Doctoral Thesis

Molecular dynamics simulations
solvent properties and solvent effects on proteins

Author(s):
Walser, Regula

Publication Date:
2001

Permanent Link:
https://doi.org/10.3929/ethz-a-004120225

Rights / License:
In Copyright - Non-Commercial Use Permitted
Molecular Dynamics Simulations: Solvent Properties and Solvent Effects on Proteins.

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Natural Sciences

presented by
REGULA WALSER
Dipl. Chem. ETH
born May 24, 1972
citizen of Wolfhalden (AR), Switzerland

accepted on the recommendation of
Prof. Dr. W. F. van Gunsteren, examiner
Prof. Dr. P. H. Hünenberger, Prof. Dr. A. E. Mark,
Prof. Dr. U. W. Suter, co-examiners

2001
Acknowledgements

I wish to thank my advisor Wilfred van Gunsteren for accepting me in his group during the time of my PhD. His broad scientific knowledge, his unshakable optimism and enthusiasm and his human style of group leading turned the time of my PhD studies into a very pleasant period of my life.

I am grateful to Alan E. Mark for all his support and for the hospitality I enjoyed during my stay in Groningen in his group as well as in his home. Thanks for reading and improving my papers.

I wish to thank Philippe H. Hünenberger for his help and the thorough correction of my manuscripts.

Thanks to the inhabitants of C62istan, Heiko, Urs and Hubi for all the good times I had at work and after work too.

Thanks to Prisca Cerutti for the efficient handling of all administrative concerns and to the system administrators Roland, Alexandre, Tomas, Fred and Martin Borer for keeping the computers up and running.

Furthermore, thanks to the group at the RUG for a great time, especially to Berk, to whom I'm also grateful for the help concerning the viscosity.

Finally, I would like to thank all the former and current members of IGC for the extremely good atmosphere and for all the activities such as ski-weekends, borrels, group outings and Palmhof visits. The time of my PhD would not have been the same without all of you.
Contents

Acknowledgements i
Kurzfassung vii
Summary ix
Publications xi

1 Introduction 1
   1.1 Statistical Mechanics 1
   1.2 Monte Carlo 2
   1.3 Molecular Dynamics 3
      1.3.1 Propagation in time 3
      1.3.2 System size 3
      1.3.3 Force Field 4
      1.3.4 Parameterisation 7
      1.3.5 The solvent in Molecular Dynamics 8

2 The Effect of Force-field Parameters on Properties of Liquids: Parameterisation of a Simple Three-Site Model for Methanol 11
   2.1 Abstract 11
   2.2 Introduction 11
   2.3 Method 13
      2.3.1 Parameterisation 13
      2.3.2 Properties 15
   2.4 Effect of the parameters 19
      2.4.1 Charge distribution 19
      2.4.2 Bond length 20
      2.4.3 Treatment of long-range forces 21
      2.4.4 Constant volume versus constant pressure 22
   2.5 Properties of model B3 23
      2.5.1 Transport properties 24
      2.5.2 Dielectric properties 24
      2.5.3 Thermodynamic properties 25
      2.5.4 Free energy 25
      2.5.5 Mixing with water 25
2.6 Conclusion ........................................................................................................ 27

3 On the temperature and pressure dependence of a range of properties of a type of water model commonly used in high temperature protein unfolding simulations 29
3.1 Abstract ........................................................................................................... 29
3.2 Introduction .................................................................................................... 29
3.3 Method ........................................................................................................... 31
3.3.1 Simulation ............................................................................................... 31
3.3.2 Analysis .................................................................................................... 31
3.4 Results .......................................................................................................... 32
3.4.1 Thermodynamic properties ..................................................................... 32
3.4.2 Structural properties .............................................................................. 34
3.4.3 Dynamic properties .............................................................................. 37
3.5 Conclusions .................................................................................................. 37

4 On the validity of Stokes’ law at the molecular level 41
4.1 Abstract ........................................................................................................ 41
4.2 Introduction ................................................................................................... 41
4.3 Method .......................................................................................................... 43
4.4 Results and Conclusions .............................................................................. 44

5 Further investigation on the validity of Stokes-Einstein behaviour at the molecular level 47
5.1 Abstract ........................................................................................................ 47
5.2 Introduction ................................................................................................... 47
5.3 Method .......................................................................................................... 49
5.4 Results and Conclusions .............................................................................. 50

6 Viscosity dependence of protein dynamics 55
6.1 Abstract ........................................................................................................ 55
6.2 Introduction ................................................................................................... 55
6.3 Models and Simulations .............................................................................. 56
6.4 Results and Discussion .............................................................................. 58
6.5 Summary and Conclusion ............................................................................ 62

7 Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal 65
7.1 Abstract ........................................................................................................ 65
7.2 Introduction ................................................................................................... 65
7.3 Method .......................................................................................................... 67
7.3.1 Simulation set-up .................................................................................. 68
7.3.2 Analysis of the simulations .................................................................... 69
7.4 Results .......................................................................................................... 70
7.4.1 Energetic properties ............................................................................. 70
7.4.2 Protein properties ............................................................................... 71
7.4.3 Ion properties as function of the scheme to evaluate electrostatic forces 71
Contents

7.4.4 Ion properties as function of the initial placement of the ions .......... 78
7.5 Conclusion .................................................................................. 79

Bibliography ..................................................................................... 81

Curriculum Vitae ............................................................................... 93
Kurzfassung


Die Eigenschaften einer Flüssigkeit hängen nicht nur von den gewählten Parametern zur Beschreibung der Moleküle ab, sondern auch von den Bedingungen, wie Temperatur und Druck, zu welchen simuliert wird. Nun werden Simulationen von Biomolekülen häufig unter ganz anderen Bedingungen ausgeführt als denjenigen, für die das Lösungsmittel ursprünglich parameterisiert wurde. So wird zum Beispiel häufig bei höheren Temperaturen simuliert, um die Atombewegungen zu beschleunigen. In Kapitel 3 werden deshalb die Eigenschaften eines Modells für Wasser bei erhöhter Temperatur untersucht, um festzustellen, wie stark dieser Einfluss ist.

Lösungsmittel höherer Viskosität nicht nur auf der Proteinoberfläche verlangsamen, sondern dass dieser Effekt bis in das Proteinzentrum hinein wirksam ist.

Summary

Molecular dynamics simulations have contributed significantly to the understanding of biomolecular systems. When simulating biomolecules in solution, the correct description of the solvent’s influences are important, as they partially determine the properties of the molecule. Therefore, in this thesis the properties of pure liquids in view of their use as solvents are investigated and also the influences of the solvent on the solvated molecule.

The first chapter gives a brief introduction into simulation of molecules. The basic theory is explained. The interaction function and the approximations applied are presented. The determination of the parameters describing the potential energy and the different possibilities to describe the solvent are briefly presented.

The correct description of the solvent requires several different parameters. Chapter 2 describes the determination of those parameters for methanol. Starting from an existing model for methanol, it was tried to improve it such, that the mixing with water and the dielectric permittivity agree closer with experiment. Doing this, the effect of changing the parameters describing molecular geometry and charge distribution on the properties of the liquid was investigated.

The properties of the liquid do not depend solely on the parameters describing the molecules but also on the simulation conditions, such as temperature and pressure. Simulations of biomolecules are often carried out at conditions very different from those for which the solvent was parameterised. For instance, simulations are often performed at higher temperatures in order to increase atom mobility. In Chapter 3 the properties of a water model at higher temperatures are investigated to determine the size of this influence.

In computer simulations, certain properties can be changed very specifically. This is used in Chapters 4 and 5 to change the viscosity of water without affecting most other properties. This allows the testing of Stokes’ law for small molecules. According to Stokes’ law the product of viscosity and diffusion constant depends only on temperature and the particle’s size. It is often used for molecules, though it is derived from the assumption of a sphere in a fluid. The same principle of changing the viscosity is also used in Chapter 6 to investigate the influence of the solvent’s viscosity of the solvated protein’s dynamics. A protein (factor Xa) was simulated in two solvents of different viscosity and the differences were investigated. It turns out that in the high viscosity solvent the atomic motions are slowed down not only on the protein surface, but also in the protein centre.

Besides the parameters and the simulation conditions, the treatment of the intermolecular interactions has an influence on the simulation. In the last chapter three different ways of treating the electrostatic interactions are investigated. As a test system a protein crystal was selected, as simulations of crystals are more sensitive to the way of treating the electrostatic interactions than simulations of solvated molecules. Additionally the influence of counter ions and, as the counter ions are not very mobile, the influence of the initial position of the ions was investigated. The
result of these investigations is that long-ranged electrostatic interactions may not be neglected in crystal simulations, but that the addition of counter ions is not required.
Publications

This thesis has led to the following publications:

**Chapter 2:**

Regula Walser, Alan E. Mark, Wilfred F. van Gunsteren, M. Lauterbach, G. Wipff,  
“The effect of force-field parameters on properties of liquids: Parameterization of a simple three-site model for methanol”  

**Chapter 3:**

Regula Walser, Alan E. Mark, Wilfred F. van Gunsteren,  
“On the Temperature and Pressure Dependence of a Range of Properties of a Type of Water Model Commonly Used in High Temperature Protein Unfolding Simulations”  

**Chapter 4:**

Regula Walser, Alan E. Mark, Wilfred F. van Gunsteren,  
“On the validity of Stokes’ law at the molecular level”  

**Chapter 5:**

Regula Walser, Berk Hess, Alan E. Mark, Wilfred F. van Gunsteren,  
“Further investigation on the validity of Stokes-Einstein behaviour at the molecular level”  

**Chapter 6:**

Regula Walser, Wilfred F. van Gunsteren,  
“Viscosity Dependence of Protein Dynamics”  
Chapter 7:

Regula Walser, Philippe H. Hünenberger, Wilfred F. van Gunsteren,
“Comparison of Different Schemes to Treat Long-range Electrostatic Interactions in Molecular Dynamics Simulations of a Protein Crystal”

Proteins in press.
Chapter 1
Introduction

1.1 Statistical Mechanics

Understanding the relationship between the microscopic properties of individual molecules and the macroscopic properties of a thermodynamic system is the domain of statistical mechanics. Statistical mechanics permits the derivation of macroscopic information, such as pressure, internal energy and entropy, for any system of microscopic particles. Using classical mechanics, a microscopic system of \( N \) particles can be described by \( 6N \) variables, as each particle has \( 3 \) degrees of freedom for the position and \( 3 \) for the momenta. Since a thermodynamic system is characterised by much fewer quantities, it is clear that an infinite number of microscopic states are represented in a given thermodynamic system. Therefore the problem in statistical mechanics is to average correctly over all microscopic states belonging to a given thermodynamic system in order to obtain the desired thermodynamic properties.

The positions and momenta of all particles in a microscopic state can be represented by a vector in a \( 6N \)-dimensional space, called the phase space. Such a vector in phase space will be denoted by \( \Gamma \). In an isolated system the specification of \( \Gamma \) and a potential energy function will determine completely the system’s past and future.

Consider an instantaneous observable \( G \), depending only on the microscopic state vector \( \Gamma \). The expectation value of the observable \( G \) can be calculated either as an integral over the whole phase space weighted by the probability of occurrence of each \( \Gamma \) (ensemble average) or by following the system over a sufficiently long time period \( t \) and averaging over the measured values of \( G \) (time average).

\[
\langle G \rangle = \int G(\Gamma)p(\Gamma)d\Gamma = \lim_{t \to \infty} \frac{1}{t} \int_{t_0}^{t+t_0} G(\Gamma(t))dt
\]

where \( p(\Gamma) \) is the probability density of the system being in the volume element \( d\Gamma \) around \( \Gamma \). The identification of the ensemble average with the time average has not been rigorously proven, but is justified by many successful comparisons between the predictions of statistical mechanics and experimental results. In a micro-canonical ensemble, i.e. an ensemble with constant energy \( E \), volume \( V \) and number of particles \( N \), the probability density is assumed to be constant in phase space (postulate of equal probability)

\[
p(\Gamma) = \frac{\delta(H(\Gamma) - E)}{\int \delta(H(\Gamma) - E)d\Gamma} = \frac{\delta(H(\Gamma) - E)}{Q_{NVE}}
\]

(1.2)
where $\delta$ is the Dirac delta function, $H(\Gamma)$ the Hamiltonian (total energy) of the system in the microscopic configuration $\Gamma$ and $Q_{NVE}$ is called the micro-canonical partition function. In a canonical ensemble, i.e. an ensemble with constant volume $V$, temperature $T$ and number of particles $N$, the probability of each state is given by the Boltzmann distribution

$$p(\Gamma) = \frac{e^{-\beta H(\Gamma)}}{\int e^{-\beta H(\Gamma)} d\Gamma} = \frac{e^{-\beta H(\Gamma)}}{Q_{NVT}}$$

where the factor $\beta$ is defined as $\frac{1}{k_B T}$, $k_B$ being the Boltzmann constant.

While in statistical mechanics some simple systems can be characterised analytically, such as the ideal gas or the two-dimensional Ising model, most are not amenable to an exact description. Consequently, the use of computers to obtain numerical solutions for complex statistical mechanics problems has become widespread. The key idea is to reduce the integral over the whole phase space in Eq. (1.1) to a summation over a sufficient but finite number of states. Focusing on the canonical ensemble, two schemes are particularly useful to achieve this goal, because they may be used to generate directly a sequence of configurations, obeying a Boltzmann probability distribution: Monte Carlo (MC) and Molecular Dynamics (MD). Since only MD is used in the present work, MC techniques will only be briefly mentioned here. When either method is used Eq. (1.1) reduces to the approximate form

$$\langle G \rangle \approx \frac{1}{S} \sum_{s=1}^{S} G(\Gamma_s) \quad (1.4)$$

where $S$ is the number of states picked.

### 1.2 Monte Carlo

A Monte Carlo simulation is started from an arbitrary initial configuration. New states are then generated in sequence by performing random atomic displacements from the previous configuration. A special scheme has to be applied to decide on acceptance or rejection of the newly generated configuration in order to ensure Boltzmann-weighted sampling. The new configuration is accepted according to a probability $p$ which depends on the potential energy difference between the two configurations.

$$p = \begin{cases} 
1 & \text{if } \Delta U \leq 0 \\
\frac{1}{e^{eta \Delta U}} & \text{if } \Delta U > 0
\end{cases} \quad (1.5)$$

where $\Delta U$ is the difference in potential energy between the old and the new configuration. In practice, when $\Delta U > 0$, $e^{-\beta \Delta U}$ is compared to a random number uniformly distributed between 0 and 1 and the configuration is accepted if this quantity is larger than the random number. Because of the importance of random numbers in this procedure, it was called Monte Carlo (MC) by its inventors [1,2]. The main advantage of this scheme is that it allows hopping through high but narrow energy barriers. The main disadvantage is the absence of information on dynamical processes.
1.3 Molecular Dynamics

In Molecular Dynamics (MD) the initial state is propagated in time according to the laws of classical mechanics. While this may seem a more obvious way to generate an ensemble than generating configurations at random, this method was developed later than MC [3].

1.3.1 Propagation in time

In Cartesian coordinates, the evolution of the system in time is described by the Newtonian equations of motion

\[ F_i = -\frac{\partial U(x)}{\partial x_i} = m_i \ddot{x}_i \]  

where \( F_i \) is the force acting on particle \( i \) with position \( x_i \) and mass \( m_i \), and \( U \) is the potential energy of the system.

There are several methods to integrate Eq. (1.6) by discretising the time into small time steps \( \Delta t \). The one implemented in the GROMOS96 program [4, 5], which was used for all simulations in the present work is called the leap-frog algorithm [6]

\[ \dot{x}_i \left( t + \frac{\Delta t}{2} \right) = \dot{x}_i \left( t - \frac{\Delta t}{2} \right) + \frac{F_i}{m_i} \Delta t \]  

Since Newton’s equations conserve the total energy, MD samples by default a micro-canonical ensemble. In practice, total energy will not be conserved due to errors introduced by approximations like the use of a cutoff radius for the interactions between particles. In order to sample the canonical ensemble a thermostat has to be introduced. A simple, commonly used thermostat is the Berendsen algorithm [7], where all velocities are scaled at each time step to reach the required temperature asymptotically.

\[ \dot{x}_{scaled}(t) = \dot{x}(t) \sqrt{1 + \frac{\Delta t}{\tau_T} \left( \frac{T_0}{T(t)} - 1 \right)} \]  

where \( \tau_T \) is an adjustable parameter called temperature relaxation time, which governs how tightly the system is coupled to the bath at temperature \( T_0 \). For sampling of the isothermal-isobaric ensemble (NPT) a similar scheme must be applied to couple the box volume to the actual pressure.

1.3.2 System size

The size of the system considered in a computer simulation is limited by the computer speed and memory requirements. To avoid that surface effects, which would be very large in the small systems simulated, disturb the simulation, periodic boundary conditions are often applied. In this case, a rectangular (or any other space-filling volume) box is surrounded by an infinite number of replicas of itself. When a particle leaves the box through one face, its periodic image
enters the box through the opposite face. In practice, the rectangular parallelepiped and truncated octahedron boxes are the most commonly used box shapes in molecular simulations. The latter have the advantage of being closer to a spherical volume compared to a cube.

In order to limit the effect of artificial periodicity in simulations of solutions, the cut-off radius must be chosen smaller than half the smallest box length (in the case of a rectangular box). This avoids a particle interacting with another particle and at the same time with the other particle’s image.

1.3.3 Force Field

In principle $U(x)$ in Equation (1.6) can be any function which is differentiable within the cut off. However, the choice is limited by the computational effort. Three-body terms for instance, which would contribute to the total energy, are very time consuming to calculate. Consequently, they are usually neglected. Instead, their mean effect is incorporated into computationally less demanding pair potentials, which thus become “effective” pair potentials. Commonly used potential energy functions contain a van der Waals term, a Coulomb term and covalent interaction terms. The van der Waals and electrostatic interactions are collectively referred to as non-bonded interactions. To save computer time, only a limited number of interacting pairs are considered, for otherwise the effort would increase as $N^2$, $N$ being the number of atoms simulated. In practice, only the non-bonded interactions between particles separated by less than a certain cut-off distance $r_C$ are calculated, which effectively reduces the computational effort to the order of $N$.

Van der Waals interactions

The pairwise van der Waals interaction term is an empirical function used to reproduce repulsion-dispersion interactions arising from quantum mechanics. It consists of a strongly repulsive component at close distances, accounting for electron shell overlap, and a longer-ranged attractive component, which results from instantaneous dipole interactions. A commonly used form for the van der Waals interaction between two atoms is the Lennard-Jones potential

$$U_{LJ} = 4\varepsilon \left( \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right)$$  \hspace{1cm} (1.10)$$

where $\varepsilon$ and $\sigma$ are constants characteristic of a specific atom pair. This equation is also commonly written as

$$U_{LJ} = \frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^6}$$  \hspace{1cm} (1.11)$$

For interactions involving two different types of atoms, the interaction parameters $C_{12}$ and $C_6$ or $\varepsilon$ and $\sigma$ are generally calculated as the mean of the parameters characteristic of the type of the two atoms involved. In GROMOS96 a geometric mean of the atomic parameters is used for the $C_{12}$ and $C_6$ terms.

In GROMOS96, no special term is used to describe hydrogen bonding. In this case the van der Waals interaction can also be used (in balance with the electrostatic interactions) to mimic hydrogen bonding. This is done by increasing the $C_{12}$ term for interactions involving hydrogen bonding.
1.3. Molecular Dynamics

Electrostatic interactions

According to Coulomb’s law, the electrostatic interaction term is

\[ U_{ele} = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}. \]  

While neglecting the interaction of pairs beyond some cut-off distance poses little problem for the short-ranged Lennard-Jones interactions, this is not the case for the farther-reaching Coulomb interactions. The corresponding error may be somewhat reduced by grouping atoms into so called charge groups with zero total charge whenever possible. Neutral charge groups interact as dipoles. Performing the cut-off truncation in terms of these charge groups reduces the cut-off error, since dipole-dipole interactions decrease faster \( (r^{-3}) \) than charge-charge interactions \( (r^{-1}) \). A further reduction of the error can be obtained by introducing a second, larger cut off (twin-range method). The interactions of particles at a distance between the first and the second cut off are calculated only every 5 or 10 steps. Assuming that the high-frequency fluctuations of interactions in this range are negligible, the effective cut off is increased with a moderate increase in computational effort.

Two main strategies have been developed to further reduce the error introduced by neglecting interactions beyond a cut-off distance, the reaction-field method and the lattice-sum methods.

**Reaction-field method:** In this method, the cut-off sphere of radius \( r_{RF} \) around an atom or charge group is considered to be embedded in a dielectric continuum. The electrostatic interaction term must be modified to [8]

\[ U_{ele} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_1} \left( \frac{1}{r_{ij}} - \frac{\frac{1}{2}C_{RF}(r_{ij})^2}{r_{RF}^3} \right) \]  

\[ C_{RF} = \frac{(2\varepsilon_1 - 2\varepsilon_{RF})(1 + \kappa r_{RF}) - \varepsilon_{RF}(\kappa r_{RF})^2}{(\varepsilon_1 + 2\varepsilon_{RF})(1 + \kappa r_{RF}) + \varepsilon_{RF}(\kappa r_{RF})^2} \]  

where \( \varepsilon_{RF} \) is the relative dielectric permittivity of the dielectric continuum and \( \kappa \) the inverse Debye screening length. This technique has the advantage of being computationally inexpensive. Since it is based on the idea of a homogeneous dielectric continuum outside the cut-off sphere, it is a priori more suited for liquid simulations than for crystal simulations.

**Lattice-sum methods:** The idea here is to compute the electrostatic interactions, as if the system were an infinite ideal crystal formed by the computational box and an infinite number of periodic copies of itself. Several methods have been developed to compute these interactions. Ewald summation [9] is the oldest and probably the best known of these. In Chapter 7 the particle-particle particle-mesh (P³M) method [10, 11] was used. Another alternative is the particle-mesh Ewald method [12]. In these three approaches, point charges are shielded by charge distributions of equal magnitude but opposite sign (see top of Figure 1.1). These charge distributions are often taken to be Gaussian, but may also take different shapes, for instance a hat shape. The screened interactions are short-ranged and may be calculated easily by real-space summation using a short cut-off distance. Charge distributions of the same sign as the original
Figure 1.1: The charge distribution used in lattice sum methods. The top part shows the original point charges with the screening charge distribution in a hat form. They are summed up in real space. The bottom part shows the cancelling charge distribution which is summed up in Fourier space.

Charges are then added so that the total potential matches the potential resulting from the original point charges (see bottom of Figure 1.1). The potential arising from this correcting charge distribution can be easily evaluated in Fourier space.

In the P^3M method, the computational effort of evaluating the reciprocal-space interactions is reduced by mapping the charge distribution onto a grid, which allows the use of Fast Fourier transform algorithms. Because lattice-sum methods are based on the idea of a perfect crystal, they are a priori more suited for the simulation of crystals, whereas in simulations of liquids, they might produce artifacts resulting from introduction of an artificial periodicity [13]. In addition, these methods are computationally more expensive than the reaction-field method.

Covalent interactions

The covalent interaction terms describe interactions between directly or indirectly connected atoms, e.g. for chemical bonds, angles between three connecting atoms or torsional angles formed by four atoms. Although a proper representation of the chemical bond requires the use of quantum mechanics, it is often described in classical MD as a classical harmonic oscillator characterised by an ideal value of the bond length and a harmonic force constant. The simplest description of a bond potential is thus

\[ U_{\text{bond}} = \frac{1}{2} k_i (b_i - b_i^0)^2 \]  

where \( b_i^0 \) is the ideal length of bond \( i \), and \( k_i \) the force constant. In GROMOS96 a slightly different, quadratic expression is used, which is computationally more efficient. Such a description of bonds does not allow for bond formation or bond breaking.
Similar descriptions involving an ideal value and a force constant are used for angles and
improper torsional dihedral angles, whereas the interaction terms for dihedral angles involve
cosine functions.

Bond vibrations are the fastest motions in a molecule and thus determine the maximum step
$\Delta t$ for accurate integration of the equations of motion. To be able to increase the time step, bond
lengths are often constrained to fixed values and not allowed to vibrate. The algorithm used in
GROMOS96 to enforce bond constraints is called SHAKE [14].

### 1.3.4 Parameterisation

All interaction functions mentioned in the last section require the definition of a number of
parameters. For the Coulomb interaction it is the partial charge $q_i$, for the Lennard-Jones inter-
action it is $\varepsilon$ and $\sigma$ (or $C_{12}$ and $C_6$), and for the bonded interactions it is the ideal bond length
$b_i^0$ and the force constant $k_i$. Determination of these parameters such that the observables cal-
culated from molecular dynamics simulation reproduce experimental data is a difficult task, but
essential to obtain reliable results.

Several problems must be addressed during force-field parameterisation. First, there is gen-
erally no simple relationship between one parameter and one simulated property. Thus, param-
eterisation involves the adjustment of several parameters, and the calculation and comparison
with experiment of several observables, since one single observable may not be sensitive to all
parameters. Second, there is not always a clear connection between a simulated observable and
an experimental observable. For instance, the simulated potential energy is related to the exper-
imental heat of vaporisation, but the exact form of the relationship is ambiguous [15]. Finally,
the observable of interest may be expensive to calculate. This may make the specific observable
unsuited for use in parameterisation, since the fitting naturally involves multiple calculations of
the property for different trial sets of parameters.

Depending on the parameter, there are several ways to obtain an initial parameter value.
Some parameters can be determined directly from experiment, for instance ideal bond lengths
from X-ray diffraction. However, for the correct reproduction of the desired experimental prop-
erties, it might be more important to reproduce condensed-phase properties (e.g. density), than
the exact molecular geometry. Thus, even parameter values obtained from experimental data
may have to be readjusted during the parameterisation. For other parameters, an initial param-
eter value can be obtained from ab initio calculation, for instance atomic partial charges. Here
also, these parameters have generally to be refined to reproduce experimental condensed-phase
properties like density, energies or transport properties.

Several methods are used to refine the parameters. If a strong relationship between one
parameter and one observable exists, the parameter can be refined using the weak coupling
method [16]. Otherwise, search techniques like MC may be used in parameter space in order to
find a good set of parameters. However, most parameter optimisations are still made by trial and
error. An example of a parameterisation of methanol is given in Chapter 2.

There are usually several parameter sets which perform equally well in reproducing exper-
imental properties, as can be seen from the many models existing for liquid water [15,17–19].
The quality of a force field cannot easily been assessed. Each force field will perform best for the
conditions under which it was calibrated. This turns the mixing of parameters from two differ-
ent force fields, optimised separately for different purposes, into a risky operation. In particular,
force-field parameters calibrated for gas phase and for condensed phases should not be mixed.

1.3.5 The solvent in Molecular Dynamics

Experimental techniques provide more and more information about proteins. However they do not allow to monitor the dynamics at the atomic level, which is in contrast possible in MD simulation. Therefore, MD is often used to study protein dynamics. However, the time scales accessible to MD are still shorter than the time scales relevant for many processes of interest in proteins (e.g. protein folding). To circumvent this problem, one may attempt to reduce the number of particles in the system.

In a system consisting of a protein in solution a large fraction of the system is represented by solvent molecules. Since usually only the protein is of interest, one would like to model the solvent in less detail, retaining only a proper representation of its mean effect on the solute.

Vacuum

The simplest approach is to completely ignore the solvent and simulate the molecule in vacuum. This usually leads to strong distortions in the properties of atoms at the surface of the system. The shape of the molecule may also be distorted, since it generally tends to minimise its surface area. Finally, the electric screening effect of a solvent with high dielectric permittivity, which also affects the simulated properties, is absent in simulations in vacuo. Thus, simulations in vacuo normally do not produce accurate results.

Implicit Solvent Models

Another solution is to mimic the effect of the solvent by applying a mean force together with random forces onto the protein atoms, with the goal of reproducing the mean effect of solvation and the collisions between the solute and the solvent atoms. The energy introduced by the random forces must be compensated by means of an additional frictional force. The frictional force may by proportional to the solvent accessible surface area of each solute atom. Modelling the solvent in this way is called Stochastic Dynamics.

However, it is difficult to include all relevant solvent effects in a such a model. For instance the solvent entropy changes during protein folding will be neglected. Protein folding is driven by a small free energy difference and it will be hard to accurately reproduce this energy difference when representing a large part of the system by a very crude approximation.

Explicit Solvent Models

Although an explicit representation of the solvent is computationally much more expensive than the afore mentioned methods, it seems to be the only one capable of reproducing the free energy surface well enough to observe correct protein motions [20]. While this method is certainly the most accurate way of treating the solvent, some points still have to be considered.

- The computational effort needed must be limited to a minimum. This requires often a simplified description of the interactions, which neglects important effects. For instance,
polarisability is neglected in commonly used water models. The omission is compensated by increasing the molecular dipole.

- Compatible force fields should be used for the solute and the solvent, since the proper solvation depends on the correct balance between solute-solute, solute-solvent and solvent-solvent interactions. These are not guaranteed to be correct if different force fields are used.

- Usually a solvent is parameterised for a certain state point (temperature and pressure) and generally not accurate at different state points. Due to the use of a simplified interaction function, it might not be possible to find models that work equally well at different state points. However, simulations of macromolecules are often performed at extreme conditions, for instance at temperatures above 100 °C in liquid water, in order to speed up the simulation. The properties of the solvent model might be quite different at these extreme conditions compared to those at the state point where it was parameterised. A study of the effect of high temperatures on the SPC water model is presented in Chapter 3.

Simulations in explicit solvent also demonstrate an advantage of computer simulations over experiment. In computer simulations it is possible to perform unphysical experiments. Sometimes this allows to specifically change a certain property and investigate the effect of this change. For instance it is possible to change the viscosity of a liquid without affecting interaction parameters. This is used in Chapters 4 and 5 to investigate Stokes' law and in Chapter 6 to investigate the influence of a change in solvent viscosity on protein dynamics.
Chapter 2

The Effect of Force-field Parameters on Properties of Liquids: Parameterisation of a Simple Three-Site Model for Methanol

2.1 Abstract

A simple rigid three-site model for methanol compatible with the simple point charge (SPC) water and the GROMOS96 force field is parameterised and tested. The influence of different force-field parameters, such as the methanol geometry and the charge distribution on several properties calculated by molecular dynamics is investigated. In particular an attempt was made to obtain good agreement with experimental data for the static dielectric constant and the mixing enthalpy with water. The model is compared to other methanol models from the literature in terms of the ability to reproduce a range of experimental properties.

2.2 Introduction

Computer simulations are increasingly used to study and understand the properties of pure solvent mixtures and solute-solvent interactions at an atomic level. For example the properties of biomolecules such as proteins are strongly influenced by the nature of the surrounding solvent or solvent mixture [21–26]. To effectively study mixed systems it is a prerequisite that the models used to describe the individual components are compatible. That is, first they have to be derived in a consistent manner and second, they have to be parameterised not only to reproduce the properties of the isolated components but also of the mixture. Unfortunately, this is not always the case. A model for one compound parameterised to reproduce one set of properties may be combined with a model for another compound parameterised on a different set of properties and the interactions between the compounds derived using simple combination rules.

This chapter describes the parameterisation and testing of a simple rigid three-site model for methanol compatible with the simple point charge (SPC) water model [17] and the GROMOS96 force field [4]. The model is also compared to other methanol models from the literature including those proposed by Jorgensen [27] (referred to as J2), Haughney, Ferrario and McDonald [28,29] (referred to as H1) and van Leeuwen and Smit [30] (referred to as L1) in terms of the
The Effect of Force-field Parameters

A number of simple models for methanol have been previously proposed. Jorgensen’s OPLS model for methanol [27] (J2), based on an earlier TIP model [31], was developed within a homologous series of alcohols by studying hydrogen-bonded methanol dimers and methanol-water complexes. The model was optimised to reproduce the heat of vaporisation and the liquid density at ambient temperature and pressure based on Monte Carlo simulations in the NPT ensemble on a system of 128 methanol molecules using a cut off of $R_c = 0.95$ nm and a correction term for Lennard-Jones interactions neglected beyond $R_c$. Amongst other properties investigated, the heat capacity and isothermal compressibility of his model agreed closely with experimental data.

Haughney et al. [28, 29] developed two models for methanol (H1 and H2), as refinements of Jorgensen’s TIP model [31], designed primarily to better reproduce the strength of the hydrogen bonding. H1 differed from TIP only in the choice of charge distribution, while H2 differed also in the values used for the Lennard-Jones coefficients. They performed MD simulations based on the generalised methods of constraints [32] in the NVT ensemble for a system of 108 methanol molecules over a temperature range of approximately 80 K (260 K to 340 K) for their models H1 and H2 as well as for Jorgensen’s OPLS and TIP models. Electrostatic interactions were treated by the Ewald method. They concluded that the models H1 and OPLS gave results for a range of properties, e.g. the self-diffusion coefficient, that overall are in good agreement with the available experimental data.

Stouten’s model (OM2) was derived within a framework of a comparative research concerning hydrogen bonding in the crystalline and liquid phase of methanol [33]. Stouten performed MD simulations on a system of 216 methanol molecules at room temperature using a cut off of 0.7 nm for the van der Waals interactions and of 1.1 nm for the Coulomb interactions, and obtained results comparable to those of Jorgensen’s OPLS model [27] and the models of Haughney et al. [29]. Stouten developed a second flexible model (OM1) for comparison with water concerning hydrogen bond properties [33].

Van Leeuwen and Smit [30] re-refined the models of Jorgensen [27, 31] and Haughney et al. [28, 29] to reproduce phase coexistence properties of methanol over a wide range of temperatures (275 K to 525 K) and densities (up to 0.8 g/cm$^3$), based on studies of the vapour-liquid equilibria using Gibbs ensemble Monte Carlo simulations [34]. This approach had previously been used by Mezei [35] to calculate the phase diagram for Jorgensen’s OPLS model [27].

In the current study we use a system of 512 molecules and long simulation times to expand the range of experimental properties previously used to compare the above models. The properties considered include: the heat of vaporisation, the density at standard pressure, the diffusion constant, the viscosity, the dielectric constant, the isothermal compressibility, the heat capacity, the thermal expansion coefficient and the Debye relaxation time. We also investigate the sensitivity of a subset of these properties to changes in the van der Waals parameters, the partial charges assigned to the atoms, the geometry of the molecules, the effect of cut-off size and the effect of using a reaction field to treat the long-range Coulomb interactions.

Apart from the pure liquids, numerous studies of methanol-water binary mixtures are found in literature. The solvation of a methanol molecule in water has been investigated by Monte Carlo simulation [36–38]. MD simulations of binary mixtures with different mole fractions of methanol are reported by Pálinkás et al. [39, 40], Stouten and Kroon [41], Ferrario et al. [42], Freitas [43] and Laaksonen et al. [44]. Wheeler and Rowley [45] have simulated ternary mixtures of water, methanol and acetone. We present a comparison of the density and heat of mixing
2.3 Method

All simulations were performed using the GROMOS96 simulation package [4]. The methanol was kept rigid by applying the SHAKE procedure [14] using a relative geometric accuracy of $10^{-4}$. The intermolecular potential energy function was represented as the pair wise sum over all pairs of different molecules of a Coulomb and 12-6 Lennard-Jones interaction term,

$$U_{ij} = \frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_{6}(i,j)}{r_{ij}^{6}} + \frac{q_i q_j}{4 \pi \varepsilon_0 r_{ij}}$$  \hspace{1cm} (2.1)

where $r_{ij}$ represents the distance between two atoms $i$ and $j$, $q_i$ the charge of atom $i$ and $\varepsilon_0$ the dielectric permittivity of the vacuum. $C_6(i,j)$ and $C_{12}(i,j)$ are the Lennard-Jones coefficients for the interaction between atoms $i$ and $j$. The potential energy $U(r_{ij})$ was calculated by using a twin-range cut off. The non-bonded interaction between molecules, where the distance between their first atoms (oxygen atoms) lies within a spherical cut off radius of $R_c = 0.9 \text{ nm}$ was calculated every step, while the interactions for molecules with distances between $R_c = 0.9 \text{ nm}$ and $R_{cl} = 1.4 \text{ nm}$ were evaluated only every fifth step. In the NVT simulations, the cubic periodic box with 3.2596 nm edge length contained 512 methanol molecules resulting in an experimental density [46] of 473.7 u/nm$^3$. During the NPT simulations the pressure was kept at 1 atm. The temperature (298 K) and pressure (NPT) were maintained by weak coupling to an external bath [7] using a coupling time of 0.1 ps for the temperature and 0.5 ps for the pressure. In the constant pressure runs, the isothermal compressibility of the system was set to $7.5 \cdot 10^{-4} \text{ mol nm}^3/\text{kJ}$.

Where stated, a reaction-field term [8,47], $U_{RF}$, was included in Equation (2.1) to approximate the long-range Coulomb forces beyond $R_{cl}$, using a dielectric constant of the continuum [48] of $\varepsilon_{RF} = 32.63$:

$$U_{RF}(r_{ij}) = \frac{q_i q_j (\varepsilon_{RF} - 1) r_{ij}^2}{4 \pi \varepsilon_0 (2 \varepsilon_{RF} + 1) R_{cl}^3}$$  \hspace{1cm} (2.2)

2.3.1 Parameterisation

The models were fitted to reproduce the experimental heat of vaporisation $\Delta H_{vap}$ at a given temperature and pressure. The intermolecular potential energy can be related to the heat of vaporisation by the expression [49]:

$$\Delta H_{vap}(T) = -U(T) + p\Delta V + Q_{int} + Q_{ext}$$  \hspace{1cm} (2.3)

where $\Delta H_{vap}$ is the experimental molar heat of vaporisation ($\Delta H_{vap}(298 \text{ K}) = 37.99 \text{ kJ/mol}$) [46], $U$ the intermolecular potential energy, $p$ the pressure and $\Delta V$ the molar volume change between liquid and gas ($p\Delta V$ is essentially equal to $pV_{gas} = RT = 2.48 \text{ kJ/mol at 298 K}$). The quantum
correction term $Q_{int}$ takes into consideration the difference in intra-molecular vibrational energy between the liquid state and the gas phase and is approximately equal to 1.69 kJ/mol [33]. The second quantum correction term $Q_{ext}$ applies only to liquids and depends on the active intermolecular vibrational modes. This term was taken equal to -1.81 kJ/mol [33].

The methanol O-H bond length corresponded to the SPC water O-H distance [33], while the starting O-Me bond lengths and the bond angle were based on structures obtained from microwave spectra [50]. The model M1a derived by Lauterbach [51] was chosen as a starting point for the parameterisation procedure. Its parameters were derived from the GROMOS87 charge values and Lennard-Jones parameters for alcoholic oxygen (O), united atom CH$_3$ group (Me) and hydrogen (H) [52]. The GROMOS force field does not contain a special hydrogen bonding term but mimics the effect of the hydrogen bonding properties of a polar atom such as an alcoholic oxygen by using different values for the van der Waals repulsion depending on the type of interaction. For non-hydrogen bonding pairs such as $C_12$(Me,O) the interaction is considered to be the geometrical mean of $C_1^1$(O,O) and $C_12$(Me,Me). For pairs of atoms involved in hydrogen bonds a second, larger value $C_2^1$(O,O) is used. The neglect of polarisability is counteracted by increasing the dipole above the experimental value in the gas phase. All models presented in Table 2.1 have a dipole between the experimental dipole in gas phase [48] of 1.7 D and the experimental dipole in liquid phase [53] of 2.9 D. For parameterisation the charge of

<table>
<thead>
<tr>
<th>Model name</th>
<th>$q_H$ (e)</th>
<th>$q_O$ (e)</th>
<th>$q_{Me}$ (e)</th>
<th>$d_{OH}$ (nm)</th>
<th>$d_{MeO}$ (nm)</th>
<th>$d_{HMe}$ (nm)</th>
<th>$\mu$ (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GR96 [4, 54]</td>
<td>0.398</td>
<td>-0.574</td>
<td>0.176</td>
<td>0.1000</td>
<td>0.1430</td>
<td>0.1988</td>
<td>1.91</td>
</tr>
<tr>
<td>H1 [28, 29]</td>
<td>0.431</td>
<td>-0.728</td>
<td>0.297</td>
<td>0.0945</td>
<td>0.1425</td>
<td>0.1944</td>
<td>2.33</td>
</tr>
<tr>
<td>J2 [27]</td>
<td>0.435</td>
<td>-0.700</td>
<td>0.265</td>
<td>0.0945</td>
<td>0.1430</td>
<td>0.1948</td>
<td>2.22</td>
</tr>
<tr>
<td>L1 [30]</td>
<td>0.435</td>
<td>-0.700</td>
<td>0.265</td>
<td>0.0945</td>
<td>0.1424</td>
<td>0.1944</td>
<td>2.22</td>
</tr>
<tr>
<td>M1a [51]</td>
<td>0.398</td>
<td>-0.574</td>
<td>0.176</td>
<td>0.1000</td>
<td>0.1430</td>
<td>0.1988</td>
<td>1.91</td>
</tr>
<tr>
<td>A1</td>
<td>0.408</td>
<td>-0.634</td>
<td>0.226</td>
<td>0.1000</td>
<td>0.1430</td>
<td>0.1988</td>
<td>2.08</td>
</tr>
<tr>
<td>A2</td>
<td>0.408</td>
<td>-0.654</td>
<td>0.246</td>
<td>0.1000</td>
<td>0.1430</td>
<td>0.1988</td>
<td>2.15</td>
</tr>
<tr>
<td>A3</td>
<td>0.408</td>
<td>-0.674</td>
<td>0.266</td>
<td>0.1000</td>
<td>0.1430</td>
<td>0.1988</td>
<td>2.22</td>
</tr>
<tr>
<td>B3</td>
<td>0.408</td>
<td>-0.674</td>
<td>0.266</td>
<td>0.1000</td>
<td>0.1530</td>
<td>0.2077</td>
<td>2.29</td>
</tr>
</tbody>
</table>

Table 2.1: Geometries and charges of the presented models. Here $q$ denotes the partial charges, $d$ the distances between atoms of the molecule and $\mu$ the dipole of the molecule.

the hydrogen atom $q_H$ was increased to 0.408 e, then the repulsive Lennard-Jones parameters $C_{12}$ as well as the partial charge of the oxygen $q_O$ were adjusted, so that the potential energy and the pressure during the MD simulations approached their target values of -35.6 kJ/mol, calculated from the heat of vaporisation using Equation (2.3), and 1 atm respectively. The effect of increasing the dipole by extending the bond between the O- and the Me-atom $d_{OMe}$ was also investigated. The attractive Lennard-Jones parameters were set to the GROMOS96 values [4, 54], and $q_{Me}$ was set equal to $-q_O - q_H$ in order to preserve electro neutrality.

The parameterisation simulations were 300 ps in length of which 50 ps was for equilibration. No reaction-field correction was applied and the simulations were performed at constant volume.
2.3 Method

2.3.2 Properties

To study the influence of different force-field parameters on the properties of liquid methanol, 4 different models were developed and compared. They are named A1, A2, A3, and B3, where the number refers to the charge distribution and the letter to the bond length of $d_{\text{MeMe}}$. These models are also compared to other models taken from the literature. The literature models are referred to as GR96, for the MeOH model of the GROMOS96 force field [4], H1 for the model of Haughney et al. [28, 29], J2 for the OPLS model of Jörgensen [27] and L1 for the model of van Leeuwen and Smit [30]. The geometries and charge distributions of these models are summarised in Table 2.1 and the Lennard-Jones interaction parameters in Table 2.2.

For each model studied, a 1 ns MD simulation with reaction-field correction at 298 K was performed. The last 900 ps of these simulations were used for analysis. Additional simulations with a different setup were performed where necessary.

Self-diffusion coefficient $D$

The diffusion coefficient was calculated from the mean-square displacement of the particles using the Einstein relation:

$$D = \lim_{t \to \infty} \frac{\left\langle (r(t) - r(0))^2 \right\rangle}{6t}$$

where $r(t)$ denotes the position vector of a molecule at time $t$.

Shear viscosity $\eta$

The viscosity was calculated as described by Tironi and van Gunsteren [55]. The off-diagonal elements $P_{xy}$, $P_{xz}$, $P_{yz}$ of the pressure tensor [56] are given by:

$$P_{\alpha\beta}(t) = \frac{1}{V} \left( \sum_i p_{\alpha i}(t) p_{\beta i}(t) \frac{m_i}{m_i} + \sum_{i<j} F_{\alpha ij}(t) b_{ij}(t) \right)$$

where $\alpha$ and $\beta$ denote $x$-, $y$- or $z$-components of the inter-particle vector $r_{ij} \equiv r_i - r_j$, $p_{\alpha i}$ the $\alpha$-component of the momentum of particle $i$ and $F_{\alpha ij}$ the $\alpha$-component of the force exerted on particle $i$ by particle $j$. The shear viscosity $\eta$ has been calculated from the 'displacement' of $P_{\alpha\beta}$:

$$\Delta P_{\alpha\beta}(t) = \int_0^t P_{\alpha\beta}(t') dt'$$

via the Einstein relation

$$\eta = \frac{1}{2k_B T} \lim_{t \to \infty} \frac{d}{dt} \langle \Delta P_{\alpha\beta}^2(t) \rangle$$

Because of poor statistics at long times only $\langle \Delta P_{\alpha\beta}^2(t) \rangle$ between 5 ps and 10 ps was used for analysis.
<table>
<thead>
<tr>
<th>Model name</th>
<th>((C_{12}(O))^{1/2})</th>
<th>((C_{12}^2(O))^{1/2})</th>
<th>((C_{12}(\text{Me}))^{1/2})</th>
<th>((C_6(O))^{1/2})</th>
<th>((C_6^2(O))^{1/2})</th>
<th>((C_6(\text{Me}))^{1/2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>GR96 [4, 54]</td>
<td>1.1250</td>
<td>1.227</td>
<td>4.5665</td>
<td>0.0476</td>
<td>0.0476</td>
<td>0.0942</td>
</tr>
<tr>
<td>H1 [28, 29]</td>
<td>1.5839</td>
<td>1.4683</td>
<td>5.7685</td>
<td>0.0521</td>
<td>0.0501</td>
<td>0.1002</td>
</tr>
<tr>
<td>J2 [27]</td>
<td>1.4124</td>
<td>1.4124</td>
<td>5.3864</td>
<td>0.0488</td>
<td>0.0488</td>
<td>0.1001</td>
</tr>
<tr>
<td>L1 [30]</td>
<td>1.3126</td>
<td>1.3126</td>
<td>5.1191</td>
<td>0.0472</td>
<td>0.0472</td>
<td>0.0979</td>
</tr>
<tr>
<td>M1a [51]</td>
<td>1.1247</td>
<td>1.2272</td>
<td>4.5665</td>
<td>0.0476</td>
<td>0.0476</td>
<td>0.0942</td>
</tr>
<tr>
<td>A1</td>
<td>1.3250</td>
<td>1.3970</td>
<td>4.5665</td>
<td>0.0476</td>
<td>0.0476</td>
<td>0.0942</td>
</tr>
<tr>
<td>A2</td>
<td>1.3250</td>
<td>1.4450</td>
<td>4.5665</td>
<td>0.0476</td>
<td>0.0476</td>
<td>0.0942</td>
</tr>
<tr>
<td>A3</td>
<td>1.5250</td>
<td>1.4570</td>
<td>4.5665</td>
<td>0.0476</td>
<td>0.0476</td>
<td>0.0942</td>
</tr>
<tr>
<td>B3</td>
<td>1.5250</td>
<td>1.5250</td>
<td>5.6812</td>
<td>0.0476</td>
<td>0.0476</td>
<td>0.0942</td>
</tr>
</tbody>
</table>

Table 2.2: The Lennard-Jones interaction parameters for the different methanol models. \((C_{6,12}(I))^{1/2}\) are the single atom van der Waals parameters for atom I. \((C^{1/2}_{12}(O))^{1/2}\) is used for combinations with Me, while \((C^{1/2}_{12}(O))^{1/2}\) is used for interactions with O-atoms.
2.3. Method

**Dielectric constant \( \varepsilon_S \)**

For most purposes the frequency-dependent relative (i.e. with respect to the vacuum value \( \varepsilon_0 \)) dielectric permittivity \( \varepsilon(\omega) \) can be modelled by the Debye equation

\[
\varepsilon(\omega) = \varepsilon(\infty) + \frac{\varepsilon(0) - \varepsilon(\infty)}{(1 + i\omega\tau_D)^p}
\]  

(2.8)

where \( \tau_D \) is the Debye relaxation time defined in (2.11) below, and \( p \) adopts values close to one, e.g. \( p \approx 0.986 \) for water. The infinite-frequency value \( \varepsilon_\infty = \varepsilon(\infty) \) is due to displacements of electrons and the zero-frequency or static value \( \varepsilon_S = \varepsilon(0) \) is due to dipolar orientation. The relative static dielectric constant for methanol was calculated from the fluctuations of the total dipole moment \( \mathbf{M} \) of the system, using the formula of Neumann [57]:

\[
(\varepsilon_S - 1) \frac{2\varepsilon_{RF} + 1}{2\varepsilon_{RF} + \varepsilon_S} = \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\varepsilon_0 V k_B T}
\]  

(2.9)

where \( \varepsilon_{RF} \) is the relative dielectric constant of the continuum employed in the reaction-field term in Equation (2.2). To avoid the suppression of the fluctuations in the total dipole moment caused by the application of an electrostatic interaction cut off, all simulations used to determine the dielectric constant were performed using a reaction field.

**Debye relaxation time \( \tau_D \)**

The Debye relaxation time \( \tau_D \) was calculated from the autocorrelation function \( \Phi(t) \) of the total dipole moment \( \mathbf{M} \) of the box which can be fitted to an exponential decay:

\[
\Phi(t) = \frac{\langle \mathbf{M}(t)\mathbf{M}(0) \rangle}{\langle \mathbf{M}^2(0) \rangle} = \frac{t}{\tau_D}
\]  

(2.10)

via [47]:

\[
\tau_D = \frac{2\varepsilon_{RF} + \varepsilon_S}{2\varepsilon_{RF} + 1} \tau_\Phi
\]  

(2.11)

**Isothermal compressibility \( \kappa_T \)**

The isothermal compressibility can be estimated from the following expression [58], relating two state points 1 and 2:

\[
\kappa_T = -\frac{1}{V} \left( \frac{\partial V}{\partial p} \right)_T = \frac{1}{\rho} \left( \frac{\partial \rho}{\partial p} \right)_T = \left( \frac{\partial \ln(\rho)}{\partial p} \right)_T \approx \left( \frac{\ln \left( \frac{p_2}{p_1} \right)}{p_2 - p_1} \right)_T
\]  

(2.12)

where \( \rho \) is the density of the system.

To calculate the compressibility, an additional simulation with a decreased volume of 33.233 nm\(^3\) was performed.
Thermal expansion coefficient $\alpha$

A similar finite difference expression for the thermal expansion coefficient is [55]

$$ \alpha = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_p \approx - \left( \frac{\ln \left( \frac{p_2}{p_1} \right)}{T_2 - T_1} \right)_p $$  \hspace{1cm} (2.13)

To calculate $\alpha$ two simulations were performed at constant pressure, one at 298 K and one at 328 K.

Heat capacity $C_p$

The constant volume heat capacity was obtained using the following equation [55]:

$$ C_V = \left( \frac{\partial E}{\partial T} \right)_V \approx \frac{U_2 - U_1}{T_2 - T_1} + 3R + C_{\text{vib}}^V $$  \hspace{1cm} (2.14)

where $E$ is the total and $U$ the potential energy. $C_{\text{vib}}^V$ is a correction term for the vibrational contribution to the heat capacity when comparing the rigid methanol model of the simulation to experiment. It can be estimated from the partition function for a harmonic quantum-mechanical oscillator using experimental normal mode frequencies [59]. Using values from Herzberg [59] and Wilson [60], it was calculated as 5.749 J mol$^{-1}$ K$^{-1}$. The constant pressure heat capacity $C_p$ can then be calculated from $C_V$ using the following relation [61]:

$$ C_p - C_V = T \frac{v \alpha^2}{\kappa_T} $$  \hspace{1cm} (2.15)

where $v$ is the molar volume, $\alpha$ the thermal expansion coefficient and $\kappa_T$ the isothermal compressibility.

Binary mixtures

A certain number out of 1000 methanol molecules was replaced by SPC water molecules [17] to obtain binary mixtures with the desired mole fraction of methanol $x_{\text{MeOH}}$. Simulations with $x_{\text{MeOH}}$ of 0.000, 0.125, 0.250, 0.375, 0.500, 0.625, 0.750, 0.875 and 1.000 were performed at constant pressure without a reaction-field correction. Each mixture was simulated for 1 ns, only the last 900 ps were used for analysis.

The molar enthalpy of mixing was calculated from the potential energy via:

$$ \Delta H_{\text{mix}} = U(\text{mix}) - x_{\text{MeOH}}U(\text{MeOH}) - (1 - x_{\text{MeOH}})U(\text{H}_2\text{O}) $$  \hspace{1cm} (2.16)

where $U(\text{mix})$ is the potential energy of the mixture and $U(\text{MeOH})$ and $U(\text{H}_2\text{O})$ the potential energy of pure methanol and pure water, respectively.
2.4. Effect of the parameters

Free energy

The free energy differences between two states A and B can be expressed as

$$\Delta F_{BA} = \int_{\lambda_0}^{\lambda_B} F'(\lambda) d\lambda = \int_{\lambda_0}^{\lambda_B} \langle \frac{\partial H}{\partial \lambda} \rangle d\lambda$$

(2.17)

where the Hamiltonian $H$ has been made dependent on a coupling parameter $\lambda$ and the angular brackets $\langle \rangle$ denote averaging over an equilibrium ensemble generated with $H(\lambda)$. The integral in Equation (2.17) was evaluated by obtaining ensemble averages at 25 discrete $\lambda$-points and determining the integral numerically. At each $\lambda$-point 20 ps equilibration and 50 ps sampling were performed. To avoid numerical instabilities as atoms were created or deleted a soft-core interaction function [62, 63] was used as described by Daura et al. [64]

The excess Helmholtz energy was obtained by changing the liquid state ($\lambda = 0$) to the gas state ($\lambda = 1$) by switching off the non-bonded interactions in a NVT simulation [55, 65]. It is compared to an experimental value of $17.9 \text{ kJ mol}^{-1} \text{ K}^{-1}$ calculated by:

$$\Delta A \approx RT \left( \ln \left( \frac{RT}{P_{vap} v} \right) - 1 \right)$$

(2.18)

where $P_{vap}$ is the vapour pressure at temperature $T$ and $v$ the molar volume of the solvent. For methanol at 298 K $P_{vap}$ is [46] $0.164 \times 10^{-25} \text{ kJ/nm}^3$.

The hydration free energy was calculated by growing a methanol molecule in a box of 999 SPC water molecules. This was done by turning on the non-bonded interactions of this molecule to the water molecules. The pressure was kept constant.

2.4 Effect of the parameters

The influence of different force-field parameters and simulation conditions on the simulated properties was investigated. The force-field parameters investigated include the charge distribution and the geometry of the methanol model. By adjusting the Lennard-Jones parameters it was possible to get agreement with the experimental values for the pressure and the potential energy for all tested models. That the models reproduce pressure and potential energy does not necessarily mean that other properties are also well reproduced. Especially the dielectric properties may deviate from the experimental value. We therefore calculated the dielectric constant and the Debye relaxation time for the different models, to examine how they depend on the geometry and charge distribution and if there exists a simple relationship between these parameters and the calculated results. Not only the model parameters, but also the exact simulation conditions influence the results of the calculation of the properties. Therefore the influence of simulating at constant pressure versus constant volume as well as different treatment of the long-range interaction forces, i.e. long-range cut off and inclusion of a reaction-field force was also examined.

2.4.1 Charge distribution

The results of increasing the dipole moment by changing the charge distribution are summarised in Table 2.3. The target value for the potential energy $U$ was reached, while the value for the
Table 2.3: The effect of increasing the dipole moment by changing the charge distribution is shown. Here \( p \) denotes the pressure, \( U \) the potential energy, \( D \) the self-diffusion coefficient, \( \varepsilon_S \) the static permittivity and \( \tau_D \) the Debye relaxation time. The other parameters are the same as in Table 2.1. The experimental values are at 298 K and 1 atm, except \( \varepsilon_S \), which is at 293 K.

<table>
<thead>
<tr>
<th>Model name</th>
<th>( q_H ) (e)</th>
<th>( q_O ) (e)</th>
<th>( q_{Me} ) (e)</th>
<th>( \mu ) (D)</th>
<th>( p ) (kJ/mol nm(^2))</th>
<th>( U ) (kJ/mol)</th>
<th>( D ) (10(^{-3}) nm(^2)/ps)</th>
<th>( \varepsilon_S )</th>
<th>( \tau_D ) (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.408</td>
<td>-0.634</td>
<td>0.226</td>
<td>2.08</td>
<td>-8.060</td>
<td>-35.5</td>
<td>2.9</td>
<td>21.3</td>
<td>33</td>
</tr>
<tr>
<td>A2</td>
<td>0.408</td>
<td>-0.654</td>
<td>0.246</td>
<td>2.15</td>
<td>-11.035</td>
<td>-35.5</td>
<td>2.9</td>
<td>24.3</td>
<td>32</td>
</tr>
<tr>
<td>A3</td>
<td>0.408</td>
<td>-0.674</td>
<td>0.266</td>
<td>2.22</td>
<td>0.283</td>
<td>-35.5</td>
<td>2.6</td>
<td>20.2</td>
<td>25</td>
</tr>
<tr>
<td>exp</td>
<td>1.7[48]</td>
<td>0.061</td>
<td>-35.6</td>
<td></td>
<td></td>
<td></td>
<td>2.4[66] 32.6[48] 49 [67]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

pressure \( p \), which is more sensitive to the exact simulation conditions, differs more from the experimental value. By gradually increasing the dipole moment \( \mu \) of the methanol model a slight decrease of the diffusion constant \( D \) was obtained, but it seems to be rather insensitive to the dipole moment of the molecule. The permittivity \( \varepsilon_S \) did not show a consistent trend and differed in all models by about 50% from the experimental value. The Debye relaxation time decreased as function of the dipole moment.

### 2.4.2 Bond length

Since increasing the dipole moment by changing the charge distribution did not result in a permittivity closer to the experimental value, the dipole moment was increased by increasing the Me-O bond. An increase of the bond length may be justified when it is considered that the centre of mass of the Me group lies not on the C-atom but is shifted 0.0073 nm outwards. This is a very small length increase, which does not increase the simulated permittivity significantly. A larger but still small bond-length increase of 0.036 nm is obtained by considering the centre of geometry of the H-atoms of the Me-group. However, to obtain a significant increase of the simulated permittivity, the Me-O bond was increased by 0.1 nm to the value of a CH\(_n\)-CH\(_n\) bond. We note that when using a united-atom representation, *i.e.* implicit aliphatic hydrogen atoms, model parameters such as the united-atom bond lengths, bond angles and charge distribution are non-physical quantities, whose values cannot be expected to have a physical meaning and should therefore not be compared to experimental values. The effect of changing the Me-O bond length is shown in Table 2.4. Though the permittivity is increased it still remains far from the experimental value. The diffusion is not affected by the change, while the Debye relaxation time shortens.

Comparing these results to that of two water models SPC [17] and SPC/E [15], where an increase in the dipole moment of only 3% from 2.28 D to 2.35 D leads to an increase in the permittivity of 15% from 54.0 to 62.3, [68] one sees that increasing the dipole moment is much less effective for the methanol models. The dipole increases from model A1 to B3 by 10%, but the permittivity increases only by 6%. The most likely reason for this is related to the fact that we are considering only non-polarisable models. M can be considered to be composed of
2.4. Effect of the parameters

<table>
<thead>
<tr>
<th>Model name</th>
<th>$d_{\text{OH}}$ (nm)</th>
<th>$d_{\text{MeO}}$ (nm)</th>
<th>$d_{\text{HMe}}$ (nm)</th>
<th>$\mu$ (D)</th>
<th>$p$ (kJ mol$^{-1}$)</th>
<th>$U$ (kJ mol$^{-1}$)</th>
<th>$D$ ($10^{-3}$ nm$^2$ ps)</th>
<th>$\varepsilon_S$</th>
<th>$\tau_D$ (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A3</td>
<td>0.100</td>
<td>0.143</td>
<td>0.199</td>
<td>2.22</td>
<td>0.283</td>
<td>-35.5</td>
<td>2.6</td>
<td>20.2</td>
<td>25</td>
</tr>
<tr>
<td>B3</td>
<td>0.100</td>
<td>0.153</td>
<td>0.208</td>
<td>2.29</td>
<td>0.523</td>
<td>-35.3</td>
<td>2.6</td>
<td>22.7</td>
<td>17</td>
</tr>
<tr>
<td>exp</td>
<td>1.7[48]</td>
<td>0.061</td>
<td>-35.6[48]</td>
<td>2.4[66]</td>
<td>32.6[48]</td>
<td>49[67]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4: The effect of increasing the dipole by enlarging the Me-O bond length. The symbols are the same as in Tables 2.1 and 2.3.

two contributions $M_I$ and $M_h$, referring to low and high frequency components respectively. For dipolar substances $M_h$ is mainly due to the displacement of the electrons and $M_I$ to the orientation of the dipoles. For non-polarisable models $M$ is in effect assumed equal to $M_I$. $M_I$ can be expressed as: [69]

$$
\frac{M_I^2}{3\varepsilon_0 V k_B T} = \frac{(\varepsilon_S - \varepsilon_\infty)(2\varepsilon_{RF} + 1)^2}{(2\varepsilon_{RF} + \varepsilon_\infty)(2\varepsilon_{RF} + \varepsilon_S)}
$$

(2.19)

thus only if $\varepsilon_\infty = 1$ is $M_I$ truly equal to $M$. The greater $\varepsilon_\infty$ becomes in comparison to $\varepsilon_S$ the more will $M_I$ differ from $M$. For water $\varepsilon_\infty$ and $\varepsilon_S$ are [48] 5.2 and 78.4 respectively, while for methanol the experimental values are [48, 67] 5.38 and 32.63 respectively. Therefore, it is to be expected that the permittivity of water can be better reproduced by an unpolarisable model than is the case for methanol.

We note that since the GROMOS96 force field [4] does not include explicit polarisability and has been parameterised without taking into account the self-polarisation energy of dipoles, it is more consistent with the SPC water model [17], which has also omitted the self-polarisation term, than with the SPC/E water model [15], which has been parameterised including self-polarisation. Since our goal is to derive a simple rigid three-site model for methanol compatible with the SPC water model and the GROMOS96 force field, we did not include a self-polarisation energy term when fitting the energy of liquid methanol to the experimental heat of vaporisation.

2.4.3 Treatment of long-range forces

The effects of different treatment of the long-range forces are summarised in Table 2.5. Two different aspects are considered, one is the inclusion of a long-range cut off, and the other the use of a reaction-field contribution to the forces due to the dielectric continuum beyond the outer cut off. The inclusion of a reaction-field contribution changes the pressure by a factor of about 10 if it is applied outside a long-range cut off $R_{cl}$ of 1.4 nm (model B3) and even more if only a single cut off $R_c$ of 0.9 nm is used (model M1a). Inclusion of long-range forces between $R_c = 0.9$ nm and $R_{cl} = 1.4$ nm also affects the pressure (model H1). The potential energy is hardly affected by the addition of a reaction-field force. The change is greater but still small when a long-range cut off is used. The permittivity is not very sensitive to the long-range cut off used.
2. The Effect of Force-field Parameters

<table>
<thead>
<tr>
<th>Model name</th>
<th>$R_{el}$</th>
<th>Reaction field</th>
<th>$p$</th>
<th>$U$</th>
<th>$D$</th>
<th>$\varepsilon_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3</td>
<td>1.4</td>
<td>no RF</td>
<td>0.065</td>
<td>$-35.2$</td>
<td>2.4</td>
<td>–</td>
</tr>
<tr>
<td>B3</td>
<td>1.4</td>
<td>RF</td>
<td>0.523</td>
<td>$-35.3$</td>
<td>2.6</td>
<td>22.7</td>
</tr>
<tr>
<td>H1 [28, 29]</td>
<td>0.9</td>
<td>RF</td>
<td>61.038</td>
<td>$-34.7$</td>
<td>20.6</td>
<td>–</td>
</tr>
<tr>
<td>H1 [28, 29]</td>
<td>1.4</td>
<td>RF</td>
<td>86.619</td>
<td>$-34.0$</td>
<td>20.4</td>
<td>–</td>
</tr>
<tr>
<td>M1a†</td>
<td>0.9</td>
<td>no RF</td>
<td>0.122</td>
<td>$-35.5$</td>
<td>3.1</td>
<td>–</td>
</tr>
<tr>
<td>M1a†</td>
<td>0.9</td>
<td>RF</td>
<td>4.393</td>
<td>$-35.3$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>M1a†</td>
<td>1.5</td>
<td>no RF</td>
<td>$-21.356$</td>
<td>$-36.2$</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

* In all simulations $R_c = 0.9$ nm
† The results for model M1a are taken from Lauterbach [51]

Table 2.5: The effect of different treatment of the long-range interactions. $R_{el}$ is the long-range cut off, RF and no RF mean that the simulations have been done with and without a reaction-field contribution to the forces respectively.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Simulation time (ns)</th>
<th>$\rho$</th>
<th>$U$</th>
<th>$D$</th>
<th>$\varepsilon_S$</th>
<th>$\tau_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>1</td>
<td>NVT</td>
<td>$-11.035$</td>
<td>473.70</td>
<td>$-35.496$</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPT</td>
<td>0.051</td>
<td>486.35</td>
<td>$-35.877$</td>
<td>2.4</td>
</tr>
<tr>
<td>B3</td>
<td>1</td>
<td>NVT</td>
<td>0.523</td>
<td>473.70</td>
<td>$-35.349$</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPT</td>
<td>0.051</td>
<td>473.07</td>
<td>$-35.351$</td>
<td>2.3</td>
</tr>
<tr>
<td>B3</td>
<td>2</td>
<td>NVT</td>
<td>0.907</td>
<td>473.70</td>
<td>$-35.348$</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPT</td>
<td>0.048</td>
<td>473.23</td>
<td>$-35.350$</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Table 2.6: Simulating at constant pressure or volume may influence the calculated dielectric permittivity. In the table $\rho$ denotes the density of the system, the other symbols are the same as in Tables 2.1 and 2.3.

2.4.4 Constant volume versus constant pressure

Analysing the different simulations we found differences in the permittivity between the simulations done at constant pressure and those at constant volume. The results for the two models simulated at both conditions are summarised in Table 2.6. While it might be possible that the differences for model A2 arise from the pressure difference between the two simulations, this is not the case for model B3, where the pressure differs only slightly. Another possibility is that the permittivity has not yet converged. The two runs of B3 were therefore continued for a further 1 ns. While the permittivity obtained from the constant pressure run approaches the result of the constant volume run, the difference is still considerable. Given a sufficiently large system the two ensembles are equivalent. Since only the value of the simulation at constant pressure changed when increasing the simulation length, the permittivity probably is converged after 1 ns.
when simulated at constant volume but takes longer to converge when simulating at constant pressure.

### 2.5 Properties of model B3

The model B3 showed the best overall agreement with experimental values. Therefore, it was evaluated further and compared to other models from the literature (See Table 2.7). The model was parameterised to reproduce the pressure and potential energy and therefore the calculated values are close to the experimental values. All the other models show large deviations in the pressure. The L1 model comes closest to the experimental value, while the H1 model deviates the most. This is consistent with the results of Haughney et al. [29] who reported a pressure of about 47 kJ mol⁻¹ nm⁻³ at 299.3 K. In the simulations at constant pressure the density of the model J2 differs from the experimental value by almost 5%. L1 having a too high value for the pressure is in contrast with the results of Mountain [70], who obtained a negative pressure when simulating at experimental density at 288 K. It is important to note that the calculated pressure is very sensitive to the method of simulation, the size of the system and the effective cut-off radius [33] (See also Table 2.5). The model J2 was parameterised using a correction term for the Lennard-Jones interactions neglected beyond the cut-off distance of 0.95 nm. In contrast, Haughney et al. [29] used Ewald summation for the electrostatic interactions when developing the model H1 (and H2) and comparing to Jorgensen’s model J2 (and J1). Such differences in methodology most likely explain the discrepancies between the values shown in Table 2.7 and the equivalent values published by Haughney et al. [29] or by Stouten [33]. Each model would be expected to perform best under the conditions it was parameterised. However, as the current aim is to develop a methanol model compatible with the SPC water model [17] and protein force fields [4, 54], we restrict ourselves to treating the long-range interactions by applying cut offs and reaction-field contributions.

<table>
<thead>
<tr>
<th>Model name</th>
<th>$\mu$ (D)</th>
<th>$p$ (kJ mol⁻¹ nm⁻³)</th>
<th>$\rho$ (g cm⁻³)</th>
<th>$U$ (kJ mol⁻¹)</th>
<th>$D$ (10⁻³ nm² ps⁻¹)</th>
<th>$\varepsilon$</th>
<th>$\tau_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVT GR96</td>
<td>1.9</td>
<td>-21.872</td>
<td>473.7</td>
<td>-36.1</td>
<td>3.1</td>
<td>17.7</td>
<td>19</td>
</tr>
<tr>
<td>H1 [28, 29]</td>
<td>2.33</td>
<td>61.038</td>
<td>473.7</td>
<td>-34.7</td>
<td>2.2</td>
<td>20.6</td>
<td>16</td>
</tr>
<tr>
<td>L1 [30]</td>
<td>2.22</td>
<td>2.203</td>
<td>473.7</td>
<td>-37.6</td>
<td>1.9</td>
<td>18.1</td>
<td>33</td>
</tr>
<tr>
<td>B3</td>
<td>2.29</td>
<td>0.523</td>
<td>473.7</td>
<td>-35.3</td>
<td>2.6</td>
<td>22.7</td>
<td>17</td>
</tr>
<tr>
<td>NPT J2 [27]</td>
<td>2.22</td>
<td>0.094</td>
<td>453.7</td>
<td>-35.3</td>
<td>2.6</td>
<td>21.0</td>
<td>18</td>
</tr>
<tr>
<td>B3</td>
<td>2.29</td>
<td>0.05</td>
<td>473.1</td>
<td>-35.4</td>
<td>2.3</td>
<td>18.6</td>
<td>14</td>
</tr>
</tbody>
</table>

**Table 2.7:** Comparison to other models from the literature. The model names are the same as used in the text.
2. The Effect of Force-field Parameters

<table>
<thead>
<tr>
<th></th>
<th>298 K</th>
<th>328 K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p ) (kJ mol(^{-1}) nm(^{-3}))</td>
<td>( p ) (kJ mol(^{-1}) nm(^{-3}))</td>
</tr>
<tr>
<td>NVT</td>
<td>0.907</td>
<td>-0.035</td>
</tr>
<tr>
<td>NPT</td>
<td>0.048</td>
<td>0.088</td>
</tr>
<tr>
<td>exp</td>
<td>0.061</td>
<td>0.061</td>
</tr>
</tbody>
</table>

*interpolated from values [46] at 323 K and at 333 K

Table 2.8: Properties of model B3 obtained at different temperatures.

2.5.1 Transport properties

The diffusion coefficient for model B3 is too high compared to the experimental value. In contrast, the diffusion coefficient of model H1 is of about the same difference too low, possibly as a result of the too high pressure. The result for model L1 is even lower than for model H1 and agrees with the result of Mountain [70], who got \(1.8 \times 10^{-3}\) nm\(^2\)/ps at a lower temperature of 288 K. The values for models H1 and J2 are higher than those published by Haughney et al. [29] or Asahi and Nakamura [71]. Haughney et al. pointed out in their paper that a large diffusion coefficient might be expected for a model that underestimates the strength of hydrogen bonding in the liquid. Large diffusion coefficients are also commonly found for united atom models. The diffusion constant increases with increasing temperature as shown in Table 2.8. This is in agreement with experiment, but the rise is too steep compared to experiment.

The value of the shear viscosity obtained for model B3 was 236 kJ mol\(^{-1}\) nm\(^{-3}\) ps, which is lower than the experimental value [72] of 326 kJ mol\(^{-1}\) nm\(^{-3}\) ps at 298 K. Mountain [70] and Wheeler and Rowley [45] both calculated the viscosity of model L1 and obtained values that were higher than the experimental values, 391 kJ mol\(^{-1}\) nm\(^{-3}\) ps and 346.65 kJ mol\(^{-1}\) nm\(^{-3}\) ps respectively.

2.5.2 Dielectric properties

The dielectric constant is smaller than the experimental value for all models, with model B3 having the highest and model GR96 the lowest value. The underestimation of the dielectric constant is most likely due to the neglect of the electronic polarisability. With increasing temperature (see Table 2.8) the dielectric constant of model B3 increases, although the experimental value decreases. However, the simulated increase of \(\varepsilon_S\) with temperature may not be significant: the values obtained from the 1900 ps trajectories have an estimated error of about 7%. This size of convergence error for \(\varepsilon_S\) was found when using 1900 ps averaging in a recent MD simulation study of liquid water extending over several nanoseconds (A. Glättli, private communication). Secondly, the contributions of the electronic and orientational polarisation to \(\varepsilon_S\) may be different at different temperatures. Since the former contributions are neglected in our model, the temperature dependence of \(\varepsilon_S\) may be incorrectly modelled.

The differences between the simulation at constant pressure and the one at constant volume
decreases with rising temperature, though the result from the constant pressure run is still lower than from the one at constant volume. The value for model H1 is lower than those of Richardi et al. [73] and of Fonseca and Ladanyi [74], who obtained 25.3 and 24 respectively. The value determined for model L1 is lower than the value given by Mountain [70], 22. The Debye relaxation time is much lower than the experimental value for all models, with model L1 coming closest to it.

2.5.3 Thermodynamic properties

The calculated isothermal compressibility for model B3 is 1.6·10⁻³ kJ⁻¹ mol nm³, lower than the experimental value [48] of 2.0·10⁻³ kJ⁻¹ mol nm³. A too low compressibility shows that the liquid is too resistant to compression when put under pressure. This may be a consequence of the chosen representation for the non-bonded interaction. The r₁² term in the Lennard-Jones function is known to cause too sharp an increase of energy at small atom-atom distances [55]. The calculated heat capacity is 95.44 J mol⁻¹ K⁻¹ to be compared to the experimental value [75] of 81.47 J mol⁻¹ K⁻¹.

Two simulations at a higher temperature (328 K), one with constant pressure and one with constant volume were performed. The simulation at constant volume is not at the experimental density but at the density that gives a pressure close to 1 atm. The calculated permittivity is not distorted by pressure effects. The results are shown in Table 2.8. From the simulations at constant pressure the thermal expansion coefficient α can be calculated via Equation (2.13). In this way an α of 1.50·10⁻³ K⁻¹ was obtained, which is close to the experimental value [48] of 1.49·10⁻³ K⁻¹.

2.5.4 Free energy

Both, the excess Helmholtz energy and the hydration free energy of model B3 are close to the experimental value. The excess Helmholtz energy was found to be equal to 17.0 kJ/mol. This may be compared to an experimental value, calculated by Eq. (2.18) of 17.8 kJ/mol. The hydration free energy was calculated as -21.4 kJ/mol which agrees very well with the experimental value of -21.4 kJ/mol [76].

2.5.5 Mixing with water

The distribution of methanol and water molecules in the mixtures is homogeneous, as is illustrated in Figure 2.1. For xₘₐₒₜ = 0.5, the first peak in the oxygen-oxygen radial distribution function for water-methanol pairs is as large as the ones for the water-water and methanol-methanol pairs.

The results for the mixing enthalpy, calculated by Eq. (2.16), are shown in Figure 2.2 together with the experimental values [77]. A negative mixing enthalpy was obtained over the whole range of mixtures. The experimental values are slightly more negative and the largest deviation from ideal mixing behaviour occurs at a methanol fraction near 0.3, whereas the largest deviation from ideality for B3 methanol mixed with SPC water occurs at xₘₐₒₜ = 0.5.

Ferrario et al. [42] reported an excess enthalpy of mixing of -1.3 kJ/mol for H1 methanol together with TIP4P water [18] (NVT) which they compared to an experimental value of
Figure 2.1: Oxygen-oxygen radial distribution functions for a $x_{\text{MeOH}} = 0.5$ mixture of methanol and water.

Figure 2.2: The mixing enthalpy at 298 K is calculated with Equation (2.16). The experimental values are taken from Landolt-Börnstein [77].
-0.9 kJ/mol. Stouten [41] obtained a value of -0.767 kJ/mol for an equimolar mixture of the flexible OM2 methanol model and SPC water (NPT), compared with the experimental value of -0.8 kJ/mol at 297.21 K. Freitas [43] performed Monte Carlo simulations of J2 methanol with TIP4P water for methanol mole fractions equal to 0.10, 0.25, 0.50 and 0.75 in the NVT and NPT ensembles. The values at NPT were systematically too high (e.g. \( \Delta H_{\text{mix}} = -0.574 \text{ kJ/mol at} \ x_{\text{MeOH}} = 0.25 \)). The enthalpies of mixing at NVT using the experimental densities were in better agreement with the experimental data.

The densities of the methanol-water mixtures are presented in Figure 2.3 together with the experimental values [72]. The calculated densities agree closely with experiment over the whole range of mixtures, except close to \( x_{\text{MeOH}} = 0 \), where the too low density of SPC water compared to experiment plays a bigger role.

For the J2 methanol model together with TIP4P water Freitas [43] found that the simulated densities are smaller than the experimental values and the error increased with increasing methanol concentration. For a methanol mole fraction of \( x_{\text{MeOH}} = 0.50 \) a density of 524 u/nm\(^3\) was reported, to be compared to the B3/SPC value of 527 u/nm\(^3\) and to the experimental value [78] of 531 u/nm\(^3\). Stouten [41] obtained a density of 529 u/nm\(^3\) at 297.21 K. Laaksonen et al. [44] got a density of 533 u/nm\(^3\) at 300 K for a mixture of Haughney’s methanol model and SPC.

### 2.6 Conclusion

We have presented a methanol model and compared its performance for a wide range of properties to models found in literature. To achieve this goal each model has been simulated under the same conditions for long simulation times. Our model B3 is shown to be in overall agreement...
with experiment and to be competitive concerning the reproduction of a variety of experimental quantities with reference models from the literature. The initial aim of getting a model that combines well with SPC water and that has a dielectric permittivity close to the experimental value is only partly fulfilled. The B3 methanol model interacts well with SPC water over the whole range of mole fractions and also reproduces well the hydration free energy. It proved not possible to reproduce the experimental dielectric permittivity with a simple 3 site rigid model for methanol. By systematically varying a wide range of model parameters we have been able to demonstrate the sensitivity of various simulated properties of liquids to the underlying model. This work clearly demonstrates that basic properties such as the density and pressure can be reproduced with a wide range of parameters and that the correct density and pressure is a necessary but not sufficient condition for a good model. The work also highlights the sensitivity of those models to the precise simulation conditions and the need for the incorporation of polarisability if dielectric properties are to be correctly reproduced.
Chapter 3

On the temperature and pressure dependence of a range of properties of a type of water model commonly used in high temperature protein unfolding simulations

3.1 Abstract

Molecular dynamics simulations of protein folding and unfolding are often carried out at temperatures (400 K – 600 K), which are much higher than physiological or room temperature to speed up the (un)folding process. Use of such high temperatures changes both the protein and solvent properties considerably, compared to physiological or room temperature. Water models designed for use in conjunction with biomolecules, such as the simple point charge (SPC) model, have generally been calibrated at room temperature and pressure. To determine the distortive effect of high simulation temperatures on the behaviour of such “room temperature” water models, the structural, dynamic and thermodynamic properties of the much-used SPC water model are investigated in the temperature range from 300 K to 500 K. Both constant pressure and constant volume conditions, as used in protein simulations, were analysed. We found that all properties analysed change markedly with increasing temperature, but no phase transition in this temperature range was observed.

3.2 Introduction

Molecular dynamics simulations are widely used to gain insight into the equilibrium properties of proteins in solution. Increasingly, simulations are also being used to study time- and environment-dependent phenomena, such as the folding or unfolding of proteins. In particular, many simulations of the process of thermal denaturation have been performed. Currently accessible time scales of approximately $10^{-9}$ seconds are, however, insufficient to simulate folding and unfolding processes at experimentally relevant temperatures. To circumvent this problem
most simulations of protein unfolding have used high simulation temperatures to shorten the

time scale of the unfolding process. For example, the thermal unfolding of hen egg white (HEW)
lysozyme in water was simulated at 500 K [79]. The destabilisation of bovine pancreatic trypsin
inhibitor (BPTI) and its reduced form in water was simulated at 423 K and 498 K [80, 81]. The
denaturation of the C-terminal fragment (CTF) of the L7/L12 ribosomal protein was simulated
at 498 K [82], that of the enzyme β-lactamase at 600 K [83] and that of the protein barnase at
498 K and 600 K [84–87]. Potato carboxypeptidase inhibitor was simulated at 600 K [88], and
cutinase was simulated at 393 K [89].

In each of these cases the proteins unfold. The question is, how relevant are the results from
a simulation of a protein at 500 K to the process of thermal denaturation close to physiological
temperatures? At a more basic level one must also ask if the molecular models and simulation
protocols developed for use at 300 K are still appropriate at elevated temperature.

The process of protein folding and unfolding is driven by the balance between protein-
protein, protein-water and water-water interactions. Each of these interactions involves some
degree of enthalpy-entropy compensation and will depend on the temperature. At nonphysi-
ological temperatures the unfolding pathways may also be quite different from that at 300 K.
For one, at high temperature the structural, dynamic and thermodynamic properties of a solvent
such as water will differ from those at 300 K, which will in turn affect the process of protein
unfolding.

In this chapter we investigate the extent to which the structural, dynamic and thermodynamic
properties of a water model commonly used in biomolecular simulations, the simple point charge
(SPC) model [17], change as a function of temperature between 300 K and 500 K.

Because protein unfolding simulations at high temperatures have been carried out at constant
volume as well as at constant pressure, the properties of liquid water are investigated under both
these conditions.

Other studies of the temperature dependence of the properties of water have been published.
However, these studies have not covered the whole range of temperatures or properties relevant
to protein destabilisation and denaturation simulations. They have either focused on the range
up to 373 K [90], on the phase equilibrium [91, 92], or on the supercritical conditions [93–96],
or considered only a limited set of properties. When Levitt et al. [97] developed and tested their
flexible water model F3C, they also examined some of its properties, i.e., energy, heat capacity,
radial distribution function and diffusion constant at higher temperatures. Brodholt and Wood
[98] investigated the behaviour of the energy, pressure, heat capacity, and radial distribution
function of TIP4P [18] as well as SPC/E [15] and a model by Watanabe and Klein [19] over
a wide temperature range (up to 2600 K). However, none of these studies looked at the free
energy and dynamical properties of water, which also might affect protein (un)folding. Here we
concentrate in particular on those properties of water that may influence the process of protein
unfolding and lead to artifacts in unfolding simulations.
3.3. Method

3.3.1 Simulation

At five temperatures, 300 K, 350 K, 400 K, 450 K, and 500 K, two simulations were performed, one at constant pressure and one at constant volume. The system consisted of a cubic periodic box containing 1000 SPC water molecules [17]. Bond lengths and angles were constrained using the SHAKE algorithm [14] with a relative tolerance of $10^{-4}$. The system was equilibrated for 50 ps at each temperature and then simulated for 250 ps for analysis. Configurations 0.05 ps apart were saved. The temperature was kept constant by a Berendsen thermostat [7] (weak coupling) with a coupling time of 0.5 ps and a compressibility of $7.5 \times 10^{-4} \text{mol nm}^3/\text{kJ}$. In the constant-volume simulations the volume was fixed at 29.9151 nm$^3$ (box length of 3.1043 nm) which corresponds to a density of 602.22 u/nm$^3$ (1.0 g/cm$^3$). All simulations were performed using the GROMOS96 simulation package [4, 5] with a time step of 2 fs. The nonbonded interactions were calculated using a twin cutoff of 0.9 nm/1.4 nm for the oxygen-oxygen distances. The interaction between water molecules with oxygen-oxygen distances between 0.9 nm and 1.4 nm were updated every 10 fs. No reaction-field correction to long-range electrostatic interactions was applied.

The excess free energy $\Delta A_{ex}$ of the water model at each temperature was determined using the thermodynamic integration method. The volume was kept constant. All intermolecular interactions were scaled as a function of the coupling parameter $\lambda$ [64]. Simulations were performed at 29 $\lambda$ points between $\lambda = 0$ and $\lambda = 1$. At each $\lambda$-point 20 ps for equilibration and 50 ps for analysis were calculated.

The hydration free enthalpy $\Delta G_{hydration}$ was calculated in the same way, except that the pressure was kept constant and the intermolecular interactions of only one molecule were scaled as a function of the coupling parameter.

3.3.2 Analysis

The presence of a hydrogen bond was determined based on a geometrical criterion. If the O···H distance was less than 0.25 nm and the O-H···O angle greater than 135°, a hydrogen bond was considered to exist between the two water molecules. The diffusion constant was estimated from the Einstein formula

$$D = \lim_{t \to \infty} \frac{\langle (r(t) - r(0))^2 \rangle}{6t} \quad (3.1)$$

where $r(t)$ is the position vector of a molecule’s centre of mass at time $t$. The thermal expansion coefficients $\alpha$ have always been calculated from two simulations at constant pressure with $\Delta T = 50$ K via [55]

$$\alpha \approx -\left( \frac{\ln \left( \frac{\rho_2}{\rho_1} \right)}{T_2 - T_1} \right)_{p}, \quad (3.2)$$

in which $\rho_1$ and $\rho_2$ are the densities at temperatures $T_1$ and $T_2$. 
The heat capacity $C_P$ was calculated using [49]

$$C_P \approx \frac{E_{tot,2} - E_{tot,1}}{T_2 - T_1} + \frac{\partial E_{int}}{\partial T} + \frac{\partial E_{ext}}{\partial T}$$  \hspace{1cm} (3.3)$$

where $E_{int}$ is the (quantum) contribution of intramolecular vibrational modes to the specific heat. $E_{ext}$ is the difference between the quantum-mechanical and the classical intermolecular vibrational energy. Those corrections have been calculated as described by Postma [49]. At 300 K their combined value is $-9.3$ J mol$^{-1}$ K$^{-1}$ [49]. The rotational correlation times $\tau_i$ were calculated by fitting to the linear part in a logarithmic plot of the Legendre polynomial correlation function of rank $l$, $\langle P_l(\cos(\theta(t))) \rangle$, where $\cos(\theta(t))$ denotes the scalar product of the corresponding dipole vectors of unit length separated by a time $t$.

3.4 Results

3.4.1 Thermodynamic properties

Figure 3.1 shows the total energy, the kinetic energy, the potential energy, the van der Waals energy, the electrostatic energy, and the heat of vaporisation for the SPC model as a function of temperature. In the simulations with constant pressure the change in the total energy is larger than in the simulations with constant volume, as the computational box is unable to relax in the latter. The heat of vaporisation $\Delta H_{vap}$ is estimated from the simulations as

$$\Delta H_{vap} = -E_{pot} + p\Delta V + Q_{int} + Q_{ext}$$  \hspace{1cm} (3.4)$$

where $p\Delta V$ can be approximated by $RT$, because $\Delta V$ is essentially equal to $V_{gas}$. $Q_{int}$ is the intramolecular (quantum) contribution for the difference in vibrational energy between the liquid state and the gas phase, and $Q_{ext}$ is its intermolecular counterpart [49]. At 300 K $Q_{int} + Q_{ext} = -0.23$ KJ/mol [49].

Comparing the vaporisation enthalpy calculated in this way to the experimental vaporisation enthalpy [99] along the liquid-vapour curve, there is good agreement, although the simulated values are higher than the experimental ones.

The pressures and densities are shown in Fig. 3.2. The decrease in density as a function of temperature in the constant-pressure simulations is greater than observed experimentally for water up to 373 K. Beyond 373 K, where water is a gas at 1 atm, the simulations clearly overestimate the density. No sudden decrease in density with temperature, which would indicate vaporisation, was observed.

The calculated free energies are shown and compared to the experimental values in Fig. 3.2 c. The experimental values of the excess free energy $\Delta A_{ex}$ are calculated from the vapour pressure by:

$$\Delta A_{ex}(T) \approx RT \ln \left( \frac{RT}{p_{vap}(T)v(T)} \right) - 1$$  \hspace{1cm} (3.5)$$

where $p_{vap}$ is the vapour pressure at temperature $T$ and $v$ the molar volume of water. The values for $p_{vap}$ and $v$ are taken from Schmidt [102]. The calculated excess free energy agrees well with the experimental one over the whole temperature range from 300 K to 500 K.
Figure 3.1: The total, kinetic, potential, van der Waals and electrostatic energies (in kJ/mol) are shown as a function of temperature for the simulations with constant pressure ($p = 1$ atm = 0.061 kJ mol$^{-1}$nm$^{-3}$) and with constant volume ($p = 1.0$ g/cm$^3$ = 602 u/nm$^3$). All quantities are given per molecule. In f the heat of vaporisation, calculated with Eq. 3.4, is compared to the experimental heat of vaporisation [99]. The lines have been added to guide the eye.

In Figure 3.3 a the results for the thermal expansion coefficient $\alpha$ are shown. The values are too high compared to the experimental values [101], as can also be seen in Figure 3.2 b, where the density decreases faster than the experimental density. Although the thermal expansion coefficient is too large, its behaviour with increasing temperature is correct, because the slope is about the same as for the experimental values. The heat capacity $C_p$, shown in Fig. 3.3 b, agrees very well with the experimental values [48] up to 373 K. Above 373 K, the results are noncomparable, as in the simulation the water does not evaporate. Jorgensen and Jenson [90] calculated $C_p$ and $\alpha$ for SPC at 298 K from the fluctuations of the energy and volume. They obtained values that are slightly higher than the values calculated here, $1.06 \cdot 10^{-3}$K$^{-1}$ [90] for $\alpha$ compared to $0.97 \cdot 10^{-3}$K$^{-1}$ and 84.5 J mol$^{-1}$ K$^{-1}$ [90] for $C_p$ compared to 76.4 J mol$^{-1}$ K$^{-1}$. 
3.4.2 Structural properties

The number of hydrogen bonds per molecule shown in Figure 3.3 c decreases almost linearly in the constant-volume simulations. The value of 3.46 hydrogen bonds per molecule under ambient conditions corresponds well to the results of Jorgensen et al. [18], who found 3.54 hydrogen bonds per molecule, although they used an energetic definition of a hydrogen bond. The results also agree well with the experimental results of Haggis et al. [104], who determined the percentage of broken hydrogen bonds by energetic considerations. However, they are completely different from the experimental results of Walrafen [103], who determined the fraction of hydrogen bonds by Raman spectroscopy. In the constant-pressure simulations the number levels off above 450 K. This could indicate clustering of molecules, but no clustering was observed from visual inspection of specific configurations. The computational box did not expand further.
3.4. Results

Figure 3.3: Thermal expansion coefficient $\alpha$ and heat capacity $C_P$ were calculated from the simulations at constant pressure with Equations 3.2 and 3.3. The experimental values for $\alpha$ were taken from Kell [101]. The values above 373 K refer to the metastable liquid at 1 atm. The experimental values for $C_P$ were taken from Weast [48]. The average number of hydrogen bonds per molecule ($\langle n_{HB} \rangle$) is shown as a function of temperature. The hydrogen bond criterion is given in the text. The experimental values obtained by Raman spectroscopy were estimated from the graph in Walrafen [103], those obtained by energetic considerations were taken from Haggis et al. [104]. For further explanation see caption of Figure 3.1.

after about 20 ps (see Fig. 3.4), and we found no indication that the liquid would evaporate on a 100-ps time scale.

Figs. 3.5 and 3.6 show radial distribution functions $g(r)$ between the oxygen atoms of different molecules at the different temperatures. With constant pressure and with constant volume, the peak height decreases with increasing temperature and the first minimum shifts towards longer distances. The second peak that is visible at 300 K disappears at higher temperatures, but there seems to be a second peak reappearing at 500 K. This agrees somewhat with the results of Brodholt and Wood [98] for the TIP4P water model, who saw the second peak disappearing at 340 K and reappearing at 771 K.
Figure 3.4: Box volume of the simulation at 500 K with constant pressure (1 atm = 0.061 kJ mol$^{-1}$ nm$^{-3}$) as a function of time.

Figure 3.5: Oxygen-oxygen radial distribution function $g(r)$ of the simulations with constant volume ($\rho = 1.0$ g/cm$^3 = 602$ w/nm$^3$) for different temperatures. The different curves are vertically shifted by one unit for better visibility.
3.5. Conclusions

Figure 3.6: Oxygen-oxygen radial distribution function $g(r)$ of the simulations with constant pressure ($p = 1$ atm $= 0.061$ kJ mol$^{-1}$ nm$^{-3}$) for different temperatures. The different curves are vertically shifted by one unit for better visibility.

3.4.3 Dynamic properties

The results for the dynamic properties are shown in Fig. 3.7. The simulated diffusion coefficient is larger than the experimental one [105, 106] up to about 400 K, but it changes less with increasing temperature than in the experiment. The increase with temperature is stronger for constant-pressure simulations, especially for temperatures above 400 K. The dipolar rotational correlation times $\tau_i$ decrease with increasing temperature. There is no significant difference between the constant-pressure and the constant-volume simulations. $\tau_1$ is compared to the experimentally measurable decay time $\tau_D$ of the macroscopic polarisation [107]. It should be between $\frac{1}{2} \tau_D$ and $\frac{3}{2} \tau_D$ [108, 109]. Although the values at 300 K lie in this range, it looks like $\tau_1$ is not decaying fast enough with temperature compared to experiment. The ratio between $\tau_1$ and $\tau_2$ is for most temperatures between 2.5 and 3, except for constant pressure at 450 K and for constant volume at 500 K, where it is 1.8 and 3.5, respectively. These deviations probably result from the way $\tau_i$ is calculated. At high temperatures the exponential part in the decay function is short. Thus few points for fitting are available.

3.5 Conclusions

In an attempt to investigate the behaviour of the solvent in simulations at higher temperatures, simulations of SPC water have been performed at temperatures up to 500 K. From the present work, the conclusion is that the structure of the solvent changes with increasing temperature. Although there is no vaporisation of the water even at temperatures up to 500 K, the number of hydrogen bonds per molecule decreases. The excess free energy and hydration free enthalpy per
water molecule change by about 10 kJ/mol over this temperature range. Both these factors would be expected to affect the folding of a protein. In addition, the dynamics of the water molecules changes quite dramatically. Of course, all dynamic properties indicate that the molecules move much faster, the diffusion increases by four- (NVT) to sevenfold (NPT) when the temperature is raised from 300 K to 500 K.

In general the properties of the NVT and NPT systems, which are effectively equivalent at 300 K, 1 atm, and at a density of 1 g/cm³, deviate widely with increasing temperature. The choice of ensemble in simulations of proteins at high temperature is thus a critical issue.

Overall, in comparison with the available experimental data, it is apparent that the SPC water model performs well over a wide range of temperature. This said, it is also clear that at elevated temperature not only does the model begin to deviate from experiment, but that the properties of
3.5. Conclusions

Water as a solvent are also very different from those under which thermal denaturation is studied experimentally. The use of temperatures beyond 400 K in simulations of proteins in water is very likely to significantly affect the (un)folding thermodynamics, pathways, and kinetics. Its results should therefore be very cautiously interpreted.

Financial support was obtained from the Schweizerischen Nationalfonds, project number 21-50929.97, which is gratefully acknowledged.
3. Temperature and pressure dependence of SPC
Chapter 4

On the validity of Stokes’ law at the molecular level

4.1 Abstract

In order to investigate the dependence of the viscosity on the mass of the molecules in a liquid, and thus check the validity of Stokes’ law for molecules, several molecular dynamics simulations of “water” molecules with different mass and different molecular mass distributions were performed. The viscosity is shown to be sensitive to the mass but less sensitive to the mass distribution. The product of diffusion coefficient and viscosity, which according to Stokes’ law should be independent of the mass, varies. We may therefore conclude that Stokes’ law is not valid for small molecules.

4.2 Introduction

More than 150 years ago, Stokes [110] proposed the drag or frictional force \( f \) for a sphere of diameter \( d \) moving with constant velocity \( v \) in a fluid of shear viscosity \( \eta \) is given by,

\[
f = -3\pi \eta d v. \tag{4.1}
\]

This result is obtained by solving the equations of motion for the translational flow of fluid around a rigid sphere assuming so-called stick boundary conditions. Using slip boundary conditions the drag force is smaller: the factor 3 in (4.1) is replaced by a factor 2 [111]. When using Langevin’s formalism [112], the drag or frictional force exerted on a particle with mass \( m \) moving with velocity \( v \) through a viscous medium

\[
f = -m \gamma v \tag{4.2}
\]

is expressed in terms of the friction coefficient \( \gamma \). The friction coefficient is related to the diffusion coefficient \( D \) of the particle through Einstein’s expression [113–115],

\[
D = \frac{k_B T}{m \gamma}, \tag{4.3}
\]
where \( k_B \) is Boltzmann’s constant and \( T \) the temperature, which is related to the mean square velocity \( \langle v^2 \rangle \) through equipartition

\[
\frac{1}{2} m \langle v^2 \rangle = \frac{3}{2} k_B T.
\]

Combining Eqs. (4.1) – (4.3) we arrive at another form of Stokes’ law,

\[
D* = \frac{k_B T}{\pi d^2}.
\]

which implies that for spherical particles of the same size at constant temperature the product of diffusion constant and shear viscosity is fixed. Thus, this product is independent of the mass of the particle.

Although Stokes’ law is derived from purely macroscopic considerations, and the Einstein expression (4.3) is based on the assumption of Brownian motion, Eq. (4.5) describes the experimental data for atomic liquids well [48,115]. Stokes’ law is also often used to obtain values for the friction coefficient \( \gamma \) from experimental data for a molecular solvent, such as water [116], which are required in stochastic dynamics simulations of a solute based on the Langevin equation [117,118].

Here we present results of molecular dynamics (MD) simulations of liquid water which show that Stokes’ law (4.5) is not satisfied at the molecular level. By changing the masses of the water molecules, the diffusion constant and shear viscosity of the liquid can be changed, but their product should remain constant if Stokes’ law (4.5) would apply.

When changing the masses of the atoms in a molecular system dynamic properties change. Thermodynamic equilibrium properties remain unaffected, since the configurational part of the statistical-mechanical partition function is independent of the atomic masses. This is true as long as no intra-molecular constraints are applied in the case of flexible molecules [119]. This result has been used to lengthen the integration time step in MD simulation by increasing the masses of the hydrogen atoms by about a factor 10 to a value comparable to that of the other atoms in the molecule [120–122]. By this trick statistical sampling is enhanced at the expense of a correct simulation of dynamical properties. The results presented here for liquid water show the extent to which the dynamic properties of liquid water are distorted by varying the mass and its distribution within a water molecule.

The shear viscosity of liquid water as a function of the mass of the water molecule is also of interest because the function of a protein in solution depends on its structure, which is unique for each particular protein, and on its dynamical properties. The latter is influenced by the viscosity of the solvent, \( i.e. \) water. In order to determine the dependence of relaxation processes in proteins on the solvent viscosity, one must change the viscosity of the solvent without changing its interaction with the protein. Experimentally, a change of viscosity at constant temperature and pressure can only be brought about by changing the molecular composition of the solvent [123,124]. This, however, also implies a modification of the protein-solvent interaction. In a computer simulation the solvent viscosity can be changed by altering the atomic masses of the solvent molecules, thereby keeping the protein-solvent interaction constant.

The viscosity and diffusivity of a liquid can be calculated straightforwardly using equilibrium or non-equilibrium MD simulation [56,125] or analytical theories [126,127]. The latter studies address the transition to non-Stokes-Einstein behaviour when liquid water is super-
4.3. Method

A total of seven molecular dynamics simulations, which only differed in the masses of the hydrogen and oxygen atoms, were carried out (Table 4.1). The water model used was the simple point charge (SPC) model [17]. Two other models with the same mass distribution as the SPC molecule, SPC$_{10}$ and SPC$_{100}$, were obtained by multiplying the masses of the atoms in the SPC model with a factor of 10 and 100, respectively, resulting in total masses of 180.154 u and 1801.54 u. Models with different molecular mass distributions were obtained by taking the mass of the oxygen atom equal to the mass of the hydrogen. In model EQ$_1$ all atom masses were set to 1 u, in EQ$_{10}$ to 10 u and in EQ$_{100}$ to 100 u. In the D2O model the mass of the oxygen atom was taken the same as in the SPC model, but the masses of the hydrogens were increased to 2.014 u. All the other parameters, such as charge distribution, Lennard-Jones parameters and molecular geometry, were as in the SPC model.

All simulations were performed with a system of 512 molecules in a cubic periodic box. The geometry of the rigid water molecules was maintained using the SHAKE algorithm [14] with a tolerance of $10^{-4}$. To keep the system at a constant temperature of 300 K, a Berendsen thermostat [7] was applied using a coupling time of 0.1 ps. The systems were first equilibrated for 50 ps at a constant pressure of 1 atm. A Berendsen manostat [7] was used with a compressibility of $2.092 \times 10^{-3}$ mol nm$^3$/kJ and a coupling time of 0.5 ps. The next 50 ps were used to calculate the average box lengths. The mean box length averaged over the seven model simulations, was 2.5 nm. In none of the simulations did the box length differ by more than 0.003 nm from this average. The systems were equilibrated for a further 50 ps at constant volume and the next 1000 ps

<table>
<thead>
<tr>
<th>Model</th>
<th>SPC</th>
<th>SPC$_{10}$</th>
<th>SPC$_{100}$</th>
<th>D2O</th>
<th>EQ$_1$</th>
<th>EQ$_{10}$</th>
<th>EQ$_{100}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{H_2O}$ (u)</td>
<td>18.0154</td>
<td>180.154</td>
<td>1801.54</td>
<td>20.0274</td>
<td>3.0</td>
<td>30.0</td>
<td>300.0</td>
</tr>
<tr>
<td>$m_O$ (u)</td>
<td>15.9994</td>
<td>159.994</td>
<td>1599.94</td>
<td>15.9994</td>
<td>1.0</td>
<td>10.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$m_H$ (u)</td>
<td>1.008</td>
<td>10.08</td>
<td>100.8</td>
<td>2.014</td>
<td>1.0</td>
<td>10.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$ρ$ (g/cm$^3$)</td>
<td>0.980</td>
<td>9.799</td>
<td>99.795</td>
<td>1.089</td>
<td>0.163</td>
<td>1.632</td>
<td>16.318</td>
</tr>
<tr>
<td>$T$ (K)</td>
<td>299.69</td>
<td>299.95</td>
<td>300.19</td>
<td>300.29</td>
<td>300.18</td>
<td>300.97</td>
<td>300.37</td>
</tr>
<tr>
<td>$E_{pot}$ (kJ/mol)</td>
<td>-41.87</td>
<td>-41.93</td>
<td>-41.85</td>
<td>-41.87</td>
<td>-41.70</td>
<td>-41.84</td>
<td>-41.84</td>
</tr>
<tr>
<td>$E_{kin}$ (kJ/mol)</td>
<td>7.48</td>
<td>7.49</td>
<td>7.48</td>
<td>7.47</td>
<td>7.48</td>
<td>7.50</td>
<td>7.48</td>
</tr>
</tbody>
</table>

Table 4.1: Masses and thermodynamic properties for the seven water simulations averaged over 50 ps. The models are defined in the text. m: atomic or molecular mass; $ρ$: density of the simulated system; T: temperature; $E_{pot}$, $E_{kin}$: average potential and kinetic energy per water molecule, respectively

cooled. Brown and Clarke [128] studied the dependence of the viscosity and diffusivity of tri-atomic model liquids on the distribution of a fixed molecular mass over the three atoms. They found Stokes’ law (4.5) to be violated for the different mass distributions, but concluded that it is not easy to find a consistent explanation for the statistically significant differences between the values of expression (4.5) for the four mass distributions used in their study.
4. On the validity of Stokes' law

Table 4.2: Masses and dynamic properties for the seven water simulations calculated over 1000 ps ($\eta$) or 500 ps ($\tau_i$). $\eta$: shear viscosity ($\sigma_\eta$: standard deviation); $D$: translational diffusion constant; $\tau_i$: dipolar rotational correlation time for the $i$-th order Legendre polynomial. Experimental values are: $\eta(H_2O) = 0.891$ cp at 298 K [130]; $\eta(D_2O) = 7.059$ cp at 298 K [131]; $D(H_2O) = 2.23 \cdot 10^{-3}$ nm$^2$/ps at 298 K [132]; $D(D_2O) = 2.27 \cdot 10^{-3}$ nm$^2$/ps at 298 K [105]; $\tau_1(H_2O) = 2.1$ ps at 303 K [133]; $\tau_1(D_2O) = 2.5$ ps at 303 K [133]. See also caption of Table 4.1.

were used for analysis. All simulations were performed with the GROMOS96™ package [4], with a cut-off distance of 0.9 nm, a time step of 2 fs and at a temperature of 300 K.

The shear viscosity $\eta$, diffusion constant $D$ and rotational correlation times $\tau_i$ were calculated as described by Tironi and van Gunsteren [55]. For calculating the viscosity the Einstein relation was used. The off-diagonal elements $P_{\alpha\beta}$ of the pressure tensor were saved every 2 fs. The time integral of these, $\Delta P_{\alpha\beta}$, was calculated every 0.1 ps. Because of poor statistics at long simulation times, the viscosity was calculated from the resulting curve between 5 ps and 10 ps [129]. Using the three independent off-diagonal components of the pressure tensor a mean value $\langle \eta \rangle$ and a standard deviation $\sigma_\eta$ were calculated. The diffusion coefficients were obtained from the slope of the mean square centre of mass displacement as a function of time over 50 ps. The rotational correlation times were obtained by fitting to the linear part in a logarithmic plot of the dipolar rotational correlation function.

### 4.4 Results and Conclusions

Table 4.1 presents the models together with the simulation results for the thermodynamic properties. As required, all thermodynamic properties tested remain unaffected by the change of the molecular mass, except of course the density.

The dynamical properties, given in Table 4.2, on the other hand, change with mass. The viscosity increases with increasing mass, though less than one would expect from the experimental increase in viscosity between water and heavy water. As expected, the diffusion constant
decreases with increasing mass and the viscosity increases. The product of viscosity and diffusion constant, which according to Eq. (4.5) should be independent of mass is not. This shows that Stokes’ law is not valid for molecules. The product of viscosity and diffusion coefficient is not correlated with the molecular mass, but it is correlated for those models with the same mass distribution. The rotational correlation times increase with increasing mass. The increase is similar for $\tau_1$ and $\tau_2$, as the ratio of $\frac{\tau_1}{\tau_2}$ is approximately 2 for all seven models.

When using equal masses for hydrogen and oxygen in a water molecule in order to lengthen the time step in a MD simulation, the dynamical properties of liquid water are significantly modified: the viscosity changes by about 50% and the translational and rotational diffusion by about 100%.

If one wishes to study the dependence of protein relaxation processes on solvent viscosity, the latter can be varied by changing the molecular mass: a change in mass by a factor 10 results in a change in viscosity of a factor 2.5.

The authors wish to thank Herman J. C. Berendsen for suggesting to seek a variation in viscosity by changing the molecular mass, and Lorna J. Smith for digging out several articles from Stokes from the Oxford University library.
4. On the validity of Stokes' law
Chapter 5

Further investigation on the validity of Stokes-Einstein behaviour at the molecular level

5.1 Abstract

Stokes-Einstein behaviour at the molecular level is investigated by simulating water at different temperatures and by simulating “water” models with different mass distributions. When combining Stokes’ law for the viscosity with Einstein’s formula for the diffusivity, an expression for the product of these quantities is obtained, which shows that the product of diffusivity and viscosity should be independent of the mass distribution and positively proportional to the temperature. Using both, equilibrium and non-equilibrium simulation techniques to compute the viscosity a slight deviation from Stokes-Einstein behaviour was found for the “water” models and temperatures investigated. Non-equilibrium simulation seems to yield systematically lower values for the viscosity than equilibrium simulation.

5.2 Introduction

Stokes’ law,

\[ \eta = \frac{m\gamma}{3\pi d} \]  \hspace{1cm} (5.1)

expresses the shear viscosity \( \eta \) of a liquid in terms of the mass \( m \), the friction coefficient \( \gamma \) and the diameter \( d \) of the particles of the liquid. The friction coefficient \( \gamma \) is related to the diffusion coefficient \( D \) of the particles through Einstein’s expression

\[ D = \frac{k_B T}{m\gamma} \]  \hspace{1cm} (5.2)

where \( k_B \) is Boltzmann’s constant and \( T \) is the temperature. Combining Eqs. (5.1) and (5.2) one obtains the relation

\[ D\eta = \frac{k_B T}{3\pi d} \]  \hspace{1cm} (5.3)
which characterises Stokes-Einstein behaviour of a liquid. Relation (5.1) is obtained by solving the equations of motion for the translational flow of a fluid around a rigid sphere of diameter $d$ assuming so-called stick boundary conditions. In the previous chapter, (Chapter 4, [134]), it was investigated if non-atomic liquids consisting of small molecules obey Eq. (5.3), i.e., show Stokes-Einstein behaviour. This was done by simulating “water” molecules with different mass and different mass distributions. Thus the diffusion constant and the viscosity of the liquid were changed without changing the intermolecular interaction or the molecular geometry.

When interpreting the results of the various simulations in Chapter 4, it was overlooked that scaling of the total mass of a system by a factor $\lambda$ at constant temperature is equivalent to a scaling of the time dimension by a factor $\lambda^{1/2}$. The kinetic energy is related to the mass and the time by

$$E_{\text{kin}} = \frac{m}{2} \left( \frac{d}{dt} r \right)^2 \quad (5.4)$$

where $m$ is the mass, $t$ the time and $r$ the distance. So, scaling the mass by a factor of $\lambda$, scales the time by a factor of $\lambda^{1/2}$ at constant kinetic energy. The viscosity should then scale by a factor of $\lambda^{1/2}$, the diffusion constant by a factor of $\lambda^{-1/2}$ and their product should be independent of $\lambda$. When scaling all masses equally, the same system is simulated, only on a different time scale. In Chapter 4 a thermostat was used to keep the kinetic energy at the equivalent of 300 K in all simulations. So, the product $D\eta$ should have been the same for those simulations that only differed in total mass. Since this product was not the same in Chapter 4 for the simulations of the different models with the same mass distribution but different total masses, we wish to examine here the reason for this discrepancy. Secondly, the occurrence of Stokes-Einstein behaviour, Eq. (5.3), at the molecular level is further investigated by simulating at different temperatures.

One reason for the dependence of the product $D\eta$ on the total mass as observed in Chapter 4 might be that all models were simulated for the same time period: 1000 ps. If one considers the scaling, the model SPC\textsubscript{100} (simple point charge water with $\lambda = 100$) for example, has therefore only been simulated for the equivalent of 100 ps, which is not sufficient to accurately determine the viscosity from fluctuations of the pressure tensor. The viscosity was calculated using the Einstein relation for the mean-square change as a function of time of the off-diagonal elements of the pressure tensor, which involves a linear least-squares fit of a chosen time period. In Chapter 4 this fitting necessary to calculate the viscosity has been done over the same time span (5 ps – 10 ps) for all models, which is effectively much shorter for the larger masses if one considers scaling. Another reason for the dependence of the product $D\eta$ on the total mass as observed in Chapter 4 might be that the same time step of 2 fs was used for all the models. So, SPC\textsubscript{100} was simulated with a time step that would correspond to only 0.2 fs when scaled back to the mass of SPC. This much more accurate integration might lead to a more exact value for the viscosity. To examine these possibilities we have extended the simulations of SPC\textsubscript{100} and SPC\textsubscript{10} to the period corresponding to 1000 ps for SPC, and in addition we have simulated SPC using reduced time steps of 1 fs and 0.2 fs for 1000 ps. Because of the scaling, the simulations of SPC\textsubscript{100} and of SPC at the reduced time step of 0.2 fs should yield the same results.

For a liquid displaying Stokes-Einstein behaviour, the product of $D\eta$ should be proportional to the temperature $T$ and independent of the mass distribution in the molecule. To further check the validity of Eq. (5.3) for liquids consisting of small molecules, we performed simulations at higher temperatures and with “water” molecules with different mass distributions.
5.3 Method

Five different models were simulated, which only differed in their total mass or mass distribution. The model masses are summarised in Table 5.1. All other model parameters, such as charge distribution, Lennard-Jones parameters and molecular geometry are taken from the SPC water model [17]. The simulations of SPC\textsubscript{10} and SPC\textsubscript{100} are continuations of those reported in Chapter 4 on the validity of Stokes' law to the period corresponding to 1000 ps for SPC, which is 3212.278 ps for SPC\textsubscript{10} and 10000.0 ps for SPC\textsubscript{100}. Two simulations of SPC with reduced step sizes of 1 fs and 0.2 fs were performed for 1000 ps. Three simulations to investigate the influence of the temperature were performed, one at 330 K, one at 360 K and one at 390 K. In order to pursue converged values for the viscosity, these simulations covered 10000 ps. Finally, two models with a mass distribution different from that of SPC were simulated. In one of those, named EQ, all atoms have (nearly) equal mass, in the other, named CPS, the H-atoms are much heavier than the O-atom. We note that the EQ model does not correspond exactly to one of the models of the EQ-series in Chapter 4. None of those EQ-models had a total mass equal to that of SPC, so extending their simulations further would have involved scaling. The simulations of the models with different mass distributions were also run for 10000 ps. To be able to compare the results of the SPC simulation at 300 K with the simulations at higher temperatures and the simulations of EQ and CPS, the SPC simulation was also extended to 10000 ps. In all simulations 512 molecules were simulated in a cubic periodic box. In the simulations at 300 K, the box size was set to 15.625 nm\textsuperscript{3} as in Chapter 4. In the simulations at elevated temperatures, 100 ps equilibration at constant pressure was performed, where the pressure was kept at 0.06102 kJ mol\textsuperscript{-1} nm\textsuperscript{-3} by a Berendsen manostat [7]. The compressibility was set to 2.092 mol nm\textsuperscript{3}/kJ and the pressure-bath coupling time to 0.5 ps. The box size averages in the second 50 ps of this equilibration period were used as the box sizes for the following constant volume simulations. The box sizes are summarised in Table 5.3. In all simulations the temperature was kept constant by a Berendsen thermostat [7] using a temperature-bath coupling time

<table>
<thead>
<tr>
<th>Model</th>
<th>SPC</th>
<th>SPC\textsubscript{10}</th>
<th>SPC\textsubscript{100}</th>
<th>EQ</th>
<th>CPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{H_2O}$ (u)</td>
<td>18.0154</td>
<td>180.154</td>
<td>1801.54</td>
<td>18.0154</td>
<td>18.0154</td>
</tr>
<tr>
<td>$m_O$ (u)</td>
<td>15.9994</td>
<td>159.994</td>
<td>1599.94</td>
<td>6.0154</td>
<td>2.0154</td>
</tr>
<tr>
<td>$m_H$ (u)</td>
<td>1.008</td>
<td>10.08</td>
<td>100.8</td>
<td>6.0000</td>
<td>8.0000</td>
</tr>
</tbody>
</table>

Table 5.1: The total mass and mass distribution of the different "water" models that have been simulated. SPC, SPC\textsubscript{10} and SPC\textsubscript{100} have the same mass distribution, but a different total mass. SPC, EQ and CPS have the same total mass, but a different mass distribution. EQ has the same mass distribution as the models of the EQ-series in Chapter 4, but is scaled to the same total mass as SPC to avoid scaling of the time. $m$ denotes the atomic or molecular mass.
of 0.1 ps. The procedure SHAKE [14] was used to constrain the bonds and H-H distance with a tolerance of 0.0001. A single cut-off radius of 0.9 nm for the intermolecular interactions was used and no reaction-field force was calculated. At the beginning of each simulation the system was equilibrated for 50 ps at constant volume.

The shear viscosity $\eta$ and the diffusion constant $D$ were calculated as described in Chapter 4 and [55]. To calculate the viscosity, the fitting to the square of the integral of the elements of the pressure tensor was again performed between 5 ps and 10 ps for the models with a total mass equal to that of SPC. For the heavier models this time span was scaled according to their mass, so it was 15.9 ps to 31.8 ps for SPC$_{10}$ and 50 ps to 100 ps for SPC$_{100}$. The error $\sigma_D$ in the diffusion constant $D$ was calculated from the diffusion constants $D_\alpha$ in the three ($\alpha = x, y, z$) dimensions via

$$\sigma_D = \sqrt{\langle D^2_{x,y,z} \rangle - \langle D_\alpha \rangle^2_{x,y,z}}. \quad (5.5)$$

The error $\sigma_\eta$ in the viscosity $\eta$ was calculated in the same way using the three off-diagonal elements of the pressure tensor. The error in the product $D\eta$, $\sigma_{D\eta}$, was calculated as

$$\sigma_{D\eta} = \left( \frac{\sigma_D}{D} + \frac{\sigma_\eta}{\eta} \right) D\eta. \quad (5.6)$$

The shear viscosities $\eta$ were, alternatively, calculated using non-equilibrium MD simulations [135]. The computational procedure largely follows the one described in [136]. All atoms were accelerated in the $x$-direction. The magnitude of the acceleration $a_x$ depends on the $z$-coordinate of the atom,

$$a_x = A \cos \left( 2\pi z L_z^{-1} \right), \quad (5.7)$$

where $L_z$ is the length of the rectangular computational box in the $z$-direction. Non-equilibrium MD simulations using Eq. (5.7) were carried out for boxes three times the size of the equilibrium boxes at the corresponding temperatures with edge lengths $L_x = L_y = \frac{1}{3} L_z$, using values of $A$ chosen such that the maximum shear rate was always about $\frac{1}{8}$ ps$^{-1}$, which is slightly larger than the inverse correlation time of the liquid. The values of $A$ for all temperatures are given in Table 5.3. The boxes contained 1536 molecules. All other simulation parameters are equal to those used in the equilibrium MD simulations.

### 5.4 Results and Conclusions

Table 5.2 gives the results of the simulations of SPC and mass scaled SPC as considered in Chapter 4, but extended in length, and of the simulations produced with reduced time steps. The results for $D$ and $\eta$ now do scale with $\lambda^{-\frac{1}{2}}$ and $\lambda^\frac{3}{4}$ respectively, where $\lambda$ is the scaling factor of the masses. The higher accuracy of integration in the simulations with a smaller time step does not affect the result for the viscosity. The result of the normal ($\Delta t = 2$ fs) SPC simulation lies between the (scaled) results of SPC$_{100}$ and SPC with a time step of 0.2 fs, which both are simulated with a higher accuracy. Increasing the simulation length, on the other hand, did improve the result. The better sampling allows fitting over the right time span, e.g., 50 ps – 100 ps for SPC$_{100}$, which is not possible with the shorter simulations as there is insufficient statistics for this time span. If in the longer simulation the fitting is still performed to the shorter time span, the results will also be off, for instance for SPC$_{100}$ the result for the viscosity is
5.4. Results and Conclusions

<table>
<thead>
<tr>
<th>Model</th>
<th>SPC</th>
<th>SPC&lt;sub&gt;10&lt;/sub&gt;</th>
<th>SPC&lt;sub&gt;100&lt;/sub&gt;</th>
<th>SPC</th>
<th>SPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta t$ (fs)</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>$t$ (ns)</td>
<td>1.0</td>
<td>3.162</td>
<td>10.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$t_{\text{fit}}$ (ps)</td>
<td>5.0 – 10.0</td>
<td>15.9 – 31.8</td>
<td>50.0 – 100.0</td>
<td>5.0 – 10.0</td>
<td>5.0 – 20.0</td>
</tr>
<tr>
<td>$\langle T \rangle$ (K)</td>
<td>300.4</td>
<td>300.5</td>
<td>300.2</td>
<td>301.7</td>
<td>302.1</td>
</tr>
<tr>
<td>$\langle D \rangle$ ($10^{-9} \text{m}^2 \text{s}^{-1}$)</td>
<td>4.074</td>
<td>1.302</td>
<td>0.397</td>
<td>4.009</td>
<td>4.202</td>
</tr>
<tr>
<td>$\sigma_D$ ($10^{-9} \text{m}^2 \text{s}^{-1}$)</td>
<td>0.173</td>
<td>0.206</td>
<td>0.030</td>
<td>0.194</td>
<td>0.345</td>
</tr>
<tr>
<td>$\langle \eta \rangle$ ($10^{-3} \text{kg m}^{-1} \text{s}^{-1}$)</td>
<td>0.635</td>
<td>1.968</td>
<td>6.610</td>
<td>0.524</td>
<td>0.580</td>
</tr>
<tr>
<td>$\sigma_\eta$ ($10^{-3} \text{kg m}^{-1} \text{s}^{-1}$)</td>
<td>0.051</td>
<td>0.142</td>
<td>0.694</td>
<td>0.043</td>
<td>0.067</td>
</tr>
<tr>
<td>$\langle D \rangle \langle \eta \rangle$ ($10^{-12} \text{kg m s}^{-2}$)</td>
<td>2.558</td>
<td>2.562</td>
<td>2.622</td>
<td>2.100</td>
<td>2.435</td>
</tr>
<tr>
<td>$\sigma_{D\eta}$ ($10^{-12} \text{kg m s}^{-2}$)</td>
<td>0.318</td>
<td>0.590</td>
<td>0.474</td>
<td>0.273</td>
<td>0.484</td>
</tr>
</tbody>
</table>

Table 5.2: Equilibrium MD simulations of the different water models covering an equivalent (scaled) time period. The results for SPC (1000 ps) with a time step of 2 fs were taken from Chapter 4. The simulations of SPC<sub>10</sub> and SPC<sub>100</sub> have been extended to the length that corresponds to 1000 ps for SPC. $\Delta t$ is the time step, $t$ the simulation length, $t_{\text{fit}}$ the time over which the fit is performed to calculate the viscosity, $D$ is the diffusion coefficient, $\eta$ the viscosity and $\sigma_x$ the error in the quantity $x$ as calculated using Equations (5.5) and (5.6).

only $4.47 \times 10^{-3}$ kg m$^{-1}$ s$^{-1}$ if the fitting is performed from 5 ps to 10 ps. So, the apparent discrepancy in Chapter 4 between the simulations that only differed in total mass, is due to the different simulation lengths and disappears when equally long simulations are compared.

The results for the simulations performed to further investigate Stokes-Einstein behaviour are summarised in Table 5.3. For all the models and temperatures the 1000 ps non-equilibrium MD simulations yield values for the viscosity that are systematically smaller than the values obtained from 10000 ps equilibrium MD simulations. The discrepancy shrinks with temperature from 24% at 300 K to 8% at 390 K. In the limit of infinite sampling time and, in the non-equilibrium case in the limit of vanishingly small acceleration $A$, which keeps the system off equilibrium, both techniques should, however, yield the same viscosity values. For a finite acceleration $A$, one can intuitively understand that the viscosity for a system in a stationary off-equilibrium state will be lower than the viscosity determined from fluctuations of a system in equilibrium. Since each method yields comparable relative differences in viscosity between the different models and between the different temperatures, we shall consider the results of the two methods separately. We note that a similar discrepancy has been observed between equilibrium and non-equilibrium MD methodology to compute diffusivity [137]. Non-equilibrium MD yields a larger diffusivity than equilibrium MD [137].

The results for diffusion and viscosity change considerably when going from SPC to EQ, but only slightly when going from EQ to CPS. Feenstra et al. [136] also found a large difference between SPC and a model corresponding to EQ. However, contrary to our results, they found a slightly smaller diffusion constant for their model with all atoms of equal mass than for their model with heavy H-atoms. This small difference in $D$ is probably not significant, since the simulations in [136] only covered 100 ps lengths. In general, Feenstra et al. [136] reported values of $D$ that are similar to ours, and values for the viscosity, which were calculated using
5. Further investigation of Stokes-Einstein behaviour

<table>
<thead>
<tr>
<th>Model</th>
<th>CPS</th>
<th>EQ</th>
<th>SPC</th>
<th>SPC</th>
<th>SPC</th>
<th>SPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{bath}$ (K)</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>330</td>
<td>360</td>
<td>390</td>
</tr>
</tbody>
</table>

**Equilibrium MD**

<table>
<thead>
<tr>
<th>$\langle T \rangle$ (K)</th>
<th>300.6</th>
<th>301.1</th>
<th>300.4</th>
<th>330.6</th>
<th>360.7</th>
<th>390.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$ (nm$^3$)</td>
<td>15.625</td>
<td>15.625</td>
<td>15.625</td>
<td>16.060</td>
<td>16.620</td>
<td>17.233</td>
</tr>
<tr>
<td>$\langle D \rangle$ ($10^{-9}$m$^2$s$^{-1}$)</td>
<td>3.102</td>
<td>3.227</td>
<td>4.058</td>
<td>6.287</td>
<td>8.829</td>
<td>12.518</td>
</tr>
<tr>
<td>$\sigma_D$ ($10^{-9}$m$^2$s$^{-1}$)</td>
<td>0.043</td>
<td>0.011</td>
<td>0.138</td>
<td>0.667</td>
<td>0.335</td>
<td>0.855</td>
</tr>
<tr>
<td>$\langle \eta \rangle$ ($10^{-3}$kg m$^{-1}$s$^{-1}$)</td>
<td>0.697</td>
<td>0.683</td>
<td>0.582</td>
<td>0.371</td>
<td>0.258</td>
<td>0.178</td>
</tr>
<tr>
<td>$\sigma_\eta$ ($10^{-3}$kg m$^{-1}$s$^{-1}$)</td>
<td>0.012</td>
<td>0.033</td>
<td>0.006</td>
<td>0.015</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td>$\langle D \rangle \langle \eta \rangle$ ($10^{-12}$kg m s$^{-2}$)</td>
<td>2.161</td>
<td>2.204</td>
<td>2.363</td>
<td>2.332</td>
<td>2.274</td>
<td>2.224</td>
</tr>
<tr>
<td>$\sigma_{Di\eta}$ ($10^{-12}$kg m s$^{-2}$)</td>
<td>0.066</td>
<td>0.114</td>
<td>0.106</td>
<td>0.343</td>
<td>0.163</td>
<td>0.260</td>
</tr>
<tr>
<td>$\langle D \rangle \langle \eta \rangle T^{-1}$ ($10^{-15}$kg m s$^{-2}$K$^{-1}$)</td>
<td>7.20</td>
<td>7.35</td>
<td>7.88</td>
<td>7.07</td>
<td>6.32</td>
<td>5.70</td>
</tr>
</tbody>
</table>

**Non-equilibrium MD**

<table>
<thead>
<tr>
<th>$A$ (nm ps$^{-2}$)</th>
<th>0.0375</th>
<th>0.0375</th>
<th>0.0375</th>
<th>0.2</th>
<th>0.14</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\langle T \rangle$ (K)</td>
<td>301.9</td>
<td>302.6</td>
<td>302.6</td>
<td>332.7</td>
<td>362.9</td>
<td>393.0</td>
</tr>
<tr>
<td>$\langle \eta \rangle$ ($10^{-3}$kg m$^{-1}$s$^{-1}$)</td>
<td>0.525</td>
<td>0.527</td>
<td>0.425</td>
<td>0.306</td>
<td>0.231</td>
<td>0.164</td>
</tr>
<tr>
<td>$\langle D \rangle \langle \eta \rangle$ ($10^{-12}$kg m s$^{-2}$)</td>
<td>1.628</td>
<td>1.701</td>
<td>1.724</td>
<td>1.923</td>
<td>2.041</td>
<td>2.057</td>
</tr>
<tr>
<td>$\langle D \rangle \langle \eta \rangle T^{-1}$ ($10^{-15}$kg m s$^{-2}$K$^{-1}$)</td>
<td>5.43</td>
<td>5.67</td>
<td>5.75</td>
<td>5.83</td>
<td>5.67</td>
<td>5.27</td>
</tr>
</tbody>
</table>

Table 5.3: MD simulations of liquid “water” models to investigate Stokes-Einstein behaviour at the molecular level. The equilibrium simulations have been performed for 10000 ps with a time step of 2 fs, and the fitting in the calculation of $\eta$ was performed between 5 ps and 10 ps. $T_{bath}$ is the bath temperature, $\langle T \rangle$ the average temperature during the simulations, and $V$ the volume of the box. The other symbols are explained in the caption of Table 5.2. The non-equilibrium MD results for the viscosity were obtained from 1000 ps MD simulations based on Eq. (5.7) using the given acceleration amplitudes $A$.

Non-equilibrium simulations, similar to our non-equilibrium results. Brown and Clarke [128] simulated a liquid of rigid tri-atomic molecules without Coulomb interactions and with van der Waals interactions that are rather dissimilar to those of SPC water, and also investigated different mass distributions for their tri-atomic model. They obtained the lowest value of the viscosity for their 'B'-model, which has a normal water distribution of the masses and the highest value for their 'A'-model where all masses are equal. For their 'C'-model they calculated a viscosity closer to their 'B'-model than to their 'A'-model which does not agree with our or Feenstra et al.’s [136] results. For the diffusion constant of their 'C'-model they get a markedly lower value than for models 'A' and 'B', whereas in our study as well as in Feenstra et al.’s [136] the differences in the diffusion constant are not that large. Since the Brown and Clarke models are rather different from ours, the difference in the dependence of $D$ and $\eta$ on the mass distribution comes to no surprise. In the SPC water model, charge or hydrogen-bonding interactions are dominant, thereby restricting the rotation of the molecules very much, which in turn will lessen the effect of a modification of the molecular mass distribution, when compared to the Brown
5.4. Results and Conclusions

and Clark models which only contain van der Waals interactions.

In simulations at higher temperatures (right hand side in Table 5.3) the product of diffusion and viscosity shows different behaviour for the equilibrium and the non-equilibrium simulations. For the equilibrium simulations, it becomes smaller with increasing temperature, it is decreasing from $2.363 \cdot 10^{-12}$ kg m s$^{-2}$ at 300 K to $2.224 \cdot 10^{-12}$ kg m s$^{-2}$ at 390 K, whereas for the non-equilibrium simulations it increases from $1.724 \cdot 10^{-12}$ kg m s$^{-2}$ at 300 K to $2.057 \cdot 10^{-12}$ kg m s$^{-2}$ at 390 K. In both cases, the product $D \eta T^{-1}$ decreases for higher temperatures. The value obtained by averaging over the four temperatures is $6.74 \cdot 10^{-15}$ kg m s$^{-2}$ K$^{-1}$ for the equilibrium simulations and $5.63 \cdot 10^{-15}$ kg m s$^{-2}$ K$^{-1}$ for the non-equilibrium simulations which lead to atom diameters through Eq. (5.3) of 0.22 nm and 0.26 nm, respectively.

When investigating the mass distribution from SPC to EQ to CPS, the product $D \eta T^{-1}$ becomes smaller, $-0.7 \cdot 10^{-15}$ kg m s$^{-2}$ K$^{-1}$ in the equilibrium simulation and $-0.3 \cdot 10^{-15}$ kg m s$^{-2}$ K$^{-1}$ in the non-equilibrium simulations. When raising the temperature from 300 K to 390 K, the product $D \eta T^{-1}$ also becomes smaller, $-2.2 \cdot 10^{-15}$ kg m s$^{-2}$ K$^{-1}$ in the equilibrium simulations and $-0.5 \cdot 10^{-15}$ kg m s$^{-2}$ K$^{-1}$ in the non-equilibrium simulations. Both tests, the simulations of models with different mass distribution, as well as the simulations at higher temperatures, give results that slightly deviate from the simple relation between $D$, $\eta$, $d$ and $T$ as embodied in the Stokes-Einstein relation (5.3). The equilibrium method, the oldest and most used one, yields larger deviations than the non-equilibrium one. It would be of interest to further investigate the observed systematic deviation between the results of the equilibrium and non-equilibrium techniques with respect to other (electric, thermal conductance) transport coefficients as a function of the system and simulation parameters (system size, box shape, temperature, size of non-equilibrium driving force, etc.).
5. Further investigation of Stokes-Einstein behaviour
Chapter 6
Viscosity dependence of protein dynamics

6.1 Abstract

The influence of the solvent viscosity on protein dynamics is investigated with molecular dynamics simulations of factor Xa in two solvents differing only in viscosity, by a factor of ten. We obtained this viscosity change by changing the masses of the solvent atoms by a factor of 100. Equilibrium properties of the protein, i.e. the average structure, its fluctuations, and the secondary structure, show no significant dependence on the solvent viscosity. The dynamic properties of the protein, i.e. atom-positional correlation times and torsional angle transitions, however, depend on the solvent viscosity. The protein appears to be much more mobile in the solvent of lower viscosity. It feels the influence of the solvent not only on the surface but even in its core. With increasing solvent viscosity, the positional relaxation times of atoms in the protein core increase as much as those of atoms on the protein surface, and the relative increase in the core is even larger than on the surface.

6.2 Introduction

The dynamics of a protein in solution depends on the type of solvent, especially on the solvent viscosity. Beece et al. [123] showed that for binding of O₂ to myoglobin, variations in pH, ionic strength, and permittivity change the binding kinetics much less than variations in viscosity. Experimentally, the influence of solvent viscosity on protein dynamics has repeatedly been studied. Apart from myoglobin, which has been well investigated in different studies [123, 138–143], an effect of solvent viscosity has also been observed for buried tryptophanes [144, 145] and for carboxypeptidase [146]. A conformational change of phenyl-benzyl-nitroxides may be induced by the solvent [147]. Kramers [148] derived an expression for the dependence of reaction rates on viscosity. This was generalised by Gavish [149] to include variations of viscosity along the reaction coordinate.

Ansari et al. [141] investigated the dependence of the rate constant for myoglobin conformational changes on the viscosity in the limit of low viscosity and observed three regimes. One was at very low viscosity, where the rate constant was solely determined by the internal friction and not by the solvent viscosity and remained, therefore, constant even with a further decrease in solvent viscosity. Then, there was an intermediate regime where the rate constant was gov-
cerned by both the protein friction and the solvent viscosity. Finally, there was a regime of high viscosity, where the rate constant was determined solely by the solvent viscosity. In the high viscosity limit, Beece et al. [123] found that the rates tend toward a finite limit. Studies by Iben et al. [139] suggest that proteins lose their flexibility when the solvent is cooled below its glass temperature.

For studying the influence of the solvent on protein properties, molecular dynamics (MD) has the advantage over experiment that particular solvent properties of interest can be changed without other properties of the solvent being affected. This is particularly true of solvent viscosity. In experiment, the viscosity is usually changed by the addition of cosolvents, although their molecular size influences the protein dynamics at a given viscosity, as was shown by Yedgar et al. [150]. A change in the composition of the solvent also implies modification of protein-solvent interactions. In MD simulations, we can obtain a change in solvent viscosity by changing the solvent mass, which does not affect other nondynamic properties of the solvent, such as pH or solvent molecule size. It also does not modify protein-solvent interactions. Changing the total mass of the solvent molecules by a factor of \( \lambda \) leads to a change in solvent viscosity by a factor of \( \sqrt{\lambda} \) (Chapters 4 and 5, [134, 151]).

A change in the solvent viscosity should not affect the equilibrium properties of the protein. The dynamic properties will change with changing solvent viscosity. It is to be expected that this influence of the solvent viscosity is stronger at the outside of the protein, which is in direct contact with the solvent, than in the core, which is shielded from the solvent. Beece et al. [123] found three barriers that an \( \text{O}_2 \) molecule must overcome when going from the solvent to its binding site in myoglobin with a decreasing viscosity dependence of the transition rates over the barriers towards the centre of the protein. Here, a sizeable globular protein, factor Xa, 235 residues long, was used to investigate by MD simulation how deep into the protein the influence of the solvent viscosity reaches.

### 6.3 Models and Simulations

To investigate the effect of the solvent viscosity on protein dynamics, we performed two simulations of the protein factor Xa with different solvents. In one simulation, the solvent consisted of regular simple point charge (SPC) water molecules [17]. In the other simulation, the solvent consisted of a “water” model in which all atoms had masses of 100 times the corresponding masses in SPC, whereas all the other parameters of the water model were taken unchanged from SPC. This model was called SPC\(_{100}\). The shear viscosity of SPC at 300 K was found to be \( \eta = 0.6 \cdot 10^{-3}\text{kg m}^{-1}\text{s}^{-1} \) and that of SPC\(_{100}\) at the same temperature was \( \eta = 6.6 \cdot 10^{-3}\text{kg m}^{-1}\text{s}^{-1} \) (Chapter 5). Both simulations started from the same equilibrated structure. All simulations were performed with the GROMOS96 [4, 5] package.

The entry \( \text{1fax} \) [152] from the Protein Data Bank (PDB) was used as a starting structure for the protein, an X-ray structure with a resolution of 0.3 nm. To reduce the system size, the L-chain was removed. Without the L-chain, factor Xa consists of 235 amino acids and a calcium ion. Hydrogen atoms attached to aliphatic carbon atoms were incorporated into these (united atoms), so the protein consisted of 2405 atoms including the Ca\(^{2+}\) ion. A schematic picture of the protein is shown in Figure 6.1. The protein was placed in a truncated octahedron box and 7863 SPC water molecules were added from an equilibrated water box by superimposing
both boxes and removing water molecules of which the oxygen was closer than 0.23 nm to any non-hydrogen protein atom. This system was then energy-minimised and equilibrated by MD for 300 ps, both with periodic boundary conditions. At this point, the simulation with the high viscosity solvent SPC$_{100}$ was branched off. For the simulation with SPC, the equilibrated structure was taken as the starting structure for the following simulation period of 3000 ps without any parameter changes. For the simulation with SPC$_{100}$, all SPC molecules were replaced by SPC$_{100}$ molecules, i.e. the atomic masses were increased by a factor 100. To keep the kinetic energy of the solvent the same, all solvent velocities were scaled by a factor of 0.1.

The time step used in the simulations was 2 fs. The temperature was set to 300 K, the protein and solvent were separately coupled to a temperature bath with a coupling time of 0.1 ps [7]. The simulations were performed with isotropic pressure coupling [7], the pressure was set to 0.06102 kJ mol$^{-1}$ nm$^{-3}$ (1 atm) and the coupling time was set to 0.5 ps. This led to a box length of about 8.17 nm. All bonds where constrained by the procedure SHAKE [14] with a relative tolerance of $10^{-4}$. A twin-range cut off of 0.8 nm and 1.4 nm was used (short-ranges forces were calculated every time step, intermediate-ranged forces were only recalculated every five time steps) and a reaction field [8] was applied beyond a cut off of 1.4 nm with a permittivity of the continuum of 54 [68].

To investigate the influence of the solvent viscosity on the protein, we analysed several structural and dynamical properties. The structural properties chosen were the root-mean-square deviation (RMSD) from the starting structure, the root-mean-square fluctuation (RMSF) around the average structure, and the secondary structure. As dynamic properties, the atom-positional correlation time and the number of dihedral angle transitions were analysed. Some of these properties were analysed as a function of distance to the solvent. The distance to the solvent
of each protein atom was calculated by averaging of the distance to the nearest solvent atom’s centre of mass along the simulation.

The RMSD from the starting structure was calculated as

$$\sigma(t) = \sqrt{\langle (r(t) - r(0))^2 \rangle} \quad (6.1)$$

where $r(t)$ is the position vector of an atom at time $t$ and $\langle \rangle$ denotes an average over all atoms except the $\text{Ca}^{2+}$ ion, which was excluded from the calculation. Before calculating $\sigma(t)$, we performed a least-squares fit of the $C_\alpha$-atom positions of the protein at time $t$ to those of the protein structure at time 0. The RMSF was calculated in the same way except that $r(0)$ was replaced by the average position of the atom during the simulation. For each simulation, the RMSD from the average structure of the other simulation also was calculated.

The secondary structure was analysed every 100 ps with the program STRIDE, which uses the method of Frishman and Argos [153].

The solvent-accessible surface area of the residues was determined for the average structure of both simulations with the program NACCESS [154] according to the method of Lee and Richards [155]. A probe size of 0.14 nm was used in these calculations.

The correlation function $C_{\Delta r}$ for the atomic displacement vector $\Delta r = r - \langle r \rangle$ was calculated as

$$C_{\Delta r}(t') = \frac{\langle \Delta r(t') \Delta r(t' + t) \rangle}{\langle |\Delta r(t')| \rangle |\Delta r(t' + t)|} \quad (6.2)$$

The correlation time $\tau_C$ was estimated from

$$C_{\Delta r}(\tau_C) = \frac{1}{e} \quad (6.3)$$

The correlation time of the residues was calculated in the same way with $r$ being the position vector of the centre of mass of each residue.

The dihedral angle transitions were counted only for the dihedral angles not involving hydrogen atoms. Only transitions of approximately 120° were counted.

### 6.4 Results and Discussion

When equilibrium is reached, the positional fluctuation of the atoms, as measured by the RMSD and RMSF, should not be affected by the different solvent viscosities. However, Fig. 6.2 shows that the protein deviates faster from the starting structure in the simulation in SPC than in the simulation in SPC100. Not unexpectedly, the protein conformation changes faster in the low-viscosity (SPC) solvent. This effect is also reflected in the slightly higher RMSF in SPC. The reason for the latter difference is that the simulations have not yet completely converged. The RMSD in SPC100 has not yet converged, and in SPC, where it looks more converged, there are still sometimes sudden increases (e.g., at 1500 ps and at 2500 ps), indicating that it may not yet have fully converged. The RMSF is for both simulations lowest around the middle. This also indicates that the simulation is not completely converged, as the average structure is closer to the structures in the middle of the trajectory than to those at the ends. However, the differences in the RMSF between the two simulations are rather small and can be assumed to disappear if the
simulations were continued. The two simulations explore slightly different parts of configuration space, as the RMSDs from the other simulation’s average structure (bottom part of Fig. 6.2) are larger than the RMSF around each simulation’s own average structure.

The evolution of secondary structures during the two simulations is shown in Figure 6.3 together with the solvent-accessible surface area of each residue in the average structures. In general, the protein has nearly the same structure in the two simulations. The biggest difference occurs between residues number 110 and 120, where the protein in SPC_{100} (right handed side) forms a turn most of the time and sometimes a β_{10} or a π-helix, whereas in SPC it forms mostly an α-helix. The direct exposure to the solvent does not increase the differences between the two simulations. The β-sheet around residue 200 has a very low accessibility but is more unstable in the simulation in SPC_{100} than in the one in SPC. However, residue 60 with a very high accessibility forms a stable turn in both simulations.

The energies and their fluctuations, as well as some other properties, i.e. temperature, volume of the box and pressure, are given in Table 6.1. As equilibrium properties, they should not be

<table>
<thead>
<tr>
<th>Property</th>
<th>SPC</th>
<th>SPC_{100}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_{\text{tot}}$ (kJ/mol)</td>
<td>$-285526$</td>
<td>$-287294$</td>
</tr>
<tr>
<td>$E_{\text{pot}}$ (kJ/mol)</td>
<td>$-350760$</td>
<td>$-351974$</td>
</tr>
<tr>
<td>$E_{\text{ele}}$ (kJ/mol)</td>
<td>$-408172$</td>
<td>$-410075$</td>
</tr>
<tr>
<td>$E_{\text{vdW}}$ (kJ/mol)</td>
<td>$44892$</td>
<td>$45351$</td>
</tr>
<tr>
<td>$T$ (K)</td>
<td>$302.1$</td>
<td>$299.5$</td>
</tr>
<tr>
<td>$T_{\text{protein}}$ (K)</td>
<td>$296.6$</td>
<td>$295.2$</td>
</tr>
<tr>
<td>$T_{\text{solvent}}$ (K)</td>
<td>$302.6$</td>
<td>$299.9$</td>
</tr>
<tr>
<td>$V$ (nm$^3$)</td>
<td>$272.6$</td>
<td>$272.5$</td>
</tr>
<tr>
<td>$p$ (kJ mol$^{-1}$ nm$^{-3}$)</td>
<td>$0.16$</td>
<td>$0.07$</td>
</tr>
</tbody>
</table>

Table 6.1: Average energies, temperatures ($T$), box volumes ($V$) and pressures ($p$) and their fluctuations (RMSF) during the two 3-ns simulations of the protein factor Xa in solution. In this table, SPC indicates the low-viscosity solvent, SPC_{100} indicates the high-viscosity solvent, $E_{\text{tot}}$ is the total energy, $E_{\text{pot}}$ is the potential energy, $E_{\text{ele}}$ is the electrostatic energy, $E_{\text{vdW}}$ is the van der Waals energy, $T$ is the total temperature, $T_{\text{protein}}$ is the temperature of the protein, $T_{\text{solvent}}$ is the temperature of the solvent, $V$ is the volume of the simulation box and $p$ is the pressure during the simulation.

affected by the changed viscosity of the solvent. It can be seen in the table that those properties are hardly affected by the solvent, only the pressure differs by more than 1% between the two simulations. However, pressure fluctuations are generally large in MD simulations, as can also be seen from its RMSF. This difference probably results from insufficient sampling for the pressure. The simulations were performed with coupling to a pressure bath, so real differences in density between the two simulations should result in different average box volumes, not in a different pressure. The box volumes, however, differ only by about 0.4%.

Typical atom-positional correlation functions are shown in Figure 6.4. From these functions, for all atoms the atom-positional correlation time of every atom was calculated and is shown in
Figure 6.2: In the top part, the atom-positional RMSD from the starting structure is shown for both simulations for all protein atoms as well as Cα atoms. In the middle part, the atom-positional RMSF around the average structure of the simulations is shown, and in the bottom part, the RMSD of one simulation from the average structure of the other simulation is shown.

Figure 6.3: The protein secondary structure before energy minimisation (to the left), during the two simulations after every 100 ps and the solvent-accessible surface area (normalised to a range of 0–1) of the average structure of each simulation. In the middle part, the structures and the accessible surface area of the simulation in SPC are shown, and in the right part, the structures and the accessible surface area of the simulation in SPC_{100} are shown.
6.4. Results and Discussion

Figure 6.5 for both simulations. The atoms with the longest correlation times (towards red in the colour scale) are usually part of a side chain on the outside of the protein. Also, the residues with a long correlation time tend to be on the outside, as can be seen in Figure 6.6.

![Graph showing atom-positional correlation functions for four representative atoms](image)

**Figure 6.4:** The atom-positional correlation functions of four representative atoms of the protein simulated in the high-viscosity solvent (SPC$_{100}$) are shown.

The distribution of the atom-positional correlation times is shown in Figure 6.7. The simulation in SPC has a large peak for correlation times below 20 ps, whereas the correlation times in SPC$_{100}$ are more widely distributed and on average longer than in SPC. This shows that the motions in SPC are faster, which was to be expected.

The atoms in the simulation with SPC as solvent generally have a shorter correlation time than those in the simulation with SPC$_{100}$ as a solvent. This effect is visible through the whole protein, not only on the outside, where the influence of the solvent is most obvious. Although atoms with a big difference in correlation time tend to lie more on the outside of the protein, as can be seen in the upper left panel of Figure 6.8, the reason for this is that more atoms are close to the solvent and that the correlation times themselves are longer on the outside of the protein. The average difference in correlation time, as shown in the lower left panel of Figure 6.8, hardly changes with increasing distance from the solvent. When considering the relative difference, i.e. the difference in correlation time divided by the average correlation time of both simulations (right panels in Figure 6.8), one sees that this difference even becomes larger towards the protein centre.

In SPC, there are 4365 dihedral angle transitions of 120° during 3 ns, but there are only 1512 in SPC$_{100}$. The dependence of the number of transitions on the distance to the solvent is shown in Figure 6.9. Most dihedral angles with many transitions are close to the solvent because the protein is more mobile on the outside. This is more evident for the results of the simulation in SPC than in SPC$_{100}$. In SPC more dihedral angles have many transitions than in SPC$_{100}$. This shows, just as the results for the correlation times showed, that the motion in SPC is faster.
6.5 Summary and Conclusion

An analysis of the two 3-ns MD simulations of the 235-residue protein factor Xa in a solvent of high viscosity and in a solvent of low viscosity showed that the equilibrium properties of the protein do not depend significantly on the solvent viscosity, as is required by statistical mechanics. Thus, the secondary structure is nearly the same in both simulations. The dynamic properties are affected by the solvent viscosity. In a solvent with a high viscosity, the protein atoms move more slowly. This is the case not only for the atoms on the surface of the protein in direct contact with the solvent but also for atoms in the centre of the protein. Although there are on the outside more atoms that are heavily affected by the solvent, the average influence is as
Figure 6.7: The distribution of the atom-positional correlation times of all atoms of factor Xa. The inset shows the distribution for the correlation times below 100 ps.

Figure 6.8: The atom-positional correlation times of factor Xa as a function of distance to the solvent. On the left side, the correlation time in SPC (low viscosity) minus the correlation time in SPC$_{100}$ (high viscosity) is shown. In the bottom part, the lines show the average of the difference $\tau_C^{\text{SPC}} - \tau_C^{\text{SPC}_{100}}$ (solid line) and the average of the absolute difference (dashed line). On the right side, the relative difference of the correlation times is shown, again together with the average and the average of the absolute relative differences.
Figure 6.9: The number of dihedral angle transitions of $120^\circ$ as a function of the distance to solvent. The distance to solvent shown is the distance of the second atom forming the dihedral angle to the nearest solvent atom.

big in the centre of the protein as on the surface. This shows that variations in solvent viscosities do influence atom dynamics in the protein centre.
Chapter 7

Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal

7.1 Abstract

Eight molecular dynamics simulations of a ubiquitin crystal unit cell were performed in order to investigate the effect of different schemes to treat the long-ranged electrostatic interactions as well as the need to include counter ions. A crystal system was chosen as the test system, as the higher charge density compared to a protein in solution makes it more sensitive to the way of treating the electrostatic interactions. Three different schemes of treating the long-ranged interactions were compared: straight cut-off, reaction-field approximation and a lattice-sum method (P3M). For each of these schemes two simulations were performed, one with and one without the counter ions. Two additional simulations with a reaction-field force and different initial placements of the counter ions were performed in order to examine the effect of the initial positions of the ions. The inclusion of long-range electrostatic interactions using either a reaction-field or a lattice-sum method proved to be necessary for the simulation of crystals. These two schemes did not differ much in their ability to reproduce the crystallographic structure. The inclusion of counter ions on the other hand seems not necessary for obtaining a stable simulation. The initial positions of the ions have a visible but small effect on the simulation.

7.2 Introduction

Adequate treatment of electrostatic interactions in computer simulations of molecular systems is a difficult problem. On the one hand, these interactions can only be treated in an approximate manner. On the other hand, since these interactions are long-ranged (proportional to the inverse distance between charges, i.e. $r^{-1}$) specific approximations have a strong influence on both the computational expense and the accuracy of a simulation. Up to recently, the most common approximation has been to neglect the effect of Coulomb interactions beyond a certain cut-off
distance $R_C$. To decrease the effect of cut-off truncation, the atomic charges can be collected into charge groups with no net charge whenever possible. If the cut-off truncation is performed on the basis of such charge groups, the interaction between neutral charge groups becomes dipolar and decays with distance much faster than the Coulomb interaction (i.e. with $r^{-3}$), thereby reducing the effect of truncation [52]. Additionally, an increase of the cut-off radius with a limited increase in computational costs can be achieved by using a twin-range cut off. In this method, interactions within a short-range cut off are calculated every simulation step, whereas the interactions within a range delimited by the short-range cut off and a long-range cut off are calculated only every 5 or 10 steps and are assumed constant in between [156]. In this case, the effective cut off can be increased with little additional computational costs, at the expense of neglecting high-frequency fluctuations in the intermediate-ranged electrostatic interactions. Many studies have demonstrated that the use of cut-off truncation induces strong artifacts in simulated properties [8, 157–161], inducing renewed interest in the use of alternative methods. Among these, one may distinguish reaction-field and lattice-sum methods.

In the reaction-field method the medium outside the cut-off sphere of an atom (or charge group) is modelled as a dielectric continuum of specified relative permittivity $\varepsilon_{RF}$. The inclusion of the polarisation of this continuum leads to a modification of the Coulomb electrostatic interactions within the cut-off range into the effective interaction [8]

$$V_{CRF} = \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon_1} \left( \frac{1}{r_{ij}} - \frac{\varepsilon_1 - \varepsilon_{RF}}{\varepsilon_1 + 2\varepsilon_{RF}} \frac{r_{ij}^2}{R_C^3} - \frac{1 - \varepsilon_{RF}}{\varepsilon_1 + 2\varepsilon_{RF}} \frac{1}{R_C} \right)$$ (7.1)

The inclusion of the last (constant) term leads to $V_{CRF}(R_C) = 0$. The representation of the medium outside the cut-off sphere as a homogeneous dielectric continuum is clearly an approximation for the simulation of heterogeneous condensed phase systems such as macromolecules in solution and crystals. The effects of this approximation are still poorly characterised.

In lattice-sum methods, the system is assumed to be a crystal formed by the computational box surrounded by an infinite lattice of periodic copies of itself. For such a system, the electrostatic interaction energies and forces can be computed exactly (within a specified numerical accuracy) through the Ewald [9, 56] or particle-particle–particle-mesh (P³M) methods [10]. However, the approximation of exact periodicity may lead to artifacts in simulations of liquids and solutions [13]. Although certainly better suited for simulations in the crystalline state, lattice-sum methods still describe an ideal crystal exempt of any static or dynamic disorder.

Besides the treatment of long-ranged electrostatic interactions, the representation of the counter ion density may have a strong influence on the outcome of condensed-phase simulations. In the case of macromolecules in solution or in crystals, the system may either be simulated without counter ions (i.e. with a net charge if the macromolecule is non-neutral) with a minimum set of counter ions (i.e. to neutralise a charged macromolecule) or with an excess of counter ions. Simulations of solvated proteins are often performed without the addition of counter ions (i.e. in the limit of zero ionic strength), since their total charge is usually small compared to the system size. However, Ibragimova and Wade [162] found that for unstable proteins, not only a minimal set of counter ions, but even an excess of counter ions to represent a ionic strength higher than the experimental one was necessary to maintain a stable structure.

Several studies have investigated the influence of the different treatments of the long-range electrostatic interactions in macromolecular simulations. The use of a straight cut-off truncation
has been found to destabilise the native (experimentally derived) macromolecular conformation compared to simulations using lattice-sum methods, in the case of ubiquitin [159, 160], RNA, and DNA [159]. Schreiber and Steinhauser [157] found that cut-off truncation distorts the dipole-dipole correlations in water, which affects conformational equilibria of peptides. On the other hand, the artificial periodicity introduced by lattice-sum methods may also perturb the relative stabilities of unfolded and folded states of peptides [13]. Finally, the inclusion of a reaction-field correction into the long-ranged electrostatic interactions has been found essential in another study of ubiquitin crystals [161].

Nearly all the above studies focus on simulation of solutions (see however [161]). In the present study, we report a detailed comparison of different treatments of long-range electrostatic interactions in a biomolecular crystal. In some respects, crystals represent excellent test systems for such an investigation due to the high charge density and reduced screening by water molecules. Furthermore, since the diffusion of counter ions in a crystal is very slow, their arbitrary placement (in the absence of experimental data) may have a strong and uncontrolled influence on the simulated properties. Therefore, in the simulations including counter ions, we also compare three alternative initial placements of the ions.

## 7.3 Method

A total of eight 2 ns-simulations (see Table 7.1) of a ubiquitin crystal unit cell (4 protein molecules) were performed with the GROMOS96 [4, 5] simulation package and force field version 43A1. Among these simulations, six were performed to investigate the effect of the treat-

<table>
<thead>
<tr>
<th>Simulation label</th>
<th>Scheme to treat electrostatics</th>
<th>Presence of counter ions</th>
<th>Initial placement of ions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>cut off</td>
<td>no</td>
<td>–</td>
</tr>
<tr>
<td>CI</td>
<td>cut off</td>
<td>yes</td>
<td>best</td>
</tr>
<tr>
<td>LN</td>
<td>lattice sum</td>
<td>no</td>
<td>–</td>
</tr>
<tr>
<td>LI</td>
<td>lattice sum</td>
<td>yes</td>
<td>best</td>
</tr>
<tr>
<td>RN</td>
<td>reaction field</td>
<td>no</td>
<td>–</td>
</tr>
<tr>
<td>RI</td>
<td>reaction field</td>
<td>yes</td>
<td>best</td>
</tr>
<tr>
<td>RI2</td>
<td>reaction field</td>
<td>yes</td>
<td>second best</td>
</tr>
<tr>
<td>RI3</td>
<td>reaction field</td>
<td>yes</td>
<td>worst</td>
</tr>
</tbody>
</table>

**Table 7.1:** Treatment of electrostatic interactions and counter-ion density for the eight simulations reported in the present study. Lattice-sum simulations employed the P3M method. The best placement of anionic counter ions refers to ions placed in turn at sites of highest (solute and solvent) electrostatic potential. The second best corresponds to determining 8 sites using the same method, and placing the four counter ions at the least favourable 4 positions. The worst corresponds to placing them in turn at sites of lowest electrostatic potential.
ment of electrostatic interactions, while the two other simulations were performed to investigate the influence of the initial ion placement.

### 7.3.1 Simulation set-up

The ubiquitin molecule consists of 76 amino acids. For the simulation, the hydrogen atoms attached to aliphatic carbon atoms were incorporated into these (united atoms) resulting in a total of 761 atoms per molecule. The \( P2_12_12_1 \) crystal unit cell contains four symmetry-related molecules. Initial coordinates were taken from the crystal structure of entry 1UBQ [163] of the Brookhaven Protein Database [164] crystalised at \( T = 277 \) K. The starting coordinates of the four ubiquitin molecules were obtained by performing the \( P2_12_12_1 \) symmetry transformation (see Figure 7.1). The four proteins were hydrated by placing 232 SPC water molecules

Figure 7.1: The crystallographic unit cell of ubiquitin contains 4 proteins. Cylinders indicate helices (residues 23–34, 38–40 and 56–59) and arrows \( \beta \)-strands (residues 2–6, 12–16, 41–45 and 66–71). The “best” (see Table 7.1) initial placement of the four chloride anions used in three of the simulations is also displayed (solid balls).

Three different schemes were applied for the treatment of the long-range electrostatic interactions: simple cut off, reaction field, and lattice sum (\( P^3M \)). For each scheme, two simulations were performed, one in the presence and one in the absence of counter ions (see below). In the two simulations with straight cut-off truncation, a twin-range method with short- and long-range cut off of 0.8 nm and 1.4 nm, respectively, was applied. The short-range non-bonded pair list was updated every 5 timesteps and the intermediate-ranged interactions re-evaluated simultaneously. The twin-range method was applied to van der Waals interactions as well as to electrostatic interactions. In the reaction-field simulations the same twin-range method was applied, but a reaction-field correction (see Equation (7.1)) [8] was applied beyond the long-range cut-off distance with a relative permittivity of 54 as appropriate for SPC water [68]. In the simulations


with $P^3M$ the twin-range method was only used for van der Waals interactions. A spherical hat charge-shaping function [166] of width 0.55 nm was used to separate real-space and reciprocal-space interactions. Real-space interactions were only effective within the short-range cut off and thus computed every step. The numbers of grid points in the $x$-, $y$-, and $z$-direction were set to $N_x = 84$, $N_y = 70$ and $N_z = 48$. Conducting boundary conditions were applied. These choices resulted in a root-mean-square force error of 0.128 kJ mol$^{-1}$ nm$^{-1}$ [167].

Since the experimental crystal structure corresponds to a pH of 5.6, the residues Glu and Asp were taken to be deprotonated, and the residues Lys, Arg and His to be protonated, leading to a charge of +1 e per ubiquitin molecule. In the simulations including counter ions, the four water molecules with the highest electrostatic potential due to solute and solvent were replaced successively by four chloride ions ("best" placement). The resulting initial placement is displayed in Figure 7.1. It turned out that this procedure leads to ion positions which are identical relative to the closest protein. Two additional simulations (employing the reaction-field method) were performed to investigate the effect of the initial ion placement. In the first one, eight water molecules with the highest electrostatic potential were determined as before, but only the four with the lowest potential were replaced by Cl$^-$ anions ("second best" placement). In the second one, the four water molecules at the lowest (least favourable) electrostatic potential were exchanged for counter ions ("worst" placement).

The simulations were performed at constant volume and temperature. Initial velocities were taken from a Maxwell distribution at 300 K. Proteins and solvent were coupled separately to a heat bath at 300 K with a coupling time $\tau_F$ of 0.1 ps [7]. The bond lengths were constrained using the SHAKE procedure [14] with a relative tolerance of $10^{-4}$. A time step of 2 fs was used to integrate the equations of motion, except in the simulation with straight cut-off truncation and counter ions. In this case, because of repeated failures of the SHAKE algorithm, the time step had to be reduced to 1 fs. The centre of mass motion was removed every 20 ps.

Prior to simulation, the initial coordinates were refined by 20 cycles steepest-descent energy minimisation, the protein atoms being harmonically restrained to their initial positions with a force constant of 25000 kJ mol$^{-1}$ nm$^{-2}$. Further equilibration was achieved by a succession of 12 MD simulation periods of 1 ps at 300 K, gradually lowering the restraining force constant to zero.

### 7.3.2 Analysis of the simulations

The different simulations were analysed in terms of the energy components, secondary structure, root-mean-square deviation (RMSD) from the X-ray structure, root-mean-square fluctuation (RMSF) around the average position for each residue, and, in the presence of Cl$^-$ anions, diffusion constant of the ions. The energy components were averaged over the last 1700 ps of the simulation, leaving the first 300 ps as equilibration. The secondary structure was analysed every 100 ps using the program STRIDE based on the algorithm of Frishman and Argos [153]. The RMSD values of all protein atoms were calculated after performing a translational and rotational fit to the crystal structure based on the $C_\alpha$-atoms of the molecule considered. For calculating the RMSF of the protein residues, a fit to the $C_\alpha$-atom positions of the analysed protein was also performed. The calculation was done by determining the RMSF of each residue’s centre of mass over successive windows of 100 ps after 300 ps equilibration, and averaging over the results of the 17 windows. No fitting was performed for the calculation of the RMSF of the Cl$^-$ ions and
the RMSF was calculated over the 1700 ns without averaging over smaller time windows.

The diffusion constant $D$ of the ions was calculated through the Einstein relation

$$D = \lim_{t \to \infty} \frac{\langle (r(t + \tau) - r(t))^2 \rangle}{6t}$$

where $r(t)$ is the position of an ion at time $t$ and $\langle \rangle$ denotes averaging over all starting times $\tau$. In practice, because the mean-square deviation of the ion as a function of time shows two regimes, two diffusion coefficients were calculated, one corresponding to $t$ up to 50 ps, and the second corresponding to $t$ up to 200 ps.

### 7.4 Results

#### 7.4.1 Energetic properties

We first investigate the influence of different electrostatic schemes on the energetic and structural properties of the system. The averages of the energy components for the eight simulations are reported in Table 7.2. The total energy displays significant variations between the different simulations. Inclusion of counter ions systematically decreases the energy by about 800 kJ/mol for straight cut-off truncation and by about 1200 kJ/mol for the reaction-field and lattice-sum schemes. This effect is almost exclusively due to a decrease in the electrostatic energy, and can be traced to the additional (favourable) electrostatic interactions of the ions with the solute and solvent molecules. Inclusion of a reaction-field correction increases the energy by about 2500 kJ/mol compared to straight cut-off truncation. The P^3M method decreases the energy by about 6500 kJ/mol compared to straight cut-off truncation. All these effects result solely from the difference in electrostatic energies. The van der Waals energies are hardly affected by the different treatments.

<table>
<thead>
<tr>
<th></th>
<th>$E_{\text{tot}}$</th>
<th>$E_{\text{vdW}}$</th>
<th>$E_{\text{ele}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>-56.5 MJ mol</td>
<td>-7.6 MJ mol</td>
<td>-71.4 MJ mol</td>
</tr>
<tr>
<td>CI</td>
<td>-57.3 MJ mol</td>
<td>-7.3 MJ mol</td>
<td>-72.2 MJ mol</td>
</tr>
<tr>
<td>LN</td>
<td>-62.9 MJ mol</td>
<td>-7.2 MJ mol</td>
<td>-77.1 MJ mol</td>
</tr>
<tr>
<td>LI</td>
<td>-64.2 MJ mol</td>
<td>-7.1 MJ mol</td>
<td>-78.4 MJ mol</td>
</tr>
<tr>
<td>RN</td>
<td>-53.8 MJ mol</td>
<td>-7.4 MJ mol</td>
<td>-68.3 MJ mol</td>
</tr>
<tr>
<td>RI</td>
<td>-55.0 MJ mol</td>
<td>-7.3 MJ mol</td>
<td>-69.5 MJ mol</td>
</tr>
<tr>
<td>RI2</td>
<td>-54.9 MJ mol</td>
<td>-7.3 MJ mol</td>
<td>-69.4 MJ mol</td>
</tr>
<tr>
<td>RI3</td>
<td>-54.9 MJ mol</td>
<td>-7.3 MJ mol</td>
<td>-69.4 MJ mol</td>
</tr>
</tbody>
</table>

Table 7.2: The averages ($\langle \rangle$) and the root-mean-square fluctuations ($\sigma$) of the energy components over 1700 ps of the eight simulations. Total energy: $E_{\text{tot}}$, van der Waals energy: $E_{\text{vdW}}$, electrostatic energy: $E_{\text{ele}}$. 
7.4. Results

7.4.2 Protein properties

The secondary structure evolution during the eight simulations is shown in Figure 7.2. The structure is stable and remains essentially native-like in all simulations. The long α-helices appear to be stable during the eight simulations, except for some temporary disruptions of the H-bonded pattern of the C-terminal region of the long α-helix in protein 3 for the CI, LN and RI simulations and in protein 4 for the RI3 simulation. Specific β-strands appear to be more labile in some simulations, such as the second β-strand of protein 1 in the LI simulation, the first and second β-strands of protein 2 in the RN simulation and the third β-strand of protein 3 in the LN simulation. Other differences can be observed around residues 38–40 and 56–59 in the four proteins, where interchanges between turn and α- and 310-helix are observed. However, besides these small differences, no specific scheme can be identified as particularly stabilising or disruptive towards secondary structure.

The root-mean-square deviation from the X-ray structure is shown in Figure 7.3 for each protein in the unit cell, as well as averaged over the four proteins. The most significant deviations are systematically observed for the simulations employing straight cut-off truncation with (CI) and, to a lesser extent, without (CN) counter ions. These two schemes also lead to the most significant differences between the behaviour of individual proteins, the deviation being larger for proteins 2 and 3 in CI and for proteins 2 and 4 in CN. For the other schemes, deviations are significantly smaller and more consistent among the four proteins. Just as in the case of straight cut-off truncation, the inclusion of counter ions in a reaction-field simulation (RN → RI) leads to larger deviations. However, the difference is much smaller than in straight cut-off simulations. Note also that the reaction-field simulation without counter ions (RN) leads to the smallest deviation among all the schemes considered. In the lattice-sum simulation, addition of counter ions (LN → LI) has only a very limited (slightly stabilising) effect. In summary, the ordering of the different methods in terms of increasing overall atom-positional RMSD from the X-ray structure is: RN < LI ≈ RI ≈ LN < CN < CI.

7.4.3 Ion properties as function of the scheme to evaluate electrostatic forces

In the second part of the analysis, we focus on the motion of the counter ions in the three simulations which started with the counter ions in the “best” (see Table 7.1) initial placement (CI, LI, RI). The motion of the ions during these simulations is displayed in Figure 7.4 a–c. The motions of the counter ions around their average positions is very limited during the three 2 ns simulations. Although these average positions differ somewhat from the initial placement, the corresponding distances remain relatively small. Since the ions only sample a small fraction of the accessible configurational space within 2 ns, their locations are not randomised during this time. Consequently, simulated properties might depend on the (partially arbitrary) initial placement of the ions (but, see below). To characterise the motion of the ions more quantitatively, Figure 7.5 a–c shows the increase in the atom-positional RMSF of the ions as a function of time. The evolution generally occurs stepwise, each jump corresponding to the ion hopping into another preferential location and undergoing only limited motions between the two jumps. No systematic trends can be found for the mobility of the different ions. For example, ion 2 is the least mobile in the CI and RI simulations and the most mobile in the LI simulation. However,
Figure 7.2: Secondary structure evolution during the eight simulations (see Table 7.1). Residues are numbered consecutively across the four proteins, e.g. the first residue of the second protein is numbered 77. The separation between successive proteins is marked by a dashed line.

Figure 7.3: Root-mean-square deviation (RMSD) from the X-ray structure for all protein atoms, a averaged over all four proteins and b–e for the four individual proteins in the unit cell (see Figure 7.1 for the numbering). The six simulations are indicated in Table 7.1.
7.4. Results

Figure 7.4: Mobility of the counter ions: The initial positions of the counter ions (large balls) together with their successive locations at 5 ps intervals during the simulations (small balls) are displayed for all five simulations that include counter ions (see Table 7.1).

The ions are generally most mobile in the LI simulation and least mobile in the CI simulation, the RI simulation being intermediate. The corresponding diffusion constants calculated from the Einstein relation (Equation (7.2)) on time scales of 50 ps and 200 ps are reported in Table 7.3. While on short time scales, the ions in the reaction-field simulation RI are the most diffusive (see Table 7.3), on the longer time scale, it is the ions in the P3M simulation. The ions in the cut-off simulation are the least diffusive in both regimes. The diffusion in the crystal is reduced by about a factor of 10 compared to the experimental value of $2.033 \cdot 10^{-3}$ nm²/ps at 298 K in water [105].

Since the motion of the counter ions is severely limited in all simulations the question arises how much the protein structure and fluctuations are affected by the (partially arbitrary) initial counter ion placement. To address this question, the atom-positional RMSD from the X-ray structure after 2 ns is displayed in Figure 7.6 a for each residue as a function of the average distance of the residue to the closest counter ion. It can be seen that the residues in close proximity to a chloride ion often display larger deviations from the X-ray structure. The effect is most significant in the straight cut-off simulation and somewhat smaller in the reaction-field and lattice-sum simulations. To investigate the correlation quantitatively, the slope and correlation coefficient $r$ of a least-squares fit to a straight line are reported in Table 7.4. The slope corresponding to the straight cut-off simulation is indeed about 60 % larger than in the two other
7. Electrostatic interactions in crystals

Figure 7.5: Development of the atom-positional root-mean-square fluctuations (RMSF) of the ions as a function of time. Results are given for all five simulations that include counter ions (For the numbering of the ions, see Figure 7.1).

schemes. Of course, this effect may also result from the fact that the residues in close proximity with the ions are located at the surface of the protein and therefore intrinsically more mobile. To address this possibility, the corresponding difference in RMSD between the simulations with and without counter ions is displayed in Figure 7.6 b as a function of the average distance to the closest counter ion. The corresponding correlation coefficient for both, the RMSD and its absolute value, are given in Table 7.4. From these results, it can be seen that close proximity with an ion always increases the deviation with respect to the X-ray structure, i.e. most residues which are close to the ions have a larger RMSD in the simulation with ions than in the simulation without ions. However, the effect (i.e. the ratio of the slopes for RMSD and ARMSD in Table 7.4) is about 5 (cut off) to 17 (P3M) to 26 (RF) times smaller than could be inferred from Figure 7.6 a. Thus, the proximity to the protein-solvent interface is probably more significant to the deviation of a residue from its X-ray position. Another effect appearing in Figure 7.6 b is also revealed by the fitting of the absolute value of ΔRMSD (Table 7.4). The proximity to a
Figure 7.6: Root-mean-square deviation (RMSD, after 3 ns) from the X-ray structure and root-mean-square fluctuations (RMSF, over 100 ps) of the centre of mass of each residue as a function of the average distance \( d_{\text{Cl}} \) to the nearest counter ion. Part a: RMSD values for the three simulations that started with the counter ions in the “best” initial position, b: the difference of the RMSD values between the three simulations with and without counter ions, c: RMSF values for the three simulations that started with the counter ions in the “best” initial position and d: difference in the RMSF values between the three simulations with and without counter ions.
counter ion increases the absolute value of the deviation of $\Delta$RMSD from the average of zero. In other words, a close counter ion may either stabilise or destabilise the native conformation of the residue, compared to the corresponding simulation in the absence of counter ions. Here again, this effect is largest in the cut-off simulation, and 6 and 4 times smaller in the lattice-sum and reaction-field simulations, respectively (compare “slope” for $|\Delta$RMSD| of LI and RI with CI in Table 7.4). These results are in line with the analysis of Figure 7.3 where it was observed that the inclusion of counter ions destabilises the native structure in the cut-off simulations (CN, CI) and has little effect in the lattice-sum simulations (LN, LI).

To investigate the second question regarding the influence of counter ion proximity on residue mobility, the analogs of Figure 7.6 a – b for the RMSF of each residue (calculated over 100 ps intervals) are displayed in Figure 7.6 c – d, and the corresponding fitting parameters are reported in Table 7.4. This figure shows a very similar influence of the ions on the mobility as was observed for the deviation from the X-ray structure. Here also, close proximity with an ion (or the solvent) appears to be correlated with increased mobility (Figure 7.6 c). However, for the fluctuations the effect is of very similar magnitude for the three schemes. One could imagine, that the ions, after provoking an initial displacement of the residues, freeze the motion of the residues in close proximity. Clearly, this is not the case, since residues close to the ions appear to have larger fluctuations. The difference between the simulations with and without ions (Figure 7.6 d), shows a very limited effect of the ion proximity on the increase in the fluctuations. For the lattice-sum simulation LI, inclusion of ions slightly decreases the mobility of residues in close proximity. The opposite is true for the other two simulations CI and RI. Nevertheless, it appears from Figure 7.6 d that the inclusion of counter ions in a cut-off simulation enhances the overall mobility of the residues, whereas in the RF and lattice-sum simulation it is essentially

<table>
<thead>
<tr>
<th>time period</th>
<th>simulation</th>
<th>$D_{Cl_1} \left(10^{-3} \text{ nm}^2/\text{ps}\right)$</th>
<th>$D_{Cl_2} \left(10^{-3} \text{ nm}^2/\text{ps}\right)$</th>
<th>$D_{Cl_3} \left(10^{-3} \text{ nm}^2/\text{ps}\right)$</th>
<th>$D_{Cl_4} \left(10^{-3} \text{ nm}^2/\text{ps}\right)$</th>
<th>$\langle D_{Cl} \rangle \left(10^{-3} \text{ nm}^2/\text{ps}\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 ps</td>
<td>CI</td>
<td>0.24</td>
<td>0.24</td>
<td>0.37</td>
<td>0.27</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>LI</td>
<td>0.40</td>
<td>0.44</td>
<td>0.22</td>
<td>0.36</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.47</td>
<td>0.33</td>
<td>0.38</td>
<td>0.37</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>RI2</td>
<td>0.12</td>
<td>0.16</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>RI3</td>
<td>0.13</td>
<td>0.14</td>
<td>0.37</td>
<td>0.30</td>
<td>0.23</td>
</tr>
<tr>
<td>200 ps</td>
<td>CI</td>
<td>0.12</td>
<td>0.06</td>
<td>0.20</td>
<td>0.08</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>LI</td>
<td>0.27</td>
<td>0.25</td>
<td>0.12</td>
<td>0.25</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>RI2</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>RI3</td>
<td>0.06</td>
<td>0.05</td>
<td>0.14</td>
<td>0.13</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 7.3: Diffusion coefficients of the four ions in the simulations with counter ions. The last column ($\langle D_{Cl} \rangle$) shows the average over the four ions. For each simulation two diffusion coefficients were calculated using Equation 7.2, one for short time periods ($t = 50 \text{ ps}$) and one for long ones ($t = 200 \text{ ps}$), because the mean-square deviation of the ions shows two different regimes.
<table>
<thead>
<tr>
<th></th>
<th>CI (- CN)</th>
<th>LI (- LN)</th>
<th>RI (- RN)</th>
<th>RI2 (- RN)</th>
<th>RI (- RN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>slope</td>
<td>r</td>
<td>slope</td>
<td>r</td>
<td>slope</td>
</tr>
<tr>
<td>RMSD</td>
<td>-0.1234</td>
<td>-0.4043</td>
<td>-0.0756</td>
<td>-0.3881</td>
<td>-0.0744</td>
</tr>
<tr>
<td>Δ RMSD</td>
<td>-0.0240</td>
<td>-0.0906</td>
<td>-0.0045</td>
<td>-0.0647</td>
<td>-0.0028</td>
</tr>
<tr>
<td></td>
<td>-0.0727</td>
<td>-0.3364</td>
<td>-0.0124</td>
<td>-0.2445</td>
<td>-0.0184</td>
</tr>
<tr>
<td>RMSF</td>
<td>-0.0177</td>
<td>-0.5632</td>
<td>-0.0174</td>
<td>-0.5105</td>
<td>-0.0186</td>
</tr>
<tr>
<td>Δ RMSF</td>
<td>0.0018</td>
<td>0.0785</td>
<td>-0.0011</td>
<td>-0.0595</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>-0.0033</td>
<td>-0.2025</td>
<td>-0.0022</td>
<td>-0.1808</td>
<td>-0.0029</td>
</tr>
</tbody>
</table>

**Table 7.4:** Least-square-fit parameters ("slope" and correlation coefficient r) for straight lines representing the RMSD from the X-ray structure after 3 ns and the RMSF over 100 ps of each protein residue as a function of the average distance to the closest Cl ion (see Figure 7.6). Δ stands for the difference between the respective RMSD and RMSF values in the simulations with ions and without ions.
unchanged. However, the correlation of $|\Delta\text{RMSF}|$ with $d_{Cl}$ shows that this effect is rather global and not really correlated with the distance of each residue to the closest counter ion.

### 7.4.4 Ion properties as function of the initial placement of the ions

Because it is obvious from Figure 7.4 that the ions only sample a fraction of the accessible configurational space during the 2 ns, two additional simulations using the reaction-field approximation (RI2 and RI3) with a different initial ion placement were performed to test the influence of this (partially arbitrary) placement. In the first simulation, the ions were placed at the “second best” positions and in the second one at the “worst” positions as determined by the electrostatic potential at water sites during the successive replacement of specific water molecules by counter ions (see Method section). The initial positions of the ions and their motions during the simulations can be seen in Figure 7.4 d – e. The atom-positional RMSD from the X-ray structure averaged over the four protein molecules is shown as a function of time in Figure 7.7. The difference in RMSD value induced by the different initial ion placements is smaller than the difference induced by the choice of a different scheme to treat the long-ranged electrostatic interactions (see Figure 7.3 a). Surprisingly, the ions that were placed in the “second best” positions (RI2) lead to smaller deviations compared to the “best” placement (RI). This finding agrees with the observations, that the atom-positional RMSF (see Figure 7.5) and the diffusion constants (see Table 7.3) of the ions are lowest in the RI2 simulation and that the ions induce the smallest effect on the residue deviations and fluctuations (see the values of the slope for $|\Delta\text{RMSD}|$ and $|\Delta\text{RMSF}|$ in Table 7.4). As expected, however, the “worst” placement leads to the largest deviations. From the absence of a systematic trend among these three simulations,
it appears that the electrostatic potential at a certain position may not be the best criterion to determine the optimal position for the ions. However, the difference between the three placements remains quite small in terms of atom-positional RMSD of the proteins from the X-ray structure (see Figure 7.7) and is not strong enough to affect the secondary structure of the proteins (see Figure 7.2).

7.5 Conclusion

In the present study, we performed a systematic investigation of the effect of the use of various schemes to treat long-ranged electrostatic forces, of including or excluding counter ions and of choosing their initial placement in simulations of a protein crystal. The following main conclusions were reached:

A: The ability of the different schemes to preserve the native protein structure (as measured by the atom-positional RMSD from the X-ray structure) is, in decreasing order: reaction field ≈ lattice sum ≫ straight cut off. Though P3M would seem to be a more appropriate scheme to simulate crystals, we did not find it to be more stabilising than the use of a reaction field. It should be stressed, however, that a small deviation is not a guarantee to accuracy since it may also be caused by a more restricted sampling of conformational space. None of the schemes induces a perturbation large enough to affect the secondary structure.

B: The inclusion of counter ions leads to an increase in the atom-positional RMSD of the proteins from the X-ray structure for both the reaction-field and the straight cut-off schemes, and to essentially no change for the lattice-sum scheme.

C: The counter ions, whenever present, sample only a fraction of the configurational space within 2 ns. Their diffusion coefficients are reduced by a factor 10 compared to bulk water solution. The position of counter ions is not randomised during such short simulations. However, the simulations with three different initial placements of the counter ions did not yield very different results.

D: The proximity of a counter ion affects the deviation of a residue from the crystal structure. The deviation of a residue close to a counter ion may be either decreased or (slightly more often) increased compared to the deviation in the absence of counter ions. Such effects are small in reaction-field and lattice-sum methods but significant in cut-off simulations.

E: The inclusion of counter ions affects the short time fluctuations of all residues. On average the fluctuations are increased in the straight cut-off scheme, and unaltered in the reaction-field and in the lattice-sum schemes. This effect is essentially uncorrelated with the proximity of a residue to a counter ion.

F: The initial placement of ions by successive replacement of water molecules on the basis of the local electrostatic potential does not result in the most stable simulation.

All together, these results suggest, that the inclusion of a correction for the long-range electrostatic interactions, either by the reaction-field or by the lattice-sum method is essential to simulate crystals. The inclusion of counter ions seems not necessary.
7. Electrostatic interactions in crystals
Bibliography


Curriculum Vitae

Personal Information

Name                      Regula Walser
Date of birth             May 24, 1972
Place of birth            Zürich, Switzerland
Citizenship              Wolfhalden, Appenzell-Ausserrhoden, Switzerland

Education

1979–1985  Primary school in Frauenfeld
1985–1987  Untergymnasium in Frauenfeld
1987–1992  Oberrealschule in Frauenfeld
1992       Matura Typus C
1992–1997  Study of Chemistry at ETH Zürich
1997       Dipl. Chem. ETH
1997–2000  Ph.D. studies at the Institute of Physical Chemistry at ETH Zürich,
           Prof. Wilfred F. van Gunsteren