical properties. Since in these simulations only three drugs were studied, it is not worthwhile to investigate here such statistical correlations, but it can be argued that the relation between FA and calculated drug permeability coefficients across lipid bilayers could be as good as that between FA and measured permeabilities across Caco-2 cell monolayers. The advantage of using computer simulations would be that drugs do not need to be synthesised, but the disadvantage is that force field parameters are required to model the drug molecules. Often real drugs have "exotic" functional groups for which parameterization becomes a difficult task. Both techniques, computer simulation and the Caco-2 assay, at the current state of the art are not suitable for high throughput screening and can be time and resource consuming. In the context of this work, it can be said that, as computer performance increases very rapidly, such simulations

may become much easier and quicker in a few years.

8.5.2 Comparison with Hexadecane Permeabilities

The alprenolol permeability coefficient from these simulations agrees favorably with its P_o(hxd). The order of magnitude is the same, but the numerical value is about three times lower. Permeability across a lipid bilayer was expected to be lower than in bulk hexadecane. As shown by the local resistance profile, the main resistance to alprenolol permeation comes from the headgroup regions 1 and 2 of the membrane, and not from the hydrocarbon core of the lipid bilayer. Moreover, as shown for the small organics in chapter 6, bilayer constraints restrict the mobility and the flexibility of permeant molecules, so that entropic factors oppose solute partition into membranes. The agreement between alprenolol P from the simulations and that in a layer of hexadecane suggests that the difference with the Caco-2 permeabilities is not due to simulation failures, but rather to intrinsic differences between the two systems. The publication[260] also finds agreement between P_o(hxd) for benzoic acid and its P across egg lecithin bilayers,[118] suggesting that P_o(hxd) experiments are reliable. This in turn can arguably suggest that the permeabilities from these simulations would also agree with experimental permeabilities across various lipid bilayers, if they had been measured for the drugs of interest.

8.6 Bilayer Structure

To investigate whether the presence of two drug molecules inside the lipid bilayer disrupted the bilayer structure, the bilayer thickness and the lipid order parameters were calculated in these simulations and compared with those obtained from a pure DPPC simulation.

The electron density along the bilayer normal can be determined with X-ray spectroscopy and the peak-to-peak distance is used as a measure of the bilayer thickness. Experimental values range between 36.4 and 39.6 Å[91] and a value of 37.3 ± 0.5 Å was obtained in the simulation of a pure DPPC bilayer described in chapter 4. When inserting the drugs, simulations yielded the following results: 38.2 ± 0.3 Å with alprenolol, 38.2 ± 0.1 Å with atenolol, 38.3 ± 0.2 Å with pindolol. Although these values are about 1 Å higher than those for a pure DPPC bilayer, they are still well within the experimental range and very close to those obtained with most of the small organics. The differences with the pure DPPC bilayer may also reside in the different length of the simulations: 1 ns for the pure lipid membrane vs. 3 or 4 ns for the drug

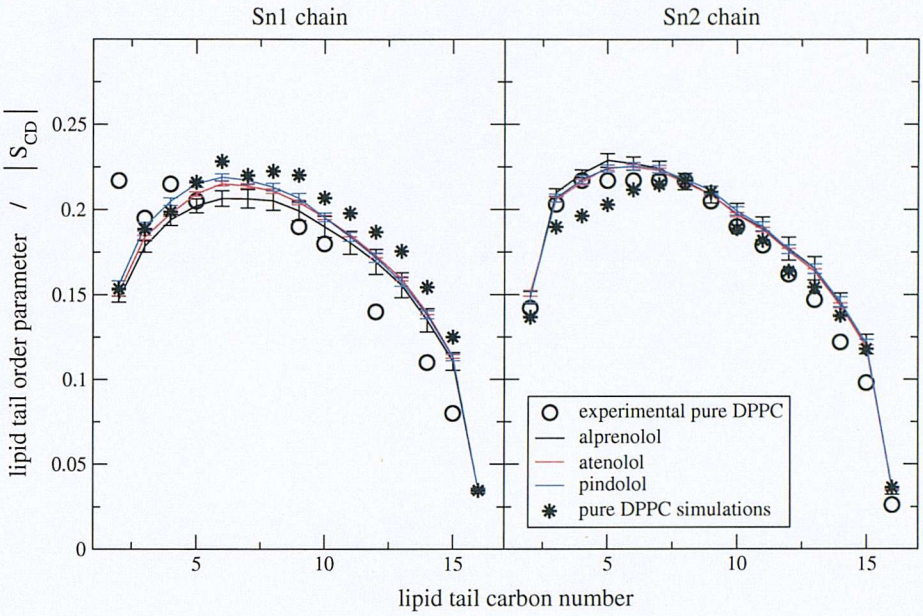


Figure 8.6: Lipid tail order parameters. Stars are values from the pure DPPC simulation and open circles are experimental values.[4–6]

permeation studies. In the original paper using the NP_NAT ensemble and the CHARMM force field, a peak-to-peak distance of 38.9 Å was obtained.[91]

The order parameter of the C-H bonds along the lipid chains can be determined by means of ^{2}H-NMR spectroscopy and it is used as a measure of the lipid order. In figure 8.6 the absolute values of the order parameters S_{CD} for the pure DPPC bilayer and for the drug/lipid systems are plotted. Stars are values from the pure DPPC simulation and open circles are experimental values.[4–6] It is clear that the presence of the solutes does not affect the lipid structure. Similar results were found when simulating a drug analogue in a bilayer of 36 DMPC molecules.[109]

8.7 Atomic Detail

8.7.1 Solute Flexibility

8.7.1.1 Population of Dihedral Torsions

Owing to the density of the lipid bilayer environment, it was thought that drug dihedral torsions could not rotate significantly during the simulations. This is why the most representative dihedral conformations were selected as described in chapter 7 and simulated separately. At the end of the drug/membrane simulations it is interesting to know the history of these torsions. It is likely that the simulations were not long enough to converge the equilibrium distribution of these angles, but some conclusions can be drawn.

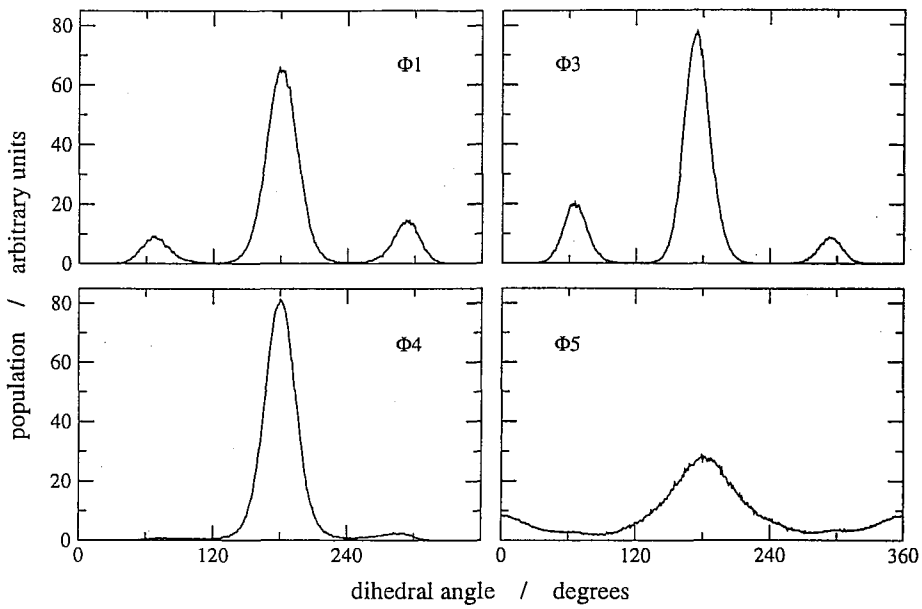


Figure 8.7: Population of drug tail dihedral torsions Φ1, Φ3, Φ4 and Φ5.

The implicit solvent studies described in the previous chapter yielded slightly different results among the three drugs regarding the most populated conformers along the drug tail, which is in all the molecules. In contrast, the drug/bilayer simulations yielded identical dihedral populations for all the drugs. The populations of dihedral torsions Φ1, Φ3, Φ4 and Φ5 along the drug tail are remarkably similar among the three drugs and the different z depths. The populations of these dihedrals are shown in figure 8.7 after averaging over the three drugs and the six z depths. Different conformers for each dihedral angle were used as starting structure for separate simulations, but in the end all conformers yielded the same population distribution in all the independent simulations. Figure 8.7 also shows that each dihedral torsion essentially prefers one angle only. This suggests that implicit solvent simulations are not sufficiently accurate to describe the angle distribution of these dihedrals. Only Φ3 was predicted to prefer a value around that which was the most populated in the lipid bilayer (i.e. 180°). For each of torsions Φ1, Φ4 and Φ5 more than one value was predicted, and these values were different from the most populated in the lipid bilayer (i.e. 180°).

Alprenolol and atenolol also have a side chain, whose first torsion, here called Φ6, yielded similar results to those in the main chain: the same distribution is found in both drugs and at all depths. This is shown in figure 8.8, after averaging results from the two drugs and the six z depths. However, in contrast to the dihedral torsions along the drug tail, Φ6 has two major populations, one around 30° and one around 210°, the former with a broader distribution and the latter with a sharper one. In this case, therefore, the implicit solvent simulations of atenolol

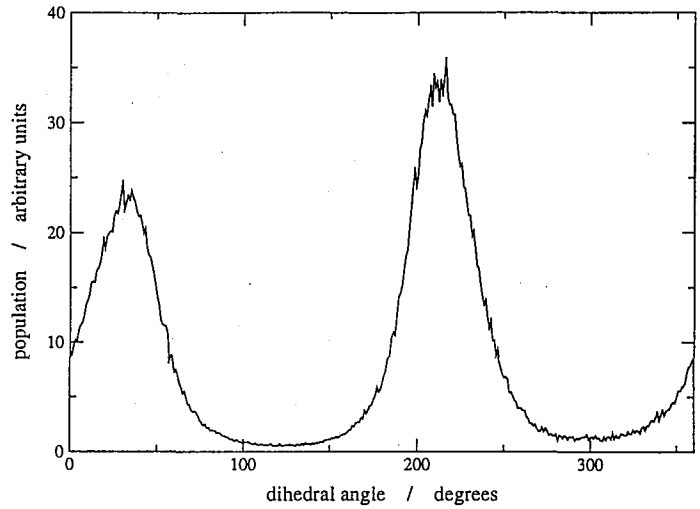


Figure 8.8: Population of drug side chain torsion Φ6.

yielded correct results in both the value of the angle and the number of preferred populations.

Along the drug tail, Φ2 has a unique behaviour in the drug/bilayer simulations. This torsion has a similar population for all three drugs, but adopts two different conformations at different z depths. This is shown in figure 8.9. Again, the population reported in the figure derives from averaging the results from the three drugs, since no significant differences exist among them. It is clear that moving from the water phase towards the bilayer centre, values of around 180° become less and less populated, while values around 300° are more and more preferred. For this torsion, implicit solvent simulation could only find populations at 80° and 300°.

Combining the above information about the dihedral torsions, two main drug configurations could be drawn in figure 8.10: one preferred closer to the water/lipid interface, and one preferred closer to the bilayer centre. The difference is in the value of torsion Φ2. It is clear that moving from the water phase to the bilayer centre these drugs become more elongated. This has also important consequences for the internal hydrogen bonds along the drug tail, as will be described later in this chapter. Another important conclusion is that there was no real need to locate important conformers in implicit solvent and simulate all of them in the lipid bilayer, since all yielded the same angle distributions in the end. Only one conformer needed to be simulated at each z-depth in the up and down orientation for each drug, reducing considerably the number of simulations required. This could not have been determined a priori, but this conclusion makes future work easier and computationally less demanding.

Simple conformational analysis of the three drugs simulated here were performed by Artursson *et al.* in vacuum[154] and in chloroform and water, using a continuum representation of

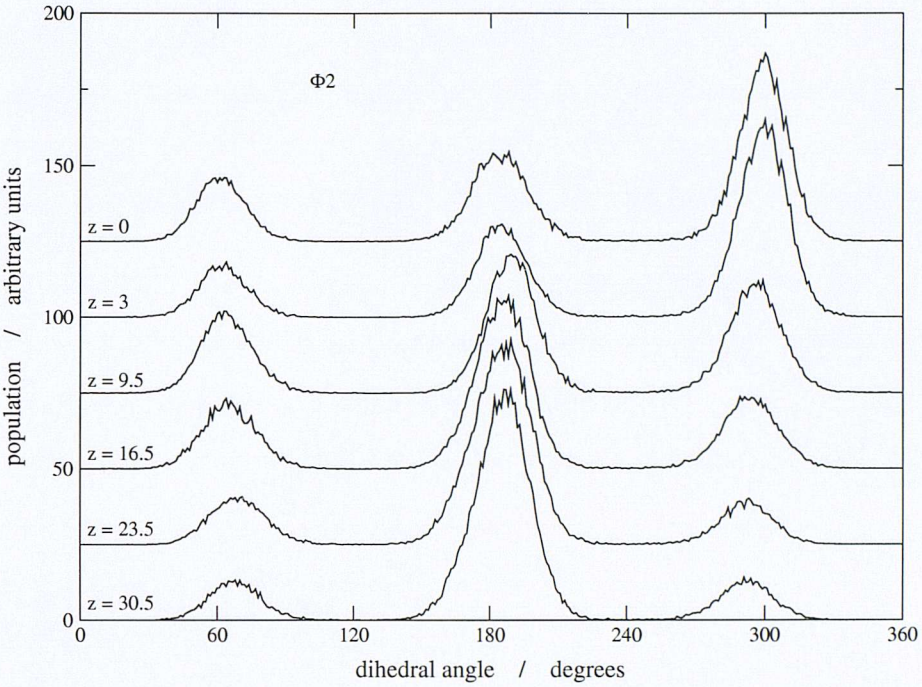
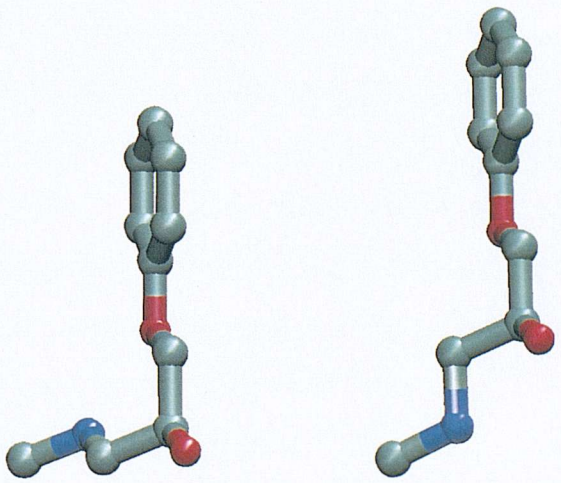


Figure 8.9: Population of drug tail torsion Φ2.



(a) Φ2 = 180° (closer to the interface)

(b) Φ2 = 300° (closer to the bilayer centre)

Figure 8.10: The two major drug configurations. Drug side chains, hydrogens and the two -CH_3 groups forming the isopropyl fragment at the end of the drug tail are not shown for clarity.

the solvent.[156] Details about some of the dihedrals are reported for the calculations in vacuum only. Only the value of torsion Φ4, which is mostly around 180°, agrees in both Artursson's publication and in this lipid bilayer. Results for Φ2 and Φ3 disagree. The difference in the value of Φ3 also has consequences for the intramolecular H-bonds, as explained later.

8.7.1.2 Interconversion Rates of Dihedral Torsions

From these simulations it was impossible to obtain any numerical information about dihedral interconversion rates, since the time scale for these motions is too long compared with the short time of the computer simulations. Using the same autocorrelation function employed for the lipid hydrocarbon chains (see equation 4.2), it was clear that the decay time would be of the order of several ns.

8.7.1.3 Drug Vector

The ability of the drug to fold on itself is studied by investigating the vector linking the centre of geometry of the drug head and the centre of geometry of the drug tail. This vector is here simply called the "drug vector". The more the drug is folded on itself, the shorter this vector becomes. Figure 8.11 plots the length of the drug vector at different z depths in the lipid bilayer. The length of this vector was also calculated for the drug in a hypothetical all-trans configuration. In this configuration the drug tail is as long as possible and the length of the drug vector is the longest. This length is also plotted in figure 8.11 with a short dashed line. This length is the same for the three drugs, since it does not depend on the drug side chain.

The length of the drug vector is quite stable in regions 1 and 2, where values are close to that in water phase. Then, it suddenly increases in region 3 and reaches its longest values in region 4. This behaviour is in perfect agreement with the reported population of dihedral torsion Φ2. However, the drug vector never reaches its maximum length as calculated from the all-trans configuration. This means that everywhere in the membrane the drug molecule is folded on itself, although to a small extent.

Another feature of these analyses is that alprenolol vector, with the exception of region 4, is always 0.2-0.3 Å shorter than atenolol and pindolol vectors. This difference is very small, particularly if compared with the 6 Å of the total vector length, and it may be a chance result. However, the possibility of alprenolol folding on itself a little more may also be one of the reasons for its lower $\Delta G(z)$ profile.

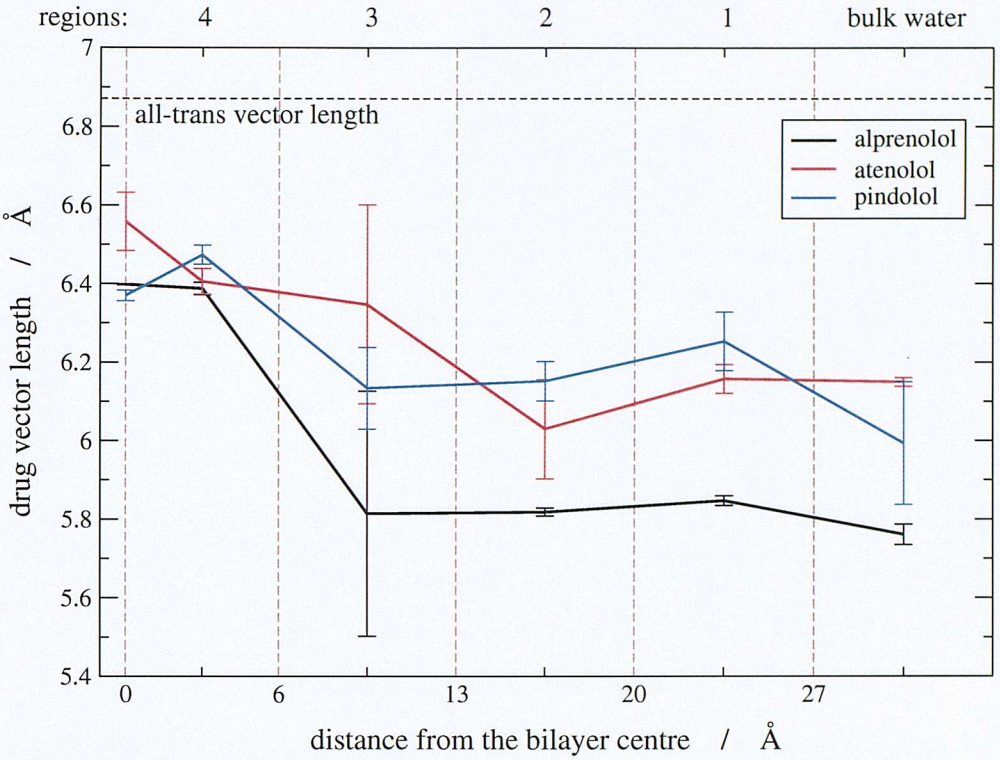


Figure 8.11: Length of the drug vector as a function of depths in the lipid bilayer. The length of the vector for the drugs in all-trans configuration is also plotted with a short dashed line.

8.7.2 Dipole moment

Drug dipole moments are plotted in figure 8.12 as a function of depth in the lipid bilayer. As was found for most of the small organic solutes, although the dipole moment plots are almost linear, there is an indication that moving from the water/lipid interface towards the bilayer centre the permeants tend to reduce slightly their net dipole moment, in accordance with the apolar nature of the membrane hydrocarbon core. The highest values are in the highly charged regions 1 and 2. Alprenolol is an exception: although its dipole moment in regions 1 and 2 is higher than in region 3, the highest value is reached in region 4.

The errors in figure 8.12 are calculated from the differences between the up and down orientations. The fact that they are small suggests that the simulations were reasonably converged. Dipole moments calculated from these simulations are in fair agreement with those calculated from optimized geometries and charges derived with more accurate models,[258,263] although values from these simulations are a little higher.

8.7.3 Hydrogen Bonds

The hydrogen bonding ability of the drug molecules is important because it is often correlated with drug partitioning in organic solvents and their permeability. The same geometrical criteria

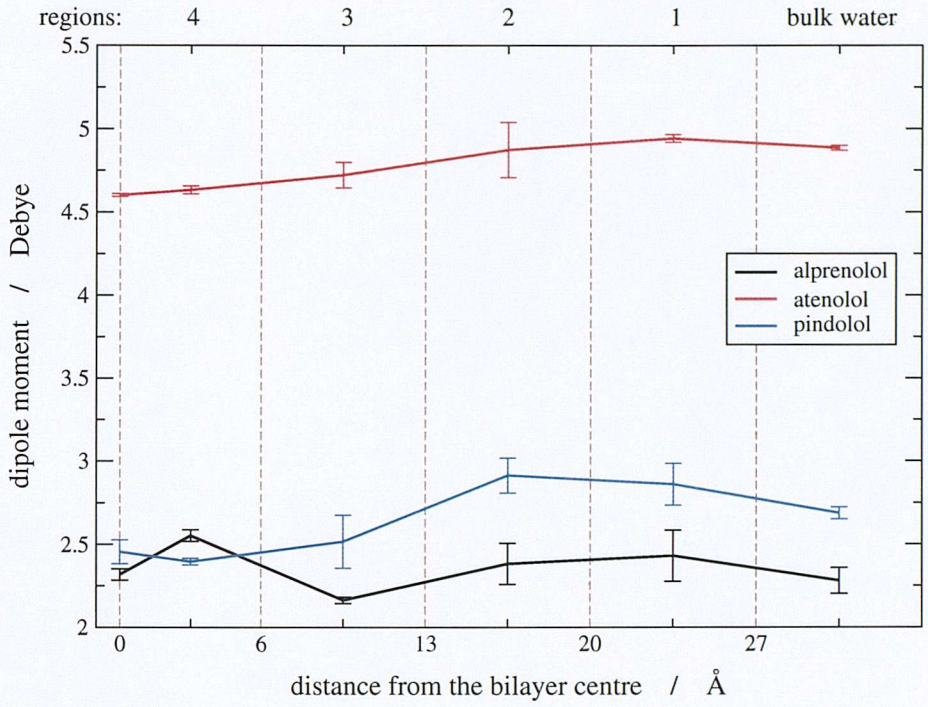


Figure 8.12: Drug dipole moments as a function of depths in the lipid bilayer.

chosen for the small organics (see figure 6.22) are also used here to identify H-bonds.

8.7.3.1 External H-Bonds

The mean number of H-bonds per simulation step between the drugs and water molecules is plotted as a function of depth in the lipid bilayer at the top of figure 8.13. For each drug, the total number of H-bonds is plotted with a solid lines. The contribution of each functional group is represented as a fraction of the total number and plotted with a dashed line. The distance between adjacent lines represents the number of H-bonds formed by the specified group. An example is given in figure 8.14.

Although the number of H-bonds is very small in region 4, hydration occurs at all depths, as was observed for the small organic solutes in chapter 6. The number of H-bonds with water involving the drug tail groups is remarkably very similar among the three drugs and therefore well represented by the number of H-bonds made by alprenolol, since this drug does not have any hydrogen bonding groups on its side chain. As expected, most of the H-bonds on the drug tail involve the -OH and -NH- groups, while the ether oxygen forms very few. Moving from the water phase towards the middle of the lipid bilayer, all the drug tail functional groups decrease the number of H-bonds with water uniformly. That is, the relative preference for a drug to form H-bonds with one of its groups does not change with depth in the bilayer. Regarding the

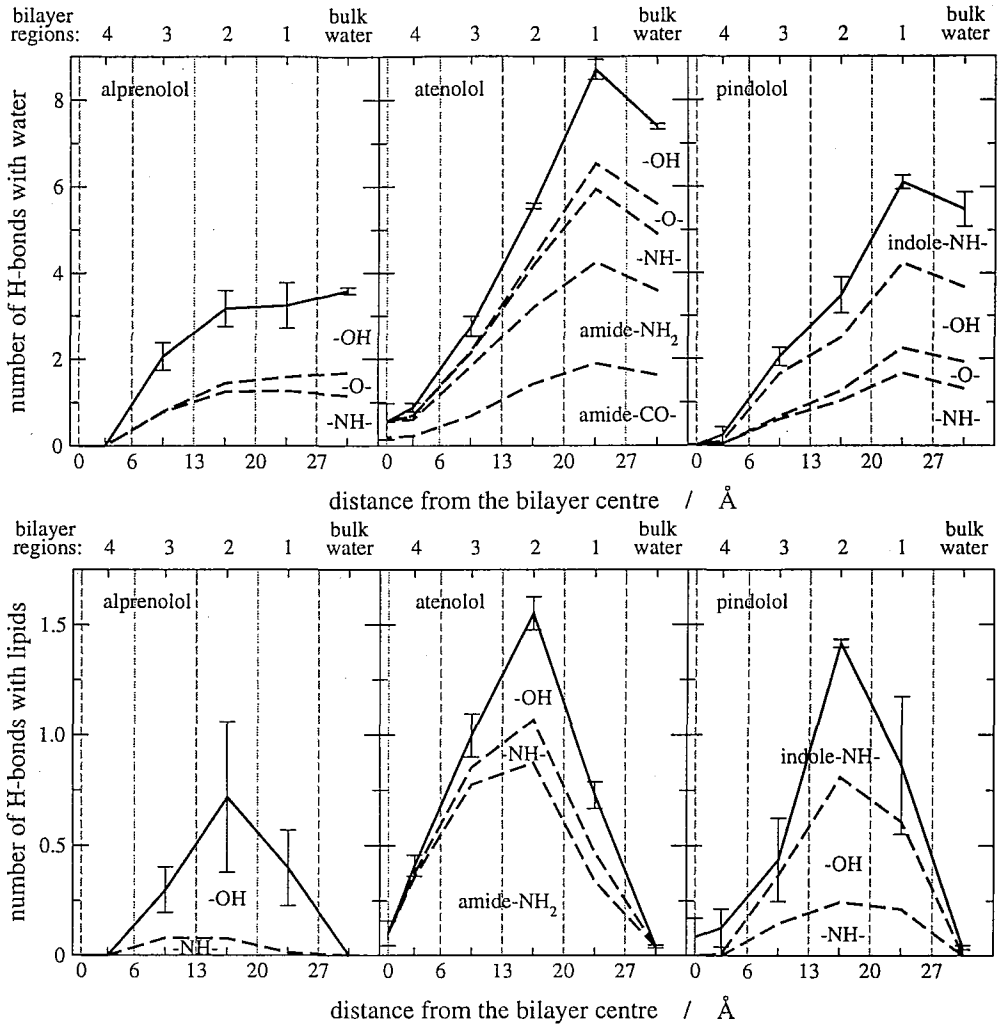


Figure 8.13: Top: mean number of H-bonds between drugs and water molecules. Bottom: mean number of H-bonds between drugs and surrounding lipids. Functional groups: -OH, -NH- and -O- are the hydroxyl, the secondary amine and the ether group respectively along the drug tail; amide and indole groups are on the side chains.

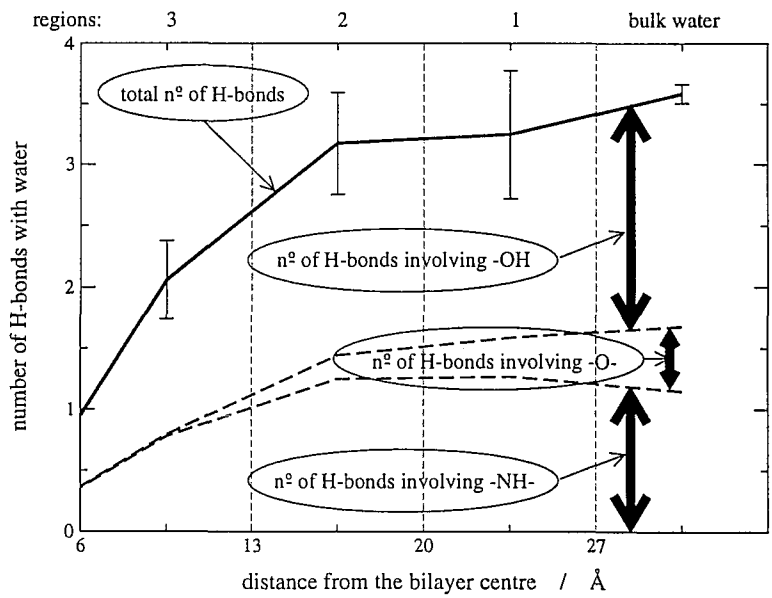


Figure 8.14: How to read H-bond plots.

side chains, the indole -NH- group in pindolol is involved in about as many H-bonds as the -NH- group in the drug tail. In total, pindolol forms about 1.5 times the number of H-bonds of alprenolol. The amide -CO- and $-NH_2$ groups together in atenolol form about as many H-bonds as the three groups in the drug tail, and in total atenolol forms approximately double the H-bonds of alprenolol.

The bottom of figure 8.13 reports the number of H-bonds between drugs and lipids. Since lipids do not have hydrogen bonding donors (see figure 6.22), the ether oxygen on the drug tail and the amide carbonyl oxygen on atenolol side chain cannot be involved in H-bonds with the surrounding lipids. As with water, the amide group in atenolol forms most of the H-bonds. On the drug tail, in contrast to what happens with water, -OH forms many more H-bonds than -NH-.

The mean lifetime of H-bonds involving drugs and waters is shown at the top of figure 8.15. Different functional groups are plotted separately. H-bonds on identical functional groups on different drug molecules last similar times, with hydroxyl and amine having longer-lived H-bonds than the ether oxygen. The side chain groups have mean lifetimes in between. The bottom of figure 8.15 shows the mean life of H-bonds between drugs and surrounding lipids. These H-bonds have lifetimes about 10 times shorter than those with water. In contrast, Stouch *et al.* found the solute interactions with lipids longer than those with water.[109] However, they did not use the same geometrical criteria employed here for defining H-bonds and, above all, the drug molecule in those simulations was not constrained to stay at specific depths.

The hydrogen bonding properties can be compared with those obtained with the small organic solutes (see figures 6.23 and 6.24). The number of H-bonds involving atenolol amide group with water and lipids is identical to that of acetamide. In contrast, the number of H-bonds involving -OH and -NH- on the drug tail are lower than those formed by methanol and methylamine, respectively. This is probably due to the fact that the small organics can freely rotate and their polar groups have no steric barrier, whereas the same groups on the larger drug molecule have restricted motions and access is sterically constrained. Another reason for lower H-bonds is that -NH- and -OH groups are also involved in intramolecular H-bonds, as described in the next section, and this decreases their availability for H-bonds with surrounding molecules.

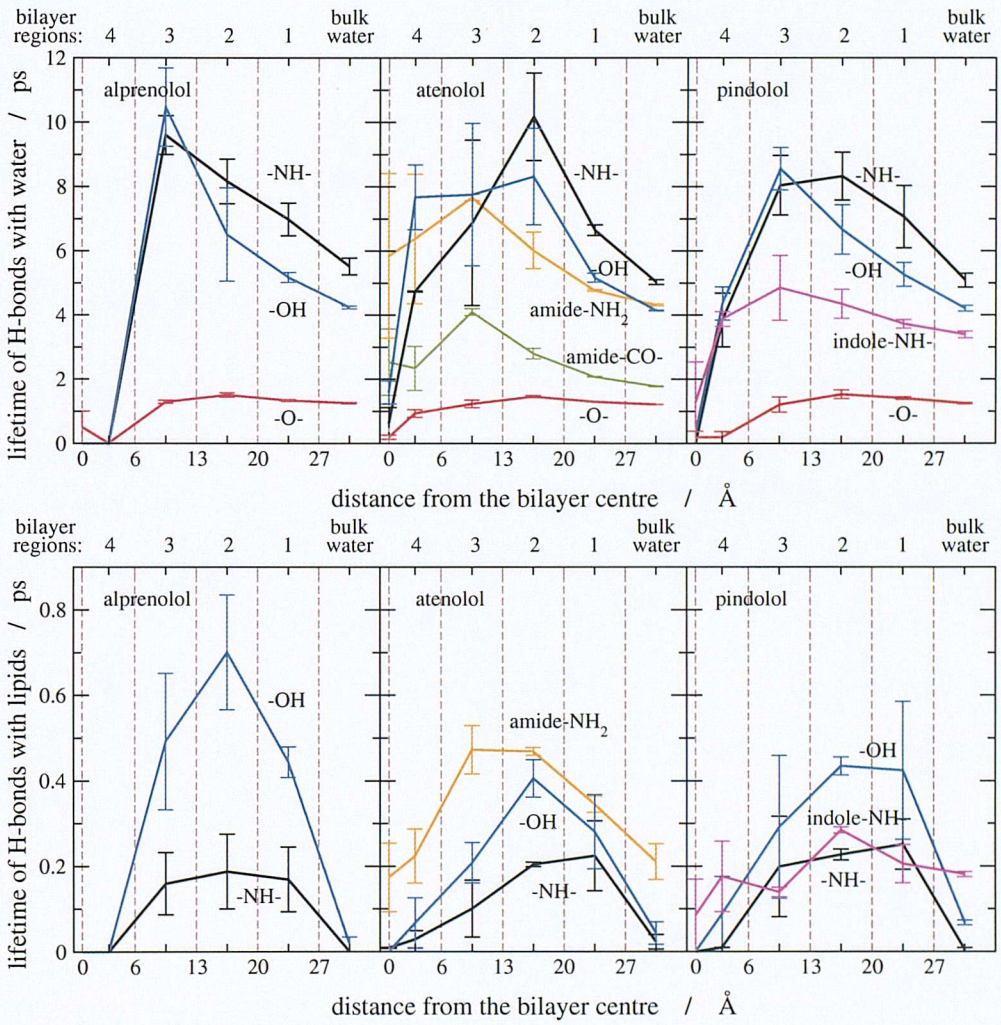


Figure 8.15: Top: mean lifetime of H-bonds between drugs and water molecules. Bottom: mean lifetime of H-bonds between drugs and surrounding lipids. Functional groups: -OH, -NH- and -O- are the hydroxyl, the secondary amine and the ether group respectively along the drug tail; amide and indole groups are on the side chains.

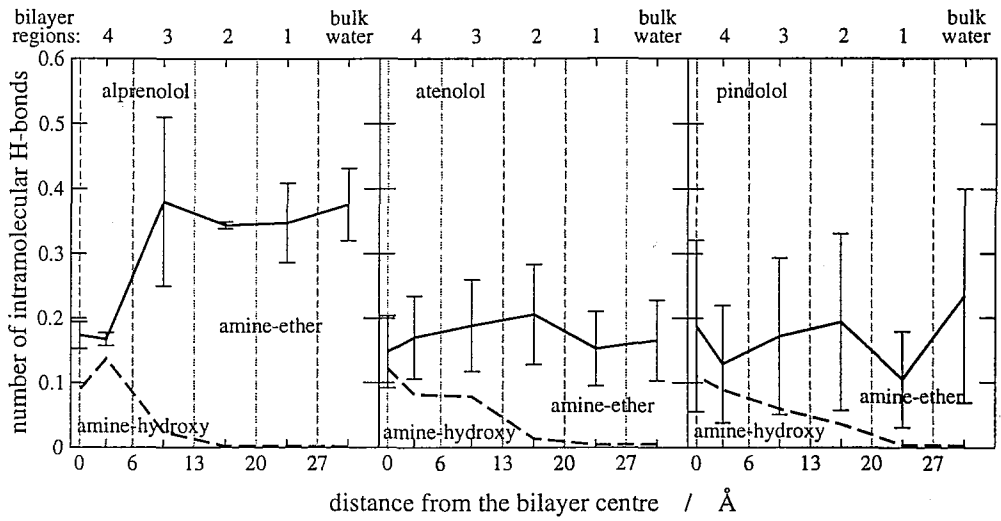


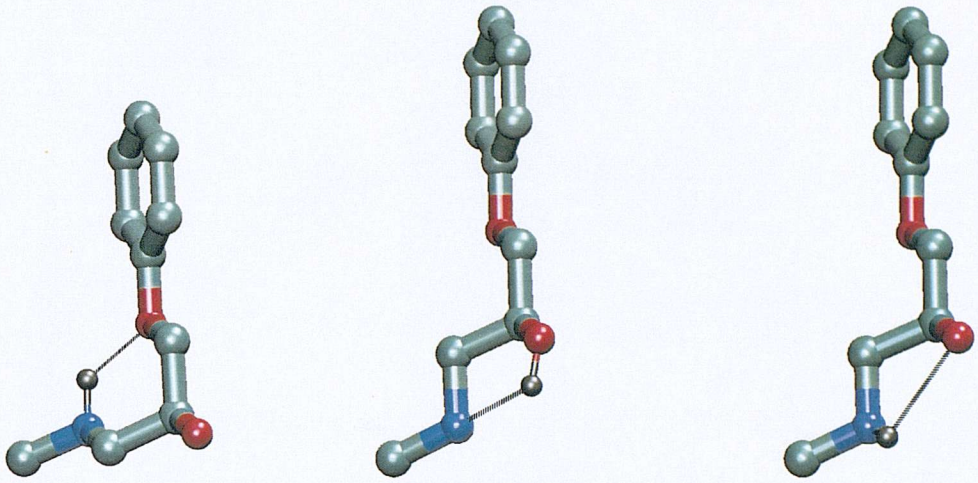
Figure 8.16: Mean number of intramolecular H-bonds between pairs of functional groups.

8.7.3.2 Internal H-Bonds

Owing to the flexibility of the drug molecules, internal H-bonds between pairs of functional groups can be formed. No H-bonds could be counted between the side chain groups (amide in atenolol and indole in pindolol) and any of the drug tail groups. On the drug tail, no H-bonds were found between the ether oxygen and the hydroxyl group. In contrast, some H-bonds involving amine-ether and amine-hydroxyl pairs were observed. These are shown in figure 8.16 at different z depths in the lipid bilayer. The total number of intramolecular H-bonds is plotted with a solilid line, and a dashed line separates the contributions from the two pairs, in the same way as described in figure 8.14.

The number of intramolecular H-bonds varies at different depths, but considering the error bars one could argue that it stays quite constant moving from region 1 to region 4. Looking at the contributions, the number of H-bonds involving -NH- and ether oxygen decreases significantly going deeper into the membrane interior, while the number of H-bonds involving -NH- and -OH increases of a similar amount. Almost no H-bonds exist between -NH- and -OH in region 1, almost no H-bonds exist between -NH- and ether oxygen in region 4. The sum of the two effects is that the total number of intramolecular H-bonds does not change significantly in different membrane regions. Therefore the conclusion is that from these simulations the lack of H-bonds between the drug and the surrounding lipids or waters when moving from the water phase towards the middle of the membrane is not compensated by an increase in the number of intramolecular H-bonds.

The preference for the -NH- group to form intramolecular H-bonds with the ether oxygen



(a) closer to interface

(b) closer to bilayer centre: involving hydroxyl H

(c) closer to bilayer centre: involving amine H

Figure 8.17: Intramolecular H-bonds: different pairs are involved when closer to the interface or to the bilayer centre. Only hydrogens involved in H-bonds are shown, in gray colour. H-bonds are drawn with stripped black lines.

when closer to the interface and with the -OH group when closer to the bilayer centre is related to the drug conformation, as mentioned at the end of section 8.7.1.1. Closer to the interface the drug is more folded on itself, the amine and ether groups point on the same side of the molecule while the hydroxyl faces the opposite side. Closer to the bilayer centre the drug is more elongated and the amine group points on the same side as the hydroxyl. This behaviour is shown in figure 8.17. Assuming that the overall drug conformation is more folded or more elongated because of steric constraints due to lipid packing, the possibility of maintaining a fairly constant number of intramolecular H-bonds by varying the pair of functional groups involved may facilitate the overall conformational change.

Simple conformational analysis of the three drugs simulated here were performed in vacuum[154] by Artursson *et al.* as also described at the end of section 8.7.1.1. They report that hydrogen bonds occur between the -OH and the ether oxygen located at the beginning of the drug tail. This is not found in these simulations. The explanation is based on the fact that in these simulations the two groups point towards opposite sides of the molecule (see again figure 8.17), whereas in the vacuum studies they face each other. This is due to the different value of torsion $\Phi 3$ found in the publication with respect to that mostly populated in these simulations.

Finally, the mean lifetime of intramolecular H-bonds is shown in figure 8.18. The results are basically identical for all three drugs. Whatever the actual number of H-bonds, the mean lifetime of each is fairly constant at all depths, with those between -NH- and ether oxygen lasting

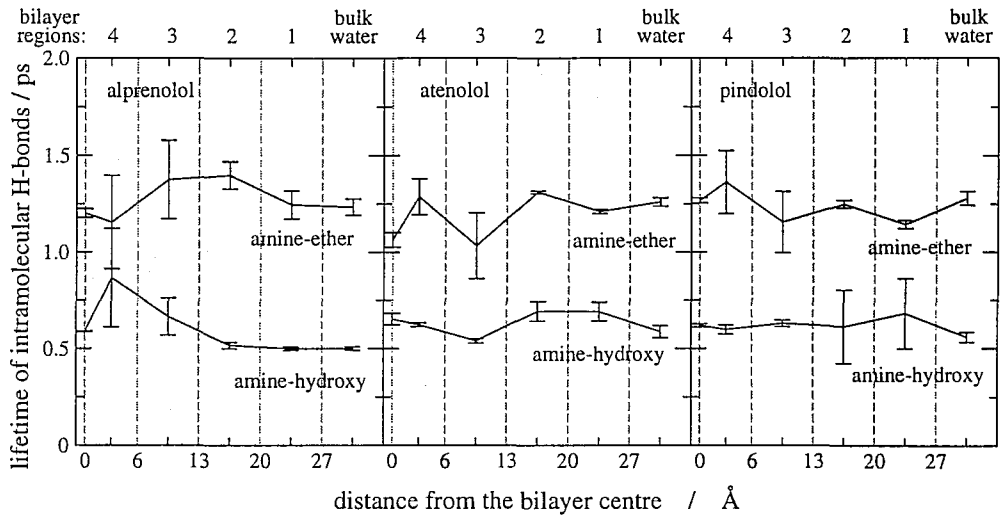


Figure 8.18: Mean lifetime of intramolecular H-bonds between pairs of functional groups.

almost twice as long as those between -NH- and -OH. The mean lifetime of intramolecular H-bonds is somewhere in between that of drug-lipids H-bonds and that of drug-water H-bonds.

8.7.4 Solute Orientation

8.7.4.1 Drug Head Orientation

The orientation of the aromatic ring representing the drug head was calculated in the same way as for the benzene molecule described in section 6.9.2 (see equation 6.17). Briefly, the Legendre polynomial of rank 2 (simply S) was calculated for the angle between the normal to the ring plane and the bilayer normal. Its average over the simulation time and eventually over the conformers sampled is plotted in figure 8.19 as a function of z depth in the lipid bilayer.

At $z = 30.5$ Å, S is zero for all the drugs. This value indicates full isotropic motion and random distribution of atoms. This value was expected, because at this position drugs are still in the water phase and can freely rotate around their centre of mass.

Deeper into the membrane S is significantly different from zero. Owing to the orientation of the lipid molecules and the tight lipid packing, negative values were expected in the membrane interior, indicating that the plane of the aromatic ring parallels the bilayer normal. In contrast, positive values were also found. For atenolol and pindolol this mainly occurs in region 2, while for alprenolol it happens in region 3. In these cases, the plane of the drug head lies perpendicular to the bilayer normal. The reason for this behaviour seems to be the possibility for the drug to form a larger number of H-bonds. The side chains of atenolol and pindolol have groups that

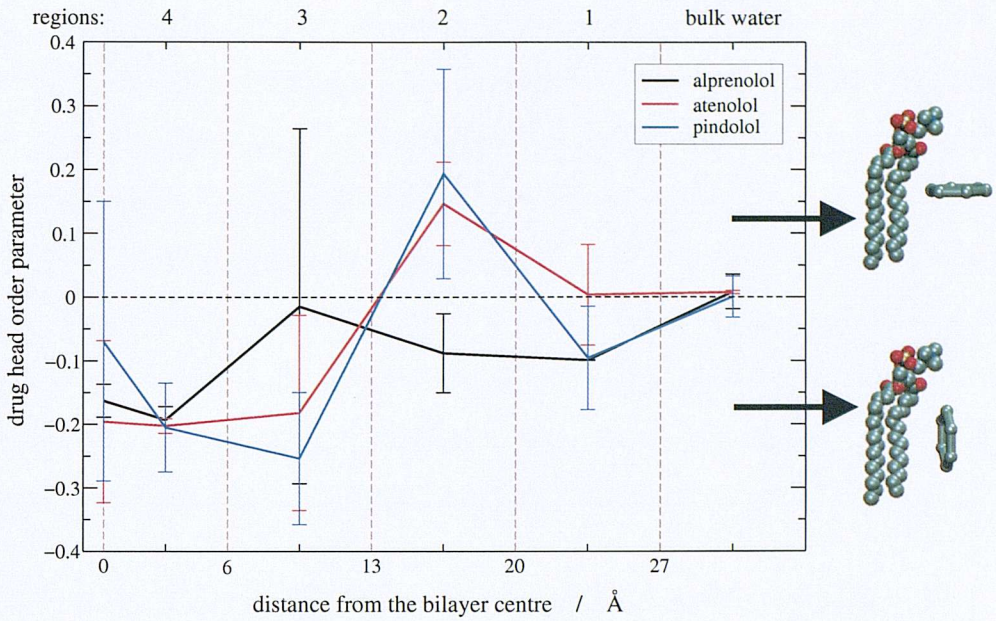


Figure 8.19: Order parameter of drug head as a function of z depths.

are able to form H-bonds, as well as the drug tail. Since the two chains point towards different sides of the drug head, the chain which is further from the interface causes the drug to rotate and place itself in such an orientation that allows the said chain to optimize its H-bonds. This effect is much stronger when the side chain of atenolol and pindolol is located further from the interface rather than the drug tail. Regarding alprenolol, its side chain is not able to form H-bonds. Therefore, this drug orients its head perpendicular to the bilayer normal when its tail, not its side chain, is further from the interface.

That drug molecules prefer to stay tilted with respect to the bilayer normal in order to optimize their hydrogen bonding with their environment was also observed by Stouch *et al.* in a simulation of a nifedipine analogue located near the interface of a DMPC bilayer.[109]

This behaviour was not observed with the small organic solutes, since they are small enough to optimally orient their polar groups while perpendicular to the bilayer normal. Moreover, in the case of benzene, which is similar to the drug head, that molecule had no groups able to form H-bonds; if phenol or similarly asymmetric and polar molecules were simulated, the aromatic ring would have probably assumed orientations similar to those observed here for the drug molecules.

The conclusion is that hydrogen bonding opportunities seem at least as important as steric restraints in determining the permeant orientation. When hydrogen bonding sites can be saturated, the solute prefers to stay parallel to the bilayer normal, because this orientation does

not require modifying the bilayer's stable structure. However, when H-bonds cannot be formed, the lipid molecules can actually rearrange themselves to make their formation possible. The experimental observation that solute permeation is reduced in those conditions where the lipid packing is higher, for example with increasing concentration of cholesterol or at lower temperatures,[127,128,135,201,262] is explained on the basis of solute exclusion, i.e. on the basis of more work required to form a cavity able to host a solute. These simulations offer an additional explanation: when lipid packing is higher solutes cannot optimize their hydrogen bonding network, because they cannot orient themselves to form the largest amount of H-bonds. This in turn reduces solute partitioning into the membrane and finally its permeation.

8.7.4.2 Drug Vector: Up and Down Orientations

The overall drug orientation is studied by investigating the vector linking the centre of geometry of the drug head and the centre of geometry of the drug tail. This vector is here simply called "drug vector". This vector was introduced in section 8.7.1.3 dealing with the drug flexibility. In that section its length was studied. Here its orientation is investigated by calculating the cosine of the angle between the vector and the bilayer normal. This is the Legendre polynomial of rank 1 of that angle, and is referred to in these analyses as the drug vector P_1. Obviously it can range between +1 and -1. A value of zero indicates either that the drug lies perpendicular to the bilayer normal, or that it can freely interconvert its orientation up→down and down→up without any restrictions (isotropic motions). An absolute value of 1 indicates instead that the drug is completely parallel to the bilayer normal. The sign has the following meaning: a positive value indicates the drug is in the down orientation, and a negative value indicates the drug is in the up orientation. Results from these simulations are plotted in figure 8.20, together with a picture of the drug in the up and down orientation. For each saved coordinate set (one every ps) P_1 is calculated and averaged over the full simulation length. This average is $\langle P_1 \rangle_t$, with $\langle ... \rangle_t$ indicating time average. For atenolol and pindolol figure 8.20 plots the further average among the individual conformers with the associated standard errors. For alprenolol, figure 8.20 plots $\langle P_1 \rangle_t$ and the errors are calculated by dividing the one simulation into frames of 200 ps. When the drug orientation in the starting structure of the simulation was UP, the average drug orientation at the end of that simulations is plotted with a solid line. When instead the drug was in the DOWN orientation in the starting structure, the average drug orientation from the simulation is plotted with a dashed line. Therefore, the terms UP and DOWN inside the

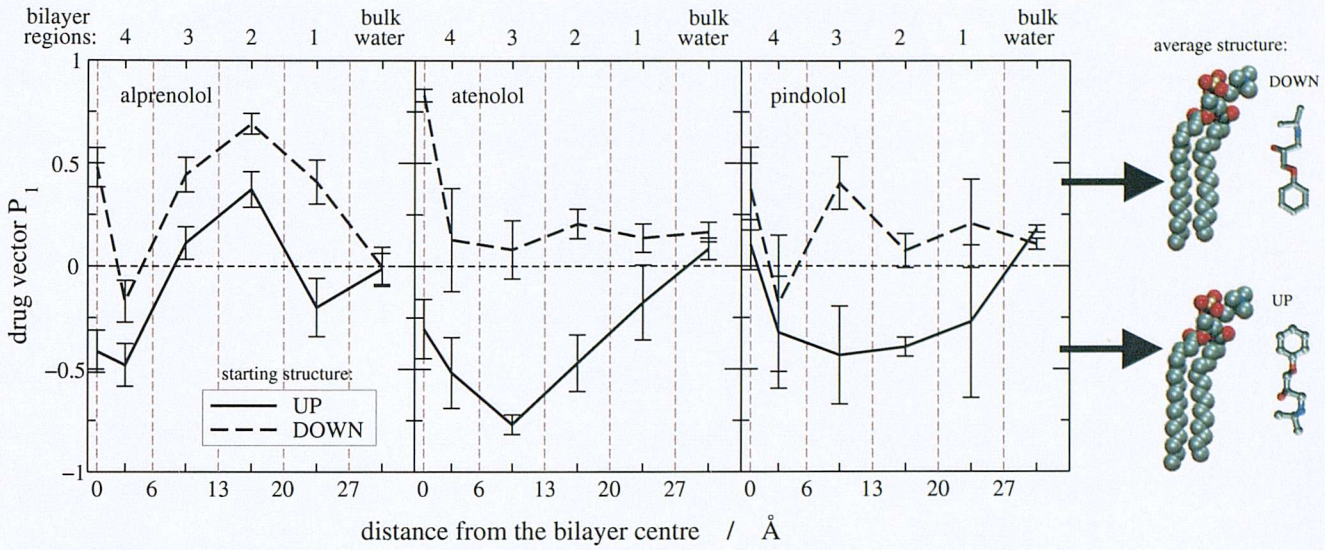


Figure 8.20: Drug vector P_1 as a function of depth in the lipid bilayer. See text for definitions.

plot legend refer to the orientation in the starting structure which yielded that plot. Instead, the terms UP and DOWN on the picture of the molecules simply show that when P_1 is positive the drug orientation is DOWN and that when P_1 is negative the drug orientation is UP.

At the furthest distance from the bilayer centre ($z = 30.5$ Å) P_1 is basically zero in all plots. At this position, drugs are still in water phase and they can freely rotate. Studying the time evolution of P_1 in the single simulations reveals that the value of zero does not derive from the drugs laying on the membrane plane, but rather the drugs have isotropic motions and can freely change their orientation, with P_1 spanning over the entire range from +1 to -1. Between 10 and 20 clear and net interconversions (up→down and down→up) could be counted for alprenolol (whose simulations lasted 4 ns), and between 5 and 10 interconversions for atenolol and pindolol (simulations lasted 3 ns).

Inside the membrane, a higher degree of order is found. Here drug orientation is difficult to study because of the long time scale of these motions with respect to the short time scale of computer simulations. In figure 8.20, if the solid lines do not reach positive values and the dashed lines do not reach negative values, it means that the drug orientation in the starting structure (either up or down) is maintained for most of the simulation and most of the conformers. However, from figure 8.20, it appears clear that this happens in the case of atenolol only. Although the starting orientation is still preferred, even for this drug up→down and down→up interconversions occurred at all z depths. For alprenolol and pindolol, the starting structure is preferred at most of the z depths but not all of them, and up→down or down→up intercon-

versions occurred at all depths. For alprenolol in particular the preferred orientation was down for both starting structures in regions 2 and 3. Although solid and dashed lines cross the zero line, starting from the up or the down orientation did not yield the same profile; the two curves still differ. This suggests that the simulations were not long enough to draw any conclusion about the equilibrium distribution of orientations of these drugs inside a lipid bilayer. This is the reason why $\Delta G(z)$ profiles from the two orientations differ. Another important conclusion is that, since some up→down or down→up interconversions can occur at all z depths during such short computer simulations, the concept of minimum resistance path, as described in the section dealing with the local resistances, probably holds in real biological membranes. The experimental observation that the solute permeation is reduced in those conditions where the lipid packing is higher, for example with increasing cholesterol concentration or at lower temperatures,[127,128,135,201,262] is explained on the basis of solute exclusion, i.e. on the basis of more work required to form a cavity able to host a solute. These simulations offer an additional explanation: when lipid packing is higher solutes cannot interconvert their orientation as easily and the minimum resistance path becomes less accessible, reducing the overall permeability coefficients.

Further important information may be obtained from these analyses. Looking again at figure 8.20 it can be seen that on average for atenolol the solid line has higher absolute values than the dashed line. This suggests that for atenolol the up orientation is more conserved than the down orientation. In other words, the drug which possesses a side chain on its head with a highly polar group able to form a large number of long-lived H-bonds prefers to stay with the head up towards the water phase rather than down with the head towards the membrane interior. The inverse situation can be described for alprenolol. This drug does not have the possibility to form H-bonds with the side chain on its head, but it still has hydrogen bonding ability on its tail. Figure 8.20 shows that the alprenolol down orientation, with the tail oriented towards the water and the head towards the bilayer centre, is indeed preferred, even when the starting structure contained an up orientation (in regions 2 and 3). The situation is somewhere in between for pindolol. It can form H-bonds with the side chain on its head, but not as many as the amide group on atenolol. In the end, the curves for pindolol are more similar to those for atenolol than to those for alprenolol.

These findings agree with experimental observables that predict a drug orientation inside biological membranes such that the hydrophobic parts of the molecules are placed towards the

membrane interior and the hydrophilic ones towards the lipid/water interface.[232–235,264] The drug vector orientation is also in agreement with the drug head orientation described in the preceding section.

8.7.4.3 Drug Vector: Parallel or Perpendicular to Bilayer Normal

Drug orientation is not however completely described by the value of P_1.

When several conformers are studied for the same drug at the same z depth and with the same starting orientation (up or down), as was done for atenolol and pindolol, in one simulation the molecule may stay in the same configuration as in the starting structure for most of the simulation and then interconvert just in the very last steps, whereas in another simulation it may interconvert just a few steps after the beginning and stay in the new orientation for the rest of the simulation. The average orientation from the two simulations would then predict $\langle P_1 \rangle \approx 0$, i.e. either that the drug lies perpendicular to the bilayer normal or that it experiences isotropic motions. The first interpretation is not true, as in both simulations the individual P_1 values were close to +1 and -1 respectively. The second interpretation is not true either, as one interconversion only occurred in the simulations and one cannot conclude that the drug molecules can freely rotate. Only the error bars can give an indication that these interpretations are too simplistic. The same considerations apply for alprenolol too, once the separate simulations of different conformers are replaced by the frames into which the one simulation was divided for the error calculation.

The conclusion is that $\langle P_1 \rangle_t$ plotted in figure 8.20 can only indicate if on average the drugs are in the up or down orientation and if they maintain the orientation of the starting structure. To know whether they tend to stay preferentially parallel or perpendicular to the bilayer normal, the order parameter of the drug vector is more appropriate. The order parameter of the drug vector is here called P_2 and is calculated in the same way as for the drug head (section 8.7.4.1), i.e. from the Legendre polynomial of rank 2 of the angle between the drug vector and the bilayer normal. A value of P_2 equal to 1 indicates that the drugs are aligned parallel to the bilayer normal, a value of -0.5 indicates that the drugs lay parallel to the bilayer surface. P_2 values are plotted in figure 8.21, together with a picture showing the meaning of the values.

At $z = 30.5$ Å, all three drugs have $P_2 \approx 0$, confirming that $\langle P_1 \rangle_t = 0$ at that z depth was really due to isotropic motions. Inside the membrane, most of the values are positive, meaning

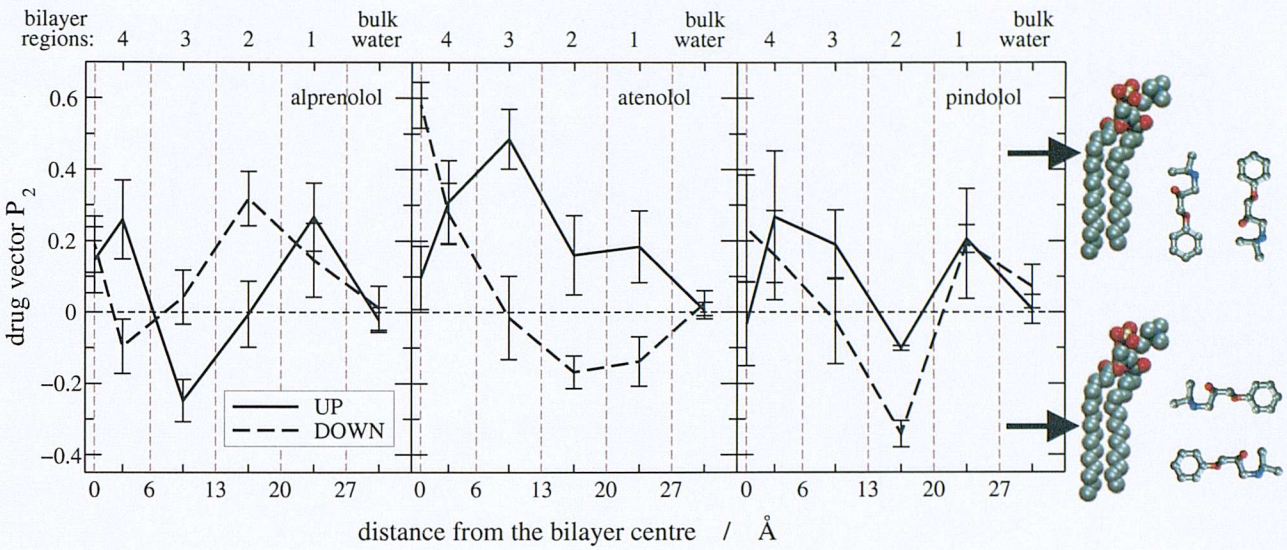


Figure 8.21: The order parameter of the drug vector as a function of depth in the bilayer.

that the drug molecules tend to parallel the bilayer normal. However, these values are small and many negative values are present in the plots. This indicates that on average the drugs are tilted with respect to the bilayer normal. Such an orientation was also found by Stouch *et al.* in a simulation of a drug analogue in a DMPC bilayer,[109] and also agrees with the behaviour observed for the drug head in section 8.7.4.1. The same explanation may apply for the drug vector: the molecules orient themselves in the membrane to optimize their H-bonds.

Considering P_2 and P_1 plots together, the conclusions are: drugs are mainly tilted with respect to the bilayer normal, up→down and down→up interconversions are possible, and the preferred drug orientation is such to maximise the number of H-bonds.

8.7.5 Correlation Between Preferred Orientation and Resistance Profile

P_1 plots show that the average preferred orientation is not the same for all the drugs. However, they also show that in the region of the membrane offering the main resistance to permeation all three drugs tend to stay in the down orientation. This occurs close to the interface for alprenolol and in the bilayer centre for atenolol and pindolol.

It appears that there is a contradiction between the preferred orientation of the drugs in the barrier region of the membrane, predominantly down, and the fact that the local resistances are smaller for the up orientation. We believe that this may be explained as follows. At any given depth in the membrane, the permeant will orient itself to minimise its free energy, for

instance by maximising the number of favourable non-bonded interactions. For these molecules, this involves orientation with the tail group pointing predominantly towards the headgroups, i.e. the down geometry. To further optimise these interactions, the drug will attempt to move towards the polar headgroups and solvating water molecules. However, the applied z-constraint will prevent this, with the result that the calculated constraint forces will be large. This in turn will increase the free energy of partitioning, and hence the local resistance. This means that arguably a more reliable estimate of the permeability is not the lowest resistance path. However, further extensive simulations that allow for convergence of the drug orientations are really required to investigate this.

8.7.6 Solute Orientational Times

The solute reorientational times were studied by investigating two types of motions: those of the aromatic ring representing the drug head and those of the drug vector representing the overall drug orientation. Since the drug models were fully flexible, these two motions can in principle be independent.

8.7.6.1 Drug Head Orientational Times

As previously mentioned, the drug head orientation was studied in terms of the angle between the normal to the aromatic ring plane and the bilayer normal. In section 8.7.4.1, the order parameter of this angle was calculated. Here, the following time autocorrelation function is studied:

$$C(t) = \langle P_2[\vec{\mu}(t) \cdot \vec{\mu}(0)] \rangle \tag{8.1}$$

P_2 is the Legendre polynomial of rank 2 and $\vec{\mu}$ is the unit vector along the normal to the ring plane. The equation turns into:

$$C(t) = \frac{1}{2}\langle 3\cos^2\alpha - 1 \rangle \tag{8.2}$$

where α is the angle between $\vec{\mu}$ at time t and itself at time 0.

Time autocorrelation functions could be fitted with a double exponential for $z = 30.5$ Å (i.e. when drugs are still in water), with two characteristic decay times τ_{long} and τ_{short}. Inside the membrane, instead, triple exponentials were required for a complete description of these motions, with a third decay time τ_{flip}. Correlation coefficients were all higher than 0.97. The results for all drugs were very similar.

τ_{short} is between 0.5 and 5 ps and it corresponds to the oscillations of the aromatic ring around a stable orientation. τ_{long} is instead related to significant orientational changes and is between 10 and 100 ps. τ_{flip} is in the order of magnitude of several hundreds of ps and in some cases a few ns, even longer than the total simulation. This decay time is difficult to interpret, but it seems to be related to drastic changes such as the interconversion of drug orientation from up to down or vice versa. Probably, it is not the case that at $z = 30.5$ Å τ_{flip} does not exist, but rather it is simply as fast as other motions and is hidden by, for instance, τ_{long}. If fitting is done with triple exponential at $z = 30.5$ Å too, it is seen that indeed τ_{long} and τ_{flip} have similar values.

From these simulations, there is no direct correlation between location inside the membrane and decay times. This suggests that either the statistics are very poor, or that drug head motions have on average similar times at all depths. The first hypothesis could be true for τ_{flip}, since it is very long or even longer than the whole simulation, but it cannot be true for τ_{long} or τ_{short}.

8.7.6.2 Drug Vector Orientational Times

The same time autocorrelation function was employed for studying the reorientational time of the drug vector. In this case, $\vec{\mu}$ is the unit vector along the drug vector itself.

The same results obtained with the drug head were also obtained with the drug vector. Time autocorrelation functions could be fitted with a double exponential at $z = 30.5$ Å, while they required triple exponentials inside the membrane. Similar values for the three decay times were found and they were related to the same type of motions.

It seems then that drug head and vector are closely related, since their motions have identical characteristic times. Looking again at figure 8.10 it is easy to understand a possible reason for this behaviour. The picture shows the preferred drug configurations. It is clear that the drug is preferentially elongated and any time its head changes orientation with respect to the bilayer normal the same change in orientation occurs for the drug vector, or vice versa. If one calculates the normalized time autocorrelation function of the difference between the angle formed by the drug head with the bilayer normal and the angle formed by the drug vector with the bilayer normal, at $z = 30.5$ Å a decay time around 50 ps is obtained, but in the membrane interior it is in the ns time scale. This confirms that the two molecular fragments have correlated motions in the membrane interior.

8.8 Conclusions

In this chapter, permeability coefficients across a phospholipid bilayer were calculated for three drugs belonging to the class of β-adrenoreceptors antagonists, by means of MD simulations. Experimental permeabilities in identical systems are not available, but permeabilities across cell monolayers are commonly measured. Although permeability coefficients from these simulations are orders of magnitude different from these experimental permeabilities, the ranking of the three drugs perfectly reproduces the experimental trend. The value for one of the drugs agrees favourably with that measured across a layer of hexadecane, suggesting that the difference in the numerical permeabilities is due to intrinsic differences between whole cells and lipid bilayers. The drug ranking in terms of permeabilities also reproduces the ranking in terms of partitioning between water and 1-octanol, although the partition coefficients themselves are different from those calculated from the free energies of drug water$\rightarrow$membrane transfer in these simulations. Partitioning between water and 1-octanol overestimates the drug ability to dissolve into the membrane. Owing to the high similarity in size and chemical structure between the three drugs studied, their diffusion coefficients are almost identical and the only contribution to their relative permeability comes from the difference in free energy of drug partition from water into the membrane.

Analysis of dihedral torsions and analysis of the vector linking the centre of geometry of the drug head with that of the drug tail show that drugs tend to stay elongated inside the lipid bilayer. Only at the very end of the drug tail is there a choice to stay more folded (closer to the interface) or more stretched (closer to the bilayer centre). This has important consequences for the intramolecular hydrogen bonds, with different pairs of functional groups involved depending on the drug conformation. The preference for the drug to stay elongated also makes the motions of its head (an aromatic ring) and its tail (a hydrocarbon chain) strongly correlated and with similar decay times. Dihedral torsion interconversion rates are on a longer time scale than these simulations and no exact conclusion can be drawn about their equilibrium distribution. However, simulations starting with different dihedral angle values yield at the end similar populations, suggesting that a few ns of simulation gives reliable sampling. This work also shows that implicit solvent studies are not required to select drug conformers, and that they are not able to yield the correct angle distributions in the condensed phase.

Inside the membrane, both the drug head and the drug tail tend to stay parallel to the lipid molecules, but perpendicular orientations do occur to make possible the formation of H-

bonds between both the drug tail and the drug side chain with the surrounding lipids and waters. Isotropic motions are observed in the water phase. In the membrane interior, however, orientational times for drug reorientation are of the order of a few ns. Up→down and down→up interconversions do occur during the simulations, but the choice of the starting structure is critical for determining free energy and permeability profiles. It is therefore easier to run simulations from both starting orientations to obtain reliable results, rather than performing one single longer simulation. From these simulations, hydrogen bonding seems to be as important as steric constraints in determining the drug behaviour inside the membrane. This may be due to the fact that the drugs studied have almost identical size and chemical structure, with the difference between them residing in the hydrogen bonding ability of a short side chain. However, the importance of H-bonds is recognised in drug design to be one of the most critical factor affecting drug absorption.[265]

CHAPTER 9

Overall Conclusions and Future Work

9.1 Conclusions

The work outlined in this thesis represents the application of computer simulations to the study of biological membranes and solute permeation. The primary motivation for this study was to further elucidate the mechanism by which drugs cross cell membranes to become available at their biological target of action. This study is thus relevant for rational drug design.

The fully hydrated DPPC membrane simulation described in chapter 4 successfully reproduced most experimental observables. Therefore, it was employed to model first small solute permeation, and then real drug permeation across biological membranes. The first permeation study is described in chapter 6 and the second in chapter 8. In both cases, molecular dynamics simulations offered an atomic detail which is not accessible from experiments.

On the methodological side, a simulation protocol was implemented to obtain equilibrium and dynamic properties involved in the permeation process. The first are described by the partitioning behaviour of the solutes between water and different regions inside the membrane, which yielded the free energy profile for membrane crossing. The second are described by the solute diffusive behaviour at different depths in the membrane interior. A simple procedure for force field parameterization was also developed for simulating real drug molecules. Eventually, free energies and diffusion coefficients yielded permeability coefficients, which agreed favourably

with experiments, or at least correctly reproduced the solute ranking. The usefulness of computer simulations was demonstrated not only by reproducing various observables but also by providing a means to understand directly the system properties.

Several insights are remarkable. These simulations confirmed that membranes do not behave like alkane slabs, primarily because of interfacial constraints that cause lipid packing and reduce solute partitioning. Diffusion appear to be less size-dependent than previously thought, since no experiments can currently perform this measure directly. Permeants tend to cross membranes aligned parallel to the membrane normal, with their polar groups oriented towards the interface, but optmization of H-bonds is as important as steric constraints in determining the permeant behaviour inside the membrane. Hydration can also occur in the middle of the membrane, which is instead usually thought to be a completely hydrophobic environment.

Results for the drug molecules simulated cannot be automatically extended to all drug entities. However, some conclusions are significant. Even large molecules can rotate around their centre of mass and orient themselves to optimize their hydrogen bonding network. Dihedral torsions along long chains can, although slowly, rotate and do show a preferred angle population such that the molecule is more stretched.

9.2 Further Research

A few issues were not completely solved in this study and are open to future research.

From a technical point of view, different simulation protocols and membrane systems could be tested. The electrostatic interactions require a careful choice of various parameters. The ones employed in this study were those described in the literature yielding the best agreement with experiments when simulating pure membranes. Different values may have been required when inserting permeating particles. There still is no consensus regarding the application of a non-zero surface tension in lipid bilayer simulations, and some investigations for the particular system under study may have been appropriate. The membrane model employed in this thesis consists of one specific lipid bilayer, but it would be interesting to examine the effect of different temperatures and different lipid compositions on the permeation process, since experimental findings show these factors significantly affect the pemeation results. Also the presence of cholesterol and proteins is known to have dramatic effects and it would be a useful challenge to perform permeability simulations on such systems.

Regarding the results of this work, it must be said that free energies and permeabilities have

large errors, the main reason being the long time scale of lipid motions with respect to the time scale which can be achieved in atomistic computer simulations. The calculation of free energies is a difficult task and approximations are always involved. Better convergence of data would be a major advance, but the resulting calculations would be too demanding. In this thesis we have already simulated for approximately 0.4 μs and the results from these simulations already agree favourably with the available experimental data. On the other hand, simpler approaches to study membrane permeation could be more attractive. Reduced representations of both lipid and solute molecules could make such computation faster and suitable for routine applications. In this case, the atomic detail would be lost, but permeability coefficients, which represent the ultimate aim of these studies, could be still obtained.

9.3 Concluding Remarks

This thesis has described the ambitious goal of modelling with an atomistic detail a representative biological membrane and the process of small solute and drug permeation across it. The simulations were successful and yielded many pleasing and interesting findings. They are proven to be useful and to allow for a better interpretation of experimental observables, because they give a deep insight into the mechanism by which molecules with different chemical structures partition and diffuse in such complex systems. This study can also help to understand why simpler theoretical approaches do or do not work and to develop some new and more correct models.

APPENDIX A

Topologies and Parameters for Drug Molecules

This appendix reports the force field parameters resulting from the work described in chapter 7 and employed in the drug/lipid simulations. Topologies and parameters are presented in the format used by the program CHARMM, as they are intended to be employed within that program and in conjunction with its force field. Topologies and parameters for lipids are available on the internet (http://www.pharmacy.umaryland.edu/faculty/amackere/research.html).

A.1 Atom Types

This is the list of all the drug atom types employed in the force field.

```
!        ! common backbone:
MASS 101 C6  12.011000 C ! aromatic C
MASS 102 Ce  12.011000 C ! aromatic C bonded to ether Oxygen (Oe)
MASS 103 C1  12.011000 C ! alyphatic C bonded to 3 heavy atoms (no nitrogen) and 1 H
MASS 104 C2  12.011000 C ! alyphatic C bonded to 2 heavy atoms (no nitrogen) and 2 H
MASS 105 C3  12.011000 C ! alyphatic C bonded to 1 heavy atoms (no nitrogen) and 3 H
MASS 106 Oe  15.999400 O ! ether O between aromatic C (Ce) and alyphatic C (C2)
MASS 107 O   15.999400 O ! alcohol O
MASS 108 N   14.007000 N ! amine N
MASS 109 Cn1 12.011000 C ! alyphatic C bonded to 3 heavy atoms (1 is nitrogen) and 1 H
MASS 110 Cn2 12.011000 C ! alyphatic C bonded to 2 heavy atoms (1 is nitrogen) and 2 H
MASS 111 HA   1.008000 H ! alyphatic H bonded to either C1, C2 or C3
MASS 112 Ho   1.008000 H ! alcohol H
MASS 113 Hn   1.008000 H ! H bonded to N in amine
MASS 114 H6   1.008000 H ! aromatic H bonded to either C6 or Cy
MASS 115 Hc   1.008000 H ! alyphatic H bonded to either Cn1 or Cn2
```

```
!        ! alprenolol and Atenolol:
MASS 116 Cc  12.011000 C ! aromatic C bonded to drug side chain (Cs)
MASS 117 Cs  12.011000 C ! alyphatic C bonded to aromatic C (Cc)
MASS 118 Hs   1.008000 H ! alyphatic H bonded to Cs
!        ! alprenolol:
MASS 119 E1  12.011000 C ! C in alkenes, bonded with 1 H and 2 C
MASS 120 Z1   1.008000 H ! H in alkenes, bonded with E1
MASS 121 E2  12.011000 C ! C in alkenes, bonded with 2 H and 1 C
MASS 122 Z2   1.008000 H ! H in alkenes, bonded with E2
!        ! Atenolol:
MASS 123 CO  12.011000 C ! amide carbonyl C
MASS 124 OC  15.999000 O ! amide carbonyl Oxygen
MASS 125 NC  14.007000 N ! amide Nitrogen
MASS 126 HZ   1.008000 H ! amide H bonded to NC
!        ! pindolol:
MASS 127 Cp  12.011000 C ! aromatic C belonging to both rings in indole
MASS 128 Cy  12.011000 C ! aromatic C in five-member ring of indole
MASS 129 Ny  14.007000 N ! Nitrogen in indole ring
MASS 130 Hy   1.008000 H ! H bonded to Ny
```

A.2 Topologies

Topologies contain the following information:

```
RESI resi_name resi_charge
GROUP
ATOM atom_name atom_type atom_charge
BOND atom_1 atom_2
```

Here, **resi_name** is the name of the residue, **resi_charge** the total residue charge, **atom_name** the name of the atom in the residue, **atom_type** the atom type in the force field parameters, **atom_charge** the partial charge on the atom, **atom_1** and **atom_2** two covalently bonded atoms.

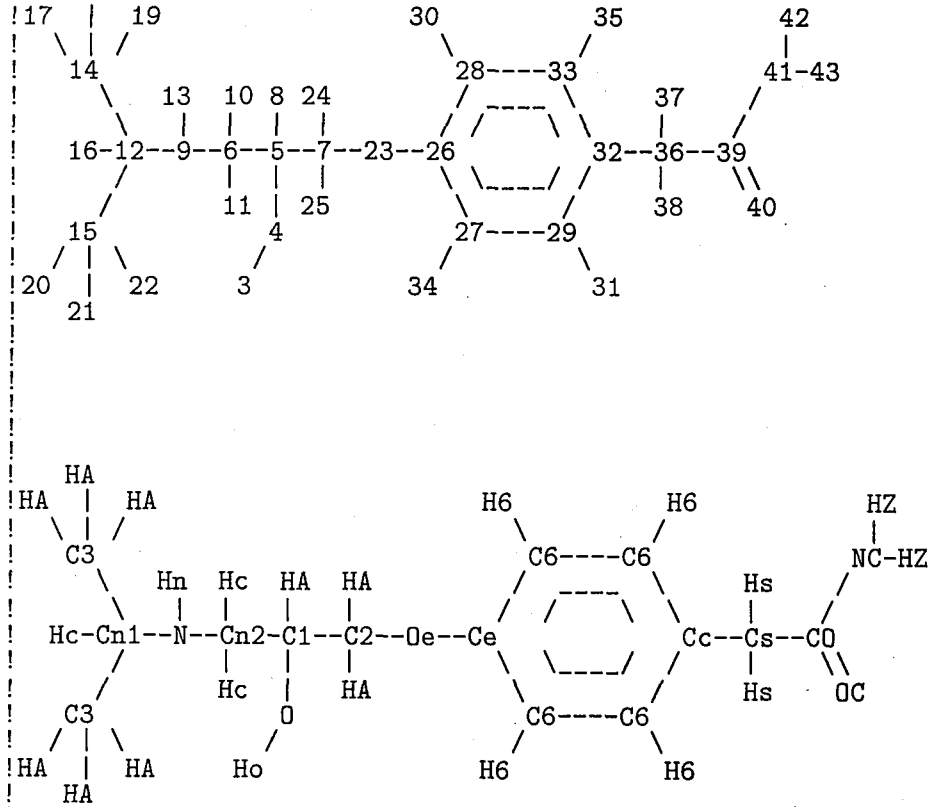
A.2.1 Alprenolol

```
RESI ALPR   0.000       ! alprenolol
GROUP                   !
ATOM C14  C3  -0.27     !    18
ATOM H17  HA   0.09     !17  |  19                30           35
ATOM H18  HA   0.09     !  \ | /                   \           /
ATOM H19  HA   0.09     !   14                      28----33
GROUP                   !     \    13  10 8 24     /  ____  \
ATOM C15  C3  -0.27     !      \   |   |  |  |    /  /    \  \
ATOM H20  HA   0.09     !   16-12--9--6--5--7--23--26        32-34
ATOM H21  HA   0.09     !      /       |  |  |     \  \____/  /
ATOM H22  HA   0.09     !     /        11 | 25      \        /
GROUP                   !   15            4     37   27----29
ATOM C12  Cn1  0.12     !  / | \          /        \  /        \
ATOM H16  Hc   0.09     !20  |  22       3          36          31
ATOM N9   N   -0.82     !   21                     /  \       42
ATOM H13  Hn   0.40     !                        38    \     /
ATOM C6   Cn2  0.03     !                            39==41
ATOM H11  Hc   0.09     !                           /     \
ATOM H10  Hc   0.09     !                         40       43
GROUP                   !
ATOM C5   C1   0.14     !
ATOM O4   O   -0.66     !   HA
ATOM H3   Ho   0.43     !HA  |  HA                      H6          H6
ATOM H8   HA   0.09     !  \ | /                         \        /
GROUP                   !   C3                           C6----C6
ATOM C7   C2   0.02     !     \    Hn  Hc  HA  HA        /  ____  \
ATOM H25  HA   0.09     !      \   |   |   |   |       /  /    \  \
ATOM H24  HA   0.09     !  Hc-Cn1--N--Cn2-C1--C2--Oe--Ce        C6-H6
ATOM C26  Ce   0.085    !      /       |   |   |       \  \____/  /
ATOM O23  Oe  -0.285    !     /        Hc  |   HA       \        /
GROUP                   !   C3             O         Hs    Cc----C6
ATOM C28  C6  -0.115    !  / | \           /            \  /        \
ATOM H30  H6   0.115    !HA  |  HA      Ho               Cs          H6
GROUP                   !   HA                        /  \      Z2
ATOM C29  C6  -0.115    !                           Hs    \    /
ATOM H31  H6   0.115    !                                  E1==E2
GROUP                   !                                 /    \
ATOM C32  C6  -0.115    !                               Z1      Z2
ATOM H34  H6   0.115    !
GROUP                   !
ATOM C33  C6  -0.115    !
ATOM H35  H6   0.115    !
!side chain:
GROUP
ATOM C27  Cc   0.000
ATOM C36  Cs  -0.18
ATOM H37  Hs   0.09
ATOM H38  Hs   0.09
GROUP
ATOM C39  E1  -0.15
ATOM H40  Z1   0.15
GROUP
ATOM C41  E2  -0.42
ATOM H42  Z2   0.21
ATOM H43  Z2   0.21
! connectivities:
BOND C33 C32
BOND C33 C28
BOND C33 H35
BOND C32 C29
BOND C32 H34
BOND C29 C27
BOND C29 H31
BOND C27 C26
BOND C27 C36
BOND C26 C28
BOND C28 H30
BOND C26 O23
BOND O23 C7
BOND C7  C5
BOND C7  H24
BOND C7  H25
BOND C5  O4
```

```
BOND C5  C6
BOND C5  H8
BOND O4  H3
BOND C6  N9
BOND C6  H10
BOND C6  H11
BOND N9  C12
BOND N9  H13
BOND C12 C14
BOND C12 C15
BOND C12 H16
BOND C14 H17
BOND C14 H18
BOND C14 H19
BOND C15 H20
BOND C15 H21
BOND C15 H22
bond C36 H37
bond C36 H38
bond C36 C39
bond C39 H40
bond C39 C41
bond C41 H42
bond C41 H43
```

A.2.2 Atenolol

```
RESI ATEN    0.000
GROUP
ATOM C14   C3   -0.27
ATOM H17   HA    0.09
ATOM H18   HA    0.09
ATOM H19   HA    0.09
GROUP
ATOM C15   C3   -0.27
ATOM H20   HA    0.09
ATOM H21   HA    0.09
ATOM H22   HA    0.09
GROUP
ATOM C12   Cn1   0.12
ATOM H16   Hc    0.09
ATOM N9    N    -0.82
ATOM H13   Hn    0.40
ATOM C6    Cn2   0.03
ATOM H11   Hc    0.09
ATOM H10   Hc    0.09
GROUP
ATOM C5    C1    0.14
ATOM O4    O    -0.66
ATOM H3    Ho    0.43
ATOM H8    HA    0.09
GROUP
ATOM C7    C2    0.02
ATOM H25   HA    0.09
ATOM H24   HA    0.09
ATOM C26   Ce    0.085
ATOM O23   Oe   -0.285
GROUP
ATOM C28   C6   -0.115
ATOM H30   H6    0.115
GROUP
ATOM C29   C6   -0.115
ATOM H31   H6    0.115
GROUP
ATOM C27   C6   -0.115
ATOM H34   H6    0.115
GROUP
ATOM C33   C6   -0.115
ATOM H35   H6    0.115
!side chain:
GROUP
ATOM C32   Cc    0.00
ATOM C36   Cs   -0.18
ATOM H37   Hs    0.09
ATOM H38   Hs    0.09
GROUP
ATOM C39   CO    0.55
ATOM O40   OC   -0.55
GROUP
ATOM N41   NC   -0.64
ATOM H42   HZ    0.32
ATOM H43   HZ    0.32
! connectivities:
BOND C33 C32
BOND C33 C28
BOND C33 H35
BOND C32 C29
BOND C32 C36 !!!
BOND C29 C27
BOND C29 H31
BOND C27 C26
BOND C27 H34 !!!
BOND C26 C28
BOND C28 H30 !!!
BOND C26 O23
BOND O23 C7
BOND C7  C5
BOND C7  H24
BOND C7  H25
BOND C5  O4
BOND C5  C6
BOND C5  H8
```



```
BOND O4  H3
BOND C6  N9
BOND C6  H10
BOND C6  H11
BOND N9  C12
BOND N9  H13
BOND C12 C14
BOND C12 C15
BOND C12 H16
BOND C14 H17
BOND C14 H18
BOND C14 H19
BOND C15 H20
BOND C15 H21
BOND C15 H22
bond C36 H37   C36 H38   C36 C39
bond C39 O40   C39 N41   N41 H42   N41 H43
```

A.2.3 Pindolol

```
RESI PIND   0.000       ! pindolol
GROUP                   !
ATOM C14  C3  -0.27     !    18
ATOM H17  HA   0.09     !17  |  19                36           35
ATOM H18  HA   0.09     !  \ | /                    \          /
ATOM H19  HA   0.09     !   14                       28----33
GROUP                   !     \    13  10 8 24      /  ____  \
ATOM C15  C3  -0.27     !      \   |   |  |  |     /  /    \  \
ATOM H20  HA   0.09     !   16-12--9--6--5--7--23--26        32-34
ATOM H21  HA   0.09     !      /       |  |  |     \  \____/  /
ATOM H22  HA   0.09     !     /        11 | 25      \        /
GROUP                   !   15             4        27----29
ATOM C12  Cn1  0.12     !  / | \          /          |      |
ATOM H16  Hc   0.09     !20  |  22      3           40-30    31-38
ATOM N9   N   -0.82     !   21                        \\  /
ATOM H13  Hn   0.40     !                              37
ATOM C6   Cn2  0.03     !                              |
ATOM H11  Hc   0.09     !                              39
ATOM H10  Hc   0.09     !
GROUP                   !
ATOM C5   C1   0.14     !
ATOM O4   O   -0.66     !   HA
ATOM H3   Ho   0.43     !HA  |  HA                    H6          H6
ATOM H8   HA   0.09     !  \ | /                       \          /
GROUP                   !   C3                         C6----C6
ATOM C7   C2   0.02     !     \    Hn  Hc  HA  HA       /  ____  \
ATOM H25  HA   0.09     !      \   |   |   |   |      /  /    \  \
ATOM H24  HA   0.09     !  Hc-Cn1--N--Cn2-C1--C2--Oe--Ce          C6-H6
ATOM C26  Ce   0.085    !     /       |   |   |       \  \____/  /
ATOM O23  Oe  -0.285    !    /        Hc  |   HA       \        /
!side chain:            !   C3             O           Cp----Cp
GROUP                   !  / | \          /            |      |
ATOM C28  C6  -0.115    !HA  |  HA      Ho             H6-Cy    Ny-Hy
ATOM H36  H6   0.115    !   HA                           \\  /
GROUP                   !                                 C6
ATOM C33  C6  -0.115    !                                 |
ATOM H35  H6   0.115    !                                 H6
GROUP                   !
ATOM C32  C6  -0.115    !
ATOM H34  H6   0.115    !
GROUP                   !
ATOM C27  Cp  -0.02     !
ATOM C29  Cp   0.13     !
ATOM C30  Cy  -0.145    !
ATOM N31  Ny  -0.61     !
ATOM C37  C6   0.035    !
ATOM H40  H6   0.115    !
ATOM H39  H6   0.115    !
ATOM H38  Hy   0.38     !
! connectivities:
BOND C33 C32
BOND C33 C28
BOND C33 H35
BOND C32 C29
BOND C32 H34
BOND C29 C27
BOND C29 N31
BOND C27 C26
BOND C27 C30
BOND C26 C28
BOND C28 H36
BOND C26 O23
BOND O23 C7
BOND C7  C5
BOND C7  H24
BOND C7  H25
BOND C5  O4
BOND C5  C6
BOND C5  H8
BOND O4  H3
BOND C6  N9
BOND C6  H10
```

```
BOND C6  H11
BOND N9  C12
BOND N9  H13
BOND C12 C14
BOND C12 C15
BOND C12 H16
BOND C14 H17
BOND C14 H18
BOND C14 H19
BOND C15 H20
BOND C15 H21
BOND C15 H22
bond C30 H40    C30 C37    C37 H39    C37 N31    N31 H38
```

A.3 Parameters

Parameters are the bond stretching, angle bending, dihedral torsions and Lennard-Jones parameters employed for the calculation of the potential energy.

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!                                        ! common backbone:
C6  C6   305.       1.3750               !
H6  C6   340.       1.0800               !
C6  Ce   305.       1.3750               !
Ce  Oe   334.3      1.4110               !
C2  Oe   428.       1.4200               !
C2  C1   222.5      1.5380               !
HA  C2   309.       1.1110               !
C1  O    428.       1.4200               !
Ho  O    545.       0.9600               !
C1  HA   309.       1.1110               !
C1  Cn2  222.5      1.5380               !
Cn2 N    255.       1.463                !
Cn1 N    255.       1.463                !
N   Hn   447.8      1.019                !
Cn1 Hc   313.8      1.098                !
Cn2 Hc   313.8      1.098                !
Cn1 C3   222.5      1.5380               !
C3  HA   322.       1.1110               !
!                                        ! alprenolol & Atenol:
Cs  Cc   230.00     1.4900               !
Cs  Hs   309.       1.1110               !
Cc  C6   305.       1.3750               !
!                                        ! alprenolol:
Cs  E1   365.00     1.502                !
E1  Z1   360.50     1.100                !
E1  E2   500.00     1.342                !
E2  Z2   365.00     1.100                !
Cc  Ce   305.       1.3750               !
!                                        ! Atenolol:
Cs  CO   200.00     1.5220               !
CO  OC   650.00     1.2300               !
CO  NC   430.00     1.3600               !
NC  HZ   480.00     1.0000               !
!                                        ! pindolol:
Cp  Ce   305.00     1.3680               !
Cp  C6   305.00     1.3680               !
Cp  Cp   360.00     1.4000               !
Cp  Cy   350.00     1.4400               !
Cp  Ny   270.00     1.3750               !
Cy  H6   350.00     1.0800               !
Cy  C6   350.00     1.3650               !
Ny  Hy   465.00     0.9760               !
Ny  C6   270.00     1.3700               !
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
!atom types     Ktheta   Theta0     Kub         S0
!                                                  ! common backbone:
```

```
C6  C6  C6      40.000   120.00     35.00      2.4162 !
H6  C6  C6      30.000   120.00     22.00      2.1525 !
C6  C6  Ce      40.000   120.00     35.00      2.4162 !
C6  Ce  C6      40.000   120.00     35.00      2.4162 !
C6  Ce  Oe      45.200   120.00                       !
H6  C6  Ce      30.000   120.00     22.00      2.1525 !
C2  Oe  Ce      61.000   108.00                       !
C1  C2  Oe      75.700   110.10                       !
HA  C2  Oe      45.900   108.89                       !
HA  C1  C2      34.500   110.10     22.53      2.179  !
HA  C3  HA      35.500   108.40      5.40      1.802  !
C2  C1  O       75.700   110.10                       !
Cn2 C1  C2      58.350   113.50     11.16      2.561  !
HA  C1  O       45.900   108.89                       !
Ho  O   C1      57.500   106.00                       !
C3  Cn1 N       67.700   110.00                       !
Hc  Cn1 N       30.500   109.70     50.000     2.140  !
Hc  Cn2 N       30.500   109.70     50.000     2.140  !
Cn1 N   Cn2     40.500   112.20      5.000     2.422  !
Cn1 N   Hn      42.100   108.90      5.000     2.029  !
Cn2 N   Hn      42.100   108.90      5.000     2.029  !
HA  C3  Cn1     33.430   110.10     22.53      2.179  !
Hc  Cn1 C3      34.500   110.10     22.53      2.179  !
N   Cn2 C1      67.700   110.00                       !
Hc  Cn2 C1      33.430   110.10     22.53      2.179  !
Cn2 C1  HA      34.500   110.10     22.53      2.179  !
Cn2 C1  O       75.700   110.10                       !
Hc  Cn2 Hc      35.800   109.00      5.400     1.802  !
HA  C2  HA      35.500   109.00      5.40      1.802  !
HA  C2  C1      33.430   110.10     22.53      2.179  !
C3  Cn1 C3      53.350   114.00      8.00      2.561  !
!                                                    ! alprenolol & Atenolol:
Hs  Cs  Cc      49.300   107.500                      !
Hs  Cs  Hs      35.500   109.00      5.40      1.802  !
Cs  Cc  C6      45.800   122.300                      !
Cc  C6  C6      40.000   120.00     35.00      2.4162 !
Cc  C6  H6      30.000   120.00     22.00      2.1525 !
!                                                    ! alprenolol:
Z1  E1  E2      42.00    118.00                       !
Z2  E2  E1      45.00    120.50                       !
Z2  E2  Z2      19.00    119.00                       !
Hs  Cs  E1      45.00    111.50                       !
E2  E1  Cs      48.00    126.00                       !
Z1  E1  Cs      40.00    116.00                       !
E1  Cs  Cc      51.800   107.500                      !
Cs  Cc  Ce      45.800   122.300                      !
Cc  Ce  C6      40.000   120.00     35.00      2.4162 !
Cc  Ce  Oe      45.200   120.00                       !
C6  Cc  Ce      40.000   120.00     35.00      2.4162 !
!                                                    ! Atenolol:
C6  Cc  C6      40.000   120.00     35.00      2.4162 !
CO  Cs  Cc      51.800   107.500                      !
Hs  Cs  CO      33.000   109.50     30.00      2.1630 !
OC  CO  Cs      15.000   121.00     50.00      2.4400 !
NC  CO  Cs      50.000   116.50     50.00      2.4500 !
OC  CO  NC      75.000   122.50     50.00      2.3700 !
HZ  NC  CO      50.000   120.00                       !
HZ  NC  HZ      23.000   120.00                       !
!                                                    ! pindolol:
Ce  Cp  Cy     160.000   130.60                       !
C6  Ce  Cp      60.000   118.00                       !
Cp  Ce  Oe      45.200   122.00                       !
Cp  Cp  C6      60.000   122.00                       !
Cp  Cp  Ny     110.000   107.40                       !
C6  Cp  Ny     160.000   130.60                       !
Cp  Cp  Cy     110.000   107.40                       !
C6  Cp  Cy     160.000   130.60                       !
Cp  Cp  Ce      60.000   122.00                       !
Cp  Ny  C6     110.000   108.00                       !
Cp  Cy  C6     120.000   107.40     25.00      2.261  !
Cp  Cy  H6      32.000   126.40     25.00      2.255  !
H6  Cy  C6      32.000   126.40     25.00      2.186  !
H6  C6  Cy      32.000   125.00     25.00      2.173  !
H6  C6  Cp      30.000   122.00     22.00      2.146  !
Hy  Ny  Cp      28.000   126.00                       !
```

```
Hy  Ny  C6      28.000   126.00                          !
H6  C6  Ny      32.000   125.00     25.00      2.177     !
Cy  C6  Ny     120.000   110.00     25.00      2.240     !
Cp  C6  C6      60.000   118.00                          !
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types     Kchi     n      delta
!
!                                        ! common backbone:
HA  C1  O   Ho  0.140    3      0.000    !
HA  C3  Cn1 Hc  0.200    3      0.000    !
HA  C3  Cn1 C3  0.200    3      0.000    !
HA  C3  Cn1 N   0.200    3      0.000    !
Ho  O   C1  C2  1.300    1      0.000    !
Ho  O   C1  C2  0.300    2      0.000    !
Ho  O   C1  C2  0.420    3      0.000    !
Ho  O   C1  Cn2 1.300    1      0.000    !
Ho  O   C1  Cn2 0.300    2      0.000    !
Ho  O   C1  Cn2 0.420    3      0.000    !
C3  Cn1 N   Hn  0.1000   3      0.000    !
C3  Cn1 N   Cn2 0.1000   3      0.000    !
Hc  Cn1 N   Hn  0.4200   3      0.000    !
Hc  Cn2 N   Hn  0.4200   3      0.000    !
Hc  Cn1 N   Cn2 0.4200   3      0.000    !
Hc  Cn2 N   Cn1 0.4200   3      0.000    !
Hc  Cn2 C1  HA  0.2000   3      0.000    !
N   Cn2 C1  HA  0.2000   3      0.000    !
Hc  Cn2 C1  C2  0.2000   3      0.000    !
N   Cn2 C1  C2  0.2000   3      0.000    !
Hc  Cn2 C1  O   0.2000   3      0.000    !
N   Cn2 C1  O   0.2000   3      0.000    !
HA  C1  C2  HA  0.2000   3      0.000    !
Cn2 C1  C2  HA  0.2000   3      0.000    !
O   C1  C2  HA  0.2000   3      0.000    !
HA  C1  C2  Oe  0.2000   3      0.000    !
Cn2 C1  C2  Oe  0.2000   3      0.000    !
O   C1  C2  Oe  0.2000   3      0.000    !
C1  Cn2 N   Hn  0.1000   3      0.000    !
C1  Cn2 N   Cn1 0.1000   3      0.000    !
C1  C2  Oe  Ce  0.820    1      0.000    !
C1  C2  Oe  Ce  0.250    3      0.000    !
HA  C2  Oe  Ce  0.270    3      0.000    !
C2  Oe  Ce  C6  1.500    2    180.000    !
C6  C6  C6  C6  3.1000   2    180.000    !
C6  C6  C6  Ce  3.1000   2    180.000    !
C6  C6  Ce  C6  3.1000   2    180.000    !
H6  C6  C6  C6  4.2000   2    180.000    !
H6  C6  Ce  C6  4.2000   2    180.000    !
H6  C6  C6  Ce  4.2000   2    180.000    !
H6  C6  C6  H6  2.4000   2    180.000    !
Oe  Ce  C6  C6  3.1000   2    180.000    !
Oe  Ce  C6  H6  4.2000   2    180.000    !
!                                        ! alprenolol and Atenolol:
Cc  C6  C6  H6  4.2000   2    180.000    !
Cs  Cc  C6  C6  3.1000   2    180.000    !
Cs  Cc  C6  H6  4.2000   2    180.000    !
Hs  Cs  Cc  C6  0.0000   6      0.000    !
!                                        ! alprenolol:
C6  C6  CC  CE  3.1000   2    180.000    !
C6  CC  CE  C6  3.1000   2    180.000    !
C6  CC  CE  OE  3.1000   2    180.000    !
CE  CC  C6  H6  4.2000   2    180.000    !
CC  CE  OE  C2  1.500    2    180.000    !
CC  CE  C6  H6  4.2000   2    180.000    !
C6  C6  C6  Cc  3.1000   2    180.000    !
C6  C6  Ce  Cc  3.1000   2    180.000    !
Cs  Cc  Ce  Oe  1.0000   2    180.000    !
Cs  Cc  Ce  C6  3.1000   2    180.000    !
Hs  Cs  Cc  Ce  0.0000   6      0.000    !
E1  Cs  Cc  Ce  0.2300   2    180.000    !
```

```
E1  Cs  Cc  C6  0.2300   2    180.000  !
E2  E1  Cs  Hs  0.1200   3      0.000  !
E2  E1  Cs  Cc  0.4000   3      0.000  !
Z1  E1  Cs  Hs  0.8700   3      0.000  !
Z1  E1  Cs  Cc  0.1200   3      0.000  !
Cs  E1  E2  Z2  5.2000   2    180.000  !
Z1  E1  E2  Z2  5.2000   2    180.000  !
!                                      ! Atenolol:
C6  Cc  C6  C6  3.1000   2    180.000  !
C6  Cc  C6  H6  4.2000   2    180.000  !
Cc  C6  C6  Ce  3.1000   2    180.000  !
OC  CO  Cs  Cc  0.0500   6    180.000  !
NC  CO  Cs  Cc  0.0500   6    180.000  !
OC  CO  NC  HZ  1.4000   2    180.000  !
OC  CO  Cs  Hs  0.0000   3    180.000  !
CO  Cs  Cc  C6  0.2300   2    180.000  !
NC  CO  Cs  Hs  0.0000   3    180.000  !
HZ  NC  CO  Cs  1.4000   2    180.000  !
!                                      ! pindolol:
Cp  Ce  Oe  C2  1.500    2    180.000  !
Cp  C6  C6  C6  3.1000   2    180.00   !
Cp  C6  C6  H6  3.0000   2    180.00   !
Cp  Ce  C6  C6  3.1000   2    180.00   !
Cp  Ce  C6  H6  3.0000   2    180.00   !
Cp  Cp  C6  C6  3.1000   2    180.00   !
Cp  Cp  C6  H6  3.0000   2    180.00   !
Cp  Cp  Ny  Hy  0.8000   2    180.00   !
Cp  Cp  Ny  C6  5.0000   2    180.00   !
Cp  Cp  Ce  Oe  3.1000   2    180.000  !
Cp  Cp  Ce  C6  3.1000   2    180.00   !
Cp  Cp  Cy  H6  2.8000   2    180.00   !
Cp  Cp  Cy  C6  4.0000   2    180.00   !
Cp  Cy  C6  H6  2.0000   2    180.00   !
Cp  Cy  C6  Ny  4.0000   2    180.00   !
Cp  Ny  C6  H6  2.0000   2    180.00   !
Cp  Ny  C6  Cy  5.0000   2    180.00   !
Ce  Cp  Cp  Ny 10.0000   2    180.00   !
C6  Cp  Cp  Cy 10.0000   2    180.00   !
C6  Cp  Cp  Ce  3.1000   2    180.00   !
Ce  Cp  Cy  H6  2.8000   2    180.00   !
Ce  Cp  Cy  C6  3.0000   2    180.00   !
C6  Cp  Ny  Hy  0.8000   2    180.00   !
H6  Cy  C6  H6  1.0000   2    180.00   !
H6  C6  Ny  Hy  0.4000   2    180.00   !
Cy  Cp  Ce  Oe  3.1000   2    180.000  !
H6  C6  Cp  Cy  3.0000   2    180.00   !
Hy  Ny  C6  Cy  0.8000   2    180.00   !
Ny  C6  Cy  H6  3.5000   2    180.00   !
Ny  Cp  C6  C6  2.8000   2    180.00   !
Ny  Cp  C6  H6  3.0000   2    180.00   !
Ny  Cp  Cp  Cy  5.0000   2    180.00   !
C6  Ny  Cp  C6  3.0000   2    180.00   !
Cy  Cp  Ce  C6  3.0000   2    180.00   !
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
NONBONDED
!
!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom       epsilon  Rmin/2       epsilon(1,4)  R/2(1,4)
!                                                         ! common backbone:
C6    0.0  -0.0700  1.9924                                !
Ce    0.0  -0.0700  1.9924                                !
H6    0.0  -0.0300  1.3582                                !
C1    0.0  -0.0200  2.2750   0.0    -0.010      1.900     !
C2    0.0  -0.0550  2.1750   0.0    -0.010      1.900     !
C3    0.0  -0.0800  2.0600   0.0    -0.010      1.900     !
```

```
Oe      0.0   -0.1521   1.7700                                        !
O       0.0   -0.1521   1.7700                                        !
N       0.0   -0.0450   2.0000                                        !
Cn1     0.0   -0.0780   1.9800                                        !
Cn2     0.0   -0.0780   1.9800                                        !
HA      0.0   -0.0220   1.3200                                        !
Ho      0.0   -0.0460   0.2245                                        !
Hn      0.0   -0.0090   0.8750                                        !
Hc      0.0   -0.0280   1.2800                                        !
!                                                                    ! alprenolol and Atenolol:
Cs      0.0   -0.0550   2.1750   0.0     -0.010      1.900           !
Hs      0.0   -0.0220   1.3200                                        !
Cc      0.0   -0.0700   1.9924                                        !
!                                                                    ! alprenolol:
E1      0.0   -0.0680   2.090                                         !
Z1      0.0   -0.0310   1.25                                          !
E2      0.0   -0.0640   2.08                                          !
Z2      0.0   -0.0260   1.26                                          !
!                                                                    ! Atenolol:
CO      0.0   -0.0700   2.000                                         !
OC      0.0   -0.1200   1.7000   0.0     -0.120      1.400           !
NC      0.0   -0.2000   1.85                                          !
HZ      0.0   -0.0460   0.2245                                        !
!                                                                    ! pindolol:
Cp      0.0   -0.0900   1.8000   0.0     -0.090      1.900           !
Cy      0.0   -0.0700   1.9924                                        !
Ny      0.0   -0.2000   1.8500                                        !
Hy      0.0   -0.0460   0.2245                                        !
```

References

[1] T. Zhu, J. Li, G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Chem. Phys.*, **109**, 9117, (1998).

[2] A. Walter and J. Gutknecht, *J. Membr. Biol.*, **90**, 207, (1986).

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

[4] A. Seelig and J. Seelig, *Biochemistry*, **13**, 4839, (1974).

[5] A. Seelig and J. Seelig, *Biochimica et Biophysica Acta*, **406**, 1, (1975).

[6] J. P. Douliez, A. Leonard and E. J. Dufourc, *Biophys. J.*, **68**, 1727, (1995).

[7] S. Balaz, *Perspectives in Drug Discovery and Design*, **19**, 157, (2000).

[8] C. Hansch, *Exploring QSAR*, (1995).

[9] H. Kubinyi, *Drug Discovery Today*, **2**, 538, (1997).

[10] K. Palm, P. Stenberg, K. Luthman and P. Artursson, *Pharmac. Res.*, **14**, 568, (1997).

[11] M. P. Allen and D. J. Tildesley, *Computer simulations of liquids*, Oxford University Press, Oxford, (1987).

[12] A. R. Leach, *Molecular modelling: principles and applications.*, Pearson Education Limited, Harlow, (1996).

[13] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comp. Chem.*, **4**, 187, (1983).

[14] R. W. Hockney, *Methods Comput. Phys.*, **9**, 136, (1970).

[15] S. Nose and M. L. Klein, *Mol. Phys.*, **50**, 1055, (1983).

[16] S. Nose, *Mol. Phys.*, **52**, 255, (1984).

[17] W. G. Hoover, *Phys. Rev. A*, **31**, 1695, (1985).

[18] H. C. Andersen, *J. Chem. Phys.*, **72**, 2384, (1980).

[19] S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, **103**, 4613, (1995).

[20] J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 327, (1977).

[21] P. Ewald, *Ann. Phys. (Leipzig)*, **64**, 253, (1921).

[22] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, **98**, 10089, (1993).

[23] L. Verlet, *Phys. Rev.*, **159**, 98, (1967).

[24] Y. Zhang, S. E. Feller, B. R. Brooks and R. W. Pastor, *J. Chem. Phys.*, **103**, 10252, (1995).

[25] S. E. Feller, Y. Zhang and R. W. Pastor, *J. Chem. Phys.*, **103**, 10267, (1995).

[26] S. Singer and G. L. Nicolson, *Science*, **172**, 720, (1972).

[27] L. Stryer, *Biochemistry*, W.H.Freeman and Company, New York, (1995).

[28] M. S. P. Sansom, *Current opinion in structure biology*, **8**, 237, (1998).

[29] V. Racansky, D. Valachovic and P. Balgavy, *Acta Physica Slovaca*, **37**, 166, (1987).

[30] G. Cevc and D. Marsh, *Phospholipid bilayers. Physical principles and models*, John Wiley and Sons, New York, (1987).

[31] T. J. McIntosh, *Current Topics in Membranes*, **48**, 23, (1999).

[32] J. F. Nagle and S. Tristram-Nagle, *Biochimica Et Biophysica Acta - Reviews On Biomembranes*, **1469**, 159, (2000).

[33] B. A. Lewis and D. M. Engelman, *J. Mol. Biol.*, **166**, 211, (1983).

[34] H. G. Buldt, H. U. Gally, J. Seelig and G. Zaccai, *J. Mol. Biol.*, **134**, 673, (1979).

[35] G. Zaccai, A. Buldt, G. Seelig and J. Seelig, *J. Mol. Biol.*, **134**, 693, (1979).

[36] G. Buldt, H. U. Gally, A. Seelig, J. Seelig and G. Zaccai, *Nature*, **271**, 182, (1978).

[37] J. F. Nagle and S. Tristram-Nagle, *Curr. Op. Struct. Biol.*, **10**, 474, (2000).

[38] J. Seelig, *Quarterly Reviews of Biophysics*, **10**, 353, (1977).

[39] J. Seelig and A. Seelig, *Quarterly Reviews of Biophysics*, **13**, 19, (1980).

[40] E. Oldfield, M. Meadows, D. Rice and R. Jacobs, *Biochemistry*, **17**, 2727, (1978).

[41] P. Meier, E. Ohmes, G. Kothe, A. Blume and H. J. Weldner, J. Elbl, *J. Phys.Chem.*, **87**, 4904, (1983).

[42] M. Brown, J. Seelig and U. Haberlein, *J. Chem. Phys.*, **70**, 5045, (1979).

[43] P. Meier, E. Ohmes and G. Kothe, *J. Chem. Phys.*, **85**, 3598, (1986).

[44] J. Seelig and W. Niederberger, *J. Am. Chem. Soc.*, **96**, 2069, (1974).

[45] J. Seelig, H. U. Gally and R. Wohlgemuth, *Biochimica et Biophysica Acta*, **467**, 109, (1977).

[46] E. J. Dufourc, C. Mayer, J. Stohrer, G. Althoff and G. Kothe, *Biophys. J.*, **63**, 1081, (1992).

[47] F. Borle and J. Seelig, *Biochimica et Biophysica Acta*, **735**, 131, (1983).

[48] H. Akutsu, *Biochemistry*, **20**, 7359, (1981).

[49] A. Blume, W. Hubner and G. Messner, *Biochemistry*, **27**, 8239, (1988).

[50] R. Mendelsohn, M. A. Davies, J. W. Brauner, H. F. Schuster and R. A. Dluhy, *Biochemistry*, **28**, 8934, (1989).

[51] R. Mendelsohn, M. A. Davies, H. F. Schuster, Z. Xu and R. Bittman, *Biochemistry*, **30**, 8558, (1991).

[52] H. L. Casal and R. N. McElhaney, *Biochemistry*, **29**, 5423, (1990).

[53] E. S. Wu, K. Jacobson and D. Papahadjopoulos, *Biochemistry*, **16**, 3936, (1977).

[54] P. F. Fahey and W. W. Webb, *Biochemistry*, **17**, 3046, (1978).

[55] S. Konig, W. Pfeiffer, T. Bayerl, D. Richter and E. Sackmann, *J. Phys. II France*, **2**, 1589, (1992).

[56] J. R. Nagle, R. Zhang, S. Tristram-Nagle, W. Sun, H. Petrache and R. M. Suter, *Biophys. J.*, **70**, 1419, (1996).

[57] R. P. Rand and V. A. Parsegian, *Biochimica et Biophysica Acta*, **988**, 351, (1989).

[58] S. Marcelya, *Biochimica Et Biophysica Acta*, **367**, 165, (1974).

[59] R. W. Pator, R. M. Venable and M. Karplus, *Proc. Natl. Acad. Sci. USA*, **88**, 892, (1991).

[60] E. Egberts, *Molecular dynamics simulations of multibilayer membranes*, Ph.D. thesis, University of Groningen, (1988).

[61] D. P. Tieleman, S. J. Marrink and H. J. C. Berendsen, *Biochimica Et Biophysica Acta-Reviews On Biomembranes*, **1331**, 235, (1997).

[62] L. R. Forrest and M. S. P. Sansom, *Curr. Op. Struct. Biol.*, **10**, 174, (2000).

[63] D. P. Tieleman, P. C. Biggin, G. R. Smith and M. S. P. Sansom, *Quart. Rev. Biophys.*, **34**, 473, (2001).

[64] M. B. Ulmschneider and M. S. P. Sansom, *Biochimica et Biophysica Acta*, **1512**, 1, (2001).

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

[66] J. J. L. Cascales, J. G. de la Torre, S. J. Marrink and H. J. C. Berendsen, *J. Chem. Phys.*, **104**, 2713, (1996).

[67] J. J. L. Cascales, H. J. C. Berendsen and J. G. Delatorre, *J. Phys. Chem.*, **100**, 8621, (1996).

[68] J. J. L. Cascales and J. G. de la Torre, *Biochimica et Biophysica Acta*, **1330**, 145, (1997).

[69] T. Husslein, D. M. Newns, P. C. Pattnaik, Q. F. Zhong, P. B. Moore and M. L. Klein, *J. Chem. Phys.*, **109**, 2826, (1998).

[70] S. W. Chiu, E. Jakobsson, S. Subramaniam and H. L. Scott, *Biophys. J.*, **77**, 2462, (1999).

[71] S. E. Feller, D. Yin, R. W. Pastor and A. D. MacKerell, *Biophys. J.*, **73**, 2269, (1997).

[72] S. E. Feller, K. Gawrisch and A. D. MacKerell, *J. Am. Chem. Soc.*, **124**, 318, (2002).

[73] T. Rog, K. Murzyn and M. Pasenkiewicz-Gierula, *Chem. Phys. Lett.*, **352**, 323, (2002).

[74] L. Saiz and M. L. Klein, *J. Chem. Phys.*, **116**, 3052, (2002).

[75] R. J. Mashl, H. L. Scott, S. Subramaniam and E. Jakobsson, *Biophys. J.*, **81**, 3005, (2001).

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

[77] U. Essmann, L. Perera and M. L. Berkowitz, *Langmuir*, **11**, 4519, (1995).

[78] K. Tu, D. J. Tobias and M. L. Klein, *Biophys. J.*, **70**, 595, (1996).

[79] R. M. Venable, B. R. Brooks and R. W. Pastor, *J. Chem. Phys.*, **112**, 4822, (2000).

[80] J. W. Essex, M. M. Hann and W. G. Richards, *Phil. Trans. R. Soc. Lond. B*, **344**, 239, (1994).

[81] A. J. Robinson, W. G. Richards, P. J. Thomas and M. M. Hann, *Biophys. J.*, **67**, 2345, (1994).

[82] M. Pasenkiewiczgierula, Y. Takaoka, H. Miyagawa, K. Kitamura and A. Kusumi, *J. Phys. Chem. A*, **101**, 3677, (1997).

[83] A. M. Smondyrev and M. L. Berkowitz, *J. Comput. Chem.*, **20**, 531, (1999).

[84] E. Lindahl and O. Edholm, *J. Chem. Phys.*, **113**, 3882, (2000).

[85] P. B. Moore, C. F. Lopez and M. L. Klein, *Biophys. J.*, **81**, 2484, (2001).

[86] E. Egberts, S. J. Marrink and H. J. C. Berendsen, *Eur. Biophys. J.*, **22**, 423, (1994).

[87] S. W. Chiu, M. Clark, V. Balaji, S. Subramaniam, H. L. Scott and E. Jakobsson, *Biophys. J.*, **69**, 1230, (1995).

[88] D. P. Tieleman and H. J. C. Berendsen, *J. Chem. Phys.*, **105**, 4871, (1996).

[89] O. Berger, O. Edholm and F. Jahnig, *Biophys. J.*, **72**, 2002, (1997).

[90] S. W. Chiu, M. M. Clark, E. Jakobsson, S. Subramaniam and H. L. Scott, *J. Comput. Chem.*, **20**, 1153, (1999).

[91] S. E. Feller, R. M. Venable and R. W. Pastor, *Langmuir*, **13**, 6555, (1997).

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

[93] S. E. Feller, D. Huster and K. Gawrish, *J. Am. Chem. Soc.*, **121**, 8963, (1999).

[94] R. W. Pastor, R. M. Venable and S. E. Feller, *Acc. Chem. Res.*, **35**, 438, (2002).

[95] T. R. Stouch, H. E. Alper and D. Bassolino-Klimas, *Supercomputer Applications and High Performance Computing*, **8**, 6, (1994).

[96] K. Tu, D. J. Tobias and M. L. Klein, *Biophys. J.*, **69**, 2558, (1995).

[97] S. E. Feller and R. W. Pastor, *Biophys. J.*, **71**, 1350, (1996).

[98] S. W. Chiu, E. Jakobsson and H. L. Scott, *J. Chem. Phys.*, **114**, 5435, (2001).

[99] D. J. Tobias, *Curr. Op. Struct. Biol.*, **11**, 253, (2001).

[100] A. D. MacKerell and S. E. Feller, *J. Phys. Chem. B*, **104**, 7510, (2000).

[101] H. I. Petrache, S. E. Feller and J. F. Nagle, *Biophys. J.*, **72**, 2237, (1997).

[102] R. S. Armen, O. D. Uitto and S. E. Feller, *Biophys. J.*, **75**, 734, (1998).

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

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

[105] S. R. Durell, B. R. Brooks and A. Ben-Naim, *J. Phys. Chem.*, **98**, 2198, (1994).

[106] W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, **79**, 926, (1983).

[107] S. J. Marrink, R. M. Sok and H. J. C. Berendsen, *J. Chem. Phys.*, **104**, 9090, (1996).

[108] T. R. Stouch, H. E. Alper and D. Bassolino, *Computer-aided Molecular Design*, **589**, 127, (1995).

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

[110] J. R. Nagle, *Biophys. J.*, **64**, 1476, (1993).

[111] T. X. Xiang, *J. Chem. Phys.*, **109**, 7876, (1998).

[112] R. M. Venable, Y. Zhang, B. J. Hardy and R. W. Pastor, *Science*, **262**, 223, (1993).

[113] R. W. Pastor, R. M. Venable, M. Karplus and A. Szabo, *J. Chem. Phys.*, **89**, 1128, (1988).

[114] R. H. Pearson and I. Pasher, *Nature*, **281**, 499, (1979).

[115] H. Hauser, I. Pasher and S. Sundell, *Biochemistry*, **27**, 9166, (1988).

[116] H. L. Scott, E. Jakobsson and S. Subramaniam, *Computers in Physics*, **12**, 328, (1998).

[117] R. C. Bean, W. C. Shepherd and H. Chan, *J. Gen. Physiol.*, **52**, 495, (1968).

[118] A. Walter and J. Gutknecht, *J. Membr. Biol.*, **77**, 255, (1984).

[119] J. M. Wolosin and H. Ginsburg, *Biochimica et Biophysica Acta*, **389**, 20, (1975).

[120] A. Finkelstein, *J. Gen. Physiol.*, **68**, 127, (1976).

[121] E. Orbach and A. Finkelstein, *J. Gen. Physiol.*, **75**, 427, (1980).

[122] J. Brunner, D. E. Graham, H. Hauser and G. Semenza, *J. Membr. Biol.*, **57**, 133, (1980).

[123] J. Gutknecht and A. Walter, *Biochimica et Biophysica Acta*, **649**, 149, (1981).

[124] Z. Bar-On and H. Degani, *Biochimica et Biophysica Acta*, **813**, 207, (1985).

[125] A. Bochain, L. Estey, G. Haronian, C. Reale, M. andRojas and J. Cramer, *J. Membr. Biol.*, **60**, 73, (1981).

[126] M. B. Lande, J. M. Donovan and M. L. Zeidel, *J. Gen. Physiol.*, **106**, 67, (1995).

[127] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **148**, 157, (1995).

[128] T. X. Xiang and B. D. Anderson, *Biophys. J.*, **72**, 223, (1997).

[129] T. X. Xiang and B. D. Anderson, *Biophys. J.*, **75**, 2658, (1998).

[130] T. X. Xiang and B. D. Anderson, *Biochimica et Biophysica Acta*, **1370**, 64, (1998).

[131] G. J. M. Bresseleers, H. L. Goderis and P. P. Tobback, *Biochimica et Biophysica Acta*, **772**, 374, (1984).

[132] J. A. Dix, D. Kivelson and J. M. Diamond, *J. Membr. Biol.*, **40**, 315, (1978).

[133] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **173**, 187, (2000).

[134] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **177**, 137, (2000).

[135] T. X. Xiang, X. Chen and B. D. Anderson, *Biophys. J.*, **63**, 78, (1992).

[136] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **140**, 111, (1994).

[137] T. X. Xiang and B. D. Anderson, *J. Chem. Phys.*, **103**, 8666, (1995).

[138] T. X. Xiang and B. D. Anderson, *J. Membr. Biol.*, **165**, 77, (1998).

[139] T. X. Xiang, *J. Phys. Chem. B*, **103**, 385, (1999).

[140] B. E. Cohen, *J. Membr. Biol.*, **20**, 205, (1975).

[141] B. E. Cohen, *J. Membr. Biol.*, **20**, 235, (1975).

[142] J. M. Diamond and Y. Katz, *J. Membr. Biol.*, **17**, 121, (1974).

[143] A. P. Todd, R. J. Mehlhorn and R. I. Macey, *J. Membr. Biol.*, **109**, 41, (1989).

[144] S. Paula, A. G. Volkov, A. N. Van Hoek, T. H. Haines and D. W. Deamer, *Biophys. J.*, **70**, 339, (1996).

[145] W. G. Hill, R. L. Rivers and M. L. Zeidel, *J. Gen. Physiol.*, **114**, 405, (1999).

[146] J. Brahm, *J. Gen. Physiol.*, **81**, 283, (1983).

[147] N. Bindslev and E. M. Wright, *J. Membr. Biol.*, **29**, 265, (1976).

[148] M. Yazdanian, S. L. Glynn, J. L. Wright and A. Hawi, *Pharm. Res.*, **15**, 1490, (1998).

[149] G. W. Caldwell, S. M. Easlick, J. Gunnet, J. A. Masucci and K. Demarest, *J. Mass Spectrom.*, **33**, 607, (1998).

[150] G. Camenisch, G. Folkers and H. van der Waterbeemd, *Eur. J. Pharm. Sci.*, **6**, 321, (1998).

[151] G. Camenisch, J. Alsenz, H. van der Waterbeemd and G. Folkers, *Eur. J. Pharm. Sci.*, **6**, 313, (1998).

[152] P. Artursson, K. Palm and K. Luthman, *Adv. Drug Del. Rev.*, **22**, 67, (1996).

[153] M. Kansy, F. Senner and K. Gubernator, *J. Med. Chem.*, **41**, 1007, (1998).

[154] K. Palm, K. Luthman, A. L. Ungell, G. Strandlund and A. P., *J. Pharm. Sci.*, **85**, 32, (1996).

[155] V. Pade and S. Stavchansky, *J. Pharm. Sci.*, **87**, 1604, (1998).

[156] K. Palm, K. Luthman, A. L. Ungell, G. Strandlund, F. Beigi, P. Lundahl and P. Artursson, *J. Med. Chem.*, **41**, 5382, (1998).

[157] W. Rubas, N. Jezyk and G. M. Grass, *Pharm. Res.*, **10**, 113, (1993).

[158] S. Yamashita, Y. Tanaka, Y. Endoh, Y. Taki, T. Sakane, T. Nadai and H. Sezaki, *Pharm. Res.*, **14**, 486, (1997).

[159] S. Yee, *Pharm. Res.*, **14**, 763, (1997).

[160] M. C. Gres, B. Julian, M. Bourrie, V. Meunier, C. Roques, M. Berger, X. Boulenc, Y. Berger and G. Fabre, *Pharm. Res.*, **15**, 726, (1998).

[161] H. Lennernas, K. Palm, U. Fagerholm and P. Artursson, *Int. J. Pharm.*, **127**, 103, (1996).

[162] K. A. Lentz, J. Hayashi, L. J. Lucisano and J. E. Polli, *Int. J. Pharm.*, **200**, 41, (2000).

[163] A. Buur, L. Trier, C. Magnusson and P. Artursson, *Int. J. Pharm.*, **127**, 223, (1996).

[164] S. Agatonovic-Kustrin, R. Beresford, A. Pauzi and M. Yusof, *J. Pharm. Biom. Anal.*, **25**, 227, (2001).

[165] P. Stenberg, U. Norinder, K. Luthman and P. Artursson, *J. Med. Chem.*, **44**, 1927, (2001).

[166] H. Z. Bu, M. Poglod, R. G. Micetich and J. K. Khan, *Rapid Commun. Mass Spectrom.*, **14**, 523, (2000).

[167] E. Liang, K. Chessic and M. Yazdanian, *J. Pharm. Sci.*, **89**, 336, (2000).

[168] P. Artursson, *J. Pharm. Sci.*, **79**, 476, (1990).

[169] A. Adson, T. J. Raub, P. S. Burton, C. L. Barsuhn, A. R. Hilgers, K. L. Audus and N. F. H. Ho, *J. Pharm. Sci.*, **83**, 1529, (1994).

[170] A. Adson, P. S. Burton, T. J. Raub, C. L. Barsuhn, K. L. Audus and N. F. H. Ho, *J. Pharm. Sci.*, **84**, 1197, (1995).

[171] V. Pade and S. Stavchansky, *Pharm. Res.*, **14**, 1210, (1997).

[172] C. A. Bailey, P. Bryla and A. W. Malick, *Adv. Drug Del. Rev.*, **22**, 85, (1996).

[173] W. Rubas, J. Villagran, M. Cromwell, A. McLeod, J. Wassenberg and R. Mrsny, *S.T.P. Pharma Sci.*, **5**, 93, (1995).

[174] S. Chong, S. A. Dando, K. M. Soucek and R. A. Morrison, *Pharm. Res.*, **13**, 120, (1996).

[175] J. T. Goodwin, R. A. Conradi, N. F. H. Ho and P. S. Burton, *J. Med. Chem.*, **44**, 3721, (2001).

[176] P. Artursson and J. Karlsson, *Biochem. Biophys. Res. Commun.*, **175**, 880, (1991).

[177] C. Pontier, J. Pachot, R. Botham, B. Lenfant and P. Arnaud, *J. Pharm. Sci.*, **90**, 1608, (2001).

[178] H. van de Waterbeemd, D. A. Smith, K. Beaumont and D. K. Walter, *J. Med. Chem.*, **44**, 1313, (2001).

[179] J. J. L. Cascales, J. G. H. Cifre and J. G. de la Torre, *J. Phys. Chem. B*, **102**, 625, (1998).

[180] K. Tu, M. Tarek, M. L. Klein and D. Scharf, *Biophys. J.*, **75**, 2123, (1998).

[181] L. Koubi, M. Tarek, M. L. Klein and D. Scharf, *Biophys. J.*, **78**, 800, (2000).

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

[183] A. Pohorille and M. A. Wilson, *J. Chem. Phys.*, **104**, 3760, (1996).

[184] A. Pohorille, M. H. new, K. Schweighofer and M. A. Wilson, *Current Topics in Membranes*, **48**, 49, (1999).

[185] B. Widom, *J. Chem. Phys.*, **39**, 2808, (1963).

[186] B. Widom, *J. Chem. Phys.*, **86**, 869, (1982).

[187] J. P. Valleau and G. M. Torrie, *Modern theoretical chemistry*, Plenum Press, New York, volume 5, pp. 169–194.

[188] P. Bopp, P. Kruger and R. Lustig, *Quick reference to molecular simulation*, in preparation.

[189] E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, *Chem. Phys. Lett.*, **156**, 472, (1989).

[190] E. Paci, G. Ciccoti, M. Ferrario and R. Kapral, *Chem. Phys. Lett.*, **176**, 581, (1991).

[191] M. Sprik and G. Ciccotti, *J. Chem. Phys.*, **109**, 7737, (1998).

[192] D. Pearlman, *J. Chem. Phys.*, **98**, 8946, (1993).

[193] W. K. den Otter and W. J. Briels, *J. Chem. Phys.*, **109**, 4139, (1998).

[194] W. K. den Otter and W. J. Briels, *Mol. Phys.*, **98**, 773, (2000).

[195] W. K. den Otter, *J. Chem. Phys.*, **112**, 7283, (2000).

[196] B. Roux and M. Karplus, *J. Phys. Chem.*, **95**, 4856, (1991).

[197] R. Kubo, *Reports on Progress in Physics*, **29**, 255, (1966).

[198] T. J. McIntosh, S. A. Simon and R. C. MacDonald, *Biochimica et Biophysica Acta*, **597**, 445, (1980).

[199] S. H. White, G. I. King and J. E. Cain, *Nature*, **290**, 161, (1981).

[200] Y. Katz and J. M. Diamond, *J. Membr. Biol.*, **17**, 101, (1974).

[201] L. R. De Young and K. A. Dill, *Biochemistry*, **27**, 5281, (1988).

[202] Y. Katz, M. E. Hoffman and R. Blumenthal, *J. Theor. Biol.*, **105**, 493, (1983).

[203] J. A. Marqusee and K. A. Dill, *J. Chem. Phys.*, **85**, 434, (1986).

[204] S. A. Simon, W. L. Stone and P. Busto-Latorre, *Biochimica et Biophysica Acta*, **468**, 378, (1977).

[205] W. L. Stone, *J. Biol. Chem.*, **250**, 4368, (1975).

[206] M. K. Jain and L. V. Wray Jr., *Biochem. Pharm.*, **27**, 1294, (1977).

[207] S. A. Simon and J. Gutknecht, *Biochimica et Biophysica Acta*, **596**, 352, (1980).

[208] R. A. Smith, E. G. Porter and K. W. Miller, *Biochimica et Biophysica Acta*, **645**, 327, (1981).

[209] S. A. Simon, W. L. Stone and P. B. Bennet, *Biochimica et Biophysica Acta*, **550**, 38, (1979).

[210] C. M. Gary-Bobo and H. W. Weber, *J. Phys. Chem.*, **73**, 1155, (1969).

[211] W. J. Albery, A. R. Greenwood and R. F. Kibbles, *Trans. Farady. Soc.*, **63**, 360, (1966).

[212] R. Niesner and H. A., *J. Chem. Eng. Data*, **45**, 1121, (2000).

[213] P. A. Witherspoon and D. N. Saraf, *J. Phys. Chem.*, **69**, 3752, (1965).

[214] D. G. Leaist and R. Lu, *Farady Trans.*, **93**, 1341, (1997).

[215] R. Millis, *J. Phys. Chem.*, **77**, 685, (1973).

[216] V. Vitagliano and P. A. Lyons, *J. Am. Chem. Soc.*, **78**, 4538, (1956).

[217] T. Tominaga, S. Yamamoto and J. Takanaka, *J. Chem. Soc., Faraday Transactions I*, **80**, 941, (1984).

[218] W. R. Lieb and W. D. Stein, *Nature*, **224**, 240, (1969).

[219] W. R. Lieb and W. D. Stein, *Curr. Top. Membr. Transp.*, **2**, 1, (1971).

[220] J. M. Wolosin, H. Ginsburg, W. R. Lieb and W. D. Stein, *J. Gen. Physiol.*, **71**, 93, (1978).

[221] W. D. Stein, *Membrane Transport* (eds. S. L. Bonting and J. J. H. H. M. de Pont), Elsevier/North-Holland Biomedical, New York, (1981).

[222] S. B. Zhu, J. Lee, G. W. Robinson and S. H. Lin, *J. Chem. Phys.*, **90**, 6335, (1989).

[223] W. Sung and P. J. Park, *Biophys. J.*, **73**, 1797, (1997).

[224] S. W. Chiu, E. Jakobsson, S. Subramaniam and J. A. McCammon, *Biophys. J.*, **60**, 273, (1991).

[225] S. W. Chiu, J. A. Novotny and E. Jakobsson, *Biophys. J.*, **64**, 98, (1993).

[226] S. Barsky, *J. Chem. Phys.*, **112**, 3450, (2000).

[227] K. S. Kostov, K. F. Freed, E. B. Webb, M. Mondello and G. S. Grest, *J. Chem. Phys.*, **108**, 9155, (1998).

[228] Y. Graziani and A. Livne, *J. Membr. Biol.*, **7**, 275, (1972).

[229] E. Overton, *Vierteljahrschr. Naturforsch. Ges. Zurich*, **40**, 159, (1895).

[230] A. Grossfield and T. B. Woolf, *Langmuir*, **18**, 198, (2002).

[231] R. A. Badley, W. G. Martin and H. Schneider, *Biochemistry*, **12**, 268, (1973).

[232] J. Frenzel, K. Arnold and P. Nuhn, *Biochimica et Biophysica Acta*, **507**, 185, (1978).

[233] L. G. Herbette, A. M. Katz and J. M. Sturtevant, *Mol. Pharmacol.*, **24**, 259, (1983).

[234] Y. Boulanger, S. Schreier and I. C. P. Smith, *Biochemistry*, **20**, 6824, (1981).

[235] D. G. Rhodes, J. G. Sarmiento and L. G. Herbette, *Mol. Pharmacol.*, **27**, 612, (1985).

[236] K. Tu, M. L. Klein and D. J. Tobias, *Biophys. J.*, **75**, 2147, (1998).

[237] A. D. MacKerell, D. Yin, B. Roux and M. Karplus, *J. Phys. Chem. B*, **102**, 3586, (1998).

[238] I. J. Chen, D. Yin and A. D. MacKerell, *J. Comp. Chem.*, **23**, 199, (2002).

[239] T. R. Stouch, *Mol. Sim.*, **10**, 335, (1993).

[240] J. G. Hardman, J. G. Hardman and L. E. Limbird, *Goodman and Gilman's The pharmacological basis of theurapetics*, McGraw Hill, (2001).

[241] R. W. Zwanzig, *J. Chem. Phys.*, **22**, 1420, (1954).

[242] W. L. Jorgensen, *BOSS, version 4.1*, Yale University, New Haven, CT, (1999).

[243] R. H. Henchman and J. W. Essex, *J. Comp. Chem.*, **20**, 499, (1999).

[244] A. Mitsutake, Y. Sugita and Y. Okamoto, *Biopolymers (Peptide Science)*, **60**, 96, (2001).

[245] R. F. Flewelling and W. L. Hubbell, *Biophys. J.*, **49**, 541, (1986).

[246] N. Ben-Tal, A. Ben-Shaul, A. Nicholls and B. Honig, *Biophys. J.*, **70**, 1796, (1996).

[247] F. Barbato, M. I. La Rotonda and F. Quaglia, *Pharm. Res.*, **14**, 1699, (1997).

[248] A. Kaliszan, R. ad Nasal and A. Bucinski, *Eur. J. Med. Chem.*, **29**, 163, (1994).

[249] G. Caron, G. Steyaert, A. Pagliara, F. Reymond, P. Crivori, P. Gaillard, P. Carrupt, A. Avdeef, J. Comer, K. Box, H. H. Girault and B. Testa, *Helv. Chim. Acta*, **82**, 1211, (1999).

[250] C. Zhu, L. Jiang, T. M. Chen and K. K. Hwang, *Eur. J. Med. Chem.*, **37**, 399, (2002).

[251] T. Osterberg, M. Svensson and P. Lundahl, *Eur. J. Pharm. Sci.*, **12**, 427, (2001).

[252] G. V. Betageri and J. A. Rogers, *Int. J. Pharm.*, **46**, 95, (1988).

[253] N. E. Tayar, R. S. Tsai, B. Testa, P. A. Carrupt and A. Leo, *J. Pharm. Sci.*, **80**, 590, (1991).

[254] P. J. Taylor, *Comprehensive Medicinal Chemistry*, Pergamon Press, Oxford, (1990).

[255] J. A. Dix, J. M. Diamond and D. Kivelson, *Proc. Nat. Ac. Sci. USA*, **71**, 474, (1974).

[256] P. V. Balimane, S. Chong and R. A. Morrison, *J. Pharm. Tox. Meth.*, **44**, 301, (2000).

[257] G. Camenish, G. Folkers and H. van der Waterbeemd, *Int. J. Pharm.*, **147**, 61, (1997).

[258] F. Yamashita, S. Fujiwara and M. Hashida, *J. Chem. Inf. Comput. Sci.*, **42**, 408, (2002).

[259] J. Karlsson and P. Artursson, *Int. J. Pharm.*, **71**, 55, (1991).

[260] F. Wohnsland and B. Faller, *J. Med. Chem.*, **44**, 923, (2001).

[261] O. Rosella, A. Sinclair and P. R. Gibson, *J. Gastroent. Hepat.*, **15**, 626, (2000).

[262] L. R. De Young and K. A. Dill, *J. Phys. Chem.*, **94**, 801, (1990).

[263] S. Fujiwara, F. Yamashita and M. Hashida, *Int. J. Pharm.*, **237**, 95, (2002).

[264] L. G. Herbette, D. W. Chester and D. G. Rhodes, *Biophys. J.*, **49**, 91, (1986).

[265] C. A. Lipinski, F. Lombardo, B. W. Domony and P. J. Feeney, *Advanced Drug Delivery Reviews*, **23**, 3, (1997).