Investigation of nanoscale structure formation and function using molecular dynamics simulations

DISSERTATION

University of Dortmund

Department of Physics

Aina Quintilla

December 2007

Contents

Chapter 1

Introduction

In the last decades, computer simulations have become an important research field and a useful tool to interpret experimental results. Increasing computer power has enabled a real cooperation between theory and experiment. By means of computational simulations virtual experiments can be carried out and complement, predict or partly substitute the laboratory trials.

Theoretical methods and computational techniques are used in a very wide range of scientific fields, from physics or material science to chemistry and biology. Within various approaches, molecular systems ranging from small chemical compounds to large biological molecules and material assemblies can be modeled [1]. Chemical reactions and electronic effects in small systems can be investigated with quantum chemical or semiempirical methods; the detailed atomic motion in molecular systems can be described (neglecting the electronic degrees of freedom) with molecular modeling techniques; and larger macromolecular systems can be studied at a more approximate level.

In particular, molecular modeling refers to theoretical methods and computational techniques to simulate the behavior of molecules. The simulated systems have a typical size in the nanometer range and are often of cross-disciplinary interest between physics, biology and chemistry [2]. To date, molecular simulations have been used mostly to understand experimental results but their predictive power is being increasingly used to design new materials with desired properties [3].

Among the molecular simulation techniques, *Molecular Dynamics (MD)* is a common computational tool to model and visualize the motion of atoms and molecules according to their interactions at the classical level. The method of MD has also been termed 'statistical mechanics by numbers' and 'Laplace's vision of Newtonian mechanics' for predicting the future by animating nature's forces and allowing insight into molecular motion at the atomic scale. It connects the macroscopic and therefore measurable characteristics of a system or state functions to their microscopic description through the ergodic hypothesis. While stochastic methods are nondeterministic and are often used to minimize the energy or optimize the structure of systems under investigation, deterministic MD simulations serve as methods to investigate the structure itself, structure formation or dynamics.

MD was originally developed within the field of theoretical physics in the late 1950's and gained popularity first in materials science and later on in biochemistry and biophysics. MD is used to examine the dynamics of atomic-level phenomena that cannot be observed directly, e.g. thin film growth, cluster formation, gas-phase structures equilibration and ion-implantation. MD constitutes an important tool for protein structure determination and refinement when using experimental techniques such as X-ray crystallography [4, 5] and NMR [6, 7]. The biological systems and processes that have been investigated using molecular simulations include protein folding, enzyme catalysis, protein stability, conformational changes associated with biomolecular function, and molecular recognition of proteins, DNA, and membrane complexes.

The multidisciplinary relevance and use of MD is illustrated in Fig. 1.1. Here, we indicate the trend in the size of elements, devices, or building blocks (system size) vs. the time. This graph does not aim to summarize the technological evolution with accuracy, it is rather meant to be a guide for the eye. Fig. 1.1 can thus be understood as visualization and guide through this thesis. The different scientific fields, such as chemistry, biology and physics are indicated in the left axis. In the last decades, the three natural sciences have started to merge. Miniaturization in technical physics from microsystems to nanotechnology, increasing functionalization of biological systems for technical purposes and complexation of chemical systems in the fields of synthetic chemistry and supramolecular chemistry have enabled the convergence of these three disciplines. MD represents a method in the cross-sectional field of nanotechnology allowing for the study of a large variety of systems at the mesoscale.

In this thesis, our main focus is on three concrete systems which we study by using MD: carbon nanotubes, proteins and molecular crystals. They may be regarded as examples coming from the individual scientific fields. Carbon nanotubes are the evidence for the miniaturization process in the field of material science. They constitute a very promising material and their possible applications include components of electrical circuits like nanotube based transistors [8], nanoelectromechanical systems [9, 10], films for use in displays for computers, cell phones, etc. [11]. However, for all these applications, a reliable method for separation of nanotubes is still under intensive investigation [12]. Proteins are

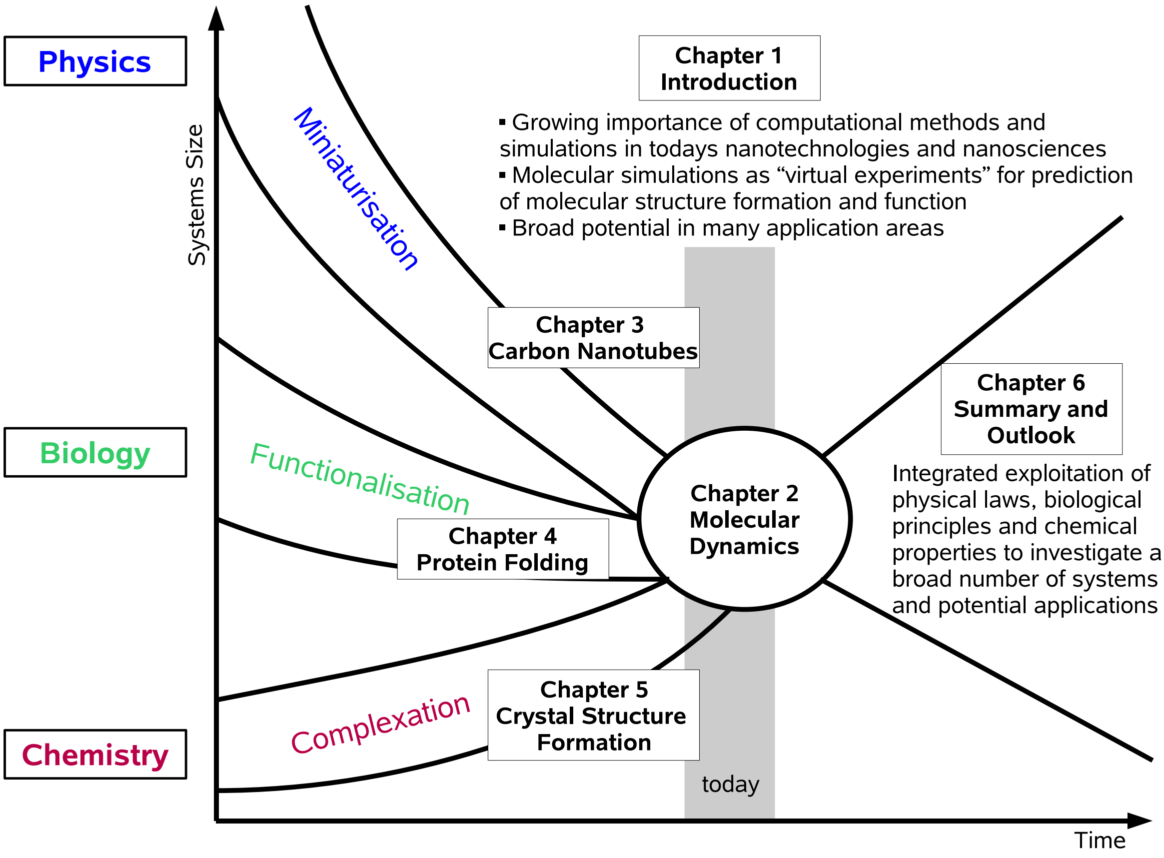


Figure 1.1: *Molecular dynamics (MD) simulations as computational method integrating chemistry, biology and physics in the cross-sectional field of nanotechnology. The present thesis investigates three exemplary systems: Carbon Nanotubes, Proteins and Crystals and discusses the power of MD to address systems at the mesoscale.*

essential biomolecules that constitute the machinery of the cell. All functions of proteins are strictly related to their three-dimensional conformation and therefore, there is an ongoing research on the complete understanding of protein structure formation [13], which might be decisive for medical development. Apart from the medical aspects, technically functionalized biomolecules constitute a promising new research field [14]. As a particular case for chemical complexation, we investigate the possible formation of molecular crystals. The growing of larger crystalline materials based on molecules will enable the fabrication of porous materials with completely tunable properties. In these three differentiated cases, the molecular structure formation and function is investigated by means of molecular dynamics simulations, where the plain exploitation of basic physical laws helps to interpretate biological principles and chemical properties.

The present thesis is organized as follows:

In **chapter 2** we give an introduction to the method of molecular dynamics. Within MD, one uses classical mechanics to describe the physics underlying the molecular models, which typically describe atoms as point charges with an associated mass. Atoms are assigned coordinates and velocities, where the atomic velocities are related to the temperature of the system which is a macroscopic quantity. MD provides information about the dynamic processes with the intrinsic inclusion of temperature effects. We recall Newton equations and present the MD algorithm for solving these equations in section 2.1.

In MD methods, the interactions between neighboring atoms are described by chemical bonds (bonded interactions) and van der Waals forces, typically using the Lennard-Jones potential, while electrostatic interactions are computed based on Coulomb's law (non-bonded interactions). The collective mathematical expression is known as a potential function and is related to the system's internal energy, again a thermodynamic quantity. The potential function computes the molecular potential energy as a sum of energy terms that describe the deviation of bond lengths, bond angles and torsion angles from their equilibrium values. Terms for non-bonded pairs of atoms describe van der Waals and electrostatic interactions. The set of parameters consisting of equilibrium bond lengths, bond angles, partial charge values, force constants and van der Waals parameters are collectively known as a force field. Different implementations of molecular mechanics use slightly different mathematical expressions, and therefore, different constants for the potential function. The common force fields in use today have been developed by using high level quantum calculations and fit to experimental data. In chapter 2.2 we discuss the physics of the force field and address the bonded and non-bonded interactions.

Molecules can be modelled either in vacuum or in the presence of a solvent such as water. Simulations of systems including the presence of solvent molecules are referred to as explicit solvent simulations. If the effect of the solvent is estimated using an empirical mathematical expression simulations are known as implicit solvation simulations. The treatment of the solvent is discussed in section 2.3.

A molecular dynamics simulation computes the behavior of a system as a function of time and solves the Newton's equation of motion by its integration. One can use different integration algorithms which lead to the atomic trajectories in space and time. In section 2.4 we briefly discuss the leap-frog and the verlet algorithm.

As experiments are normally performed at constant temperature and constant pressure, a thermodynamic ensemble with a fixed number of particles, temperature and/or pressure is used for MD simulations to reproduce thereby the experimental conditions. Chapter

2.5 addresses the temperature coupling and pressure coupling.

In **chapter 3** we investigate methods for the separation of carbon nanotubes as a first example arising from material sciences and physics. Our main interest in these systems arises from the lack of methods to massive production of collections of carbon nanotubes with desired properties, making methods for a reliable separation an important challenge. To date, separation techniques using dielectrophoresis or density gradient techniques are studied both, experimentally and theoretically. In chapter 3.1 the characteristic parameters and properties defining carbon nanotubes are introduced.

In chapter 3.2 we give an overview of experiments on carbon nanotubes. We discuss the state of the art of the experiments as the basic input and motivation for our calculations. The current experimental results which come from ongoing work by Dr. Frank Hennrich, at the Institute of Nanotechnology, are based on the density gradient technique, also called equilibrium density-gradient centrifugation. Straightforward theoretical models fail to explain the experiments, for this reason we look at these systems at the atomistic level.

As most recent experimental and theoretical works predict the formation of water inside the carbon nanotubes, we study the effect of water on the density and diameter of carbon nanotubes in chapter 3.3. Our parameterization as well as results of the MD simulations are discussed and the dependence of the density on the structure parameters of the system nanotube-water is analyzed.

In chapter 3.4 we extend our calculations by including the surfactant molecules to the nanotube-water complex, as it turns out to be physically more adequate, and crucial in order to explain the experiments. Again our MD simulations and an adjusted formula for the system´s density are discussed.

We summarize our main findings on carbon nanotubes in chapter 3.5.

In **chapter 4** we study the process of protein folding. The function of proteins is directly related to their three dimensional structure, which results from their chemical composition, or amino acid sequence. As it is experimentally easier to determine the sequence than the structure, many efforts have been made to better understand the folding process of proteins, and to predict their structure from their sequence. The complete understanding of the folding process would allow, for example, to better understand and treat misfolding diseases. In our approach we combine a stochastic simulation method and MD in order to first explore the energy surface of the protein by stochastic optimization and then perform MD simulations to investigate its stability in the dynamical folding process. In chapter 4.1 we give an introduction to proteins, where we discuss protein structure and explain the problem of protein folding.

In chapter 4.2 we analyse the free energy landscape of a protein consisting of 23 amino acids, called 1WQE. The energy is calculated according to the PFF01 force field. Stochastic methods and the basin hopping technique are used to explore the low energy conformations of the protein. The energy landscape of the protein does not have many competing local minima, making it a good candidate to perform MD on and study the kinetics of protein folding.

We complement the structure optimization results with MD simulations and describe the folding kinetics in chapter 4.3. We describe our simulations and the results and discuss the implications of the obtained results.

Again, in chapter 4.4 we summarize our main findings on protein folding.

Chapter 5 addresses the question of crystal structure formation as an example motivated from synthetic or supramolecular chemistry. Molecular crystals are very porous materials whose properties can be tuned according to the building elements. Such crystals could be used as filters for nanoscale particles among other applications. We study the structure formation of two types of molecular crystals based on organic molecules with a carbon based core (adamantan) or based on larger biomolecules: DNA, respectively.

In chapter 5.1 we investigate the feasibility of molecular crystal formation, construction and growth based on organic molecules and hydrogen bonding. We first describe the building blocks for these organic crystals based on an adamantan core and then present the results of our simulations on small complexes forming these crystals.

In chapter 5.2 we focus on molecular crystals based on DNA and organic molecules. In this part, the connectivity between building blocks is due to a DNA self assembly. Many core molecules have been proposed, but the experimental realization is still ongoing. We consider tetrahedral building blocks where the core is an organic molecule and to which DNA single strands are covalently bonded. According to a brownian dynamics study, tetrahedral building blocks would be able to form a crystal if they would be flexible enough. We define the basic building blocks of these DNA crystals. The results of the brownian dynamics study are presented and the MD simulations complementing this study are reported and discussed.

In chapter 5.3, we summarize our main findings on crystal structure formation.

In **chapter 6** we summarize the implications of our findings and briefly review the main results of the thesis on carbon nanotubes, proteins and crystals, independently.

Finally, further challenges and promising candidate systems, which might have relevance for potential future applications, and which may be of interest for further investigation are discussed.

Chapter 2

Molecular Dynamics

Theoretical models are the basis for understanding natural phenomena observed in experiments. However, systems measured in experiments usually contain a large number of atoms ($N_{Avogadro} = 10^{23}$) and in general cannot be solved exactly if theoretically studied. Therefore, theoretical results rely on analytical and numerical approximations. Computer simulation methods, such as molecular dynamics, can provide a picture of the dynamical evolution of a system in case of time dependent events or can be used instead of analytical approximations to generate representative conformations of a molecular system. The computation is not merely intended to recomputate a known or expected result, it rather aims to be a virtual laboratory in which the behavior of a system can be described and predicted.

In the present chapter we summarize the main ideas of molecular dynamics (MD). MD is a computational method based on classical analytical mechanics for solving and studying the motion and the conformational space of an N-particle system. It is the *computer representation of statistical mechanics* [15] as it relates microscopic features of a system to its state functions (like temperature, pressure, internal energy, etc.). In the first section we introduce the principal scheme for calculating the system's dynamics by iteratively solving the Newton equations. In section 2, we discuss in more detail the origin and nature of the particle-particle interactions, which as a whole contribute to the effective potential energy: the force field. In order to save computational time, the environment (i. e. solvent) of the main molecular system can be reduced to an effective field. In section 3, we describe how the environment of the system is treated. In the fourth section, we take a closer look at the integration of the equations of motion. In standard equilibrium MD the energy and volume of the system are conserved, which corresponds to the microcanonical ensemble in statistical mechanics. However, to reproduce experimental conditions, constant temperature and/or pressure might be required. In section 5 we discuss ensemble simulations, where temperature and pressure are kept constant in the MD algorithm.

2.1 Solving the Newton equations

The dynamics of an interacting N particle system is studied by solving the Newton equations of motion:

$$m_i \frac{\partial^2 \vec{r_i}}{\partial t^2} = \vec{F_i} \qquad i = 1, 2 \ldots \mathrm{N}, \tag{2.1}$$

where m_i is the mass of the i^{th} particle, $\vec{r_i}$ its position and $\vec{F_i}$ the force acting on the i^{th} particle

$$\vec{F_i} = -\vec{\nabla}_i V\left(\vec{r}^N\right). \tag{2.2}$$

Here $V(\vec{r}^N)$ is the potential to model the interactions between the N particles in the system, called the **force field**.

Fig. 2.1 illustrates the structure of the MD algorithm:

Input: The starting conformation is described by the initial N-particle positions $\{\vec{r_i}(t_0)\}$ and velocities $\{\vec{v_i}(t_0)\}$ at the initial time t_0. The initial velocities are generated in a Maxwell-Boltzmann distribution:

$$P(|\vec{v_i}(t_0)|) \propto \exp\left[-\frac{m_i \vec{v_i}(t_0)^2}{kT}\right], \tag{2.3}$$

where $P(|\vec{v_i}(t_0)|)$ is the probability of the velocity $\vec{v_i}(t_0)$, T is the temperature and k is the Boltzmann constant.

Processing: The N-particle system's potential $V(\vec{r}_i^N)$ is a function of the positions of every particle. The forces on every atom derived from the potential (Eq. 2.2) and possible constraints are computed. The Newton equations (Eq. 2.1) for each particle are solved and the positions and velocities are updated. This process is repeated for many time steps.

The length of the integration time step is in the order of femtoseconds. The maximum time step in MD simulations is limited by the smallest oscillation period that can be found in the simulated system. The shortest oscillation period as measured from a simulation

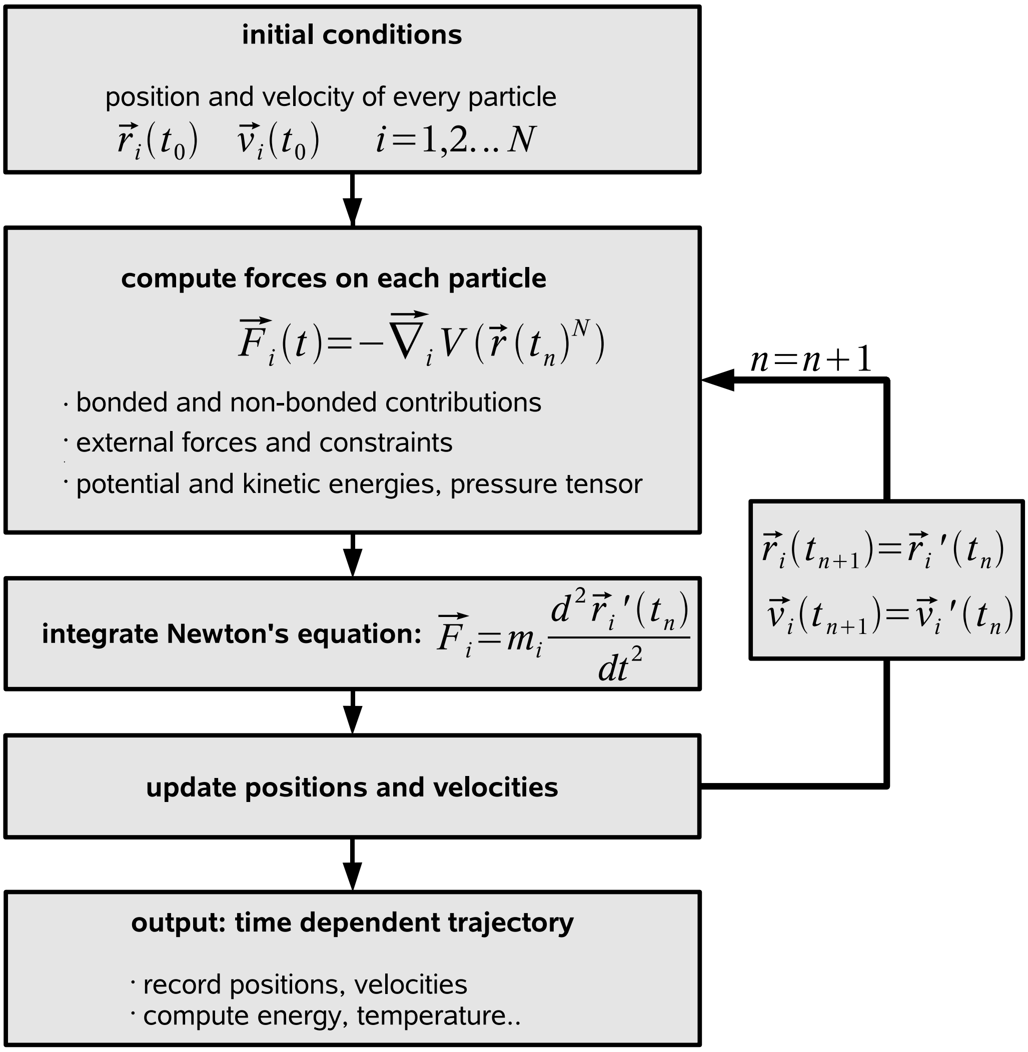


Figure 2.1: *Molecular Dynamics algorithm*

is 13 fs for bond-angle vibrations involving hydrogen atoms. If a minimum of 5 numerical integration steps should be performed per period of a harmonic oscillation in order to integrate it with reasonable accuracy, the maximum time step will be about 3 fs. For typical MD calculations where systems are simulated over 10 to 100 nanoseconds, 10^7 to 10^8 integration steps are performed.

Output: Positions and velocities are recorded regularly in a *trajectory file* that entirely describes the motion of the molecular system. However, not each time step's output is significant. Only every 100 to 1000 steps the coordinates and velocities are copied to the trajectory file.

2.2 Force field

Larger systems than a few atoms in an equilibrium state cannot be handled at a quantum mechanical level. Therefore, to simulate the dynamics of macromolecules one has to make severe approximations and resort to parameterizations based on empirical and quantum calculations. The approximate models resulting from plausible considerations constitute the so called classical forcefields.

Following the Born-Oppenheimer approximation, the electrons are not simulated explicitely at all in classical force fields, but their effects are included in an effective description of the atoms or groups of atoms. Forces depend only on the nuclear positions and in many cases very simple models for the interactions between atoms are used.

In addition to the Born Oppenheimer approximation there are two further essential assumptions: The first one is that the total energy can be written as a sum of potentials of simple physical forces. The second assumption states that the parameterized values for atoms (charge, electronegativity, van der Waals parameters, etc.) or the characteristics of entire molecular parts (bond lengths, angles and torsions) can be applied to similar molecules or larger systems containing similar functional groups.

A force field is built up from two distinct components:

1. The mathematical form of the potential energy as a function of coordinates of the system and
2. the parameters used in this set of functions.

In classical force fields, atoms are classified into atom types depending on the element and the chemical environment (this incorporates the effective description of electronic effects). Each force field has a database with *atom types* characterized, among others by mass, charge or electronegativity, van der Waals radius and van der Waals well depth. The parameterization of the interactions between atoms is force field dependent but the functional form of the potential energy due to interatomic interactions is shared by all classical force fields used for simulations in this thesis.

There are 3 classes of force fields: In Class I force fields, the particle-partice interactions include only first order terms. Additionally to the most basic formulation of particle-partice interactions, higher order terms when treating bonded interactions or polarization terms to describe molecules more accurately can be included. Class II force fields include cross terms in the bonded interactions terms. In Class III force fields, polarization effects are simulated.

A universal force field in which all elements of the periodic table and their interactions in different environments are represented has been developed by Rappé et al. during the 90's [16, 17, 18, 19]. However, an exhaustive parameterization is detrimental to a good performance and many force fields therefore focus on smaller sets of atoms instead of trying to cover the whole periodic table. Nowadays, there are many force fields avaliable and each of them is appropriate for different types of molecules. Many of the Class I classical force fields derive from CHARMM (CHemistry at Harvard Macromolecular Mechanics) [20, 21] that was developed for the simulation of biomolecules.

The classical force fields that we have used for the simulations of biomolecules and other molecular systems are listed below:

- AMBER stands for Assisted Model Building with Energy Refinement [22, 23]. The AMBER force field has been developed essentially for biomolecules. There are several versions of AMBER (parm94, parm95, parm99, parm02 and parm03) that include ongoing new parameterizations, a united atom representation (where methyl groups are simulated as 'one atom' and hydrogen effects are included implicitely), all-atom descriptions and even a polarization term. After a refinement of the AMBER parm99 force field for DNA, AMBER has become the most reliable force field to simulate nucleic acids [24]. AMBER is also a simulation package that includes different modules and allows for a manipulation of the force field parameters and for a possible parameterization of new compounds [25]. It is also possible to simulate organic molecules other than DNA or proteins with the AMBER package, since it includes a more general force field GAFF (General Amber Force Field) [26]. GAFF is compatible to the AMBER force field but has parameters for many additional organic molecules made of C, N, O, H, S, P, F, Cl, Br and I.

- OPLS stands for Optimized Potentials for Liquid Simulations. This force field was developed by Jorgensen and coworkers at Purdue University and later at Yale University [27]. Several sets of OPLS parameters have been published. OPLS was originally formulated as a united force field, OPLS-ua (where ua stands for united atoms). OPLS-aa (all atom) parameterizes every atom explicitly [28]. Later versions include parameters for other specific functional groups and types of molecules such as carbohydrates. A distinctive feature of the OPLS parameters is that they were optimized to fit experimental properties of liquids, such as density and heat of vaporization, in addition to fitting gas-phase torsional profiles.

- GROMOS stands for GROningen MOlecular Simulation. It was developed at Gröningen university and ETH Zurich by Berendsen and van Gunsteren [29]. The GROMOS force field is a united atom force field, optimized with respect to the condensed phase properties of alkanes. The most widely used versions are GROMOS96 (43a1, 43b1 for vacuum simulations, 43a2 with improved alkane dihedrals, 45a3 force field, 53a5 and 53a6 [30]). GROMACS is a GROMOS derived force field as well as a simulation package [31, 32].The GROMACS simulation package includes a growing number of versions of the GROMOS and GROMACS force field, the OPLS all-atom and the Encad force field [33]. The GROMACS simulation package is the fastest MD program avaliable but it does not include implicit solvation.

In a general way, the potential energy can be subdivided into bonded and non-bonded contributions:

$$V_{tot} = V_{bonded} + V_{non-bonded}$$

These contributions are described in the following sections.

2.2.1 Bonded interactions

Interactions between chemically bonded atoms can involve two (1-2 interactions or bond stretching), three (1-3 interactions or angle bending) or four atoms (1-4 interactions or dihedral angle torsion). The simplest possible interactions between atoms that are covalently bonded are represented in Fig. 2.2.

The contribution of the bonded interactions to the total energy of the system is:

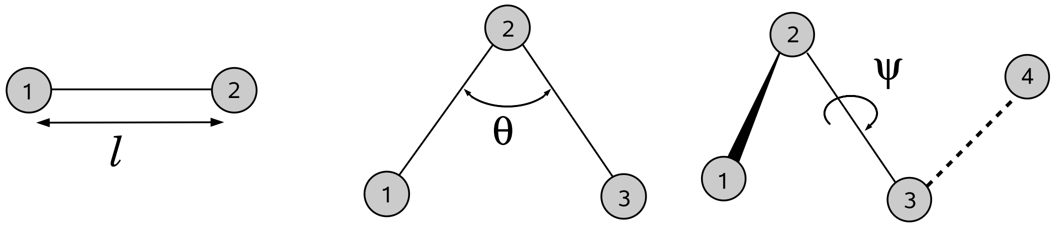
$$V_{bonded} = \sum_{1-2} V_b(r_{ij}) + \sum_{1-3} V_a(\theta_{ijk}) + \sum_{1-4} V_{tor}(\psi_{ijkl}) \tag{2.4}$$

(a) Bond stretching

Bond stretching is generally treated with a harmonic potential $V_b(r_{ij})$ as a function of the distance r_{ij} between bonded atoms i and j

$$V_b(r_{ij}) = \frac{1}{2}\, k_{ij}^b \,(r_{ij} - b_{ij})^2\,, \tag{2.5}$$

for small deviations from equilibrium values and first order vibrational states. Here k_{ij}^b and b_{ij} are the force constant and the reference bond length between particles i and j respectively (In Table 2.1 some values for these constants are given).



(a) Bond stretching (b) Angle bending (c) Torsional rotation

Figure 2.2: *Schematic representation of the different first order bonding interactions: bond stretching (or 1-2 interactions), angle bending (or 1-3 interactions) and rotation around a covalent bond with a torsion angle (or 1-4 interactions).*

For larger deviations from equilibrium, the Morse potential $V_{morse}(r_{ij})$ reproduces vibrational levels better [34]:

$$V_{morse}(r_{ij}) = D_{ij} \left\{1 - exp\left[-\beta_{ij}(r_{ij} - b_{ij})\right]\right\}^2 , \tag{2.6}$$

where D_{ij} is the dissociation energy of the bond between atoms i and j and β_{ij} is related to the well width of the potential. The Morse potential can be explicitely implemented instead of using only the harmonic approximation, but it is rather more time consuming than generating much better results.

In Fig. 2.3 the functional form of the Morse potential and the harmonic approach are represented. The harmonic potential reproduces the more realistic morse potential only for very small deviations around the equilibrium distance.

However, since forces between bonded atoms are very large (see table. 2.1), a large amount of energy would be needed to cause deviations from the equilibrium value. For this reason and because of the fact that classical force fields do not allow bond breaking the use of the harmonic potential is justified when simulating macromolecules at room temperature. Reference values for chemical bonds (b_{ij}) can be experimentally obtained or can be calculated with quantum mechanical methods.[1]

Because bond vibrations occur at very high frequencies, often distances between atoms are just constrained to their equilibrium values. This facilitates the computation of the interatomic interactions, and a longer time step can be used in MD.

---

[1] Since real molecules are vibrating and the Morse potential is not symmetric, the experimentally measured bond length is an average over vibrating states and does not correspond to the minima of the Morse potential [35].

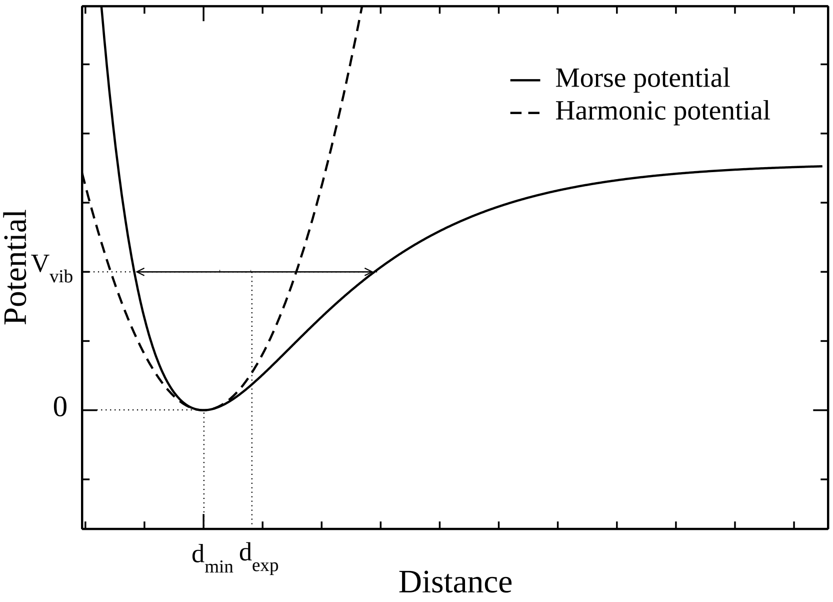


Figure 2.3: *Schematic illustration of the Morse potential (full line) and the harmonic approximation (dashed line) used in classical force fields. The harmonic simplification reproduces the behavior of the bond for particle-particle distances around the minimum of the potential function,* d_{min} *(when* r_{ij} *equals* b_{ij} *in eq. 2.5 and eq. 2.6).* d_{exp} *represents the experimentally measured distance when the bond is in the vibrational state* V_{vib} *indicated by the horizontal dotted line. All the possible distances that the bond adopts corresponding to the energy of this state are indicated by the horizontal arrow.*

(b) Angle bending

Angle bending is also described using an harmonic potential $V_a(\theta_{ijk})$ as function of the angle θ_{ijk} between the atoms i,j and k:

$$V_a(\theta_{ijk}) = \frac{1}{2}k^a_{ijk}\left(\theta_{ijk} - \theta^0_{ijk}\right)^2 \tag{2.7}$$

where k^a_{ijk} is the force constant and θ^0_{ijk} the equilibrium angle. Energy contributions due to angle bending are much lower than bond stretching energies; some values for the equilibrium angle and for the force constant parameterized in the AMBER force field are shown in the table 2.1.

(a) bond

$atom_i$	$atom_j$	$k^b_{ij}(Kcal/mol)$
C	C	310.0
CA	CA	469.0
n1	o	617.5
nf	s	463.5

(b) angle

$atom_i$	$atom_j$	$atom_k$	$k^a_{ijk}(Kcal/mol)$
CT	N3	CT	50.0
C	CA	CA	63.0
o	c	sy	61.9
n2	c	n2	71.8

Table 2.1: *Bonding and angle constants in the Amber99 force field (atom types in capital letters) and General Amber Force Field (atom types lower case) between atom types* $atom_i$ *and* $atom_j$.

(c) Torsional terms

Energy barriers of rotation about chemical bonds were believed to be a consequence of the Pauli exchange interaction and electrostatic repulsion, also called steric clashes. Recently, it has been demonstrated that torsional conformations are mainly due to hyperconjugative interactions, or electron transfer between occupied and unoccupied orbitals [36, 37, 38]. Minimal energy configurations are stabilized by a σ orbital resonance, in the same way, but quantitatively weaker, as benzene rings are stabilized by the π orbital resonance.

Changes in the dihedral angles parameterization involve much more important structural changes than bond stretching and angle bending. Dihedral parameters, together with electrostatic and van der Waals interactions account for the most important changes of the configuration and the relative energy.

The torsional term in the potential function describes the energy needed to rotate around a bond. This energy contribution is not properly represented if only non bonded interactions are taken into account. A third bonded term has to be added to the force field to reproduce the conformational behavior of the molecules. Torsional potentials are often expressed as a cosine of the torsional angle serie's expansion:

$$V_{tor,ijkl} = \sum_n \frac{V_n}{2} \left[1 + cos(n\psi - \gamma)\right] \tag{2.8}$$

where ψ is the torsional angle around the chemical bond between atom2 and atom3 (see Fig. 2.2) and γ is the phase factor that determines the angle for which the torsional potential becomes minimal. V_n indicates the relative energy contribution of the n harmonic. In Fig. 2.4, the first three terms of a cosine serie and the resulting potential are represented. The three dimensional conformations corresponding to some representative angles are shown as well.

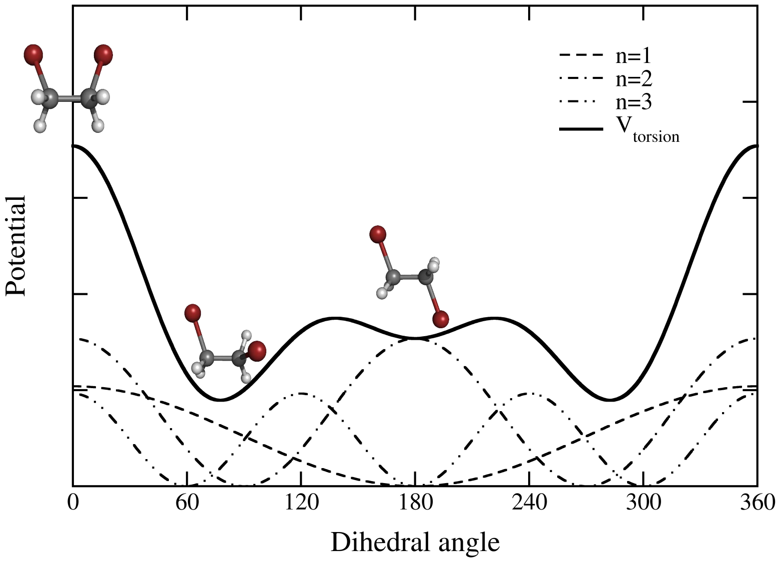


Figure 2.4: *Representation of the torsional function vs. the dihedral angle (in deg.) resulting from considering the first three terms in the cosine expansion. Each harmonic that contributes to the torsional functions is individually represented (dashed lines, as indicated, for* $n = 1, 2, 3$*). The molecular conformations cis and trans corresponding to the 0 degree and 180 degrees torsion angles and the configuration giving minimal energy are illustrated in the graph.*

In the AMBER force field, many of the torsional terms include less than three terms of the cosine expansion but still the energy profile around the torsional bond is accurately enough reproduced together with the non-bonded interactions contributions; only a few require considering higher terms in the summation series.

The GROMOS force field represents the torsion energy with a simple cosine function or with the Ryckaert-Bellemans potential (V_{RB}) for alkanes:

$$V_{RB,ijkl} = \sum_{n=0}^{5} C_n \left[cos(\psi)\right]^n . \tag{2.9}$$

The Ryckaert-Bellemans potentials are also used in the OPLS force field.

To parameterize the torsional parameters, the energy profile, extracted from quantum mechanical calculations or from experimental data, is ‘separated’ into a bonded dihedral energy term and a non-bonded contribution for 1-4 atoms.

Improper dihedrals

Improper dihedrals are meant to keep planar groups planar (*e.g.* aromatic rings) or to prevent molecules from flipping over to their mirror images. They are again represented by an harmonic potential on the torsional angle that defines the planarity of the molecule.

Cross terms

Internal coordinates such as bond stretching and angle bending are coupled. If for example the angle between $\widehat{ijk}$ atoms decreases, atoms i and k approach to each other. The bonds $j - i$ and $j - k$ might stretch to reduce the interaction between the atoms i and k. In a higher order representation, cross term interactions like bond-bond, bond-angle, etc. might be considered as a correction to the harmonic representation of isolated bonded interactions. Force fields including cross terms are called Class II force fields.

2.2.2 Non-bonded interactions

Non-bonded interactions involve atoms that do not belong to the same molecule or atoms of the same molecule that are separated by more than two bonds. While atoms separated by one or two covalent bonds are modelled to interact exclusively through bonded interactions, interactions between atoms that are separated by three bonds (1-4 atoms interactions) have a torsional term but also a non-bonded contribution. Non-bonded 1-4 interactions are reduced by a scaling factor between zero and one to reproduce physical interactions between 1-4 atoms in reality.

Whereas bonded interactions depend on a fixed number of atoms, non-bonded atoms interact principally according to their distance. For this reason non-bonded interactions are computationally more difficult to calculate.

Lennard Jones Interactions

Forces that give rise to deviations from ideal gas behavior are referred to as van der Waals forces. Such forces have an attractive contribution due to *dispersive forces* at long range distances and a repulsive contribution at short distances. The attractive contribution is due to instantaneous dipoles which arise during fluctuations in the electron clouds [39]. The Drude model [40, 41, 42] predicts that the interaction (without considering higher order terms like dipole-quadrupole and so on) is proportional to $-r^{-6}$. The repulsive

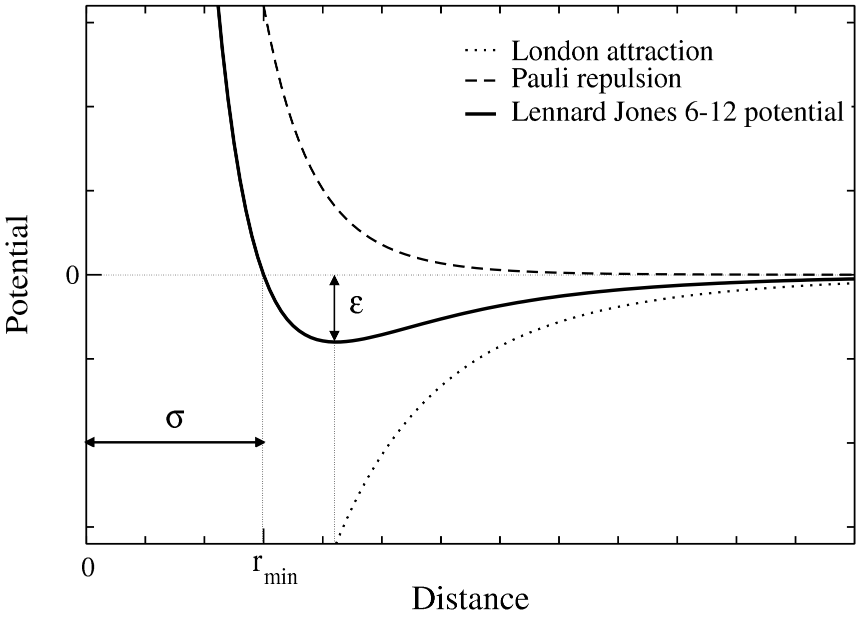


Figure 2.5: *Illustration of the Lennard Jones potential vs. the distance between atoms and the attractive (dispersive) and repulsive (exchange) contributions separately.* r_{min} *is the distance between interacting atoms where the van der Waals potential is minimal,* $V_{LJ}(r_{min}) = -\varepsilon$*, called van der Waals well depth and* σ*, also known as van der Waals radius is the distance for which* $V_{LJ}(r_{min}) = 0$.

contribution at low distances is related to the nuclear repulsion for very short distances and to the overlap forces for short distances; it behaves like r^{-1} and like $e^{\frac{2r}{a_0}}$ respectively, where a_0 is the Bohr radius.

The best known function to model van der Waals forces is the Lennard-Jones 12-6 function $V_{LJ}(r_{ij})$ of the distance r_{ij} between atoms i and j :

$$V_{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} + \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right], \tag{2.10}$$

where ε_{ij} and σ_{ij} are adjustable parameters and represent the well depth of the potential profile and collision diameter respectively. Fig. 2.5 shows the Lennard Jones potential and repulsive and attractive contributions. There are no theoretical arguments for the repulsive r^{12} but this functional form reproduces the physical behavior of rare gases and it is not as difficult as to compute as an exponential function. In the parameterization of a force field every atom type has two Lennard Jones parameters (ε_{ii} and σ_{jj} or $r_{min,ii}$). The relation between these coefficients and the potential interaction coefficients is given by the *combining rules*, equations 2.11 and 2.12 for AMBER, CHARMM and GROMOS, while OPLS applies the geometrical mean for both well depth and radii.

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj}\right) \tag{2.11}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{2.12}$$

Another potential form used to simulate van der Waals interactions from theoretical assumptions more realistic form is the *Buckingham potential*:

$$V_{Buckingham}(r_{ij}) = \varepsilon\left[\frac{6}{\alpha - 6} exp\left[-\alpha\left(r_{ij}/r_{ij,m}\right)\right] - \frac{\alpha}{\alpha - 6}\left(\frac{r_{ij,m}}{r_{ij}}\right)^6\right], \tag{2.13}$$

where α, $r_{ij,m}$ and ε are adjustable constants. Implementing the Buckingham potential is more computational time consuming since an exponential function has to be calculated.

Van der Waals interactions between atoms that are three covalent bonds away are scaled in a different way in each force field[2].

Especially because of this multiplicity of procedures to model 1-4 interactions, it is not possible to exchange parameters from one force field to another.

Van der Waals interactions are usually computed only within a cut-off radius: the function representing the interactions is set to zero for interparticle distances bigger than a certain cut-off, where the magnitude of the interaction is anyway neglectable

Hydrogen bonding

Hydrogen bonding is the interaction between an electronegative atom (H-bond acceptor) and a hydrogen atom that is bonded to an electronegative atom (H-bond donor). It can be visualized as two electronegative atoms 'sharing' one hydrogen atom. The energy of such interaction varies between 5 and 20 kJ/mol according to the environment and to

---

[2] In the AMBER parm94 force field, Lennard Jones parameters are scaled by 0.5 and in AMBER parm99 force field by 0.83 (or 1/1.2).

GROMOS force field uses a special 1-4 Lennard Jones interaction when the torsion energy is represented by a simple cosine function. With Ryckaert-Bellemans potential (see Eq. 2.9) for alkanes, the 1-4 interactions are excluded from the non-bonded list.

In the OPLS force field, van der Waals interactions involving atoms that are only separated by three bonds are scaled with 1/1.2.

the electronegativity of the donor and acceptor. Hydrogen bonds are very important for stabilizing the three dimensional structure of biomolecules; for instance, secondary structure of proteins and base pair complementarity in double helical structure of DNA are mainly explained by hydrogen bonding.

Originally, many force fields included an extra term in the potential function to represent hydrogen bonding [43]. In the recent versions of AMBER and GROMOS force fields, the behavior of hydrogen bonds is modeled by van der Waals and electrostatic potentials.

Electrostatics

Atoms with higher electronegativities are able to attract electrons stronger than atoms with lower electronegativities. This results in a non-uniform distribution of the charge in the molecule. The simplest approach to model uneven charge distribution is to associate *partial charges* to every atom of a molecule depending on its type, its environment and the geometry of the molecule. These partial charges reproduce the electrostatic properties of the molecules.

For simple molecules, the partial charges can be exactely calculated to reproduce the electric moments if the geometry is known. For larger molecules, there are several ways to determine the partial atomic charges as these are not experimentally observable. Population analysis associates a charge with every atom, but this method does not yield very good results because it focuses on the charge conformation rather than on the resulting potential. The most widely used solution is to derive the partial charges that best reproduce the quantum mechanical electrostatic potential [44] at the molecular surface [3]. Charges used by AMBER are derived with such methods.

In class I or additive classical force fields, electrostatic interactions are calculated as Coulomb interactions (eq. 2.15) between point charges q_i and q_j located in the center of

---

[3] The electrostatic potential at every position i is calculated from the wave function ϕ_i^0, and for a given charge distribution ϕ_i^{calc}. The error function R:

$$R = \sum_{i}^{N_{points}} w_i \left(\phi_i^0 - \phi_i^{calc}\right)^2 , \tag{2.14}$$

is minimized (with an iterative least-squares minimization method or with Lagrange multipliers) and a set of equations on the partial charge on every atom k, q_k, is obtained.

The points $i = 1, 2...N_{points}$ where the potential is fitted are between the surface of the molecule, at the van der Waals radius, and 1 or 2 Å beyond, where interactions have to be modelled accurately.

every atom i and j:

$$V_{el}(r_{ij}) = \frac{q_i\, q_j}{4\pi\varepsilon_0\, r_{ij}}, \tag{2.15}$$

where ε_0 is the vacuum electrical permittivity.

The electrostatic interactions between partial charges are often computed within a cut-off distance, with the Lennard Jones interactions, but since they decay as r^{-1} (i.e. much slower than the Lennard Jones interactions) the discontinuity of the potential and the force near the cut-off radius originating non conservation of the energy will be more dramatic. Moreover, charges accumulate at the cut-off radii as an artifact of truncating the electrostatic potential. To avoid all the problems resulting of truncating the electrostatic force at a relative short distance, electrostatic interactions are treated in the reciprocal space, where all the contributions are added after being Fourier transformed, also called Ewald summation [45]. Particle Mesh Ewald (PME) [46] is implemented in GROMACS and permits to calculate accurately the electrostatic interactions without introducing any artifact due to a cut-off.

However, the electronic density is not spherically symmetric around the nucleus. There are molecules that contain lone pairs (like oxygen in a water molecule) or π electron clouds in aromatic ring systems, for example. This aspect is not well reproduced by partial charges centered in the middle of the atoms. Virtual charged particles or charged extra points located on electron donating atoms (O, N, S) reproduce the presence of lone pairs or unequal charge distributions.

Polarizable or nonadditive forcefields, like AMBER(parm02) [47] include electronic polarization in the treatment nonbonded interactions. Electronic polarizability is included in the forcefield potential as a term that describes the energy associated with the polarization of the charge distribution on the i^{th}atom:

$$V_{pol} = \frac{1}{2}\, \vec{\mu}_i \vec{E}_i, \tag{2.16}$$

where $\vec{\mu}_i$ is the induced atomic dipole on the i^{th}atom and $\vec{E}_i$ is the electric field on the i^{th}atom position. Polarization is a cooperative effect, therefore it is modelled using a set of coupled equations which can be solved iteratively. The induced dipoles are first set to zero and calculated in a first approximation as the induced dipoles due to an electric field created only by static charges. The electric field due to this induced dipoles is then recalculated and the induced dipoles deriving from the refined value of the field are calculated. This iterating process continues until the discrepancy between the subsequent values obtained for the dipoles is low.

Class III force fields are based on atom-centered dipole polarizabilities. The atomic dipole is expressed in terms of atomic polarizability of the i^{th}atom (eq. 2.17), which is a parameter that is based on the atom type.

$$\vec{\mu}_i = \alpha_i \vec{E}_i \tag{2.17}$$

Polarizable force fields seem to better reproduce the orientation of hydrogen bonds [48] than additive force fields. AMBER can include the above described polarization in AMBER parm02.

2.3 Treatment of the solvent

Water or salt solutions are the most relevant environments for the systems of interest studied in this thesis. In order to simulate a bulk environment a large number of water molecules is needed. Explicit simulation of all water molecules implies a strong increase of computational time as full trajectories are calculated for all water molecules although we are only interested in the solute behavior. A more efficient method in order to reduce the computational effort is to simulate only the effects of the solvent on the solute by using implicit solvent models [49]. In chapter 4 the dynamics of proteins is studied by making use of this model. In section 2.3.1 we explain models that treat the solvent contribution without having to compute its degrees of freedom.

However, if we are interested explicitely in the behavior of water itself, it has to be explicitely simulated. In chapter 3 we study the behavior of water surrounding carbon nanotubes. A detailed description of solvent-solute interactions is most important, for instance in the simulation of hydrogen bonds. Thereby an explicit representation of solvent molecules is required, since implicit solvent models cannot account for specific solvent-solute hydrogen bonding. In chapter 5 we study the formation of DNA and adamantan cores crystals in which molecules are not preparameterized and where hydrogen bonding is very important for the structure formation. In chapter 3 and in chapter 5, explicit water simulation required. The modeling of explicit water is described in more detail in section 2.3.2.

2.3.1 Implicit solvent models

In molecular dynamics, the key quantity that needs to be computed is the total energy of the molecule or the molecular assembly in its natural environment. If all molecules in the environment are simulated, the total energy E_{tot} is simply the sum of all contributions, as we have detailed before in the force field section. However in general, we are not interested in the detailed properties of the environment. Formally the total energy of a solvated molecule can be also written as

$$E_{tot} = E_{vac} + \Delta G_{solv}, \tag{2.18}$$

where E_{vac} represents the molecule's energy in vacuum (gas-phase), and ΔG_{solv} is the free energy of transferring the molecule from vacuum into solvent, which is called 'solvation free energy'.

When using an implicit solvent model, the solvent degrees of freedom are not considered explicitly and the interactions between solvent and solute are described as a function of the solute coordinates [49]. All solvent effects are accounted for by the term:

$$\Delta G_{solv} = \Delta G_{non-el} + \Delta G_{el}. \tag{2.19}$$

ΔG_{non-el} includes van der Waals interactions and entropy costs for cavity formation. It is taken to be proportional to the total Solvent Accessible Surface Area (SASA) of the molecule, with a proportionality constant derived from experimental solvation energies of small non-polar molecules. ΔG_{el} is the electrostatic contribution to the solvation energy. It can be calculated solving the Poisson Boltzmann (PB) equation (eq. 2.20) for the electrostatic potential $\Phi(\vec{r})$ produced by a molecular charge distribution $\rho_m(\vec{r})$.

$$\vec{\nabla}\varepsilon(\vec{r})\vec{\nabla}\Phi(\vec{r}) = -4\pi\rho_m(\vec{r}) - 4\pi\sum_i c_i^{\infty} z_i \lambda(\vec{r}) q e^{\frac{-z_i q \Phi(\vec{r})}{kT}} \tag{2.20}$$

Here, $\varepsilon(\vec{r})$ represents the position-dependent dielectric constant which equals that of bulk water far away from the molecule and is expected to decrease fairly rapidly across the solute/solvent boundary, $\rho_m(\vec{r})$ includes only molecular charges, the i-summation runs over all ion types. Each term corresponds to the local concentration of ion i, c_i^{∞} is the concentration of ion i at infinite distance from the molecule, z_i is its valency, q is the proton charge, k is the Boltzmann constant, T is the temperature and $\lambda(\vec{r})$ describes the accessibility to ions at position $\vec{r}$.

If the potential is small, the PB equation can be linearized:

$$\vec{\nabla}\varepsilon(\vec{r})\vec{\nabla}\Phi(\vec{r}) = -4\pi\rho_m(\vec{r}) + \kappa^2\varepsilon(r)\Phi(r). \tag{2.21}$$

The electrostatic screening effects enter via the second term on the right-hand side, with the Debye screening parameter

$$\kappa^2 = 8\pi\frac{\sum_i c_i^{\infty} z_i^2 q^2}{2\varepsilon kT}. \tag{2.22}$$

$\kappa \approx 0.1$ Å^{-1} at physiological conditions.

Once the potential $\Phi(\vec{r})$ is computed, the electrostatic part of the solvation free energy is a combination of the classical electrostatic energy, a term originating from mixing mobile species and a term arising due to the solvent [50]. The free energy due to electrostatic effects can be written as:

$$\Delta G_{es} = \frac{1}{2} \int\limits_V \rho^f(\vec{r})\Phi(\vec{r})dV \tag{2.23}$$

where ρ^f contains only the fixed charges and $\Phi(\vec{r})$, the electrostatic potential, is obtained by solving the linearized PB equation.

The PB approach as described above has traditionally been used in calculating electrostatic properties of 'static' configurations. However, in molecular dynamics applications, the associated computational costs are often very high, as the PB equation needs to be solved every time the conformation of the molecule changes. An alternative approach to obtain a reasonable, computationally efficient estimate of the electrostatic part of the solvation energy ΔG_{el} is the analytic generalized Born (GB) method. The GB implicit solvent method is an approximation to ΔG_{el}. Each atom in a molecule is represented as a sphere of radius ρ_i with a charge q_i at its center, the interior of the atom is assumed to be filled uniformly with a material of dielectric constant $\varepsilon = 1$. The molecule is surrounded by a solvent of a high dielectric ε_w ($\varepsilon_w = 80$ for water at 300 K).

The GB model approximates ΔG_{el} by the analytical formula:

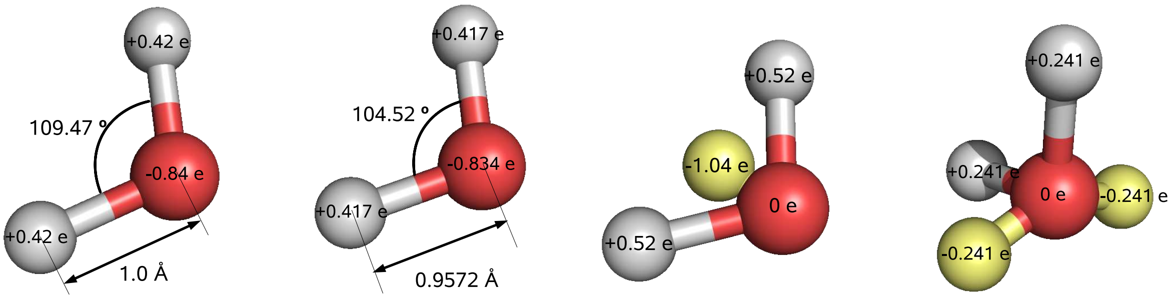
$$\Delta G_{el} \approx \Delta G_{gb} = -\frac{1}{2} \sum_{ij} \frac{q_i a_j}{f_{GB}(r_{ij}, R_i, R_j)} \left(a - \frac{e^{-\kappa f_{gb_{ij}}}}{\varepsilon_W} \right), \tag{2.24}$$

where R_i and R_j are the *effective bond radii* of atoms i and j and f_{GB} is a smooth function of r_{ij}, R_i and R_j [51]. The various GB methods differ in the way of computing the effective bond radii [52, 53]. The effective radii depend on the molecule's conformation, and have to be re-computed every time the conformation changes.

2.3.2 Explicit solvent models

In this work we have always used rigid models for simulating explicit water, but there are also flexible and polarizable water models. The flexible models allow for conformational changes in the molecule and reproduce experimental spectra. The polarizable models include explicit polarization and can reproduce different phases of water.

Rigid water models widely used in MD calculations have either three sites, as in the case of the model SPC (Single Point Charge) [54] and the TIP3P model (Transferable Intermolecular Potential 3 Points) [55], four sites, as the TIP4P model (Transferable Intermolecular Potential 4 Points) [55] or five sites, as the TIP5P model (Transferable Intermolecular Potential 5 Points) [56] for electrostatic interactions. The charge distribution of the water molecule is generally modelled by point charges on the nuclei or on one or two additional fictive sites located in the plane of the molecular frame (TIP4P) of out of it (lone pairs in TIP5P). Van der Waals interactions involve a single interaction point per molecule situated at the oxygen atom. In fig. 2.6 the rigid water models and their main characteristics are represented.



(a) SPC (b) TIP3P (c) TIP4P (d) TIP5P

Figure 2.6: *'Simple' water models. Oxygen atoms are red and hydrogen atoms white. The figure illustrates the ficticious points where the charge is localized (in yellow). Local point charges are indicated on every atom or ficticious point. The distance between the oxygen atom and the hydrogen atoms and the angles formed by the atoms H-O-H for SPC and TIP models are also indicated.*

Water models are parameterized such that the outcome of the computer simulation reproduces as good as possible the properties of real water [57]. The *liquid density* is properly reproduced by TIP4P and TIP5P models whereas SPC and TIP3P models yield smaller values than the experimental density. The *heat of vaporization* is properly reproduced by all rigid models; it quantifies importance of hydrogen bonding. The *diffusivity* is overestimated by all rigid models mentioned above except TIP5P; it expresses the mobility of water molecules, the influence of hydrogen bonds on the translational motion. The *microscopic structure, density maximum, critical parameters* and *dielectric constant* computed using rigid models agree within a reasonable interval with experimental values. To reproduce liquid state properties well, all these models have a dipole moment around 2.3 D instead of the experimental gas-phase value of 1.85 D.

The different classical force fields that we use in our simulations have been developed together with a specific water model. Choosing the water model depends on the force field used in the simulation. OPLS simulations in aqueous solution typically use the TIP4P or TIP5P water model. The SPC model was developed to work together with the GROMOS force field. AMBER force fields are parameterized with TIP3P.

Periodic boundary conditions

When explicit water molecules are used, the molecule we are interested in is normally centered in a virtual box and surrounded by a layer of water between 10-20 Å thick. Boundary conditions at the edge of the box are not physically meaningful, therefore *periodic boundary conditions* are used. The constructed box is treated like a unit cell and is replicated in all directions such that the whole space is filled and particles close to opposite edges interact with each other. In Fig. 2.7, periodic boundary conditions for a two dimensional system are illustrated.

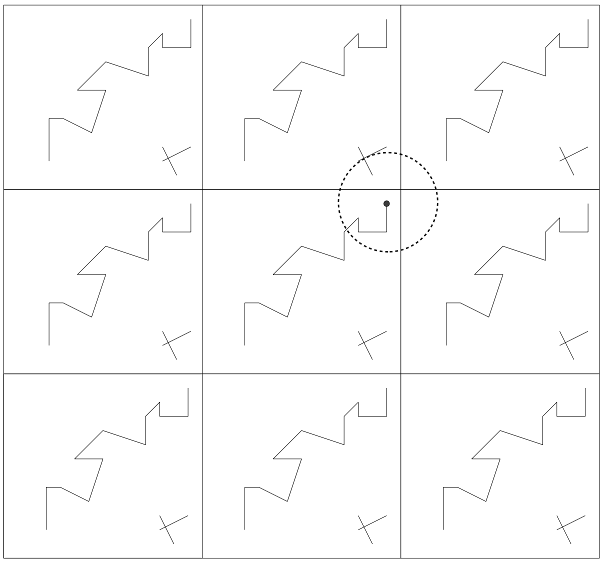
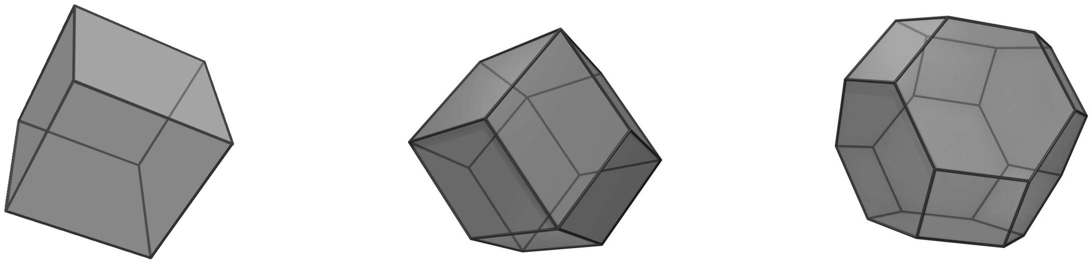


Figure 2.7: *Periodic boundary conditions. The system box is replicated in all directions.*

There are different geometries for this box (see fig. 2.8) in order to avoid the simulation of more water molecules than necessary.



(a) Standard cubic box

(b) Rhombic dodecahedron

(c) Truncated octahedron

Figure 2.8: *Different box geometries avaliable in AMBER and GROMACS. The volume of the standard cubic box is reduced by a factor of* $\frac{1}{2}\sqrt{2}$ *with a rhombic dodecahedron and by a factor of* $\frac{4}{9}\sqrt{3}$ *for a truncated octahedron.*

2.4 Integrators

The most popular MD algorithms [58] are the *Verlet* algorithm [59] (eq. 2.25) and the *Leap-frog* algorithm [60] (eq. 2.26,2.27).

The Verlet algorithm uses the atomic positions $r(t)$ and the accelerations $a(t) = \frac{F(t)}{m}$ at time t and the positions from the preceeding step $r(t-\Delta t)$ to determine the new positions at $(t+\Delta t)$[4]:

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + \frac{F(t)}{m}\Delta t^2 + O(\Delta t^4) \tag{2.25}$$

The advantages of the Verlet algorithm are that it is straightforward, and that the storage requirements are modest. The disadvantage is that the algorithm is only of moderate precision.

The leap-frog algorithm uses positions $r(t)$ and accelerations $\frac{F(t)}{m}$ at time t and velocities $v(t-\frac{\Delta t}{2})$ at time $t-\frac{\Delta t}{2}$ to compute positions $r(t+\Delta t)$ at time $(t+\Delta t)$ and velocities at time $(t+\frac{\Delta t}{2})$. The velocities are first calculated at time $(t+\frac{\Delta t}{2})$ and are used to calculate the positions at time $(t+\Delta t)$. In this way, the velocities leap over the positions, then the

---

[4] To derive the Verlet algorithm, the Taylor expansions until third order of $r(t+\Delta t)$ and of $r(t-\Delta t)$ are added:

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + O(\Delta t^4)$$
$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + O(\Delta t^4)$$

positions leap over the velocities.

$$v(t+\frac{\Delta t}{2}) = v(t-\frac{\Delta t}{2}) + \frac{F(t)}{m}\Delta t \tag{2.26}$$

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

The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time t can be approximated by the average:

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

2.5 Thermodynamic ensembles

The natural ensemble of MD is the microcanonical ensemble, where the number of particles (N), the volume (V) and the energy (E) are constant, since MD is Hamiltonian dynamics and energy is conserved. However, experiments are normally performed at constant temperature and/or constant pressure. An ensemble where the number of particles, the temperature (T) and/or the pressure (P) are constant is therefore a better representation of real systems.

To achieve this, the equations of motion need to be modified to include a thermostat and/or barostat. MD can then reproduce the canonical ensemble (NVT), if only the average temperature is kept constant, and the (NPT) ensemble, if the average temperature and pressure are kept constant. In this section we describe how temperature and pressure averages are kept constant during the MD simulations performed for different systems.

2.5.1 Temperature coupling

The absolute temperature of the N-particle system is related to the kinetic energy of the system E_{kin}:

$$\frac{1}{2}N_{df}kT = E_{kin}, \tag{2.29}$$

where $N_{df} = 3N - N_c - N_{com}$ is the number of degrees of freedom, N_c refers to the constraints applied on the system and N_{com} to the center of mass motion[5]. E_{kin} can be

---

[5] the system is invariant under translation of the center of mass, N_{com} equals then 3. In vacuum, the rotation of the center of mass can also be removed, N_{com} is set to 6.

obtained by adding the kinetic energy of every particle of the system:

$$E_{kin} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2. \tag{2.30}$$

There are several mechanisms to control the temperature. We can couple the system to a bath with the reference temperature we want our system to have as Berendsen thermostat [61] proceeds or we can introduce a friction term in the Hamiltonian of the system in order to control the temperature as in the Nosé-Hoover thermostat [62] approach.

Berendsen thermostat

The Berendsen thermostat is a global thermostat that couples the system weakly to a bath of reference temperature T_0. The deviation of the system temperature T from T_0 is corrected according to eq. 2.31 and decays exponentially with time constant τ.

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.31}$$

The time coupling constant τ can be chosen to be small for equilibration purposes or larger if the system is already equilibrated. Heat flows in/out of the system if the temperature is lower/higher than T_0 by scaling all velocities with the factor:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left\{\frac{T_0}{T(t-\frac{\Delta t}{2})} - 1\right\}\right]^{1/2} \tag{2.32}$$

at every time step Δt. τ_T (eq. 2.32) is related to the time constant τ (eq. 2.31), the total heat capacity of the system C_V and the total number of degrees of freedom N_{df} as $\tau = 2C_V\tau_T/N_{df}k$. τ_T is the parameter that appears in the dynamic options or input.

The Berendsen weak coupling algorithm is convenient for relaxing the system to a target temperature but does not generate a correct canonical ensemble.

Nosé Hoover thermostat

To perform canonical ensemble simulations, the system Hamiltonian is extended with a thermal reservoir and a friction term [63, 64]. The friction force is proportional to the product of each particle's velocity and a friction parameter ξ. The force $\vec{F_i}$ on every i^{th}particle is modified with a friction term proportional to the i^{th}particle's linear momentum $m_i\ \vec{v_i}$ as:

$$\vec{F_i}' = \vec{F_i} - \xi\ m_i\ \vec{v_i}. \tag{2.33}$$

ξ is a fully dynamic quantity with its own equation of motion: the time derivative is calculated from the difference between the current kinetic energy and the reference temperature and the strength of the coupling is determined by Q.

$$\frac{d\xi}{dt} = \frac{1}{Q}(T - T_0) \tag{2.34}$$

The Nosé Hoover thermostat generates a correct NVT ensemble but can induce oscillations in the dynamics. τ_T is the parameter that appears in the dynamic options or input: it represents the period of the oscillations of kinetic energy and is related to Q and T_0 via $Q = \frac{\tau_T^2 T_0}{4\pi^2}$

2.5.2 Pressure coupling

A macroscopic system maintains constant pressure by changing its volume. A simulation in the isothermal-isobaric ensemble also maintains constant pressure by changing the volume of the simulation cell. The amount of volume fluctuation is related to the isothermal compressibility.

The methods used for pressure control are similar to those used for temperature control. The pressure can be maintained at a constant value by simply scaling the volume. The pressure of the system can also be coupled to a fixed value (Berendsen barostat) or when fluctuations in pressure or volume are important, the equations of motion for the particles need to be modified (Parrinello-Rahman barostat).

The pressure of a N-particle system is calculated as:

$$P = trace(\mathbf{P})/3 \tag{2.35}$$

where $\mathbf{P}$ is the pressure tensor:

$$\mathbf{P} = \frac{2}{V}(E_{kin} - \Xi) \tag{2.36}$$

and Ξ the virial:

$$\Xi = -\frac{1}{2}\sum_{i<j}^{N} r_{ij} \otimes F_{ij}. \tag{2.37}$$

Berendsen barostat

The Berendsen algorithm rescales the coordinates and box vectors every step towards a given reference pressure P_0. The rate of change of the pressure is given by:

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p}, \tag{2.38}$$

where τ_p is the coupling constant. The volume of the simulation box is scaled by the scaling matrix μ is given by:

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau p}\beta_{ij}\left\{P_{0ij} - P_{ij}(t)\right\}, \tag{2.39}$$

where β_{ij} is the isothermal compressibility of the system (for water at 1atm and 300 K, $\mathrm{Tr}(\beta_{ij}) = \beta = 4.6.10^{-10}\ Pa^{-1}$), $P_{ij}(t)$ the pressure tensor at time t and P_{0ij} the reference pressure tensor. Again the Berendsen barostat is appropriate for equilibrated systems but does not reproduce the (N, P, T) ensemble.

Parrinello-Rahman barostat

The Parrinello-Rahman approach [65, 66] is similar to the Nosé-Hoover temperature coupling, it reproduces the correct thermodynamic ensemble and the equations of motion for the particles are changed. The box vectors are subject to an equation of motion to which the equation of motion for the atoms is coupled. As for Nose-Hoover temperature coupling the time constant τ_p is the period of pressure fluctuations at equilibrium. The force $\vec{F_i}$ on every i^{th}particle is modified with a friction term proportional to the i^{th}particle linear momentum $m_i\ \vec{v_i}$ as:

$$\vec{F_i}' = \vec{F_i} - \mathbf{M}\ m_i\ \vec{v_i}, \tag{2.40}$$

where $\mathbf{M}$ is a tensor related to the matrix representation of the box vectors $\mathbf{b}$ as:

$$\mathbf{M} = \mathbf{b}^{-1}\ \left[\mathbf{b}\frac{\mathrm{d}\mathbf{b}'}{\mathrm{d}t} + \frac{\mathrm{d}\mathbf{b}}{\mathrm{d}t}\mathbf{b}'\right]\ \mathbf{b}'^{-1}. \tag{2.41}$$

In the Parrinello-Rahman barostat, the matrix $\mathbf{b}$ follows the equation of motion:

$$\frac{\mathrm{d}\mathbf{b}^2}{\mathrm{d}t^2} = V\mathbf{W}^{-1}\mathbf{b}'^{-1}(\mathbf{P} - \mathbf{P}_{ref}). \tag{2.42}$$

The volume of the simulation box is called V and $\mathbf{W}$ determines the strength of the coupling: its inverse $\mathbf{W}^{-1}$ determines how the box can be deformed; it is related to the pressure time constant via:

$$\left(W^{-1}\right)_{ij} = \frac{4\pi^2\beta_{ij}}{3\tau_p^2 L}, \tag{2.43}$$

where L is the longest length of the box.

Chapter 3

Carbon nanotubes

Carbon nanotubes (CNT) are unique large macromolecules with notable physical properties. They were discovered in 1991 by S. Iijima [67] and are still the object of ongoing research focusing on the understanding of their physical properties. CNT have a very broad range of mechanical, electronic, thermal, optical, structural, etc. properties depending on their characteristic parameters, the diameter, the length, and the chirality or twist. These properties make them interesting candidates to study fundamental physical phenomena and also to objects which might be relevant as building blocks in future applications.

Despite from the huge potential for a broad use of CNT, a mass production of CNT with well defined structure is still far out of reach. The synthesis of CNT does not allow to grow them with 100% purity in terms of well defined length, diameter and chirality. Therefore, methods have to be developed, which can help to reliably separate and selectively organize CNTs, according to their functionality or structure.

As a consequence, the investigation of techniques to separate CNT according to their characteristic properties have become an important research field in the last few years. Single walled carbon nanotubes (SWCNTs) can for example be separated according to their electronic properties [12], which are determined by their chirality. This is possible by making use of a method called dielectrophoresis [1]. Other methods focus on the separation

---

[1] Separation of nanotubes with dielectrophoresis is achieved by applying a non uniform electric field that results in different effective forces for semiconducting or for metallic nanotubes. The dielectrophoresis force

$$F_{DEP} \propto \varepsilon_m Re\left\{\frac{\varepsilon_t^* - \varepsilon_m^*}{\varepsilon_m^*}\right\} \nabla|\vec{E}|^2, \tag{3.1}$$

depends on the gradient of the applied electric field ($\nabla|\vec{E}|^2$) and the complex electrical permittivity of CNTs according to their density [68, 69], which relates to CNT length and diameter. Experiments carried out by Frank Hennrich at the Institute for Nanotechnology focus on the separation of CNT by centrifugation, where nanotubes are sorted according to their density. The understanding of the relation between the density and the diameter of CNTs is required to develop a methodology for sorting the CNTs by diameter. This may help to design a method for large scale production of CNTs with well defined diameter.

In combination with other separation methods it may become possible to sort and provide CNTs with clearly defined structural, electronic, etc. properties.

In the present chapter we aim to develop a method for separating CNTs by their diameter. In aqueous solution, CNTs tend to cluster together in order to avoid being exposed to water due to their hydrophobic nature. Thus, amphibilic molecules, or surfactants, have to be added to separate nanotubes from each other. Therefore, experiments sorting the nanotubes according to their density are not exclusively influenced by the CNTs diameter but by many other factors, such as the surfactant concentration or the water conformation around or inside the nanotube. By making use of MD simulations we study the behavior of CNTs in the environment of water and surfactant molecules. With our MD simulations, we are able to explain the experimental results qualitatively and furthermore provide a theory for the relation between the density of the 'CNT-surfactant' composite and the CNT diameter for a given surfactant molecule.

In section 3.1 we give a brief introduction to CNTs and explain their characteristic parameters. In section 3.2 we report the experimental methods and results for separating CNTs by density. In section 3.3 we study the interactions between the carbon nanotubes and the water molecules and present their effects on the total density of the system. We show the relevance of considering the surfactant in order to properly explain the experiments in section 3.4. We also explain the experimental findings.

3.1 Carbon nanotubes

Structure - Chirality

Carbon nanotubes are hollow cylinders made of carbon atoms bound together in a hexagonal network. A single walled carbon nanotube (SWCNT) structure can be obtained by wrapping a graphene sheet in a certain direction into a cylinder with a continuous surface.

$\varepsilon^*_{m,t} = \varepsilon_{m,t} - i\omega\sigma_{m,t}$ of the media m and the nanotube t, respectively.

The direction of rolling the graphene sheet, i.e. the chiral vector $\vec{C_h}$, determines the (geometric and electronic) properties of the SWCNT. The hexagonal structure of graphene and the chiral vector are illustrated in Fig. 3.1. The chiral vector is defined as:

$$\vec{C_h} = n\vec{a}_1 + m\vec{a}_2 \tag{3.2}$$

where $\vec{a}_1$ and $\vec{a_2}$ are the unit vectors of the hexagonal lattice and n, m are two integers called indices.

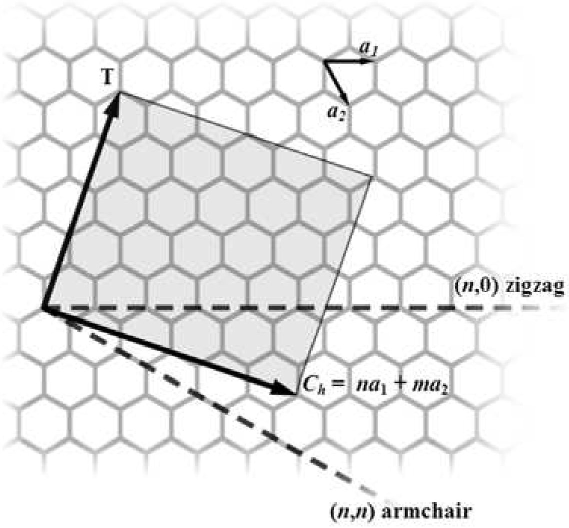


Figure 3.1: *Illustration of the chiral vector $\vec{C_h}$ on a hexagonal lattice representing the structure of graphene. $\vec{a_1}$ and $\vec{a_2}$ are the unit vectors of the hexagonal lattice. $\vec{T}$ is the vector orthogonal to the chiral vector and parallel to the axis of the nanotube. The dashed lines correspond to the special cases m=0 also called zigzag and n=m called armchair.*

Diameter

The unit vectors can be expressed using 2-D Cartesian coordinates:

$$\vec{a_1} = (\frac{3}{2}a_{c-c}, \frac{\sqrt{3}}{2}a_{c-c}) \tag{3.3}$$

$$\vec{a_2} = (-\frac{3}{2}a_{c-c}, \frac{\sqrt{3}}{2}a_{c-c}) \tag{3.4}$$

where a_{c-c} is the bond length between carbon atoms, and since the length $|\vec{a_1}| = |\vec{a_2}| = \sqrt{3}a_{c-c}$, and the chiral vector length is the peripheral length of the nanotube, the diameter

$D_{n,m}$ of the nanotube is related to the (n,m) indices as:

$$D_{n,m} = \frac{\sqrt{3}\, a_{c-c}}{\pi} \sqrt{n^2 + nm + m^2}. \tag{3.5}$$

Synthesis and length

In the experiments reported in this section, SWCNT are synthesized by the high-pressure carbon monoxide (HiPco) process, in which SWCNT grow from a thermal decomposition of $Fe(CO)_5$, or alternatively by pulsed laser vaporization (PLV), in which a laser pulse evaporates a solid target of graphite into a plasma. The particles in the plasma selfassemble to form a fraction of SWCNTs. The SWCNTs generated by HiPco have diameters ranging from 0.6 nm to 0.9 nm and the SWCNTs generated with PLV have diameters ranging from 0.9 nm to 1.5 nm. A typical length of a SWCNT is in the range of micrometers, however, it is very variable and depends on the synthesis technique in general.

Physical properties

The chirality, defined by the (n,m) indices, of the SWCNT not only determines the diameter but also the electronic properties: a SWCNT can be metallic if $\frac{2m+n}{3} \in \mathbb{N}$ or semiconducting if $\frac{2m+n}{3} \notin \mathbb{N}$. CNTs are a very stiff material in terms of tensile strength and elastic modulus respectively [70] and very good thermal conductors [71].

3.2 Experimental results

3.2.1 Centrifugation process

In 2005, Arnold et al. showed that separation of SWCNTs by diameter can be achieved through ultracentrifugation of SWCNT wrapped into DNA or suspended with surfactants in aqueous density gradients [72, 68].

A density gradient technique, called equilibrium density-gradient centrifugation, is used to separate materials on the basis of their buoyant density (see Fig. 3.2). In this case a mixture of materials with different buoyant densities is centrifuged through a steep density gradient that contains a high concentration of a medium with a preformed density gradient or a medium which generates a density gradient by itself during centrifugation. In these gradients, the materials being studied have a density somewhere in between the highest and lowest densities of the 'calibration' substances. The components of a sample begin to move along this gradient. When a component of the mixture reaches a point where the density of the solution is equal to its own density, it stops moving further and forms a distinct band. The position of the band in the tube is characteristic for the buoyancy of that component. The density-gradient ultracentrifugation protocol is similar to that of Arnold et al. described in ref. [68].

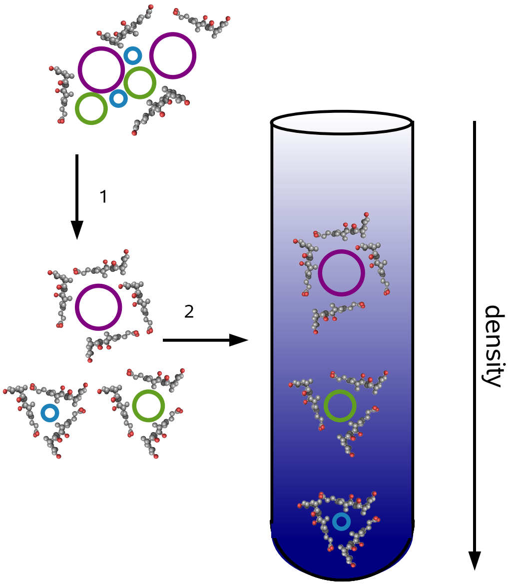
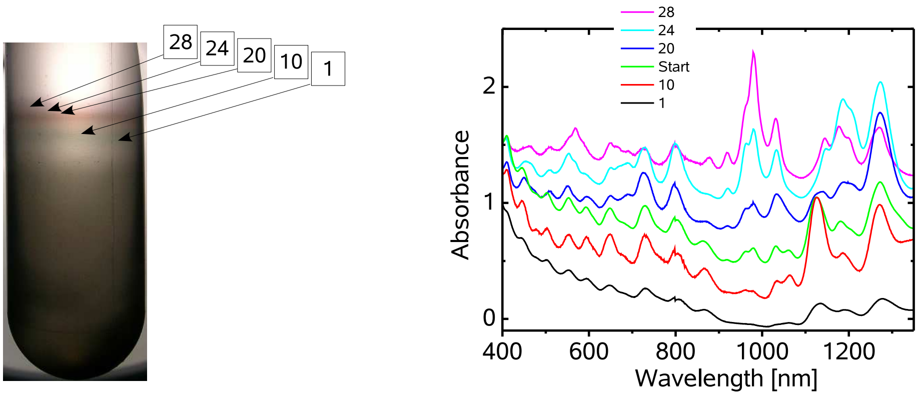


Figure 3.2: *In step 1, mixtures of surfactants are used to disperse individual nanotubes. In step 2, these are loaded into a centrifuge tube containing a liquid density gradient. When centrifuged, layers of material with similar density can be separated.*

Frank Hennrich and coworkers used self-generated gradients, i.e. a solution of uniform density at one atmosphere that forms a gradient under the influence of the centrifugal field. Once the solute begins to sediment through the solvent a concentration gradient is formed which is opposed by back-diffusion of the solute. With a sufficiently high spinning velocity, at equilibrium, the sedimentation of the solute is exactly balanced by the diffusion and the gradient is stable.

3.2.2 Results of the experiment

2-D Photoluminescence (PL) contour maps of tube samples in surfactant solutions were derived from individual PL emission spectra measured in the range of 850-1750 nm with



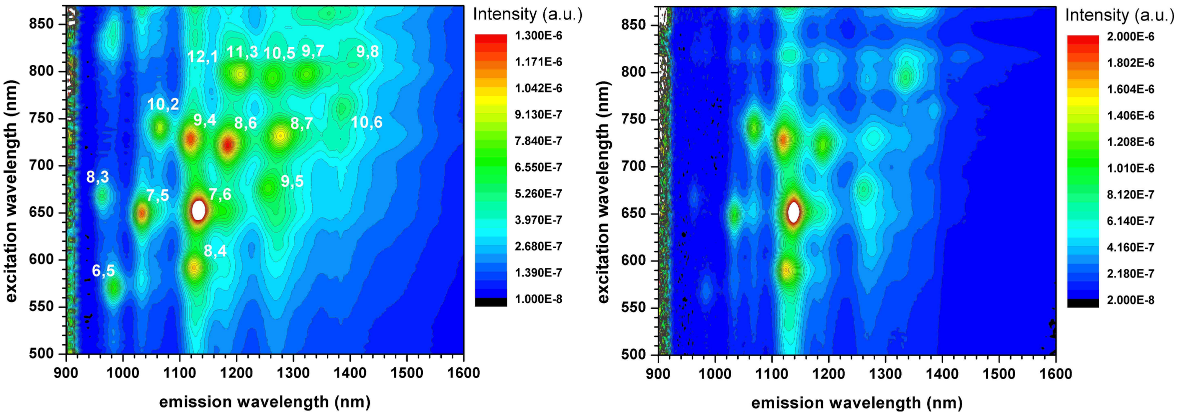
(a) Experimental sample (b) Absorption spectra

Figure 3.3: *(a) Photograph of an experimental sample after 18 hours of centrifugation. The analyzed fractions corresponding to decreasing densities are indicated by arrows. (b) Absorption spectra of the different fractions indicated in the picture.*

a similar setup and procedures as described in ref. [73]. PL is a spectroscopic technique to identify which nanotubes are present in a sample. The density of states of SWCNTs has a gap between allowed states of density (called van Hove singularities), which depends on the chirality of the SWCNT. For a similar diameter range and for increasing energy, the first and second singularities correspond to semiconducting SWCNT, and the absorption/emission energy is labelled $E_{11}S$ and $E_{22}S$ and the third singularity corresponds to a metallic SWCNT and is labelled $E_{11}M$.

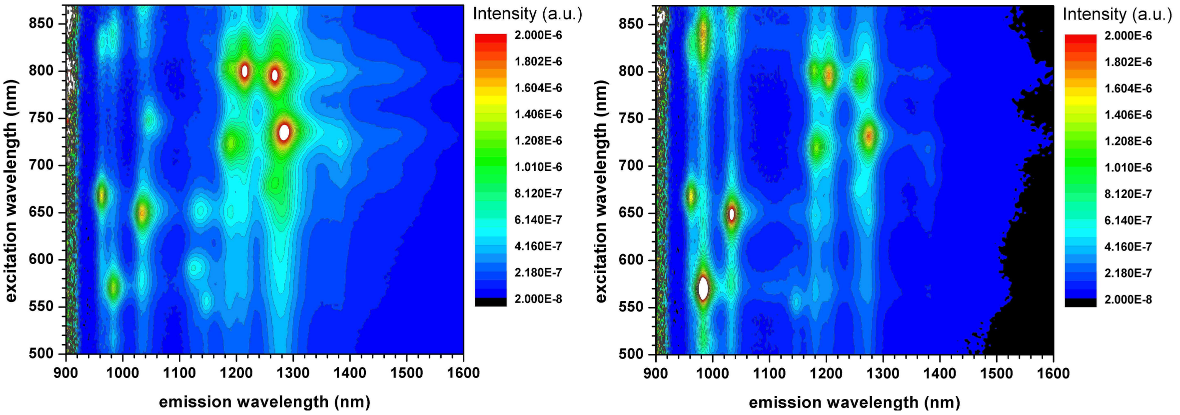
Fig. 3.3(a) shows the photograph of a sample in a separation experiment after the density-gradient ultracentrifugation. In the centrifuge tube a density gradient with higher density at the bottom than at the top has emerged after $\sim$ 18 hrs. Visually, the separation of isolated SWCNTs is evident by the formation of colored bands on top of a black region which contains bundles, aggregates and insoluble material that sediments at higher density in the gradient. The domain populated with nanotubes is separated into 28 fractions. Fractions labeled with a small number correspond to higher densities and fractions labeled with higher numbers correspond to lower density sections.

In optical absorption spectra ratios of absorbance peaks in first- ($E_{11}S$, 900-1500 nm) and second-order ($E_{22}S$, 500-900 nm) interband transitions differ between the fractions, indicating the achievement of sorting tubes by diameter/bandgap (Fig. 3.3(b)). Photoluminescence mapping allows to identify tubes by their (n,m) indices and can therefore be used to study relative changes of the abundance of different (n,m)-species in dependence of the experimental conditions.



(a) starting material PL contour map

(b) fraction 10 PL contour map



(c) fraction 22 PL contour map

(d) fraction 28 PL contour map

Figure 3.4: *Photoluminescence countour maps of the starting material (a) and of three different fractions: (b) fraction 10, (c) fraction 22 and (d) fraction 28, after the centrifugation process. The maps show the PL intensity as a function of the emission wavelength (x axis) and the excitation wavelength (y axis)*

Fig. 3.4 shows the PL contour maps of the starting material before the density-gradient ultracentrifugation and of three different fractions after the density-gradient ultracentrifugation (fraction 10, 22 and 28) cut within the colored bands. The n,m-distributions of fractions number 10, 22 and 28 which can be derived from the PL contour maps are very different to the n,m-distribution of the starting material. The following two fundamental observations are made in experiment:

1. fraction 10 is strongly enriched in tubes with diameters around 0.88 nm, with corresponding chirality: (10, 2), (9, 4), (8, 6) and (8, 4);

2. in fraction 22 and 28 tubes with diameter around 0.88 nm are almost completely missing.

The buoyant density of SWCNTs in aqueous solution depends on multiple factors, including the mass and volume of the carbon nanotube itself, its surface functionalization, the hydration layers, the surrounding molecules, and its filling.

According to a simple, geometric model treating a nanotube as a hollow cylinder, the intrinsic mass per volume of a nanotube is expected to vary inversely with diameter. This inverse relationship with diameter cannot be explained by the experimental observations in the original work of Arnold and co-workers [72]. They tried to correct the trend by including an encapsulating layer of hydrated DNA with constant thickness ($\sim$2-3 nm), but they did not take into account the filling of the tube.

Fluids confined in nanometric scales are known to behave very differently from their bulk counterparts. The strong hydrophobic nature of the nanotube material and the comparatively large exposed surface area of the inside of the tube in relation to its relatively small volume are the reason for unusual arrangements of water molecules inside nanotubes and large deviations in the hydrogen bonding network of such molecules in comparison with the bulk solvent [74]. The density and other properties of water trapped inside the nanotube will therefore deviate significantly from that of the bulk solvent under the same macroscopic conditions.

3.3 Interaction of CNT with water

In 1998, Dujardin et al. presented experimental evidence of SWCNT wetting [75]. In 2004, Rossi et al. used environmental scanning electron microscopy to demonstrate that water condensates easily inside SWCNTs and that their inner surface wetting is a function of their diameter rather than the surface details [76]. In 2005, Byl et al. presented experimental vibrational spectroscopy data providing direct evidence of a water phase with an unusual hydrogen bonding pattern due to confinement inside a SWCNT [77]. In 2006, Matsuda et al. measured with NMR the existence of inner rings of water molecules inside the SWCNTs [78]. De Souza et al. combined neutron scattering and MD simulations to investigate the dynamics of water confined in a SWCNT [79] and obtained evidences for the wetting of the inside of the SWCNT. Moreover, they predicted the formation of water shells using MD.

Many other theoretical studies complement the experimental hydration data for the inside of the SWCNT: In 2000, Gordillo and Marti permformed MD simulations on SWCNTs and water and found water inside the CNT [80]. In 2001, Hummer et al. studied the dynamic behavior of water inside a SWCNT as an hydrophobic channel and concluded that water flows very easily through the tube [74]. In 2002, Noon et al. found helical water sheets inside the SWCNTs [81] and Brovchenko et al. determined the presence of water in SWCNTs [82] in 2003. In 2004, Wang et al. performed MD simulations under ambient conditions for nanotube segments of various diameters submerged in water and obtained that single-file water chains were formed in narrow nanotubes (0.676 to 0.811 nm diameter) and that layered structures could be formed in larger size SWNTs. According to these studies it seems that (5,5) SWCNT is the smallest tube whose interior could be wetted by water. The minimal diameter of a SWCNT that can accommodate n layers of water molecules was also estimated [83]. Similar findings were also obtained by Huang et al. in 2005. Inside the tubes they studied, water molecules formed a cylindrical layer around a central water core, in which water molecules are randomly distributed due to thermal movements. However in addition, by charging the SWNTs, water molecules adsorbed by the SWNTs can also form regular structures [84].

Summarizing the experimental and theoretical results, the following questions arise: How is the structure of water inside CNT with diameter ranging from 0.5 to 1.5 nm? Are the CNT wetted inside, and if yes, in which way? How does the presence of water inside the CNT influences the results of centrifugation experiments?

To answer these questions, we perform exhaustive MD simulations comparing different coupling and solvation protocols, water models and Lennard Jones parameterizations. By this, we study the behavior of water inside all SWCNTs with the same characteristics as used in the centrifugation experiments summarized in the previous section.

3.3.1 MD simulations

We first generate the carbon nanotubes atomic representation for different diameters ranging from 0.5 to 1.5 nm. Before starting with ‘production’ simulations, we investigated the effects of the tube length and the solvation protocol on equilibration.

During the simulations, the nanotube atoms are constrained to their experimental equilibrium position and only the interactions between the carbon atoms and the water molecules and the interactions between the water molecules with each other are modeled.

The SWCNT are solvated in a rectangular water box with periodic boundary conditions such that the edges of the box are 12 Å away from the tube. We use the OPLSAA force field since it was originally parameterized for liquid simulations. The simulations are carried out with the GROMACS package, where the OPLSAA force field is implemented because it is twice as fast as than AMBER (see section 2.2).

Parameterization for the simulations

We perform MD simulations on the different systems in an NVT ensemble and in an NPT ensemble. The average temperature is set to 300 K and the pressure (for the simulations in the NPT ensemble) to 1 atm. For the temperature coupling, the Nosé-Hoover thermostat is used (see section 2.5.1). For the pressure control, Berendsen and Parrinello-Rahman barostats are used (see section 2.5.2).

To treat electrostatic interactions, the Particle Mesh Ewald (PME) method is employed and Van der Waals interactions are computed within a simple cut-off of 14 Å. The integration time step is 2 fs. The total length of the simulations is 2 ns. This is a very short simulation time if compared to the time scale in which molecular conformations change. However, since we are only interested in the behavior of water and there are no macromolecular motions, 2 ns is a large enough time scale for water to be equilibrated. The stabilization of the system is reached within the first 200 ps, as can be seen in Fig. 3.5.

We model the carbon atoms as uncharged particles with a Lennard Jones (LJ) collision diameter σ_{CC} and a well depth ε_{CC}. The collision diameter and well depth corresponding to the interaction potential between oxygen atoms and carbon atoms are denoted by σ_{CO} and ε_{CO} respectively. In Table 3.1, the Lennard Jones parameters for the carbon - water interactions are indicated. These parameters are in accordance with former studies [83, 74, 85, 84][2].

The van der Waals parameterization of C_A corresponds to an aromatic sp^2 hybridized carbon. C_A (σ_{C_A} = 0.355 nm, ε_{C_A}=0.29288 kJ.mol^{-1}) is derived from the benzene or diphenyl parameterization with the OPLSAA force field.

---

[2] Hummer et al. [74] studied the water conduction through CNTs using the TIP3P water model and a carbon-oxygen LJ potential corresponding to the AMBER96 force field LJ interactions between sp^2 hybridized carbon and water oxygen; with σ_{CO} = 0.32751 nm and ε_{CO}=0.47847 kJ.mol^{-1}. Noon et al. [81] parameterized the LJ interactions as σ_{CO} = 0.3296 nm and ε_{CO}=0.5781 kJ.mol^{-1}. Wang et al. [83] performed MD simulations with σ_{CO} =0.3275 nm and ε_{CO}=0.4785 kJ.mol^{-1} and Huang et al. [84] using $R_{min,CC}$ =0.1992 nm, $R_{min,OO}$ =0.17682 nm, ε_{CC}=0.0700 kcal.mol^{-1} and ε_{OO}=0.1521 kcal.mol^{-1}

	SPC	TIP3P	TIP4P	TIP5P
σ_{C_AO}	0.3352 nm	0.3344 nm	0.3346 nm	0.3328 nm
ε_{C_AO}	0.4364 kJ.mol^{-1}	0.4317 kJ.mol^{-1}	0.4358 kJ.mol^{-1}	0.4428 kJ.mol^{-1}
σ_{C_BO}	0.3352 nm	0.3344 nm	0.3346 nm	0.3328 nm
ε_{C_BO}	0.5661 kJ.mol^{-1}	0.5601 kJ.mol^{-1}	0.5654 kJ.mol^{-1}	0.5744 kJ.mol^{-1}
σ_{C_CO}	-	0.3224 nm	-	0.3222 nm
ε_{C_CO}	-	0.7480 kJ.mol^{-1}	-	0.7483 kJ.mol^{-1}

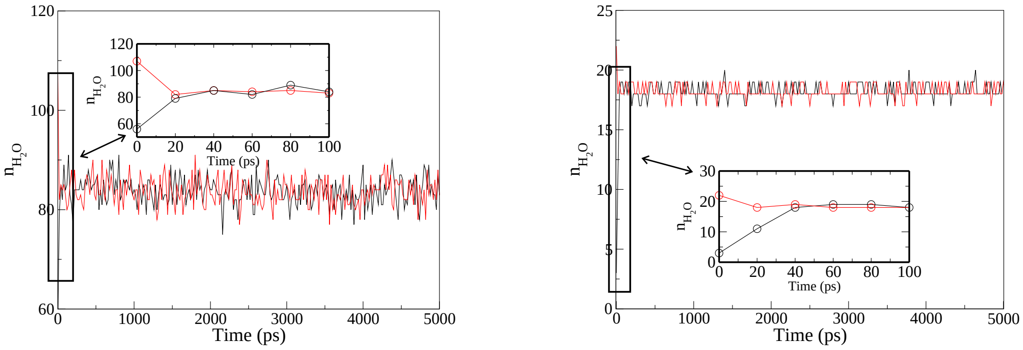
Table 3.1: *Parameters of the LJ interaction between the carbon atoms and the water oxygen atoms. The different water models (SPC, TIP3P, TIP4P and TIP5P explained in section 2.3.2) and different carbon atoms parameterizations (labeled A, B and C) are represented.* C_A *with* σ_{C_A} = *0.355 nm and* ε_{C_A}=*0.29288* kJ.mol^{-1}, C_B *with* σ_{C_B} = *0.355 nm and* ε_{C_B}=*0.49288* kJ.mol^{-1}, *and* C_C *with* σ_{C_C} = *0.33 nm and* ε_{C_C}=*0.87864* kJ.mol^{-1}.

Carbon parameterization of C_B (σ_{C_B} = 0.355 nm, ε_{C_B}=0.49288 kJ.mol^{-1}) is more hydrophobic and agrees in a better way with the LJ parameters used by Nun et al.

Carbon C_C (with σ_{C_C} = 0.33 nm and ε_{C_C}=0.87864 kJ.mol^{-1}) is an extreme case for a very pronounced well depth of the Lennard Jones potential. It does not correspond to a pre-parameterized aromatic carbon, but since inside the nanotube energetically favorable π-systems are formed, it might be a way to bias water behavior in such an environment.

First, we compare two solvation protocols: solvating the nanotube in equilibrated water at 300 K and 1 g.cm^{-3}, which results in an almost empty SWCNT, and a simulated annealing protocol implemented in MMTK [86] to pack the water as densely as possible inside the tube. In Fig. 3.5 we show, for two different nanotubes (15,1) and (8,3), how the equilibrated state is reached when starting from a densely filled nanotube (MMTK for filling the tube and GROMACS for solvation) and when starting from an almost empty nanotube (only GROMACS solvation) for two different nanotubes.

We also analyze the mobility of water molecules inside the tube to check if the starting configuration has any influence in the dynamics of the system. The simulations show that water molecules are not trapped in the tube but are constantly exchanged with the surrounding bulk water. In Fig. 3.6 we plot the evolution of the axial coordinate of three selected water molecules that are in the tube at a given time, at which the system was already equilibrated, as the simulation proceeds. It was taken from the simulation with the TIP3P water model of a (8,3) nanotube, with a 0.77 nm diameter. The figure clearly demonstrates (see snapshots on the right panel) that the water is highly mobile inside the



(a) 1.21 nm diameter and chirality (15,1) (b) 0.77 nm diameter and chirality (8,3)

Figure 3.5: *Comparison of the influence of the initial configuration. The number of water molecules inside the CNT n_{H_2O} is plotted vs. the simulation time for the MMTK solvation protocol (in red) and for the standard solvation protocol (in black). n_{H_2O} is quickly stabilized after starting the simulation and remains constant for the rest of the 5 ns run. The empty nanotubes are filled already after 40 ps and the excess of water molecules in the highly dense filled nanotubes is leaving the tube after 20 ps simulation.*

tube and is exchanged with the bulk water molecules frequently.

Here, we can already answer part of the questions formulated above and justify our simulation protocol. The SWCNTs are wetted inside and the initial conditions do not influence the stable conformation of water inside the SWCNT. Moreover, after equilibration we observe that the same water density is reached for SWCNT having the same (n,m) indices for lengths between 25 Å and 200 Å. Therefore, it is justified to perform all the simulations with 50 Å nanotubes and extrapolate to the experimental relevant case, in which much longer tubes are used.

Results of the MD simulations

In the following section we summarize the results of the MD simulations and we analyze how the results of the simulation are influenced by the choice of the LJ parameters, by using different water models (SPC, TIP3P, TIP4P and TIP5P), by the ensemble choice in which the simulations are permformed (constant pressure or constant volume) and the underlying pressure and / or temperature coupling algorithms.

However, we can already say that for CNTs with diameters below a threshold that

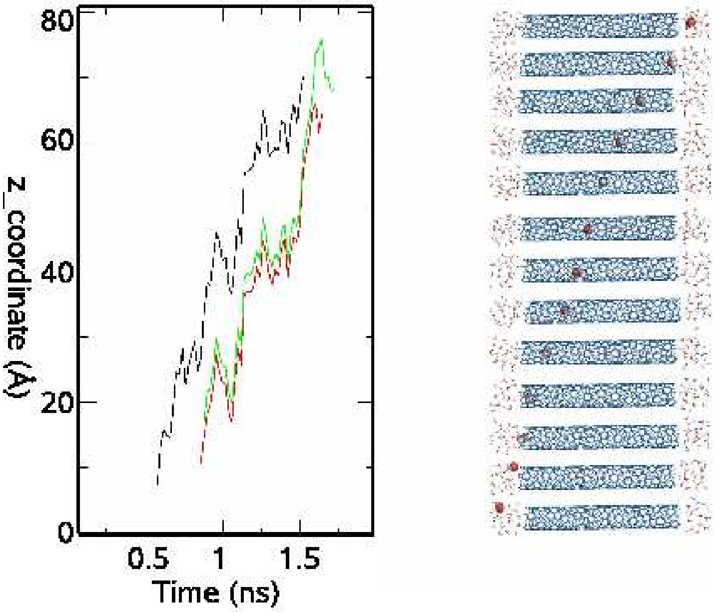
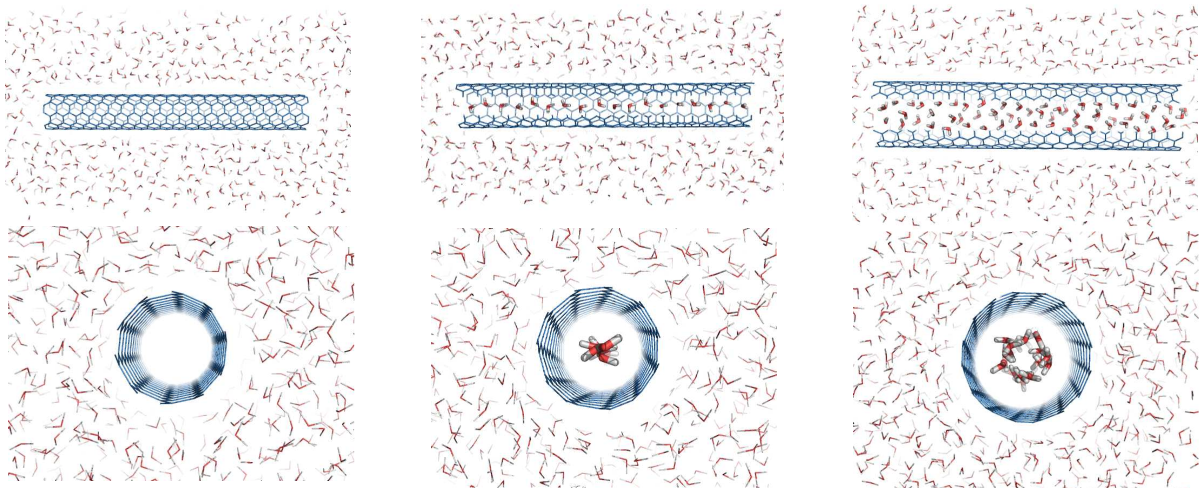


Figure 3.6: *Position of an individual water molecule along the tube axis (in Å) as a function of time during the simulation(left panel). Snapshots corresponding to the position of the water molecule as it passes the tube (right panel).*

ranges from 0.67 to 0.74 depending on the LJ parameterization, there are no water molecules inside the CNT. In CNTs with diameter above this threshold, a single chain of water molecules in formed inside the CNTs and, as we consider CNTs with larger diameters, an inner water shell is formed inside the CNTs. The respective threshold value for the diameter of a CNT where water forms an inner tube depends on the LJ parameterization and on the water model used, and typically lies between 1 nm and 1.15 nm. For CNTs with even larger diameters, a water channel is formed inside a water shell, this occurs for nanotubes with diameters larger than 1.2 to 1.3 nm, again, depending on the the LJ parameterization and the choice of the water model. These water special configurations are illustrated in Fig. 3.7.

Due to this layering behavior of the water discussed above (see Fig. 3.7) the water density inside the nanotubes depends strongly on the nanotube diameter. For SWCNTs having a diameter bigger than 2 nm the water density reaches already the bulk value.

There are several degrees of freedom that are independent from each other: the water model that best represents the behavior of confined water, the pressure and temperature coupling algorithm and the Lennard Jones parameters involved in the non-bonded interactions between the Carbon atoms and the water molecules.



(a) 0.67 nm diameter (b) 0.74 nm diameter (c) 1.08 nm diameter

Figure 3.7: *Water layered structures inside different SWCNT. Results corresponding to C_A LJ parameterization and TIP3P water model.*

We first analyze the influence of the water model in the NPT ensemble and in the NVT ensemble simulations. In Fig. 3.8, the number of water molecules 'contained' in the nanotube is plotted versus the diameter.

The number of water molecules contained in the CNT is independent of the pressure coupling algorithm. We compare simultaneously how the solvent model affects the water behavior inside the CNT. The three-point water models, SPC and TIP3P, yield equivalent results for both thermodynamic ensembles NPT (with Parrinello- Rahman and weak coupling of the pressure) and NVT. All water models agree in the formation of the first 'water channel' inside the CNT. For this LJ parameterization (corresponding to C_A), the first water channel is formed in the (6,5) CNT, having a diameter of 0.74 nm. With increasing diameter, water molecules are able to form a cylinder inside the CNT. According to the water model used, this layer is formed for 1.03 nm diameter for TIP4P and TIP5P models whereas three point water models allow the formation of this first 'shell' for CNT with larger diameters. This may be due to the fact that in five-point water, polar contacts can occur between more points and more ordered structures (closer to ice) are possible.

The water models developed for the OPLS force field (c.f. 2.3.2) are TIP4P and TIP5P. All the previous studies where confined water inside CNT was simulated were performed with three-point water models. In our further study we consider three point and five point water models for every set of Lennard Jones parameters.

We proceed our study for different Lennard Jones interaction parameterizations.

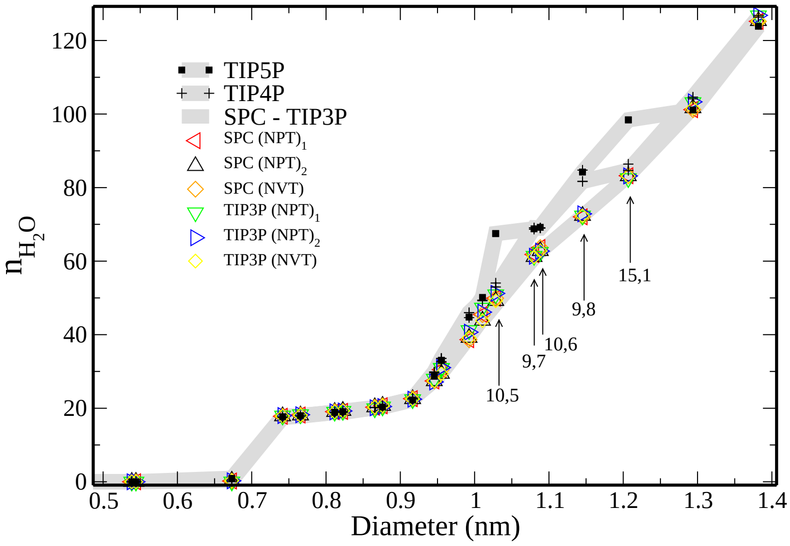


Figure 3.8: *Evolution of the number of water molecules* n_{H_2O} *inside the CNT vs. the CNT diameter.* $(NPT)_1$ *corresponds to Berendsen pressure coupling and* $(NPT)_2$ *corresponds to Parrinello-Rahman pressure coupling. The grey lines represent the agreement of all possible combinations of simulation degrees of freedom when three point rigid water models are used, when the TIP5P water model is used (with squares) and when the TIP4P model is used (with crosses). The values given by the TIP4P and the TIP5P water models diverge from the three-point water models values for some CNT. In such cases, the (n,m) indices are indicated.*

Fig. 3.9 demonstrates that the LJ parameterization influences the water density inside the CNT slightly, but no significant differences arise: the first water channel, the inner water shell and the second water channel are always formed, but in CNT with a C_C LJ parameterization these water structures are formed in CNTs with a 0.1 nm smaller diameter than in CNT with a C_A LJ parameterization.

We now turn to the analysis of the density of the CNT as a function of their diameter. The results of the MD simulations describe without notable ambiguities a stepwise filling of the SWCNTs for a range of diameters that is not much influenced by the choice of the parameters defining the carbon atoms (within a meaningful range) or of the water model.

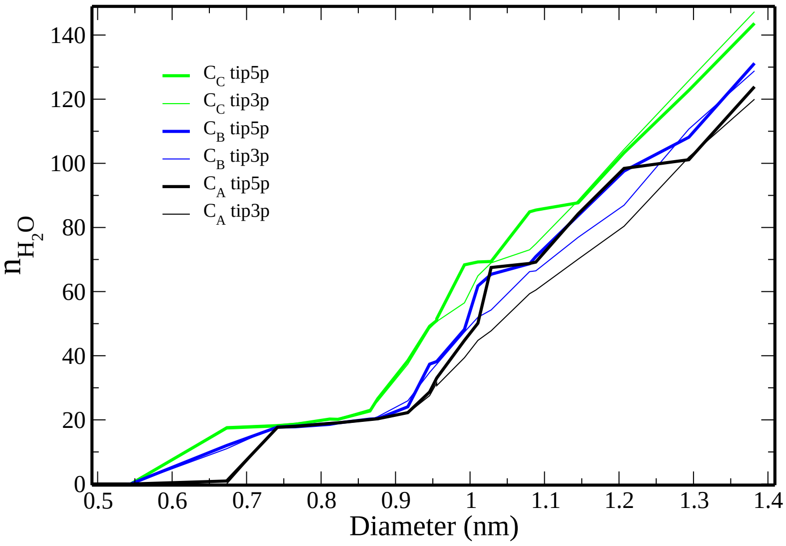


Figure 3.9: *Number of water molecules stabilized inside the CNT vs. diameter. The different Lennard Jones parameterizations are illustrated with different colors as shown in the legend. The three-point water models (thick lines) and the five-point water model (thin lines) are simultaneously compared.*

3.3.2 Density

The expected density of the combined system SWCNT-water ρ_{T1} is given by:

$$\rho_{T1} = 4\ \frac{(m_{H_2O}\ n_{H_2O}) + \sigma_{graphene}\ \pi\ d\ l}{\pi\ (d + \Delta d)^2\ \ l}, \tag{3.6}$$

where m_{H_2O} = 18 u is the mass of each water molecule in atomic mass units and n_{H_2O} the average number of water molecules inside the CNT obtained from the simulations. The diameter of the SWCNT is d and Δd=0.6 nm is the thickness of the graphene sheet, which is related to the van der Waals radius of the carbon atoms. $\sigma_{graphene}$ is the surface density of graphene[3] $\sim$ 465 u.nm^{-2}, l = 5nm is the length of the CNT. In Fig. 3.11 the parameters involved in the density are illustrated.

---

[3] The area of the unit cell of graphene is A

$$A = \frac{3\sqrt{3}}{2}\ a_{cc}^2, \tag{3.7}$$

where a_{cc} is the distance between carbon atoms and in graphene its value is around 0.141-0.144 nm.

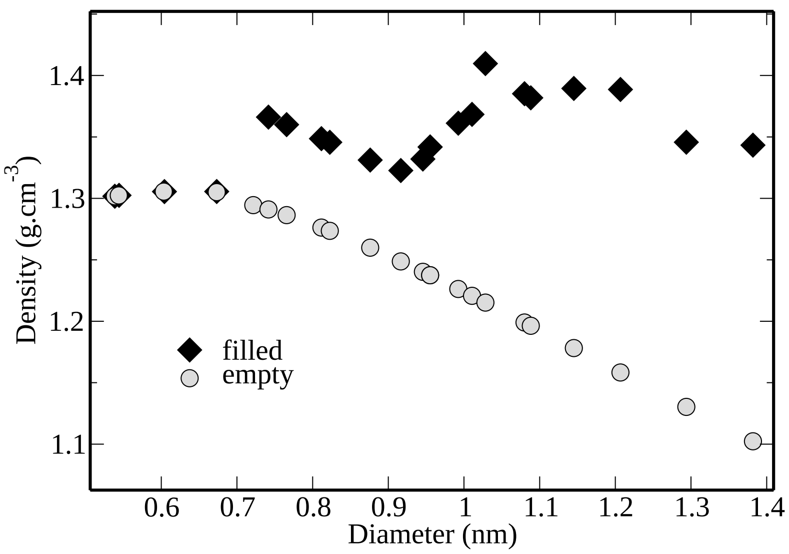


Figure 3.10: *Density of the SWCNT vs. diameter. The grey circles represent the contribution of the carbon atoms of the CNT and the black diamonds take the filling of the CNT with water into account.*

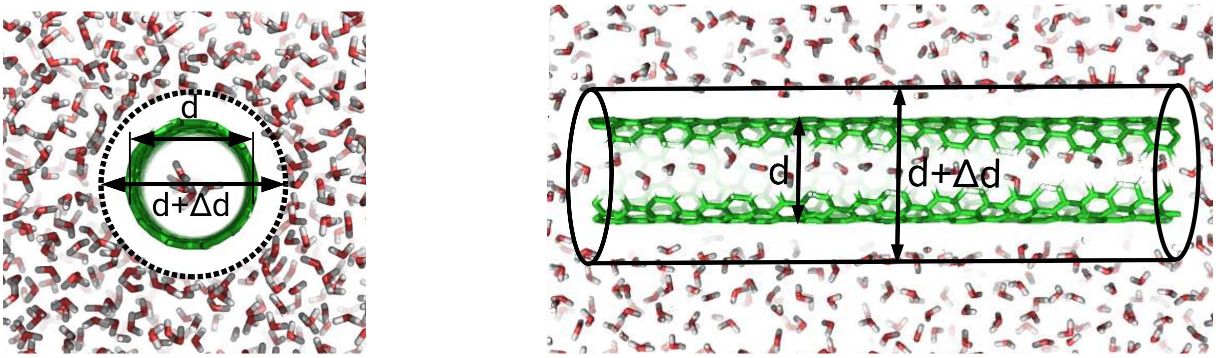
In Fig. 3.10, the density corresponding to eq. 3.6 is illustrated, without considering the filling of the nanotube with water (grey circles) and considering the filling of the tube (black diamonds).

Combining eq. 3.6 with the outcomes of our MD simulations we observe the minimal diameter of the CNT at which water can enter the tube. We observe a stepwise formation of a water channel for a SWCNT of 0.74 nm diameter and then the formation of a water shell for a SWCNT of 1.03 nm diameter is reflected in the density. However, these results do not agree with the experiments, where the high density section (fraction 10), was populated exclusively with 0.88 nm diameters SWCNTs.

There are two atoms per unit cell, therefore the superficial density of graphene is:

$$\sigma_{graphene} = \frac{2\ m_C}{A} \tag{3.8}$$

where m_C is the mass of the carbon atom.



(a) front view

(b) side view

Figure 3.11: *Snapshot of a solvated nanotube. The diameter of the nanotube and the diameter of the cylinder considered to calculate the occupied volume are indicated by d and d+Δd*

3.4 Interaction of CNT with water and surfactant

This discrepancy may arise because in the experiment the CNTs are in solvated by means of amphibilic molecules which were not taken into account so far. Obviously, the surfactant strongly influences the density-diameter relation and can not be neglected.

In this section we therefore analyze the effects when considering surfactant molecules in the density calculation.

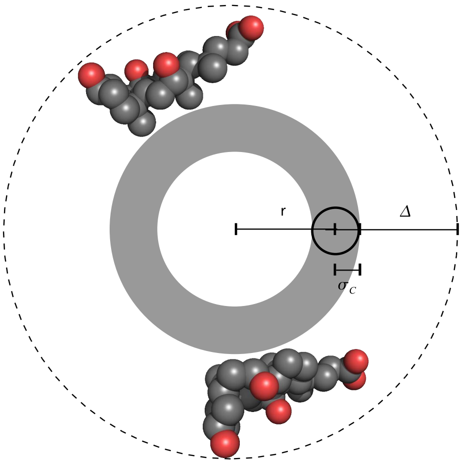


Figure 3.12: *Schematic representation of a transversal section of a SWCNT surrounded by cholate molecules.*

SWCNTs are hydrophobic. They tend to cluster together to avoid being exposed to water to minimize the surface accessible by the water molecules. In order to solvate CNTs at all, they must be coated with amphibilic molecules. The amphibilic molecules used in the experiments here are sodium cholate, $C_{24}O_5H_{39}$.

In the presence of the surfactant, the expression for the density of the complex SWCNT-water-cholate ρ_{T2} is:

$$\rho_{T2} = \frac{(m_{H_2O}\ n_{H_2O}) + \sigma_{graphene}\ 2\pi\ r\ l + \rho_{surfactant}\ \pi\ \left((r+\sigma_C+\Delta)^2 - (r+\sigma_C)^2\right)\ l}{\pi\,(r+\sigma_C+\Delta)^2\ l}, \tag{3.9}$$

as a function of the radius of the SWCNT r (r=d/2, to be consistent with the previous section), where $m_{H_2O} = 18$ u is the mass of each water molecule in atomic mass units and n_{H_2O} the average number of water molecules inside the SWCNT obtained from the simulations and $\sigma_{graphene}$ is the surface density of graphene, as in the previous section. $\rho_{surfactant}$ is the density of the surfactant molecules, σ_C is the LJ radius of the carbon atoms ($\sigma_C = \Delta d/2$, to be consisten with the previous section)) and Δ is the 'thickness' of the surfactant shell around the SWCNT (see Fig. 3.12).

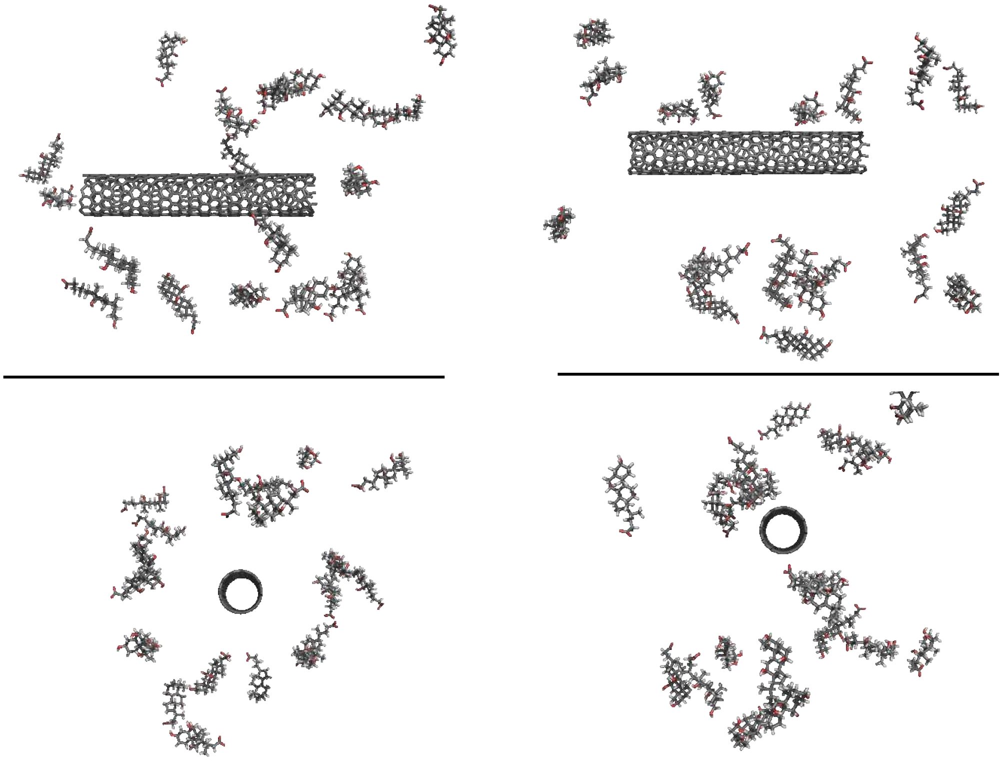
3.4.1 MD simulations

We have performed an MD simulation in the same fashion as before of a SWCNT, in particular (10,2), with C_A Lennard Jones parameterization for the carbon atoms, to obtain a theoretical value for Δ and $\rho_{surfactant}$.

We first performed a MD simulation in the NPT ensemble at ambient conditions (300 K and 1 atm) of a SWCNT in water and sodium cholate molecules to observe the behavior of surfactant molecules. In Fig. 3.13 we illustrate how the amphibilic molecules approach the SWCNT: starting from a configuration where the surfactant molecules are randomly distributed, we observe how, after 2 ns simulation time, 3 or 4 molecules have approached the SWCNT and remain stable at a short distance of the SWCNT.

From this first simulation we obtain only qualitative evidence for the micelle formation around the SWCNT driven by hydrophobic forces; the whole process takes too long to be simulated via MD and its simulation does not provide further insight or any substantial progress in the understanding of the system.

We performed a second MD simulation, in the NPT ensemble (at 300 K and 1 atm) as well, of a preformed micelle, where the SWCNT is 'wrapped' by cholate molecules to



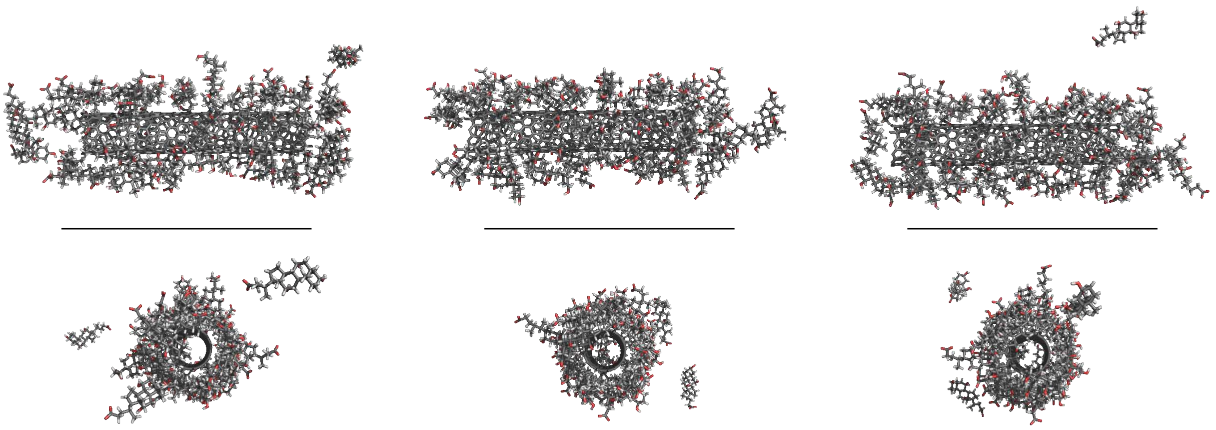
(a) initial configuration

(b) after 2 ns

Figure 3.13: *Snapshots of the first MD simulation of randomly located cholate molecules and the SWCNT in water at time t=0 ns at the left (a) and at time t=2 ns at the right side (b).*

investigate the stability of such a conformation. We studied the influence of the surfactant on the water inside the SWCNT dynamics and obtained a theoretical value (or range of values) of the density of the surfactant molecules and of the thickness of the cholate shell. The initial conformation is artificially extrapolated from the equilibrated interaction cholate-nanotube. In Fig. 3.14, we illustrate the behavior of such a simulated micelle within a simulation time of 5 ns and in Fig. 3.15 we illustrate the distribution of surfactant molecules around the SWCNT for a stabilized system after 2 ns of simulation for the CNT (6,5) and for the CNT (10,2).

The presence of surfactant does not influence the filling of the CNT. After 2ns of simulation, the surfactant being equilibrated, the average number of water molecules inside the CNT (10,2) and inside the CNT (6,5) are about 20 and 18 respectively, and they form a channel inside the CNTs in the same fashion as in the simulations without surfactant.



(a) after 2 ns (b) after 3 ns (c) after 4 ns

Figure 3.14: *Snapshots of the second MD simulation starting from a preformed conformation where the CNT is surrounded by surfactant molecules. (a), (b), and (c) correspond to snapshots of the system after 2, 3 and 4 ns of simulation respectively.*

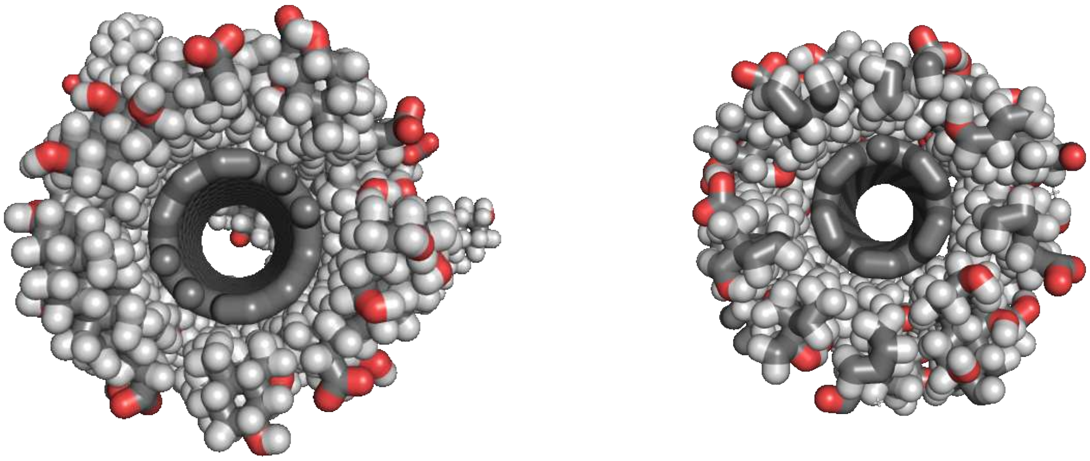
3.4.2 Density

According to the equilibrated configuration illustrated in Fig. 3.15, the average value of Δ is around 0.7 nm. To calculate the density of the surfactant molecules, we do not take into account the water around, but also not the space filled by the water since in the experiment the whole system SWCNT-solvate molecules exclusively migrates to the environment having exactely its density. The volume of the Δ thick shell around the SWCNT is, according to the MD simulation, 80% occupied by surfactant and the rest is water. This has an effect in the way we calculate the total density ρ_{T3}, the volume occupied by the surfactant is 80% the outer shell volume and therefore only this volume is taken into account. The density of the system SWCNT-surfactant results:

$$\rho_{T3} = \frac{(m_{H_2O}\; n_{H_2O}) + \sigma_{graphene}\; 2\pi\; r\; l + \rho_{surfactant}\; X\; \pi\; \left((r+\sigma_C+\Delta)^2 - (r+\sigma_C)^2\right)\; l}{\pi\,(r+\sigma_C)^2\; l + \; X\; \pi\; \left((r+\sigma_C+\Delta)^2 - (r+\sigma_C)^2\right)\; l}, \tag{3.10}$$

where X is the proportion of surfactant in the Δ-thick shell around the SWCNT.

In Fig. 3.16, the density corresponding to eq. 3.10 is illustrated. The grey circles represent again the density without considering the filling of the nanotube with water. The density considering the filling of the tube is illustrated by the black diamonds. The values for n_{H_2O} have been taken from the previous simulations, since the filling of the tube



(a) (10,2) nanotube after 2 ns

(b) (6,5) nanotube after 2 ns

Figure 3.15: *Transversal section of the stabilized SWCNT-surfactant systems (for the CNT (10,2) and for the CNT (6,5)) solvated in water (water and counter ions are not represented)*

is not influenced by the presence of the surfactant around the nanotube for the (10,2)-CNT and (6,5)-CNT, with which we permformed the simulations with the surfactant.

From the experimental data, we can only say that SWCNT of diameter around 0.88 nm are found in section 10, i.e. high density section, and lower densities sections contain a mixture of SWCNT of larger and smaller diameters. Experiments realized using another type of surfactant molecules (with an hydrophilic head and a long hydrophobic tail) are not able to separate the CNTs via ultracentrifugation.

Analysis of the density function

According to eq. 3.10, the density of the whole system SWCNT-surfactant-water depends on the thickness Δ of the surfactant shell around the SWCNT, the density $\rho_{surfactant}$ of the surfactant, and the surfactant proportion X in the shell. The density for the system SWCNT-surfactant-water is computed while accounting for the number n_{H_2O} of water molecules inside the CNTs and integrating the results of our MD simulations (see Fig. 3.17(a)).

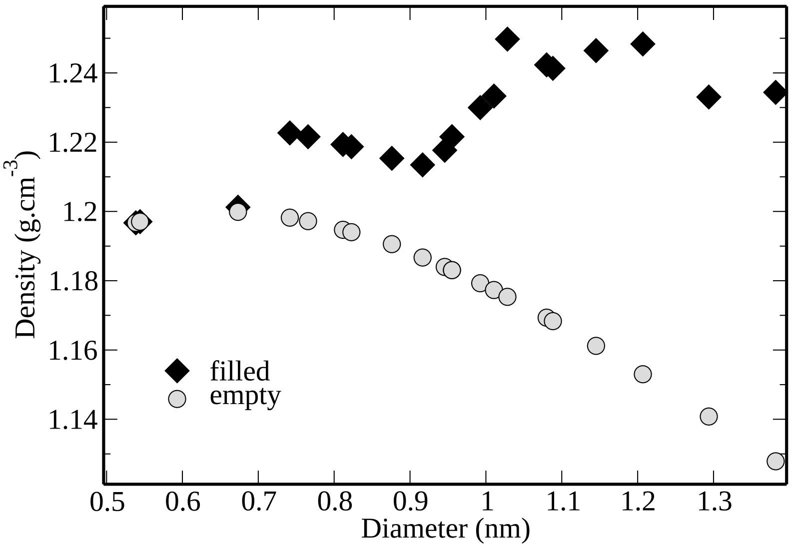
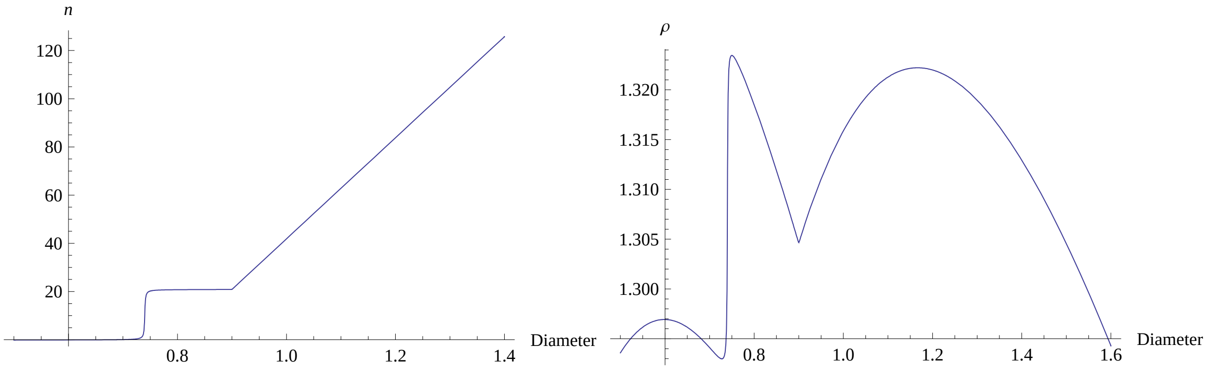


Figure 3.16: *Density of the system SWCNT-surfactant vs. SWCNT diameter. The grey circles correspond to de density computed according to eq. 3.10 where 80% of the outer shell is occupied by the surfactant molecules and the black diamonds take into account the filling of the SWCNT*

In Fig. 3.17(b), the density profile of the whole system is illustrated for Δ=0.7 nm, X=0.5 and $\rho_{surfactant}$=1.3 g.cm^{-3}. The choice of these parameters is motivated by the outcomes of our simulations. The resulting density profile serves as starting point for further analysis of how it changes with variations of the parameters Δ, $\rho_{surfactant}$ and X. The irregular filling of the CNT with water leads to the presence of two local maxima. The first one shows up for diameter values around 0.74 nm and is due to the formation of the first water shell. The second maximum is much broader and located around 1.2 nm. It originates from the the successive chains contributing to the inner water shell.

We study how the density profile behaves in different conditions for given Δ, $\rho_{surfactant}$, and X. In Fig. 3.18, we illustrate the diameter corresponding to the maximal density as a function of these parameters. The abrupt change for the diameter corresponding to the two identified maxima from Fig. 3.17(b) is due to a switch between these local maxima, which dominate depending on the above parameters.

Fig. 3.18(a) illlustrates where the first or second maximum dominates depending on Δ and X, while keeping the density $\rho_{surfactant}$ at a fixed value of 1.3 g.cm^{-3}. We can see



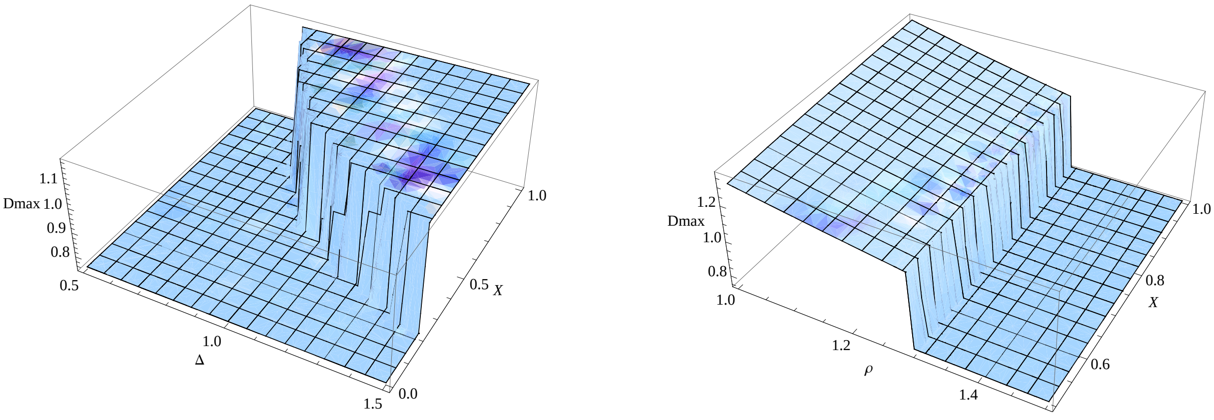
(a) Number of water molecules inside the CNT (b) Density of the system (g.cm^{-3})

Figure 3.17: *(a) Simplification of the number of water molecules inside the CNTs vs. diameter. The simplification accounts for the characteristics as obtained from the MD simulations. (b) Density of the SWCNT-surfactant-water system according to eq. 3.10 in g.cm^{-3} vs. diameter for Δ=0.7 nm, $\rho_{surfactant}$=1.3 g.cm^{-3} and X=0.5 with n_{H_2O} as given by (a).*

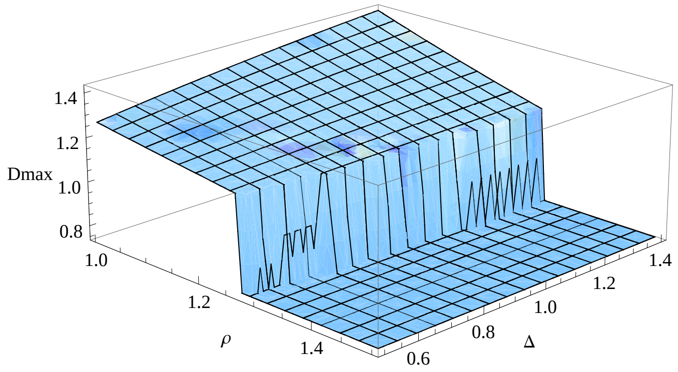
that the variation of the thickness of the surfactant shell and of the volume proportion of surfactant molecules does not influence the location of the local maxima (flat behavior of D_{max}) but their absolute contribution to the total density (switching between first and second maximum as absolute maximum). In Fig. 3.18(b) we observe that, for a fixed Δ=0.7 nm, there is no significant effect of the volume proportion X on the location of the absolute maximum of the density. However, it decreases continously with increasing $\rho_{surfactant}$ and switches to the constant value of 0.74 nm above $\rho_{surfactant}$ ~1.3 g.cm^{-3}. In Fig. 3.18(c) we study the localization of the second maximum for fixed X=0.5. We find that it decreases with decreasing Δ and increasing $\rho_{surfactant}$. Here, for values of $\rho_{surfactant}$ between 1.25 and 1.35 g.cm^{-3} the switch between the maxima happens.

In summary, equation 3.10 combined with the outcomes of our MD simulations allows us to estimate a value around 1.3 g.cm^{-3} for the density of the system. However, we do not find a maximum of the density for CNTs with diameter around 0.88 nm. Instead there is an invariant maximum at 0.74 nm followed by an invariant minimum at 0.9 nm and a slightly varying second maximum around 1.1 to 1.3 nm. If we perform the same study in the case without considering inner wetting of the CNTs, which is in contradiction to our computational findings, we find a density maximum for CNTs with diameter around 0.88 nm. This happens for a very low density of the surfactant around the CNT at $\rho_{surfactant} \sim 0.75 g.cm^{-3}$.

We conclude that first, more detailed experimental results are needed to better un-



(a) $D_{max}(\Delta,X)$ for $\rho_{surfactant}$=1.3 g.cm^{-3} (b) $D_{max}(\rho_{surfactant}, X)$ for Δ=0.7 nm



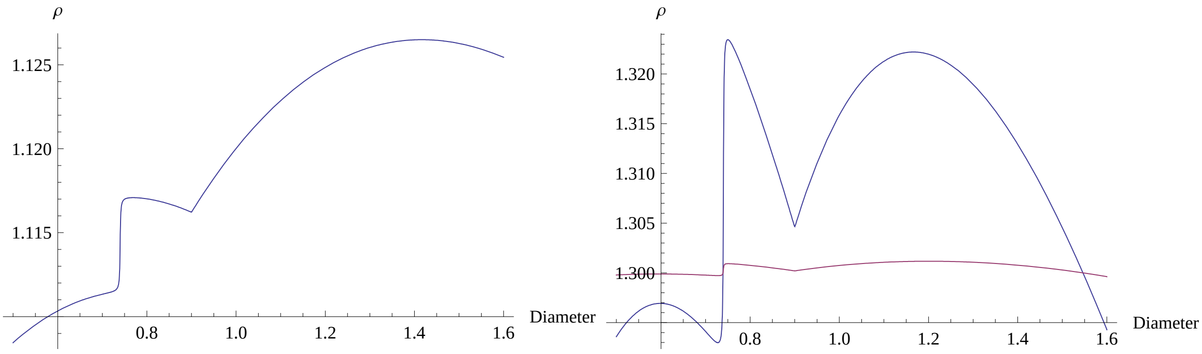
(c) $D_{max}(\Delta, \rho_{surfactant})$ for X=0.5

Figure 3.18: *Illustration of the Diameter D_{max} of the CNT for which the density of the CNT-surfactant-water is maximal as a function of Δ, $\rho_{surfactant}$, and X.*

derstand the density profile in the theoretically studied diameter regime. Second, an exhaustive MD study of the surfactant behavior around the CNTs for all the CNTs with diameter within the studied range could reveal some surfactant effects that can not be simply extrapolated from one result.

On the other hand, experiments carried out by Arnold et al.[72] found that, for CNTs with diameter within a range between 0.76 and 1.01 (at 0.76 nm, 0.78 nm, 0.84 nm, 0.88 nm, 0.97nm and 1.01nm), the density has the value 1.17 $g.cm^{-3}$ with an error of 0.02 and increases with the diameter. The CNTs in these experiments were wrapped with DNA, which relates to a thickness of the shell around the CNT between 2 and 3 nm. We find that our model reproduces very well such a behavior as illustrated in Fig. 3.19(a).

If another surfactant is used, no possible separation via centrifugation takes place. This can also be interpreted using this model: as the surfactant shell around the SWCNT



(a) Δ=2.5nm $\rho_{surfactant}$=1.15$g.cm^{-3}$ (b) blue:Δ=0.7nm magenta:Δ=7nm

Figure 3.19: *The graphic on the left side illustrates the density vs. the diameter for values of Δ=2.5nm and $\rho_{surfactant}$=1.15g.cm^{-3} corresponding to the experiment of Arnold et al. [72]. The graphic on the right side shows the influence of the thickness of surfactant shell.*

becomes larger, the density difference between SWCNTs with different diameters becomes smaller: The density considering a $\Delta \sim 0.7nm$ takes values within a range 1.29 to 1.32 nm but if the surfactant shell becomes ten times larger as it would happen for the second surfactant used, then the variations of the density of the system SWCNT-surfactant-water are smaller than 0.002 and therefore not detectable by the experimental setup. In Fig. 3.19(b), we plot the resulting densities for the same values of $\rho_{surfactant}$ and X and for Δ = 0.7 nm (in blue) and Δ = 7 nm (in magenta).

3.5 Conclusion

In this chapter, we have briefly introduced carbon nanotubes and presented recent experiments that aim to separate carbon nanotubes by centrifugation. We have taken the results of these experiments as starting point for our simulations and proposed a formulation for the density of the complex nanotube-surfactant.

Experimental results demonstrate that the CNTs with maximal density correspond to 0.88 nm diameter CNTs. Based on former experiments and simulations that confirm the presence of water inside the CNTs, we tried to explain this unusual density profile as an effect or consequence of the water structure inside the CNTs. We performed MD simulations with different parameterizations for the non-bonded interactions between the carbon atoms and the water molecules of SWCNTs solvated in water and studied the behavior of water molecules inside the CNTs.

Our results with this simple model do not agree with the experimental results, since in experiments other molecules than only the CNT and the water are present in the solute. These are the surfactant molecules that cluster around the CNTs and allow for their separation. Moreover, the separation of CNTs with other surfactant molecules as the ones used in the described experiments (sodium cholate) are not possible. Thus, the effect of the environment is decisive for the explanation of the experiment rather than the behavior of the water inside the CNTs.

We therefore considered the presence of surfactant molecules and performed MD simulations of 2 different SWCNTs in a ‘real’ environment. From the MD simulations, we extracted a first try values for the surfactant characteristics, density and conformation around the SWCNTs and calculated the density of the complex SWCNT-water-surfactant with the obtained values. We observed that the density behavior vs. the SWCNT diameter does not present an absolute maximum for SWCNTs having a diameter of 0.88 nm, but its value is on the same order of magnitude as in the experiments and moreover, agrees with the recent experimental results of Arnold and coworkers [72].

We conclude that a more accurate description of the experimental environment is needed in a next step. The experimental conditions are not completely reproduced in our simulations, since experimentally, the system SWCNT-surfactant is not embedded exclusively in water but in a solution of variable density, corresponding to the one of the studied system. The fact that the real environment is not simply water could affect the behavior of the surfactant molecules around the SWCNT. It raises then the idea of an ‘designing’ experiments to sort a particular type of CNT, by choosing a solvent and a surfactant that fullfils the density and dimensions required, and maybe a solvation medium with a given polarizability.

Chapter 4

Protein folding

In the present section, we investigate how proteins adopt their native (folded) conformation starting from a random unstructured configuration. We combine two simulation techniques with 'complementary' characteristics: deterministic MD and stochastic optimization methods, to analyze the folding process. MD protein folding events have been very rare; for the first time in 2002, Simmerling et al. [87] observed folding of the twenty amino acids miniprotein trpcage using MD. The reason for the scarcity of proteins being folded with all-atom MD is the amount of computational time needed. The simulation of a macromolecular system as a protein and its environment over a time slot comparable to protein folding scales requires a huge computational effort and the existence of metastable states or kinetic traps slow down the folding process. Our strategy consists of first exploring the energy surface of the protein to decide if a concrete protein is a good candidate for detailed MD simulations on and then making a dynamics study on a protein only if its energy surface is not much rugged.

The present chapter is structured as follows. In the section 4.1 we briefly describe the essentials of protein structure; we introduce their building blocks, the amino acids, describe their basic structure, and summarize the problem of protein folding, or how the native structure is reached. In section 4.2 stochastic energy optimization methods are described, and the results on a particular peptide are presented. Finally, in section 4.3, the MD simulations are detailed and the results leading to a protein folding picture are interpreted.

4.1 Proteins

Proteins are the nanoscale molecular machinery of the cell and essential parts of organisms. They catalyze biochemical reactions and have structural or mechanical functions, they are important in cell signaling, in immune responses, in cell adhesion, and in the cell cycle. All functions of proteins are strictly related to their three-dimensional conformation, which is determined by the amino acid sequence.

The question of how sequence and structure are related arises, as well as how the functional conformation of a protein is achieved. We briefly discuss the main characteristics of protein structure and the principal problem of protein folding.

4.1.1 Protein structure

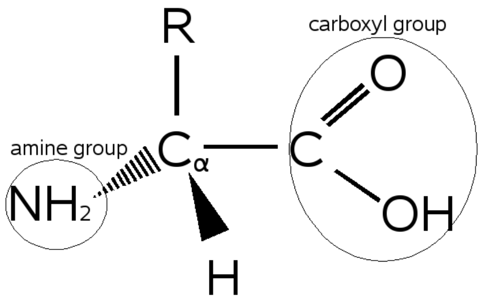
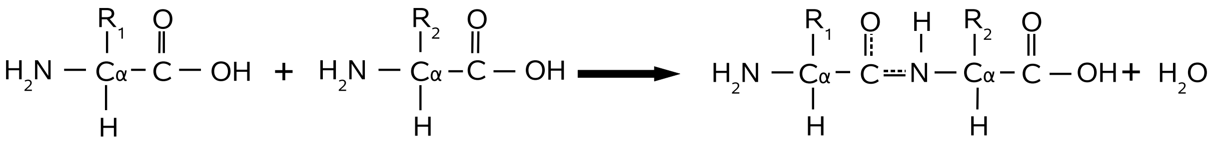


Figure 4.1: *Amino acid basic structure. The side chain, represented by R, is bonded to the* C_α.

Proteins are large organic molecules made of amino acids (see Fig. 4.1) arranged in a linear chain and bonded together by a peptide bond (see Fig. 4.2) that joins the carboxyl group and the amino group of neighboring amino acids.



$$H_2N - \underset{H}{\overset{R_1}{C_\alpha}} - \overset{O}{\overset{\|}{C}} - OH + H_2N - \underset{H}{\overset{R_2}{C_\alpha}} - \overset{O}{\overset{\|}{C}} - OH \longrightarrow H_2N - \underset{H}{\overset{R_1}{C_\alpha}} - \overset{O}{\overset{\|}{C}} = \overset{H}{\overset{|}{N}} - \underset{H}{\overset{R_2}{C_\alpha}} - \overset{O}{\overset{\|}{C}} - OH + H_2O$$

Figure 4.2: *Peptide bond formation between two amino acids. A peptide bond is a covalent bond between the amine and the carboxyl group of amino acids.*

Amino acids structure consists of the backbone naturally occurring in all amino acids, and of a side chain. There are twenty different side chains that determine the properties of the amino acids: amino acids are hydrophilic if their side chain is polar and hydrophobic if their side chain is non-polar. Polar side chains have various functional groups such as acids, amides, alcohols, and amines. At cellular pH the carboxyl group in the acidic side chains and and the amino group in the basic side chains acids are dissociated and those amino acids are then ionic. Side chains that contain alcohols or amide functional groups are polar even if they are not ionic. Non-polar amino acids have side chains which have pure hydrocarbon alkyl groups (alkane branches) or aromatic (benzene rings).

Protein structure is classified in different levels (see Fig. 4.3). The **primary structure** is the amino acids sequence. The **secondary structure** refers to the local conformation of some part of the protein. It mostly focuses on common regular folding patterns of the protein backbone. Fig. 4.4(a) shows the representation of a polypeptide chain. Because of the delocalized π orbital of the carboxyl group the peptide bond has a partial double bond character and rotating around it requires high energy. Rotation around the other covalent single bonds of the backbone is much less costly. The torsional angles ϕ, around the bond between C_α and N, and ψ, around the bond between C_α and C, determine the secondary structure of the peptide chain completely. A few types of secondary structure are particularly stable and occur widely in proteins. The most prominent are α-helix and β configurations (see Fig. 4.4(b), 4.4(c)).

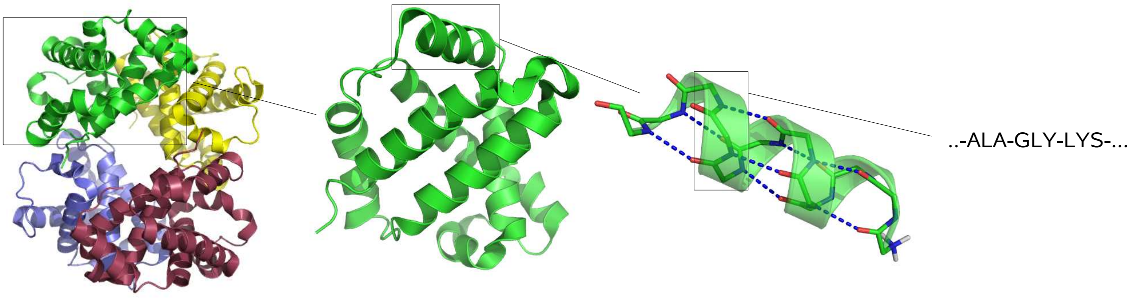
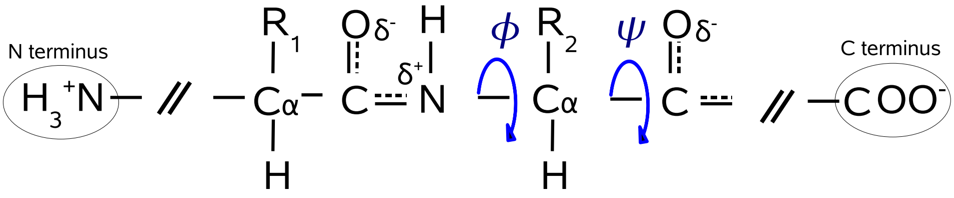
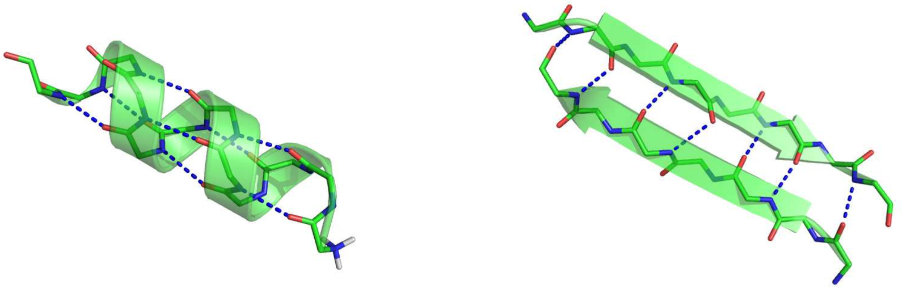


Figure 4.3: *Protein structure. From quaternary structure to primary structure.*

The overall three-dimensional arrangement of amino acids is referred to as the protein's **tertiary structure**. Whereas the term secondary structure refers to the spatial arrangement of amino acid residues that are adjacent in the primary structure, tertiary structure includes longer range aspects of the amino acid sequence. Some proteins contain two or more separate polypeptide chains or subunits, which may be identical or different. The arrangement of these protein subunits in three-dimensional complexes constitutes the **quaternary structure**.



(a) Torsional angles.

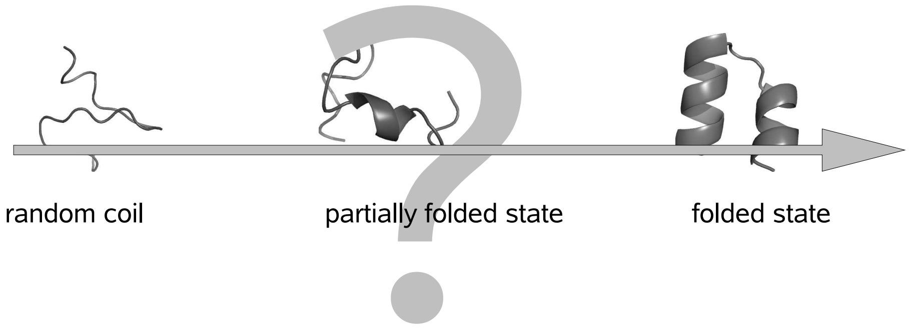


(b) Alpha helix ϕ and ψ

(c) Beta sheet ϕ and ψ

Figure 4.4: *Secondary structure of peptides. (a) shows the representation of a ploypeptide chain. (b) and (c) show the most prominent secondary structure patterns. The backbone of the protein is represented with sticks and the polar contacts responsible for stabilizing the secondary structures are depicted in blue.*

4.1.2 The problem of protein folding



Most proteins fold into a unique structure. How does a polypeptide chain achieve its native conformation? In the folding pahway of a polypeptide chain not all the principles that guide the physical process have been worked out. The investigation of how proteins reach their functional conformation is motivated by the fact that defects in protein folding may be the molecular basis for a wide range of genetic disorders. Improved understanding of protein folding may thus lead to new therapies for misfolding diseases.

The time scales for protein folding range between 10^{-7}s for small peptides or fast folders [88] and 10^4s. If proteins would fold spontaneously by a random process in which they try out all possible sterically and energetically allowed conformations, the folding process would take many orders of magnitude longer than it does. (A protein having n amino acids can adopt $O(3^n)$ possible conformations, if each amino acid can adopt 3 different configurations, α-helix, β-sheet or random coil.) Therefore there must be 'shortcuts' in protein folding rather than being a completely random trial and error process. This problem was first pointed out by Cyrus Levinthal in 1968, and is often called Levinthal's paradox [89].

In a hierarchical model of protein folding, local secondary structures form first and the process continues until complete domains form and the entire polypeptide is folded. In an alternative model, folding is initiated by a spontaneous collapse mediated by hydrophobic interactions among nonpolar residues of the polypeptide into a compact state that may have a high content of secondary structure, but many amino acid side chains are not entirely fixed.

Thermodynamically, the folding free energy landscape can be viewed as a kind of free energy funnel [90]. The unfolded states are characterized by a high degree of conformational entropy and relatively high internal energy. As folding proceeds, the narrowing of the funnel represents a decrease in the number of conformational species present. Small depressions along the sides of the free-energy funnel represent metastable intermediates that can briefly slow the folding process. At the bottom of the funnel, the ensemble of folding intermediates has been reduced to a single native conformation.

4.2 Analysis of the free energy landscape

According to the thermodynamic hypothesis [91], the native structure of a protein can be predicted by locating the free energy minimum of the protein and its environment. Using stochastic (Monte Carlo) methods to explore the energy landscape of biomolecules takes less computational effort and yields reproducible results compared to the deterministic approach (Molecular Dynamics) for locating the native state.

In the stochastic approach, the energy landscape is explored by generating randomly changes in the geometry of the molecule and accepting the 'moves' if the energy change is favorable, or with decreasing probability if the energy change is slightly uphill. To compute the energy of the randomly generated conformations, a free energy protein force field, such as PFF01 [92] is required.

4.2.1 Stochastic optimization methods

In Monte Carlo (MC) simulations, the relation between intermediate states in the simulation is not deterministic as in MD but based on a stochastic process. Monte Carlo simulations generate Markov chains [1] of three-dimensional conformations of the protein.

Based on energy criteria, configurations are accepted or rejected: When the energy calculated according to the forcefield [92] E' is lower than the energy of the previous configuration E or fullfils $e^{-(E'-E)/kT} < random[0;1]$ the new configuration is accepted, otherwise the process is repeated for the previous configuration (see Fig. 4.5).

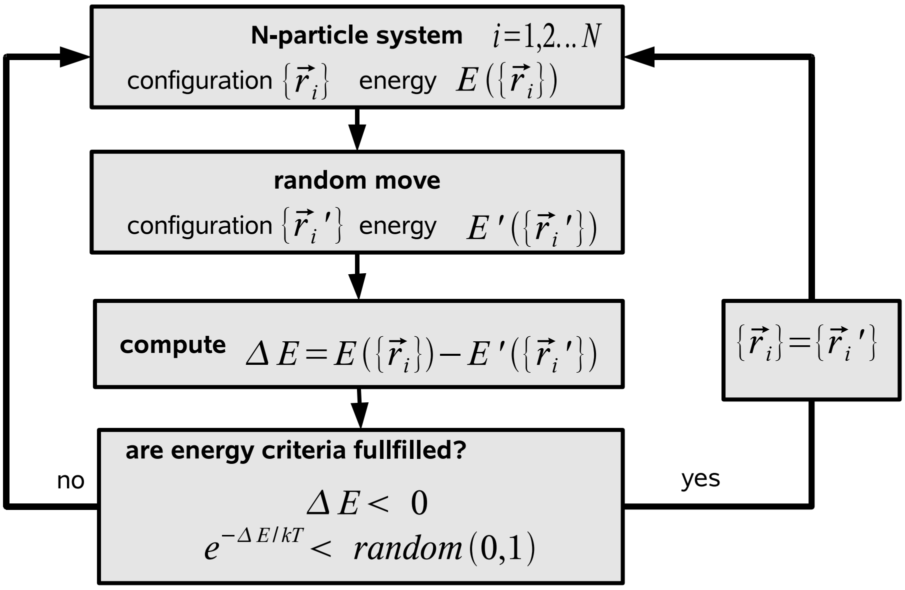


Figure 4.5: *Stochastic optimization methods.*

Each configuration differs from the previous one by rotations about single bonds. In the folding process under physiological conditions the degrees of freedom of a peptide are confined to rotations about single bonds. In our simulation we therefore consider only moves around the sidechain and backbone dihedral angles, which are attempted with thirty and seventy percent probability respectively. The moves for the sidechain angles are drawn from an equally distributed interval with a maximal change of 5 degrees. Half

---

[1] A Markov chain is composed of discrete stochastic processes such that for each process the knowledge of the previous states is irrelevant for predicting the probability of subsequent states. In this way a Markov chain has ‘no memory’ and no given state has any causal connection with a previous state.

The transition probability $T(\{\mathbf{X}_2\},\{\mathbf{X}_1\})$ for the system to go to state $\mathbf{X}_2$ from state $\mathbf{X}_1$ is normalized to the unity, and for systems in equilibrium the transition probability often obeys the detailed balance criterion: $T(\{\mathbf{X}_2\},\{\mathbf{X}_1\})\rho(\mathbf{X}_1) = T(\{\mathbf{X}_1\},\{\mathbf{X}_2\})\rho(\mathbf{X}_2)$

of the backbone moves are generated in the same fashion, the remainder is generated from a move library that was designed to reflect the natural amino-acid dependent bias towards the formation of α-helices or β-sheets. The probability distribution of the move library was fitted to experimental probabilities observed in the PDB database [93]. While driving the simulation towards the formation of secondary structure, the move library introduces no bias towards helical or sheet structures beyond that encountered in nature.

4.2.2 Basin Hopping technique

The low-energy part of the free energy landscape of proteins is extremely rugged due to the comparatively close packing of the atoms in the collapsed ensemble. Rugged potential energy surfaces are characterized by the existence of many low-lying minima, which are separated by high energy barriers. For this reason, the global optimum of such a surface is difficult to obtain computationally. Simple methods, such as steepest descent [2] or simulated annealing[3], are almost always trapped in metastable conformations.

Efficient optimization methods must therefore speed up the simulation by avoiding high energy transition states, by adapting large scale moves wherever possible or by accepting unphysical intermediates. One of the simplest ideas to effectively eliminate high energy transition states of the free-energy surface is the basis of the basin hopping technique [94, 95], also known as Monte-Carlo with minimization (see Fig. 4.6). This method simplifies the original potential energy surface by replacing the energy of each conformation with the energy of a nearby local minimum. This replacement eliminates the high energy barriers that are responsible for the freezing problem in simulated annealing. In many cases the additional minimization effort to find a local minimum for each starting configuration is more than compensated by the increase of efficiency of the stochastic search on the simplified potential energy surface. While each simulated annealing run is typically much more expensive than a local minimization using gradient based techniques, it can nevertheless be competitive for very rugged potential energy surfaces, or when the computation of the gradient of the potential is prohibitive. The number of moves in each individual basin hopping cycle, is increased with the square root of the cycle number m as N=10000$\sqrt{m}$.

---

[2] Steepest descent is an optimization algorithm to find a local minimum of a function. It takes steps proportional to the negative of the gradient of the function at the current point.

[3] Simulated annealing consists of slowly decreasing the temperature of the system to find the minimal energy configuration.

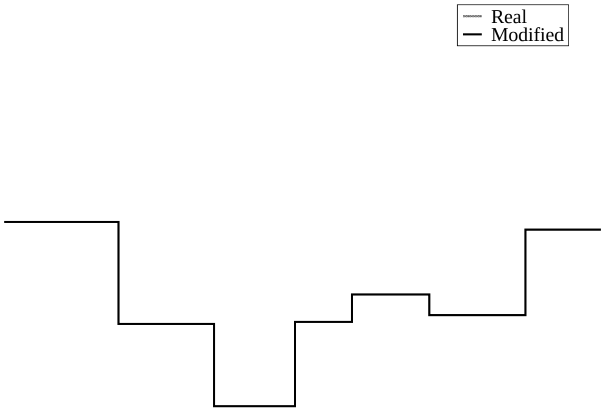


Figure 4.6: *Schematic graph to illustrate the basin hopping technique: the real potential energy surface is simplied by replacing the energy of each conformation with the energy of the nearby local minimum.*

4.2.3 Free Energy protein force field PFF01

The all-atom (with the exception of apolar CH groups) free-energy protein forcefield (PFF01) models the low-energy conformations of proteins with minimal computational demand [96, 97, 98]. The forcefield parameterizes the internal free energy of the protein (excluding backbone entropy) and contains the following non-bonded interactions:

$$V(\vec{r}_i^N) = \sum_{ij} V_{ij}\left[\left(\frac{R_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{ij}\frac{q_i q_j}{\varepsilon_{g(i)g(j)} r_{ij}} + \sum_{i} \sigma_i A_i + \sum_{\text{hbonds}} V_{hb} \quad (4.1)$$

Here r_{ij} denotes the distance between atoms i and j and $g(i)$ the type of the amino acid i. The Lennard-Jones parameters (V_{ij}, R_{ij} for potential depths and equilibrium distance) depend on the type of the atom pair and were adjusted to satisfy constraints derived from a set of 138 proteins of the PDB database [99, 96, 100]. The non-trivial electrostatic interactions in proteins are represented via group-specific dielectric constants ($\varepsilon_{g(i)g(j)}$ depending on the amino-acids to which the atoms i and j belong). The partial charges q_i and the dielectric constants were derived in a potential-of-mean-force approach [101]. Interactions with the solvent were first fit in a minimal solvent accessible surface model [102] parameterized by free energies per unit area σ_i to reproduce the enthalpies of solvation of the Gly-X-Gly family of peptides [103]. A_i corresponds to the area of atom i that is in contact with a fictitious solvent. Hydrogen bonds are described via dipole-dipole interactions included in the electrostatic terms and an additional short range term for backbone-backbone hydrogen bonding (CO to NH) which depends on the OH distance, the angle between N,H and O along the bond and the angle between the CO and NH axis [98].

4.2.4 Energy profile of 1WQE

The protein studied here, 1WQE (see Fig. 4.7) has 23 amino acids. Its secondary structure has two α-helix spanning from amino acids 2 to 12 and 15 to 22. Its tertiary structure is stabilized by two covalent bonds between the sulfur atoms of the cystein residues, *disulfide bridges* involving residues 2-22 and 8-18.

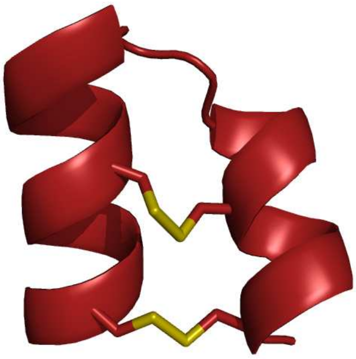


Figure 4.7: *Cartoon representation of 1WQE. The sidechains of the cystein residues are explicitely represented with sticks and sulfur atoms are yellow. Two disulfide bridges between 1:cystein4 and cystein22 and 2:cystein8 and cystein18 stabilize the tertiary structure consiststing of two α-helices.*

Structures for the peptide 1WQE with 23 amino acids, were retrieved from the PDB database [104] and unfolded by setting all backbone dihedral angles to random values until non-clashing conformations were obtained. The starting conformations had backbone root mean square deviations (RMSB)[4] of 9.7 Å to the native conformation of 1WQE; they had no secondary structure.

We performed 20 independent basin hopping simulations with 200 cycles each. In order to avoid any bias towards the native conformation, there was no potential representing the disulfide bridges in these simulations. We use a simulated annealing process [105] for

---

[4] Root Mean Square Distance is the most common measure of the fit between two structures A and B:

$$RMSD_{A-B} = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} (\vec{r}_{iA} - \vec{r}_{iB})^2}{N_{atoms}}} \tag{4.2}$$

the minimization step. The temperature is decreased geometrically from its starting to the final value of $T_f = 2$K. Following an optimized protocol [106] the starting temperature T_s is drawn randomly from a distribution $p(T_s) \propto \exp(T_s/T_0)$. The performance of the method is only weakly dependent on the choice of T_0, which was chosen as 750K. At the end of one annealing step the new conformation is accepted if its energy difference to the current configuration is not higher than a given threshold energy ε_T, an approach proven optimal for certain optimization problems [107]. Throughout this study we use a threshold acceptance criterion of 1 kcal/mol.

Table 4.1 summarize the energies, RMSB deviations and secondary structure for the final population of these simulations. There are thirty NMR models for the peptide 1WQE, which differ in the unstructured tail-fragments after amino-acid 20 of the sequence. The table therefore reports the RMSB deviation to the closest model and the RMSB deviation to the structurally conserved part.

Predictive folding is achieved, when near-native structures dominate the low-energy spectrum of the simulated ensemble. In a free-energy forcefield the native conformation is selected on the basis of its estimate of internal free-energy in comparison to other conformations with well defined secondary and tertiary structure. This is in contrast to MD or Replica Exchange Methods (REM)[5] investigations, where occupation probability determines the thermodynamically stable conformation. Thus finding a particular conformation repeatedly with the lowest energy, as was observed for here, predicts the native conformation. Not in all simulations that reach the native conformation all stabilizing tertiary interactions are fully formed. As a result there may be many more near-native conformations that are slightly higher in energy. We found the lowest 18 of twenty simulations to converge to near-native conformations of 1WQE. 90 % of the simulations converge to conformations with Root Mean Square deviation of the Backbone (RMSB) deviations of less than 2.2 Å to the native conformation.

In Fig. 4.8(a) we show the overlay of the lowest energy conformation with the respective experimental model. The figure demonstrates the high degree of similarity of the folded and experimental conformations.

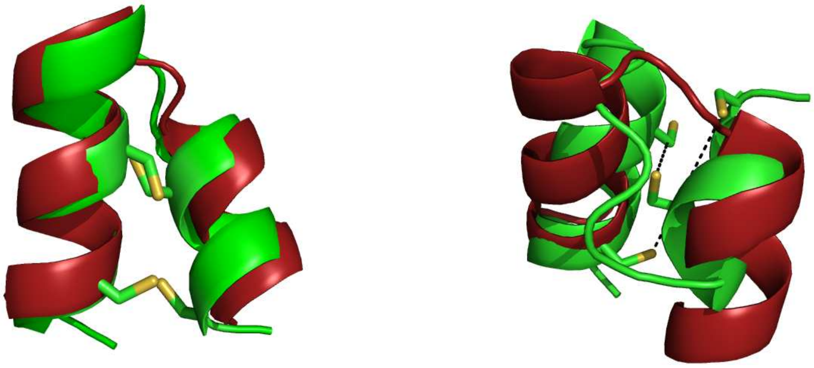
Next we turn to the surface of the internal free-energy (excluding backbone entropy) of 1WQE. Fig. 4.9 shows energy versus RMSB for all accepted configurations at the end of basin hopping cycles (from all simulations). The triangles indicate the terminal configurations of the individual simulations. We clearly see two broad funnels of conformations,

---

[5] REM consists of parallel MD simulations of several non-interacting copies (or replicas) of the original system at different temperatures, that are exchanged during the simulations.

RMSB	RMSB$_{1-20}$	Energy	Secondary Structure
1.90	1.84	-57.00	CHHHHHHHHHHHHTCHHHHHHHHHC
1.64	1.63	-56.90	CHHHHHHHHHHHHTCHHHHHHHHHC
1.66	1.66	-56.90	CHHHHHHHHHHHHTCHHHHHHHHHC
1.70	1.67	-56.80	CHHHHHHHHHHHHTSHHHHHHHHHC
1.70	1.69	-56.50	CHHHHHHHHHHHTCCHHHHHHHHHC
1.68	1.67	-56.50	CHHHHHHHHHHHTCCHHHHHHHHHC
2.13	2.10	-56.50	CHHHHHHHHHHHTCCHHHHHHHHHC
1.74	1.72	-56.50	CHHHHHHHHHHHTCCHHHHHHHHHC
1.73	1.70	-56.40	CHHHHHHHHHHHTCCHHHHHHHHHC
1.72	1.70	-56.40	CHHHHHHHHHHHTCCHHHHHHHHHC
1.68	1.67	-56.30	CHHHHHHHHHHHTCCHHHHHHHHHC
1.74	1.71	-56.30	CHHHHHHHHHHHTCCHHHHHHHHHC
1.69	1.66	-56.30	CHHHHHHHHHHHTCCHHHHHHHHHC
2.12	2.14	-56.20	CHHHHHHHHHHHTCCHHHHHHHHHC
1.69	1.66	-56.20	CHHHHHHHHHHHTCCHHHHHHHHHC
1.69	1.67	-56.10	CHHHHHHHHHHHTCCHHHHHHHHHC
1.71	1.68	-56.10	CHHHHHHHHHHHTCCHHHHHHHHHC
2.18	2.14	-55.00	CHHHHHHHHHHHCSCHHHHHHHHHC
2.04	2.02	-54.70	CHHHHHHHHHHHCSCHHHHHHHHHC
5.73	4.54	-54.40	CHHHHHHHHHHTCCSCHHHHHHHHC
4.71	3.84	-53.50	CHHHHHHHHHHCTTSCHHHHHHHHC

Table 4.1: *Final population of decoys of the basin hopping simulations for 1WQE. We computed the minimal RMSB deviation (in Å) to the thirty experimental models of the full protein and of amino-acids 1-20 respectively. The secondary structure was computed with DSSP [108]: H,T,S,C designate helix, turn, strand and coil conformations respectively.*



(a) RMSB = 2Å (b) RMSB = 6Å

Figure 4.8: *Overlay of the structures obtained after stochastic energy minimization (green) and the native structure (red). (a): folded configuration. (b): configuration corresponding to the non-native funnel. The black dashed lines indicate the distance between sulfur atoms.*

which terminate into low-energy structures with 2 Å and about 6 Å RMSB deviation to the native conformation respectively. The configuration corresponding to the non-native funnel is shown in Fig. 4.8(b). This conformation is inconsistent with the formation of the correct number of native disulfide bridges of this peptide. There is only one, very broad folding funnel consistent with the native disulfide bridge topology. For this reason, the proteins studied here may be ideal examples to follow the kinetics of protein folding with molecular dynamics or replica exchange methods [109, 110, 111].

4.3 Folding Kinetics

The free energy surface of 1WQE, as illustrated in Fig. 4.9, is much more simple than that encountered for other proteins. However the internal free-energy estimate does not contain backbone entropy, stabilization of one particular conformation with respect to all others does not mean that this conformation is stable with respect to the unfolded ensemble. To settle this question kinetic or thermodynamic simulations must be performed. We have therefore performed all-atom implicit water molecular dynamics simulations for this protein.

In the present section, we give a detailed description of the simulation conditions and we present the results obtained from MD simulations. Finally we draw some conclusions about the folding pathways of this protein.

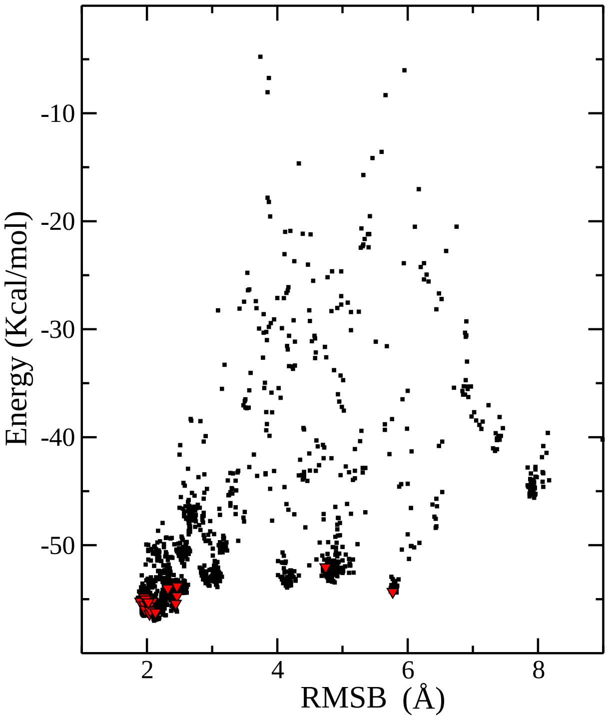
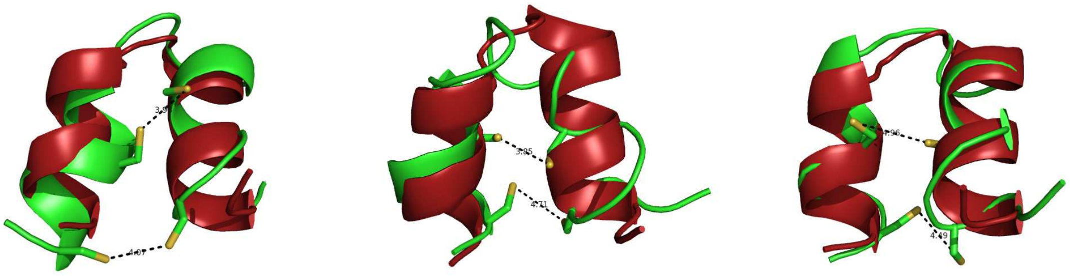


Figure 4.9: *Plot of the energy vs. the RMSB in all accepted conformations in the simulations for 1WQE, the red triangles show the best conformations of the 21 simulations. There are only two structural clusters in the free-energy landscape with characteristic RMSB deviations of around 2 and 6 Å to the experimental model.*

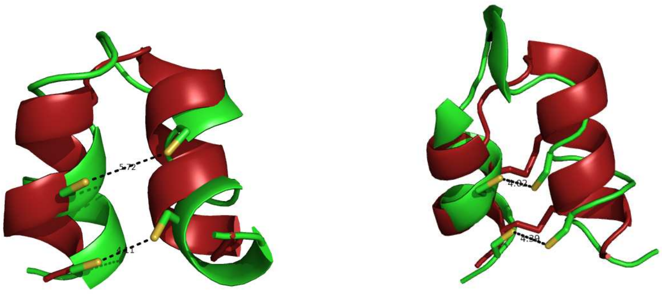
4.3.1 MD simulation

Starting from the same unfolded conformation as in the MC run, we perform all-atom implicit water molecular dynamics simulation using the AMBER8 simulation package [25] with the AMBER99 forcefield using the Born/SASA solvation model [112, 113, 114, 115].

The system is first minimized by steepest descent. Starting from 'frozen' conformations, we increase the temperature slowly, from 0 K to 300 K and from 0 K to 325 K for different systems. The temperature is coupled following weak coupling algorithm. We generate five trajectories with 50 ns total simulation time each, three at 300K and two at 325K. For a better representation of electrostatic interactions, no cutoff is defined, thus no artificial effect can be originated by a step in the electrostatic potential.



(a) RMS 3.5 Å (b) RMS 3.3 Å (c) RMS 3.0 Å



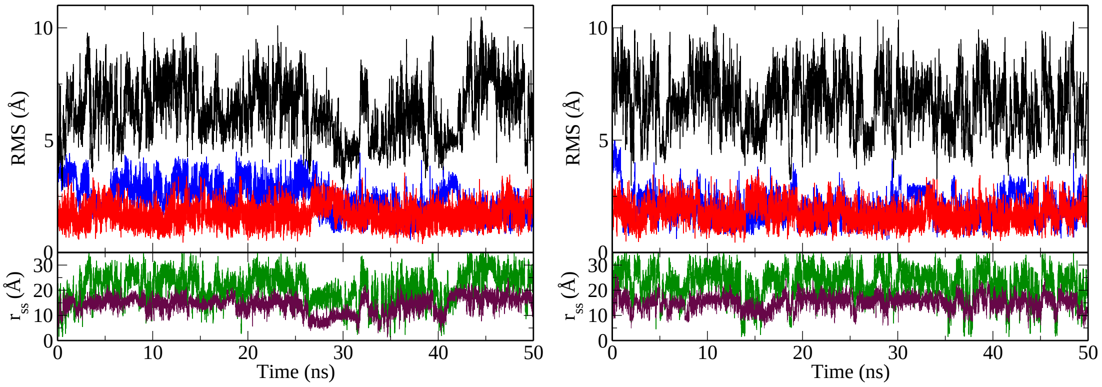
(d) RMS 5 Å (e) RMS 5Å

Figure 4.10: *Overlay of the structures obtained after molecular dynamics (green) and the native structure (red). (a), (b) and (c) correspond to snapshots from a simulation at 325 K after 40 ns, 43.75 ns and 43.76 ns of simulating respectively. (d) and (e) correspond to different simulations at 325 K as well after 46.2 ns and 48.8 ns of simulation*

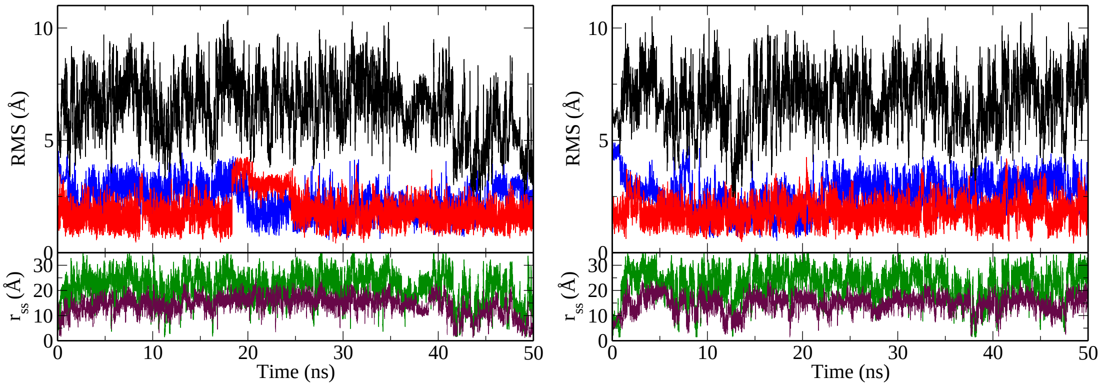
4.3.2 Results

The results for the deviation of the actual conformation from the native structure and the two helices are shown in Fig. 4.11. The simulations equilibrate quickly into a rapidly fluctuating ensemble with an average overall Root Mean Squared Deviation (RMSD) deviation between 5 and 8 Å. When we analyze the RMSD deviation of the helical segments however (Helix 1: 1-11, Helix 2: 15-21), we find that the entire simulation is dominated with conformations that are within 1-2 Å of the respective fragment of the protein.

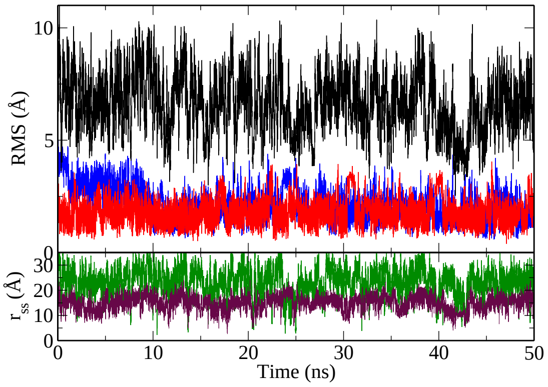
We have also analyzed the helix propensity as a function of time for each amino acid as a function of time, as measured by DSSP. Fig. 4.12 demonstrates a very strong helical content for both segments, but the propensity of helix formation may be forcefield dependent. The figure illustrates very nicely that numerous folding and unfolding events occur for each helix. Both helices dissappear completely for short time windows during the simulation, only to form again on a 10 ps timescale.



(a) 300 K (b) 300 K

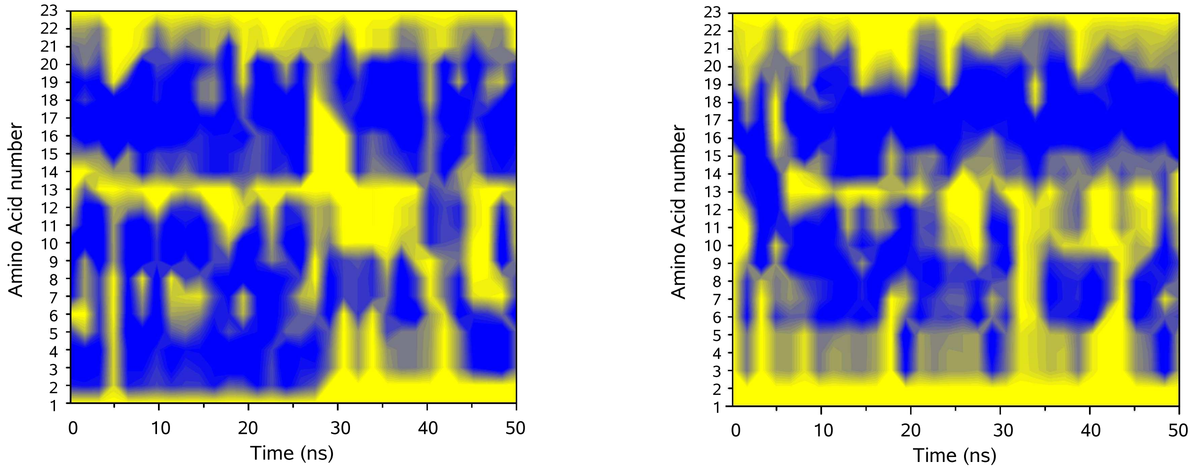


(c) 325 K (d) 325 K



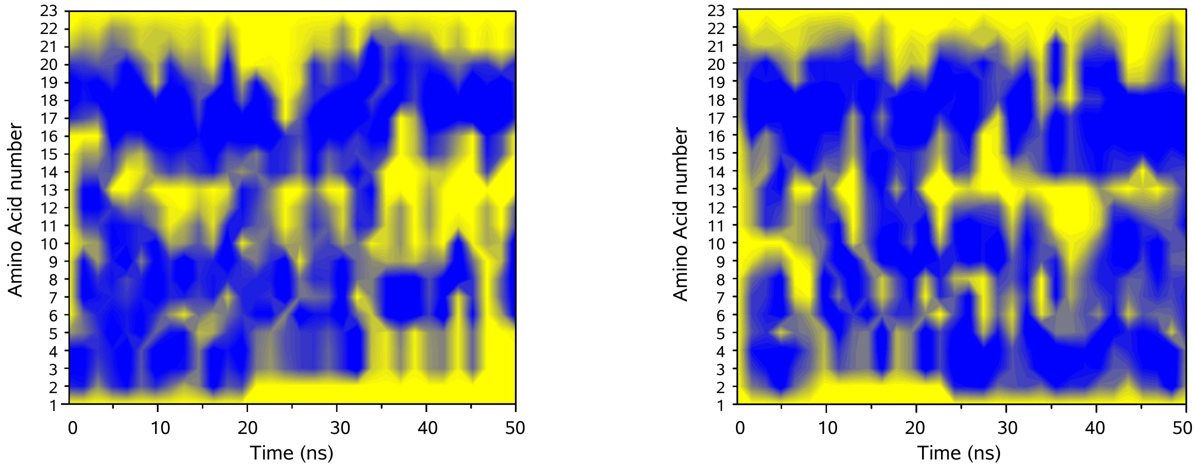
(e) 325 K

Figure 4.11: *Analysis of the molecular dynamics trajectories at 300 K (a) and (b) and at 325 K (c), (d), and (e) as a function of simulation time. The top panel of each graph shows the RMSD of the actual conformation to the native conformation (black) and for the helical fragments only (red: helix 1-11, blue: helix 15-21). The lower panel always shows the deviation of the sulfur-sulfur distance for a potential disulfide bridge (at 2 A distance) for the amino acids forming the first (green, CYS8-CYS18) and the second disulfide bridge (brown, CYS4-CYS22).*



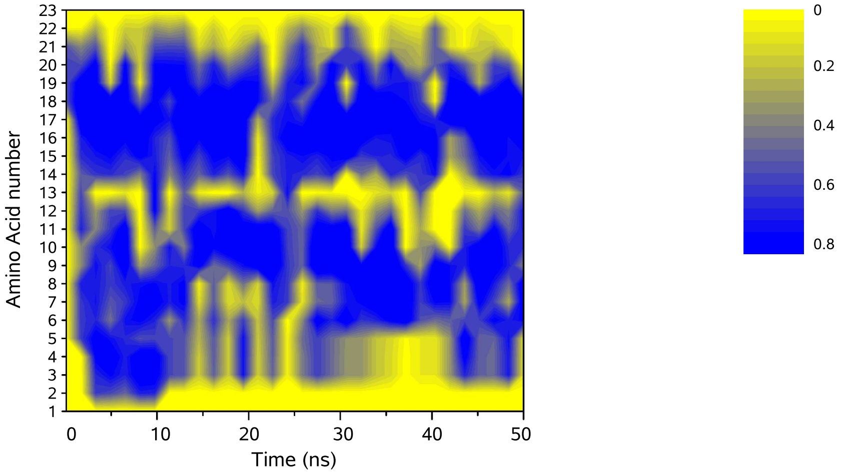
(a) 300 K

(b) 300 K



(c) 325 K

(d) 325 K



(e) 325 K

Figure 4.12: *Time average over a 100-ps moving window of the helix propensity of each amino acid in the molecular dynamics simulations at temperature 300 K (a) and (b) and 325 K (c), (d), and (e). In the native conformation the first helix spans amino acids 1-11, and the second helix spans amino acids 14-21, respectively. The scale for helix probability is the same in all panels.*

Next we analyze the sulphur-sulphur distance between CYS8-CYS18 and CYS4-CYS22 as a function of time (lower panels in Fig. 4.11). These distances also fluctuate strongly, averaging more then 10 Å during the simulations. On occasion, however, some of the sulphur atoms approach each other to within 3-4 Å, i.e. close enough for a disulphide bridge to form. On isolated instances, which occur in three of the five simulations (in one simulation two times independently), folding events occur in which both pairs of sulphur atoms approach one another, while both helices are preformed. In those occurrences (which last several ps), the simulations attain all-atom RMSDs to native of 3.43 Å, 3.80 Å and 3.47 Å, respectively. The intra-helix RMSD vary between 2.1-2.5 Å for helix 1 and between 0.8-1.0 Å for helix 2 in this time-frame.

4.4 Conclusion

In this chapter we briefly introduced the problem of protein folding. Motivated by a stochastic study of a protein with simple energy landscape, we performed MD simulations to explore its folding dynamics.

From this analysis emerges a picture of the folding process for 1WQE: the low-energy part of the folding funnel is characterized by fluctuating conformations in which both helices are preformed. Both helices fold and unfold repeatedly during the simulation. As the protein explores this landscape it occasionally visits conformations that can lead to the formation of the correct disulphide bridges that would stabilize the native conformation. We note that neither the MD simulations nor the free-folding simulations in PFF01 produced conformations that are consistent with a non-native disulphide bridge pairing. These events can occur in sequence on a timescale below 1ns, but happen even concurrently on a timescale of the order of 100 ns.

The free-energy model also predicts the existence of an exclusively helical low-energy ensemble, which collapses into the native conformation at the bottom of the free-energy funnel. Because the free-energy model contains no backbone entropy, the native conformation is found with high probability in the free-energy approach, even though it is not stable (without disulphide bridges) under physiological conditions.

These results are best put into perspective in the context of the framework [116, 117, 118] or diffusion-collision [119, 120] approach of protein folding, where secondary structure fragments of the protein assemble first, which then assume their final tertiary structure by docking into one another. The folding process may thus conceptually divided into two steps: the assembly of the secondary structure precursors and the final collapse into the tertiaray structure. For the proteins described here the latter requires the formation of disulfide bridges for enthalpic stabilization. We note that both simulations suggest the existence of preformed helical sections, for the specific systems in question the failure to connect the disulfide bridges in any other than the native topology from the low-energy conformations further supports the idea that the helices must be largely formed for the assembly of tertiary structure. However, in accordance with the funnel paradigm of protein folding 3,4, there is no single unique intermediate conformation that must be passed in the folding process to the native state. Instead our simulations suggest the existence of a wide ensemble of two-helix structures that precede the final collapse to the native conformation. In this final collapse there is a huge loss of configurational entropy that is apparently not compensated by the weak hydrophobic free-energy gain (increase in solvent entropy) afforded by this small system. Therefore an enthalpic contribution from the disulphide bridges is required to stabilize the native conformations.

In this picture the folding time is determined by the rate of helix formation and the rate of disulfide bridge formation. Many present molecular dynamics forcefields, in particular those with implicit solvent models [115] , may contain a bias towards a particular secondary structure. The combination of Amber99/GBSA that was used in the present study was reported to overemphasize helical secondary structure elements [121, 122, 123, 124]. This would influence the frequency with which the helical precursors for disulfide bridge formation are visited and further studies are required for quantitative results.

Folding of the protein thus proceeds by a diffusion-collision mechanism [120] where of pre-formed helices approach one another occasionally to form the disulphide bridges which ultimately stabilize the native conformation. Our results indicate that for these peptides secondary structure formation precedes hydrophobic collapse[125, 126], in contrast to most standard folding scenarios. This raises the intriguing question whether it is possible to substitute the cysteine residues to hydrophobic residues leading to hydrophobic collapse of the preformed helical ensemble into a well defined tertiary structure that requires no stabilization by disulfide bridges. Such design exercises may help to guide the design of stable hydrophobic cores for such small proteins which would have implications for important challenges in protein design, e.g. for zinc-finger design [127, 128, 129, 130, 131].

Chapter 5

Crystal structure formation

In nature, solid state material is often found in a crystalline structure, where atoms are located on the points of a periodic three dimensional lattice network. These can be defined by a minimal unit reproducing the whole periodic structure, the unit cell. The way in which the atoms are ordered in this solid state structure defines many of the bulk properties, such as the optical and electronic properties. Therefore, the ability to control this ordering would afford control over these properties.

The crystal units can be atoms, ions or molecules. Crystals made of atoms or ions arise from covalent or ionic bonding, and their lattice properties are determined by the atoms or ions forming the crystal. Molecular crystals in contrast, are made of organic or inorganic molecules and based on interactions between molecules instead of covalent bonding. Their properties depend on the nature of their molecular components and of the interaction or driving force that holds the periodic structure together. Molecules enable therefore the realization of new materials with completley novel properties.

The idea of designing materials with desired properties, and synthesizing molecular solid state structures with predefined properties based on an understanding and exploitation of intermolecular interactions emerged in 1989. Desiraju [132] defined crystal engineering as *'the understanding of intermolecular interactions in the context of crystal packing and the utilization of such understanding in the design of new solids with desired physical and chemical properties'*. Crystal engineering relies on noncovalent bonding and self-assembly to achieve the organization of molecules and ions in the solid state and it often involves an interaction between complementary hydrogen bonding faces. The two main strategies currently in use for crystal engineering are based on hydrogen bonding (as an example, we discuss in section 5.1, the case of adamantan based crystal) and coordination complexation (in section 5.2, we discuss DNA based crystals). Today, a community of 150 independent research groups is working in the filed of crystal engineering [133]. In a collaboration with experimental groups at the university of Karlsruhe, we have investigated the possibilities of crystal formation in two different systems.

We first analyze simple organic molecular building blocks, where the interaction between molecules is hydrogen bonding. In the second section of this chapter we study bigger complexes where the molecular recognition of DNA strands accounts for the stabilizing interaction. The molecular crystals envisaged here have a larger lattice constant than the length of covalent bonds in atomic crystals. They are therefore extremely porous and can be used to filter and sort by size nanoscaled particles.

A first investigation into crystal formation is presented in section 5.1. We focus on organic molecular building blocks that are small organic molecules and that interact with each other via hydrogen bonds. The advantage of such a system is that the formation process is reversible, i.e. defects that occur during the building process can be easily corrected.

In section 5.2 we study DNA based crystals. The interaction that accounts for the stability of such crystals is DNA duplex formation. A large thermodynamic impulse or force is needed during the crystallization process to compensate the loss of entropy during the binding of complementary DNA strands to each other. On the other hand, the selectivity of DNA binding would principally allow for the formation of chiral or anisotropic crystals.

5.1 Organic crystals

In this section, we study theoretically the feasibility of hydrogen bonded molecular crystals based on organic molecules. The molecules here have been synthesized by Tobias Grab, in the group of Prof. Bräse. Fig. 5.1 illlustrates the envisioned diamond structure consisting of tetrahedral units.

When designing a hydrogen bond based structure two practical problems must be taken into account. The probability of the formation of an hydrogen bond between a donor hydrogen and an acceptor varies with the involved groups and also depends on the specific environment. Moreover, due the possible spatial combinations of hydrogen bonding, several conformations, including some that yield an amorphous structure, are possible. We have therefore performed MD study calculations for the possible crystal structure candidates made of an adamantan core to form an organic crystal based on hydrogen bonding interactions.

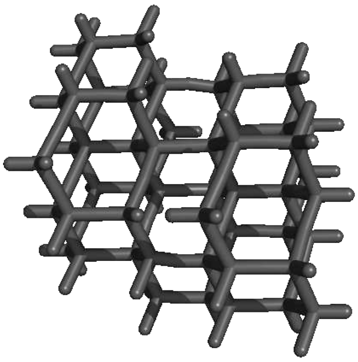


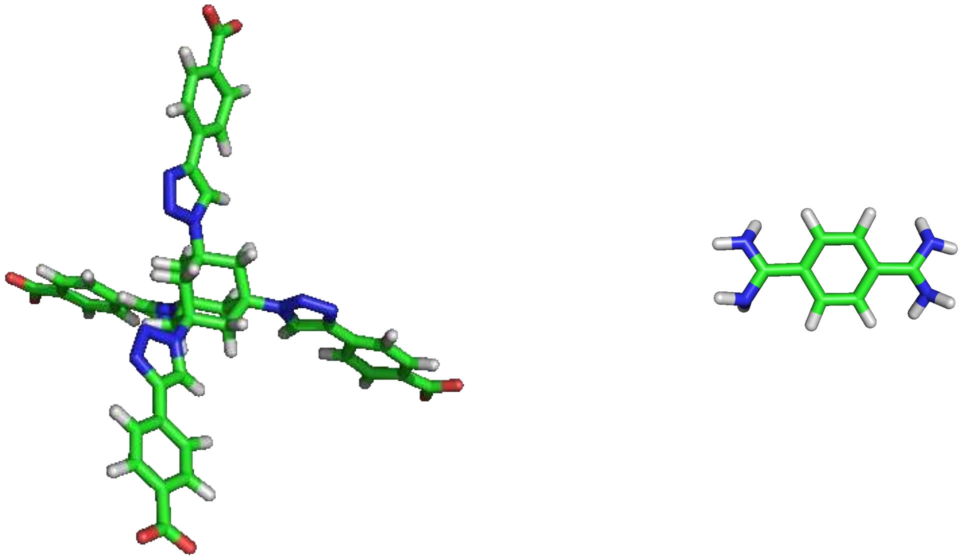
Figure 5.1: *Diamond structure: example of a crystalline structure resulting from tetrahedral units.*

5.1.1 Building blocks I

Fig. 5.2 illustrates the basic building blocks for an adamantan based crystal and its bonding geometry. The core molecule (a) has a tetrahedral geometry and is made of an adamantan core and of four 'arms' of aligned aromatic rings which end with a polar carboxyl group. Adamantan is a compound made of carbon atoms in sp3 hybridization. It is a cycloalkane, the simplest diamondoid and the most stable isomer of $C_{10}H_{16}$. The linker molecules (b) aim to link the core molecules. They are linear and almost planar and have two polar groups at both sides. The stability of such a crystal relies on hydrogen bonding between the polar groups of the molecular building blocks. In solution, the polar groups of the core are deprotonated and therefore negatively charged and the polar groups of the linkers are protonated and positively charged. Two possible geometries of the hydrogen bonding pattern are depicted in Fig. 5.2(c) and Fig. 5.2(d). (c) illustrates the 'ideal' geometry required to build a perfect crystal while (d) illustrates a physically possible competing geometry resulting in deviations from the expected crystal.

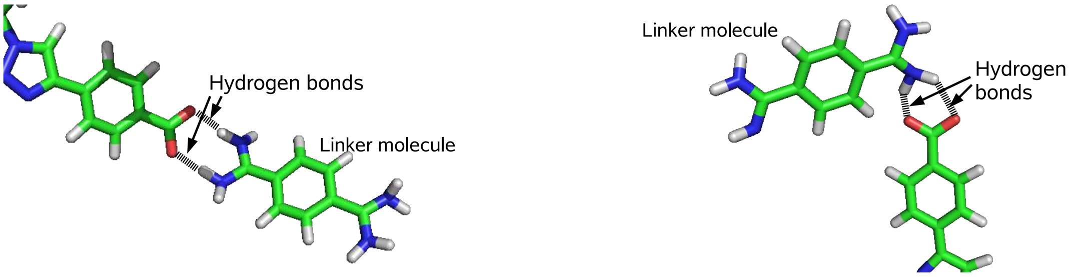
MD simulations

To study the stability and the feasibility of an adamantan core based crystal, we perform several MD simulations of the crystal building blocks introduced above starting from different initial conformations. Stability is a first condition for further studies and investigations or the spontaneous formation of a unit cell. First, we check the stability of the crystal in the MD simulation by starting from a perfect crystal conformation and observe the evolution of the hydrogen bonding pattern between complementary building blocks. Then, we investigate if the formation of such a crystal is probable at all by starting from



(a) Tetrahedral core molecule

(b) Linker molecule



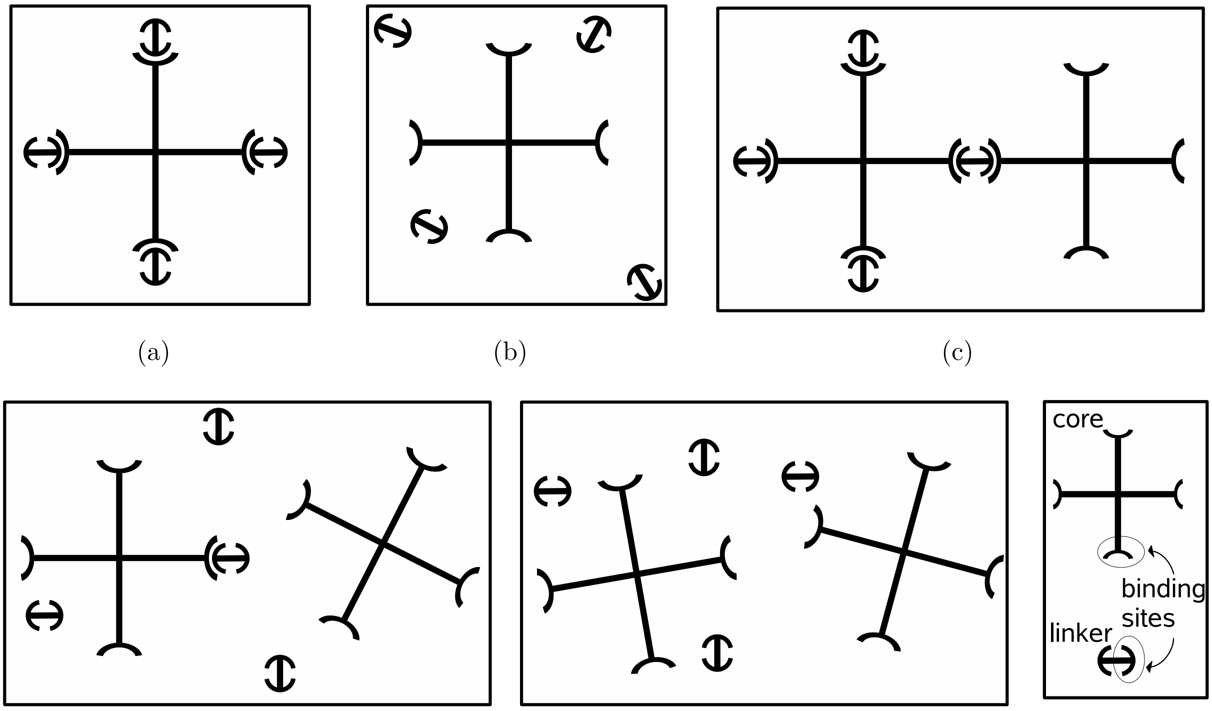
(c) Hydrogen bond, ideal configuration

(d) Hydrogen bond, possible geometry

Figure 5.2: *Building blocks in stick representation where carbon is green, oxygen is red, nitrogen is blue and hydrogen is white. The tetrahedral adamantan core molecule (a) couples to a neighbouring molecule via a linker molecule (b) and hydrogen bonds in a linear way (c) or under a different geometry (d).*

a randomly distributed building blocks conformation.

The simulations have been carried out for 10 ns with the AMBER simulation package, using the General Amber Force Field (GAFF) [26] in the (NPT) ensemble. Explicit water is simulated with the TIP3P model, periodic boundary conditions are applied, temperature is fixed at 300 K and pressure at one atmosphere using Langevin dynamics. The molecules investigated here do not belong to the standard set of already parameterized molecules, like proteins or DNA. We therefore have to parameterize the molecules in the GAFF force field with the Antechamber module in AMBER [134] with the protonation of the side groups according to the experimental conditions: the polar groups in the linker molecule are always protonated and the carboxyl groups in the core are deprotonated.



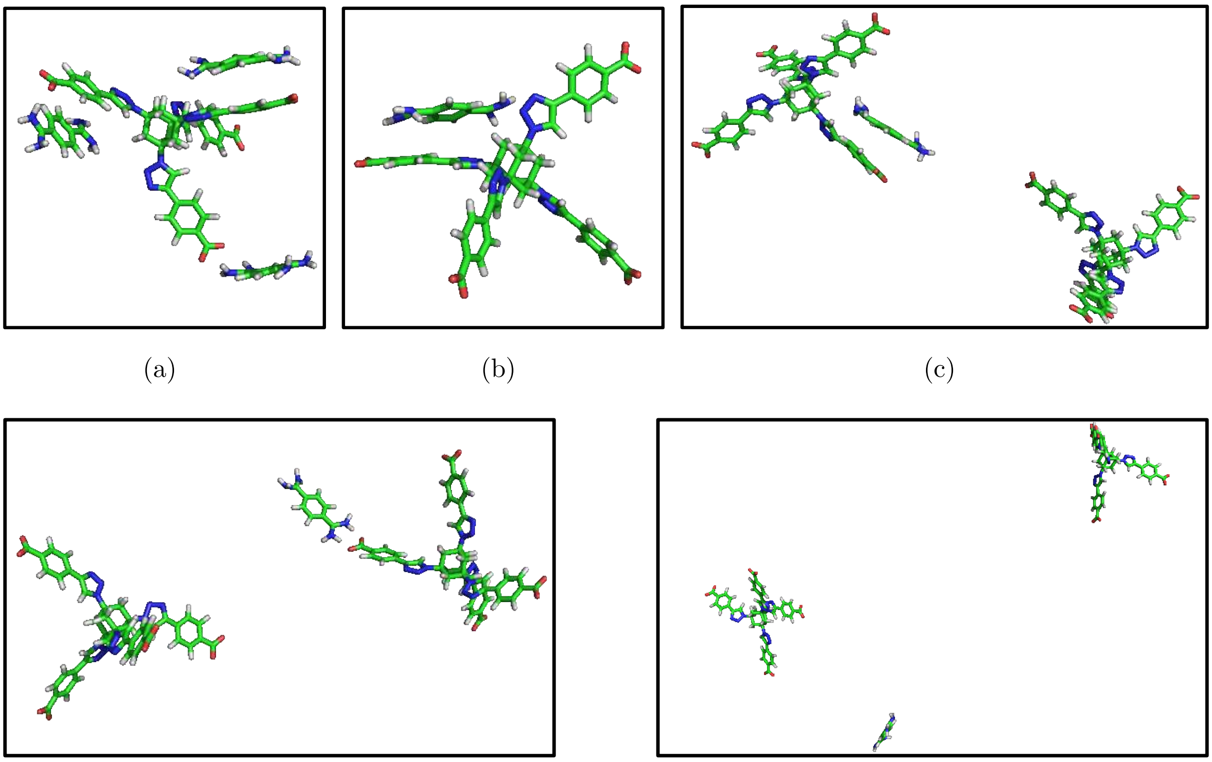
(a) (b) (c)

(d) (e)

Figure 5.3: *(a)-(e): Schematic representation of the initial conformations of core and linker molecules used as starting configurations for our MD simulations. (f):Legend*

During this parameterization, an atom type is assigned to every atom according to its chemical environment. The model used to assign the charges was the AM1-BCC model (AM1 [135] with bond charge correction), that generates atomic charge that resemble the RESP charges [136, 137] but are actually Mulliken charges. In this case, we assign the charges of the linker groups (COO^- and $N_2H_4^+$) by analogy to already parameterized functional groups (i.e. carboxyl group and guanidinium group) in solution, not in vacuum.

Our simulation strategy is illustrated in Fig. 5.3. We perform a set of simulations starting from different conformations:
(a) a single core molecule with the linker molecules at the ‘ideal’ position, to check for the stability of such a molecular entity,
(b) a single core molecule and randomly distributed linker molecules, to investigate if the correct conformation is spontaneously formed,
(c) two core molecules connected with a linker molecule, at the ‘ideal’ position, to check the stability of a bigger molecular entity belonging to the planned crystal,
(d) two separated core molecules and a linker molecule close to one of the core molecules,



(a) (b) (c)

(d) (e)

Figure 5.4: *Snapshots of the simulated systems during the simulations. (a), (b), (c), (d), and (e) correspond to the same labels as in the Fig. 5.3.*

to study if the correct conformation forms, and
(e) two core molecules and some linker molecules randomly distributed, to investigate if the correct conformation is spontaneously formed.

In Fig. 5.4, we illustrate intermediate states of the simulated systems. After 5 ns of simulation time, we observe that:
The system (a) is not completely stable. The π-stacking interactions between aromatic rings, i.e. the attractive interaction between perpendicular π orbitals to the aromatic systems, are energetically as favorable as the hydrogen bonding interactions. They compensate the molecular recognition between the carboxylate anions and the aminidium cations and yield a very different geometry than the expected to build the crystal.
The simulation of the system (b) does not show the complete spontaneous formation of a basic unit of the crystal, but one of the four randomly distributed linkers approaches the core molecule. However, even if the driving force for the first crystal formation stage is the electrostatic attraction between the cations and the anions, the linker stabilizes due

to the hydrophobe π-stacking interactions.
During the simulation of the system (c), the dynamics of the linker molecule is again driven by the hydrophobic interactions rather than by hydrogen bonding.
In the simulations of (d) and (e), aiming to investigate if the expected macromolecular structure is formed, longer simulation times are needed.

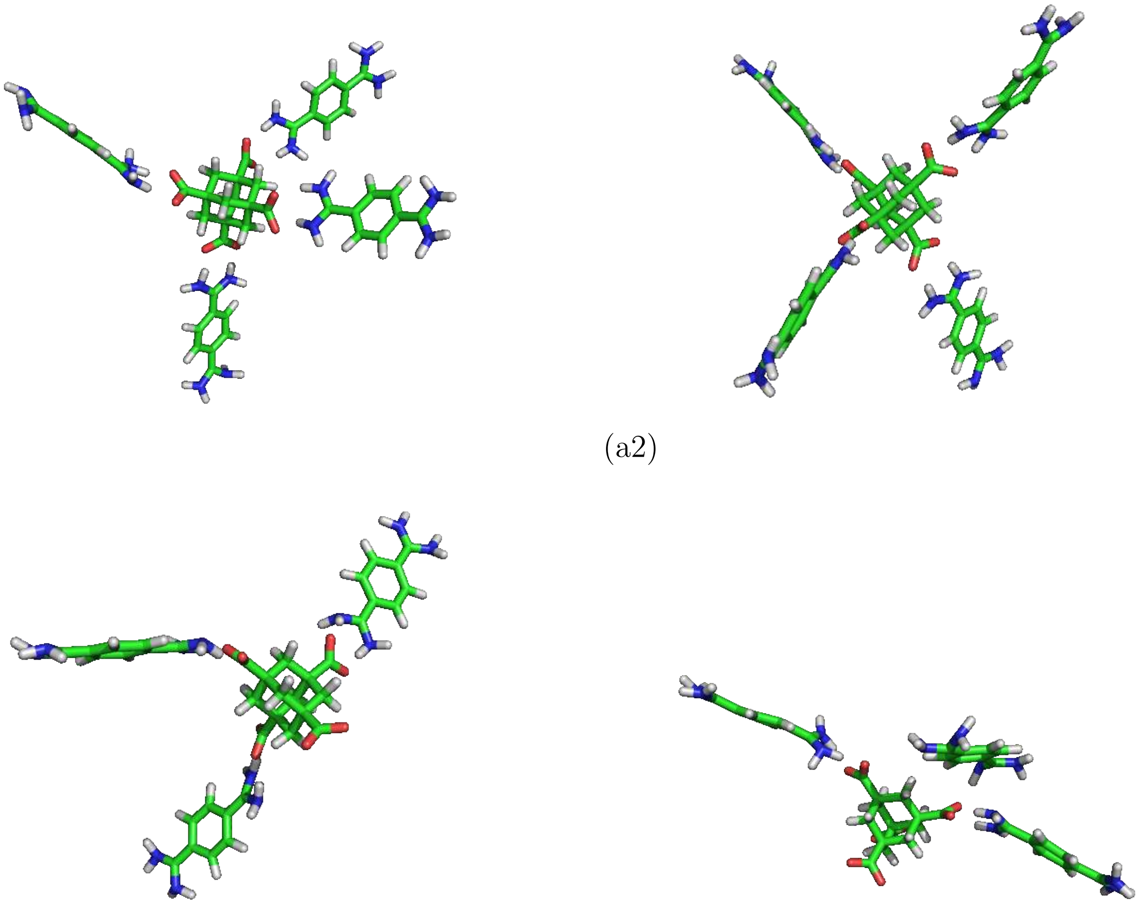
According to our MD simulations, the above studied building blocks are not well suited for crystal engineering. This is due to the competing π-stacking interactions between the aromatic groups in the core molecule and in the linkers. However, the charges used for all other atoms that are not the linking functional groups COO^- and $N_2H_4^+$, are derived from Mulliken charges, and these do not take into account the presence of the solvent. As a next step, a better parameterization of the partial charges might be necessary, to represent with more accuracy the charge distribution in presence of a solvent. In reality, the hydrophobic interactions leading to defects of this crystal in our simulations are weaker that hydrogen bonding. Following these simulations we suggested to consider a core molecule where the binding sites (the carboxylate anions) are directly attached to the adamantan core. We expect a more stable crystal structure, since no hydrophobic interactions can take place now.

5.1.2 Building blocks II

In this section we focus on a simpler system where the tetrahedral building blocks are only made of an adamantan center and the binding sites are directely attached to it (see Fig. 5.5). The linker molecules are not modified. We investigate the stability of macromolecular building cells of the crystal. According to Fig. 5.3, our starting conformations correspond to (a) and (c). Rather than trying to mimic the dynamics of the crystal formation, our simulations concentrate on the stability study of these structures in an aqueous environment at ambient conditions.

MD simulations

We follow the same simulation protocol as for the larger core molecule and we focus first on the conformation (a). We perform four independent MD simulations starting from an 'ideal' configuration where every linker molecule is close to a different carboxyl group of the core molecule. In Fig. 5.5 we illustrate the conformation of each independent run after 10 ns of simulation time. We observe that, in two of the simulations, (a1) and (a2),



(a1)

(a2)

(a3)

(a4)

Figure 5.5: *Snapshots of the configurations after 10 ns of simulation starting from conformation (a). Four independent simulations (a1)-(a4) have been carried out.*

the optimal geometry is preserved. In the other two, (a3) and (a4), three linker molecules are stabilized close to the core but one diffuses away.

If we consider the simulated trajectories of these systems, we observe that the conformations illustrated in Fig. 5.5 are not frozen or completely fixed in time, since the system undergoes constant changes and attains general intermediate states. However, these intermediate conformations are not stable and do not dominate, such that the systems returns to the stable configurations as in Fig. 5.5.

As a quantitative analysis of these simulation, in Fig. 5.6 we plot the evolution in time of the distances between the binding sites: between the functional groups COO^- in the core molecules and $N_2H_4^+$ in the linker molecules. The two increasing values correspond

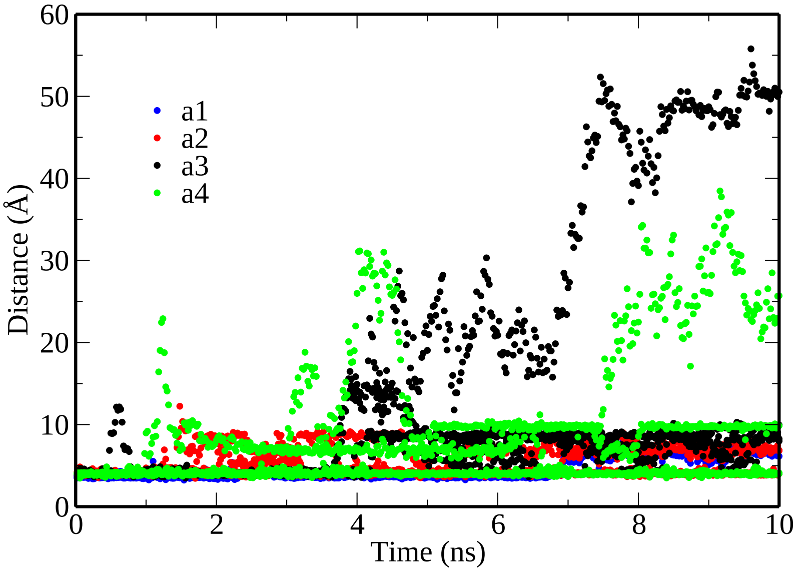
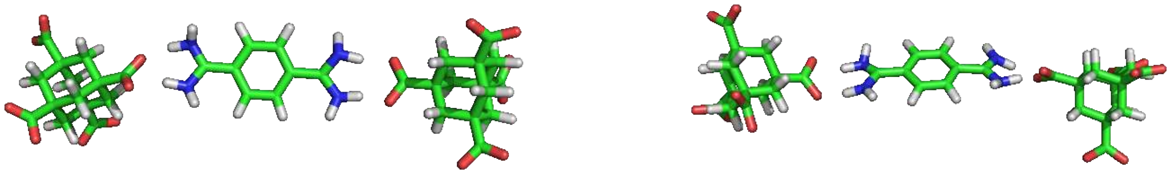


Figure 5.6: *Distances vs. time between the carbon atom of the COO^- group and the carbon atom to which the $N_2H_4^+$ are bound. Each simulation representation is shown in the legend, there are four distance values for every simulation, one for each binding site*

to the simulations a3 and a4, where one linker molecule diffuses away from the core. An interesting feature of this plot is that the values of the distance between the binding sites are stabilized at 4 Å and at 8-10 Å. Comparing the snapshots of the simulation at different instants, we observe that the linker molecules do not have a fixed orientation. They either rotate and form an hydrogen bond in the right configuration with the other side of the molecule since they are symmetric, or move from one binding site to the other. These conformational changes lead to the stable positions that yield the optimal geometry for building a molecular crystal for 14 of 16 free linkers.

We can therefore conclude for the reduced core molecule which is made only of adamantane and binding sites, that the conformation (a) of Fig. 5.3 is stable (almost 90%).

Now we turn to the conformation (c) of Fig. 5.3. We performed four independent MD simulations starting from an 'ideal' configuration where one linker molecule is placed between two core molecules. In Fig. 5.7 we illustrate two snapshots of the simulations corresponding to the initial conformation and the configuration after 10 ns of simulation time.



(a) initial conformation

(b) after 10 ns of simulation

Figure 5.7: *Snapshots of the initial conformation (a) and the configuration after 10 ns simulation (b) for the system shown in 5.3(c).*

To check the stability of such a system, we study the behavior of the hydrogen bonds accounting for the stability of the structure. In Fig. 5.8 we plot the evolution of the distance between the binding sites vs. time during the simulation. We observe that during the 10 ns of simulation, the complex core-linker-core remains stable in a conformation that yield the proper crystalline structure.

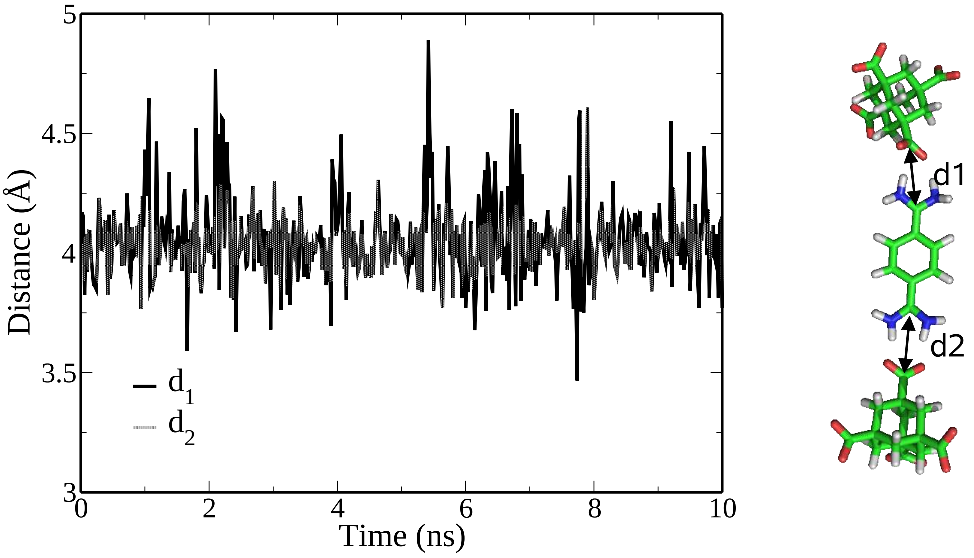


Figure 5.8: *Distances between binding sites, indicated in the left panel, vs. time.*

5.2 DNA crystals

In this section, we focus on molecular crystals composed of an organic core and DNA single strands. DNA (Deoxyribonucleic Acid) is the molecular repository of genetic information but beyond its biological function, this polymer has many properties (e.g. selfassembly, availability) making it interesting for further technical applications. DNA is made of four different nucleotide units, which bind selectively to each other: adenine (A) binds specifically to thymine (T) and guanine (G) to cytosine (C).

The selfassembly properties of DNA between complementary strands constitute the binding interactions in the crystals studied in the present section. The DNA strands are covalently bonded to the core molecules and interact with each other. Several core molecules, from methane or adamantan based molecules to fullerenes have been proposed as candidates for building molecular crystals. These molecules are synthesized in the organic chemistry department at the university of Karlsruhe by the collaborators of Prof. Richert.

In the present section, we investigate if the formation of a DNA based crystal is possible combining brownian dynamics and MD technique. According to brownian dynamics studies, the formation of a crystalline structure starting from identical tetrahedral building blocks depends on the stiffness of these molecules and their interactions. In section 5.2.1 we summarize the method of brownian dynamics and present results on crystal formation obtained by Konstantin Klenin that motivate our MD simulation. We investigate the conformational properties of a molecular complex made of a core molecule to which four identical DNA single strands are bonded by performing MD simulations in section 5.2.2. In section 5.2.3 we discuss the results on core molecules that have been synthesized and investigate their spatial compatibility with DNA duplex formation.

5.2.1 Brownian Dynamics results

Brownian dynamics describes the movement of solute interacting particles with an implicit model for the solvent. For aqueous solution, the main effect on the solute is the viscosity, i.e. a frictional force on a diffusing solute. The collisions of the particles with water molecules add a random component to a solute's motion. The technique of brownian dynamics allows for realistic simulations of the diffusion of molecules in solution without needing to include any explicit solvent molecules.

The computational implementation follows the equations 5.1 and 5.2. For a sphere of radius a in a solvent of viscosity η and at temperature T, having a translational diffusion coefficient $D_{tr} = \frac{kT}{6\pi\eta a}$ and a rotational diffusion coefficient $D_{rot} = \frac{kT}{8\pi\eta a}$, the translational displacement $\Delta\vec{r}$ of the particles in time Δt is given by:

$$\Delta\vec{r} \;=\; \frac{D_{tr}}{kT}\,\vec{F}\,\Delta t \;+\; \vec{R}, \tag{5.1}$$

and the rotational displacement $\Delta\vec{\omega}$ by:

$$\Delta\vec{\omega} \;=\; \frac{D_{rot}}{kT}\,\vec{T}\,\Delta t \;+\; \vec{\phi}. \tag{5.2}$$

$\vec{R}$ and $\vec{\phi}$ are random translational and rotational displacements added to mimic the effects of collisions with the solvent molecules. They fullfill the conditions $\langle\vec{R}\rangle = 0$ and $\langle\vec{\phi}\rangle = 0$. $\vec{F}$ and $\vec{T}$ are the force and the torque acting on the particle.

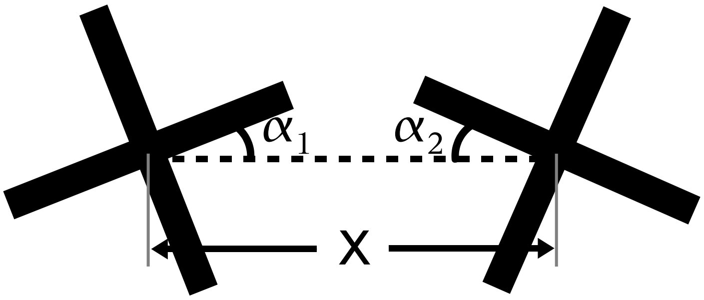


Figure 5.9: *Schematic 2-D representation of two tetrahedral building blocks. The interaction potential U(x,α_1,α_2) between the building blocks is defined in eq. 5.3 as a function of the distance x between the centers of the building blocks, and a function of the relative angles α_1 and α_2 as indicated in the figure.*

The building blocks of the crystal are considered as spheres for solving eqs. 5.1 and 5.2. The radius a was chosen to be 3 nm according to the characteristic size of the molecules under consideration. The brownian particles interact via the potential:

$$U(x,\alpha_1,\alpha_2) = \begin{cases} 0\;, & if \quad |x-A| > B,\ or\ \alpha_1 > \alpha_0,\ or\ \alpha_2 > \alpha_0 \\ -\frac{U_0}{B\,\alpha_0^2}\cdot|B-|x-A||\cdot|\alpha_0-\alpha_1|\cdot|\alpha_0-\alpha_2|, & else, \end{cases} \tag{5.3}$$

where A, B, α_0 and U_0 are constants. We used $A = 4nm$, $B = 2nm$, $\alpha_0 = 30^{\circ}$ and U_0= 10 kT. The above potential is only a phenomenological approximation to the real interaction between the DNA single strands attached to the core. It is most attractive for a distance $x = A$. It is zero for x greater and lesser than $A \pm B$ if at the same time α_1 and α_2 are below a threshold value α_0. $U(x,\alpha_1,\alpha_2)$ is related to the force $\vec{F}$ and the torque $\vec{T}$ included in the translational and rotational displacements (eq. 5.1 and eq. 5.2) as:

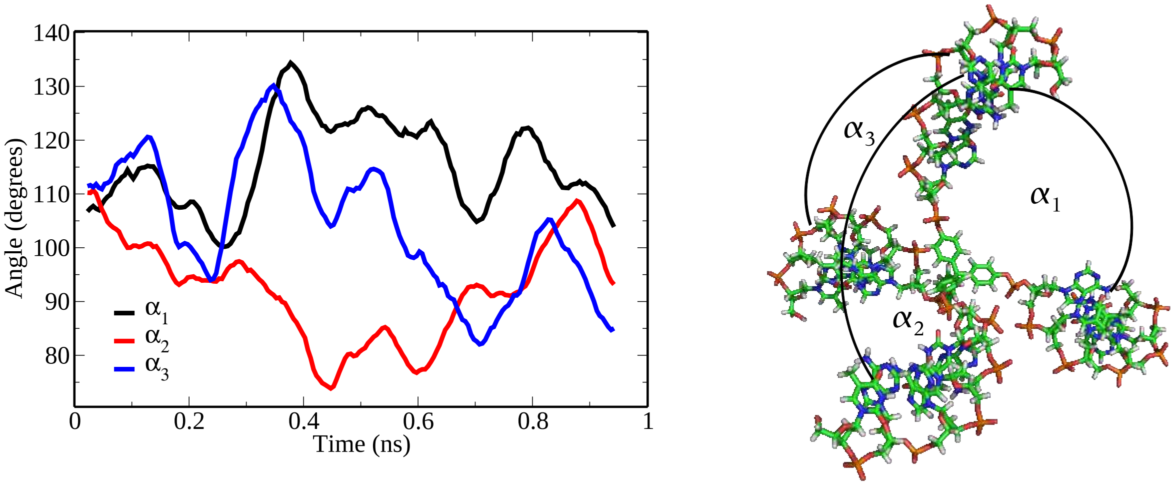
$$\vec{F} = -\frac{\partial U(\vec{r},\vec{\omega})}{\partial \vec{r}} \quad \text{and} \quad \vec{T} = -\frac{\partial U(\vec{r},\vec{\omega})}{\partial \vec{\omega}}. \tag{5.4}$$

The differential equations 5.1 and 5.2 are solved for every time step Δt and, as no interatomic interactions are modeled, the time steps can be much longer than the steps in MD simulations. The output of a brownian dynamics study is also a trajectory of the solute particles over a longer time scale than in MD. In this study, an annealing process with time step Δt=1 ps starting from the gas phase in a closed volume is performed. Building blocks with four strands and with six strands were simulated. In the case of tetrahedral building blocks, two cases were considered: (i) all the particles interact with each other and (ii) two types of particles selectively interact only with the ones that belong to the own group type.

According to the results of the brownian dynamics simulations, tetrahedral building blocks are able to form a crystal with hexagonal diamond or ice structure if two types of particles that only bind among themselves are considered. If only one type of building blocks is used, which would be preferable from an experimental point of view, the tetrahedral building blocks form a pentagonal structure, that cannot form an extended crystalline structure. Only if these basic units or building blocks are flexible, the angle difference between pentagonal and hexagonal structures can be compensated and crystals can be built. To investigate the flexibility of the crystal building blocks properly it is necessary to consider the molecular structure of the particles themselves. We therefore simulate them at the atomic level using MD.

5.2.2 MD simulations

We perform a MD study of a tetrahedral building block to investigate the stiffness of the crystal building blocks. The core molecule linking the four DNA strands consists of a methane core in which the hydrogen atoms are substituted by phenyl rings and to which the DNA strands are attached. Single DNA strands made of six nucleotides are covalently bonded to the phenyl rings. In Fig. 5.10(b) we illustrate the basic building block of these DNA based crystals. The obtained values for the distance from the center of the core to the last nucleotide, approximatively 3 nm, and the length of the DNA strands, approximatively 2 nm, are consistent with the input values in the brownian dynamics simulations. We simulate this basic unit solvated in water and counter ions to



(a) (b)

Figure 5.10: *(a) Time evolution of the angles indicated in (b). (b) Illustration of the tetrahedral building block.*

neutralize the whole system in the (NPT) ensemble. The core molecule is parameterized with the General Amber force field [26] and the DNA with the AMBER99 force field. The temperature is 300 K and the pressure corresponds to the experimental conditions, 1 atm.

In Fig. 5.10(a), we plot the time evolution of the angles α_1, α_2, and α_3 as indicated in the illustration of Fig. 5.10(b). We observe that the angles between DNA single strands take values within 110 ± 30 degrees.

Our MD simulations show that the formation of a crystal of tetrahedral building blocks made of a core molecule based on methane and one aromatic ring per strand and selfcomplementary DNA is possible. The variations of the angles between the DNA strands are very large and would therefeore allow formation of mixed crystals with pentagonal unit cells, even if only one type of DNA is used. However, if the core molecule is too small, steric clashes will prevent the DNA double strands to form.

We next present two core molecules that are possible candidates to link the DNA single strands. Experimentally, it is more convenient to have small core molecules, since less synthesization steps are needed. Thus, an compromise between experimentally favorable and theoretically feasibl molecule size has to be made. We evaluate the smallest core molecules proposed by the experimentalists and check if they are large enough to allow for structure formation. In Fig. 5.11 we illustrate the atomic structure of two suggested core molecules. The core molecule illustrated in (a) is based on a methane molecule where

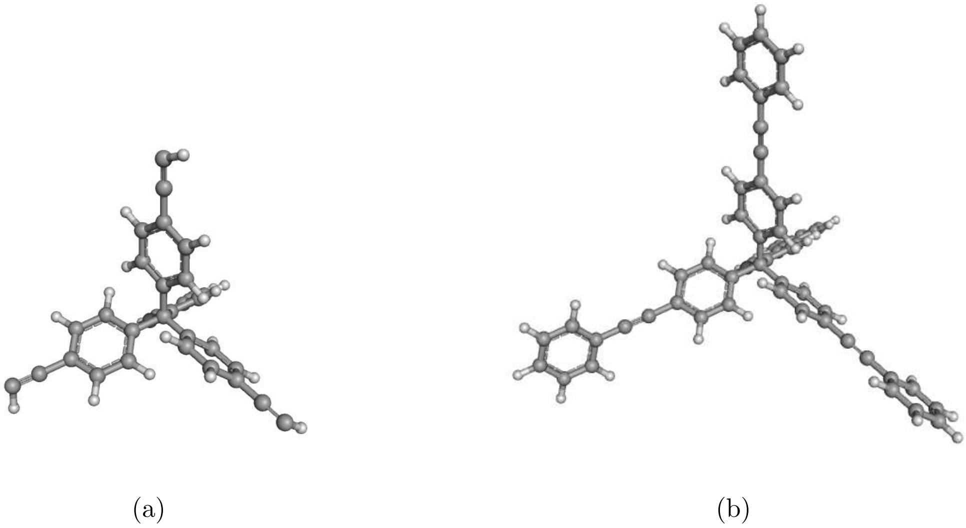


Figure 5.11: *Proposed molecules to link four DNA strands and build thereby a tetrahedral building block.*

the hydrogen atoms are substituted by a phenyl ring and a further carbon bonded throug a triple bond. The core molecule (b) has a further phenyl ring.

We perform an MD simulation starting from a conformation where a double DNA helix is bonded to the core molecule. We first minimize the system and raise the temperature gradually to 300K, and during 1 ns, we analyze the stability of such a complex. Both systems are stable and sterically possible. The molecule illustrated in Fig. 5.11(a) is compatible with four preformed DNA strands but it is then very stiff, it might be then used as core if two types of non-complementary DNA are used. The molecule illustrated in Fig. 5.11(b) fullfils completely the conditions for molecular crystals to be build as it allows for a little more spatial freedom.

5.3 Conclusion

In this chapter, we have introduced the concept of crystal engineering. We have first analyzed the feasability of simple molecular crystals based on an adamantan core to be built exploiting hydrogen bonding as the interaction responsible for the stability of the crystal. Second, we have analyzed more complex molecular crystals based on DNA, where the interaction for crystal stability is DNA duplex formation.

According to our MD simulations, in the case of organic molecular crystals with adamantan cores, hydrophobic interactions between aromatic groups constitute a competing force to the hydrogen bonding and therefore, the intended simple organic crystals are not stable. To avoid hydrophobe interactions, the molecules were further simplified by avoiding aromatic groups. The basic unit cell of a molecular crystal with the simpler molecule turns out to be stable in our MD simulations and the growth of such a crystal is theoretically possible. However, a more reliable parameterization might be necessary to reproduce with more accuracy the effects of the solvent on the charge distribution.

In our simulations on molecular crystals based on DNA selfassembly, such systems are theoretically possible. A brownian dynamics study predicts the formation of an amourphous solid without periodicity and thus no crystalline structure if the building blocks are very stiff. However, the study envisages the correct formation of a crystalline structure if a little spatial freedom is availiable. Since brownian dynamics simulations do not allow for a study at the atomistic level, we perform MD simulations on a single building block including the core molecule and the DNA. Our atomistic simulations complementing the brownian dynamics simulations show that the single building blocks of the envisioned crystals are flexible enough and would allow for the formation and growth of such molecular crystals.

Chapter 6

Summary and Outlook

The unifying element of the investigations presented in this thesis is the prediction of nanoscale structure formation and function using molecular dynamics simulations. As pointed out in the introduction, molecular dynamics simulations have become an increasingly important tool which is used in many diverse scientific fields, such as chemistry, biology, physics or material science to study phenomena at the mesoscale in the nanometer regime. By means of molecular dynamics simulations in combination with complementary simulation methods we have investigated the problem of structure formation in three systems of interest. The studied systems accounting for the different scientific fields of material science, biology and macromolecular chemistry are carbon nanotubes, proteins and molecular crystas respectively. In the following we summarize the main achievements for these systems independently and address open questions, challenges and perspectives for further interesting systems to be studied in an outlook.

As an example of material science or technical physics application, we have focused on carbon nanotubes and their targeted systematic separation according to their density by means of a centrifugation process. Our interest in carbon nanotubes separation processes arises from the scarcity of techniques to provide mass production of nanotubes with tailored properties. To date carbon nanotubes can be separated only by centrifugation or by the dielectrophoresis technique, which separates nanotubes by electronic properties. A detailed understanding of the centrifugation process for nanotubes is crucial to quantitatively describe such systems and to predict conditions in which specific nanotubes can be separated. We have proposed a theoretical model for the density distribution of carbon nanotubes according to their diameter that partially agrees with the ongoing experiments. Our calculations qualitatively reproduce the presence of a maximum in the density profile vs. the nanotube diameter, but still fail to give quantitative agreement with the experiments. Within our simulations we were able to confirm previous experimental and theoretical studies that predicted the presence of water inside the nanotubes. Furthermore, we have found that there is a strong influence of the surfactant molecules that plays an important role when sorting the nanotubes by density. We have analyzed the behavior of the surfactant molecules around the carbon nanotubes and derived approximate values for the calculation of the density of the complexes formed by nanotube-surfactant-water. However, the description of the real experimental environment is still not complete, since these complexes are not embedded simply in water but in a much denser solution that might have an effect in the atomic structure of the whole complex. Determining with more accuracy the real environment of the simulated systems, i.e. considering the specific solvent used in the experiments, is still an open challenge. It might constitute the next step towards designing experiments to sort a precise range of carbon nanotubes and enabling thereby the mass production of single types of carbon nanotubes.

As a second example, we have addressed the problem of protein folding, an important outstanding question in molecular biology. We have proposed a mechanism for the folding process of a small 2-helix peptide with disulfide bridges, which is of interest for better understanding of misfolding diseases and of protein function. The process of protein folding for the so-called fast folders, which have no metastable intermediates in the folding path, is not directly observable with experiments. The elucidation of the detailed behavior of a protein during its folding process is of general interest from a scientific point of view. However, to date very few protein folding processes could be simulated so far with atomistic resolution and kinetic methods. This is due to the amount of computer power that is necessary to reproduce the dynamical process of protein folding. Consequently, only proteins that fold very fast, i.e. without many metastable states in their free energy surfacs, are relevant candidates for computational studies. We have first studied the energy surface of a small protein with stochastic methods. After obtaining indications suggesting a very simple energy landscape, we have performed a dynamic study on this protein. Our main findings are that the secondary structure is initially formed and precedes the tertiary structure formation, which is stabilized by covalent disulfide bridges. The combined use of stochastic methods and molecular dynamics simulations has shown to be a promising tool to identify adequate proteins for studying the process of protein folding.

Motivated by ongoing experimental investigations in supramolecular chemistry, we have studied molecular crystal formation as a third example of structure formation at the nanoscale. Many interesting questions regarding nanoscale structure formation and growth have emerged with the synthesization of more and more complex structures in the recent years. Molecular crystals are increasingly studied due to their potential applications as filters for nanoscale particles with predefined lattice constants, for example. Since the synthesis of such complex crystalline structures is very difficult and elaborate, theoretical feasibility studies guiding the experiments can help to effectively select the most promising building blocks and protocols. In the case of simple molecular crystals based on adamantan and hydrogen bonding as stabilizing interaction, we have found that competing interactions between aromatic systems can prevent formation of the crystalline structure. In contrast, a more simplified structure without aromatic rings make crystal formation much more likely. The advantage of exploiting hydrogen bond formation is the simplicity and reversibility of the bonding process which may help to reduce the frequence of the occurrence of impurities or defects. A more complicated structure with a larger number of degrees of freedom can be achieved by using DNA complementarity as binding interaction. The use of DNA for crystal formation is of interest because DNA has self-assembly properties unmatched by many other materials, is available and stable. Moreover, DNA allows for crystal formation with predefined and controlable lattice constants without having the need to perform many synthesis steps. Again, we have combined two simulation methods: a brownian dynamics study that predicts the possibility of growing a crystal if the building blocks are not too rigid, and a molecular dynamics study that provides the brownian dynamics simulations with effective parameters derived on the basis of detailed atomic information on the building blocks. We have been able to predict the minimal necessary molecular core molecules needed to enable crystal growth and to guide the experiments towards the development of building blocks and protocols which make the growth of these complex crystals possible.

The systems and structures studied in this thesis constitute only three examples out of a large number of possible and interesting questions to be addressed with molecular dynamics method. The results of our investigations suggest that these computational methods will gain further importance in the forthcoming years and provide increasingly quantitative complementary tools for experimental research.

Bibliography

[1] European Computational Science: the Lincei Initiative "from computer to scientific excellence", Technical report, European Science Foundation, 2007.

[2] P. Westmoreland, P. Kollman, A. Chaka, P. Cummings, K. Monkuma, M. Neurock, E. Stechel, and P. Vaushishta, Applications of Molecular and Materials Modelling, Technical report, International Technology Research Institute, 2002.

[3] C. K. et al., Advanced Silicon Device Technologies, Fujitsu Scientific & Technological Journal **39**, 106 (2003).

[4] A. Patterson, A Direct Method for the Determination of the Components of Interatomic Distances in Crystals, Zeitschrift für Kristallographie **90**, 517–542 (1935).

[5] A. Ducruix and R. Giegé, *Crystallization of Nucleic Acids and Proteins: A Practical Approach.*, Oxford: Oxford University Press, 1999.

[6] K. Wuthrich, Protein structure determination in solution by NMR spectroscopy., J Biol Chem. **265**, 22059 (1990).

[7] G. Liu, Y. Shen, H. Atreya, D. Parish, Y. Shao, D. Sukumaran, R. Xiao, A. Yee, A. Lemak, A. Bhattacharya, T. Acton, C. Arrowsmith, G. Montelione, and T. Szyperski, NMR data collection and analysis protocol for high-throughput protein structure determination., Proc Natl Acad Sci U S A. **102**, 10487–10492 (2005).

[8] H. W. C. Postma, T. Teepen, Z. Yao, M. Grifoni, and C. Dekker, Carbon Nanotube Single-Electron Transistors at Room Temperature, Science **293**(5527), 76–79 (2001).

[9] R. Singh, D. Pantarotto, D. McCarthy, O. Chaloin, J. Hoebeke, C. Partidos, J.-P. Briand, M. Prato, A. Bianco, and K. Kostarelos, Binding and Condensation of Plasmid DNA onto Functionalized Carbon Nanotubes: Toward the Construction of

Nanotube-Based Gene Delivery Vectors, Journal of the American Chemical Society **127**(12), 4388–4396 (2005).

[10] C. J. Gannon, P. Cherukuri, B. I. Yakobson, L. Cognet, J. S. Kanzius, C. Kittrell, R. B. Weisman, M. Pasquali, H. K. Schmidt, R. E. Smalley, and S. A. Curley, Carbon nanotube-enhanced thermal destruction of cancer cells in a noninvasive radiofrequency field, Cancer (2007).

[11] T. Simmons, D. Hashim, R. Vajtai, and P. Ajayan, Large Area-Aligned Arrays from Direct Deposition of Single-Wall Carbon Nanotube Inks, Journal of the American Chemical Society **129**(33), 10088–10089 (2007).

[12] R. Krupke, F. Hennrich, H. von Löneysen, and M. M. Kappes, Separation of Metallic from Semiconducting Single-Walled Carbon Nanotubes, Science **301**, 344–347 (2003).

[13] J. N. Onuchic and P. G. Wolynes, Theory of protein folding, Current Opininon in Structural Biology **14**, 70–75 (2004).

[14] M. Mickler and T. Hugel, Molekulare Motoren künstliche Nanomaschinen, Physik in unserer Zeit **39**, 14–20 (2008).

[15] T. Schlick, *Molecular Modeling and Simulation*, Springer New York, 2002.

[16] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. G. III, and W. M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, J. Am. Soc. **114**, 10024–10035 (1992).

[17] C. J. Casewit, K. S. Colwell, and A. K. Rappé, Application of a Universal Force Field to Organic Molecules, J. Am. Soc. **114**, 10035–10046 (1992).

[18] C. J. Casewit, K. S. Colwell, and A. K. Rappé, Application of a Universal Force Field to Main Group Compounds, J. Am. Soc. **114**, 10046–10053 (1992).

[19] C. J. Casewit, K. S. Colwell, and A. K. Rappé, Application of a Universal Force Field to Metal Complexes, Inorg. Chem. **32**, 3438–3450 (1993).

[20] A. D. MacKerell, J. Wiorkiewicz-Kuczera, and M. Karplus, An all-atom empirical energy function for the simulation of nucleic acids, Journal of the American Chemical Society **117**(48), 11946–11975 (1995).

[21] A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, Journal of Physical Chemistry B **102**(18), 3586–3616 (1998).

[22] D. A. Pearlsman, D. A. Case, J. W. Calwell, W. S. Ross, T. E. III, S. DeBolt, D. Ferguson, and G. S. adn Peter Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, Comp. Phys. Comm. **91**, 1–41 (1995).

[23] D. A. Case, T. E. CheathamIII, T. Darden, H. Gohlke, R. Luo, K. M. M. Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, The Amber biomolecular simulation programs, J. Comp. Chem. **26**, 1668–1688 (2005).

[24] A. Perez, I. Marchan, D. Svozil, J. Sponer, I. Cheatham, Thomas E., C. A. Laughton, and M. Orozco, Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alpha/gamma Conformers, Biophys. J. **92**(11), 3817–3829 (2007).

[25] D. Case, T. Darden, T. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, K. Merz, B. Wang, D. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. G. J. Mongan, V. Hornak, G. C. P. Beroza, C. Schafmeister, J. Caldwell, W. Ross, , and P. Kollman, AMBER 8, Technical report, University of California, San Francisco, 2004.

[26] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Development and Testing of a General Amber Force Field, Journal of Computational Chemistry **25**, 1157–1174 (2004).

[27] W. L. Jorgensen and J. Tirado-Rives., The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin., Journal of the American Chemical Society **110**, 1657–1666 (1988).

[28] W. Jorgensen, D. Maxwell, and J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Or-

ganic Liquids, Journal of the American Chemical Society **118**(45), 11225–11236 (1996).

[29] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 manual and user guide*, Hochschuleverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.

[30] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, A Biomolecular Force Field Based on the Free Enthalpy of Hydra tion and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6, J. Comp. Chem. **25**, 1656–1676 (2004).

[31] H. Berendsen, D. van der Spoel, and R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation., Comp. Phys. Comm. **91**, 43–56 (1995).

[32] L. E. H. B. van der Spoel D., GROMACS 3.0: A package for molecular simulation and trajectory analysis., J. Mol. Mod. **7**, 306–317 (2001).

[33] M. Levitt, M. Hirshberg, R. Sharon, and V. Daggett, Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution., Computer Physics Communications **91**, 215–231 (1995).

[34] P. M. Morse, Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels, Phys. Rev. **34**(1), 57–64 (Jul 1929).

[35] B. Ma, J.-H. Lii, H. Schaefer, and N. Allinger, Systematic Comparison of Experimental, Quantum Mechanical, and Molecular Mechanical Bond Lengths for Organic Molecules, Journal of Physical Chemistry **100**(21), 8763–8769 (1996).

[36] L. Goodman, V. Pophristic, and F. Weinhold, Origin of Methyl Internal Rotation Barriers, Accounts of Chemical Research **32**(12), 983–993 (1999).

[37] F. Weinhold, A new twist on molecular shape, Nature **411**, 539.541 (2001).

[38] V. Pophristic and L. Goodman, Hyperconjugation not steric repulsion leads to the stagered sturcture of ethane, Nature **411**, 565–568 (2001).

[39] F. London, Zur Theorie und Systematik der Molekularkräfte, Zeitschrift für Physik **63**, 245–279 (1930).

[40] A. R. Leach, *Molecular Modelling: Principles and applications*, Pearson Education EMA, 2001.

[41] M. Rigby, E. B. Smith, W. A. Wakeham, and G. C. Maitland, *Intermolecular Forces: Their Origin and Determination.*, Oxford, Clarendon Press, 1981.

[42] M. Rigby, E. B. Smith, W. A. Wakeham, and G. C. Maitland, *The forces between molecules*, Oxford, Clarendon Press, 1986.

[43] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, J. Profeta, Salvatore, and P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, Journal of the American Chemical Society **106**, 765–784 (1984).

[44] S. R. Cox and D. E. Williams, Representation of the Molecular Electrostatic Potential by a New Atomic Charge Model., Journal of Computational Chemistry **2**, 304–323 (1981).

[45] P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale., Ann. Phys. **64**, 253–287 (1921).

[46] T. Darden, L. Perera, L. Li, and L. Pedersen, New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations., Structure **7**, 55–60 (1999).

[47] Z.-X. Wang, W. Zhang, C. Wu, H. Lei, P. Cieplak, and Y. Duan, Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides, Journal of Computational Chemistry **27**, 781–790 (2006).

[48] J. A.D. Mackerell, Empirical Force Fields for Biological Macromolecules: Overview and Issues, J. Comput. Chem. **25**, 1584–1604 (2004).

[49] B. Roux and T. Simonson, Implicit solvent models, Biophysical Chemistry **78**, 1–20 (1999).

[50] F. Fogolari, A. Brigo, and H. Molinari, The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology, Journal of Molecular Recognition **15**, 377–392 (2002).

[51] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics., Journal of American Chemical Society **112**, 6127–6129 (1990).

[52] G. D. Hawkins, C. J. Cramer, and D. G. Thrular, Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium., J. Phys. Chem. **100**, 19824–19839 (1996).

[53] M. Feig, A. Onufriev, M. Lee, W. Im, D. A. Case, and I. C. L. Brooks, Performance Comparison of the Generalized Born and Poisson Methods in the Calculation of the Electrostatic Solvation Energies for Protein Structures., J. Comput. Chem. **25**, 265–284 (2004).

[54] W. G. H.J.C.Berendsen, J.P.M.Postma and J. Hermans, *Interaction models for water in relation to protein hydration*, D. Reidel Publishing Company, 1981.

[55] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, Comparison of simple potential functions for simulating liquid water, The Journal of Chemical Physics **79**(2), 926–935 (1983).

[56] M. W. Mahoney and W. L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, The Journal of Chemical Physics **112**(20), 8910–8922 (2000).

[57] B. Guillot, A reappraisal of what we have learnt during three decades of computer simulations on water., Journal of Molecular Liquids **101**, 219 (2002).

[58] H. J. C. Berendsen and W. F. van Gunsteren, Practical algorithms for dynamics simulations, Technical report.

[59] L. Verlet, Computer experiments on classical fluids. i. thermodynamical properties fo lennard-jones molecules., Phys. Rev. **159**, 98–103 (1967).

[60] R. W. Hockney, S. P. Goel, and J. Eastwood, Quiet highresolution computer models of a plasma., J. Comp. Phys. **14**, 148–158 (1974).

[61] H. J. C. Berendsen, J. P. M. Postma, W. F. V. Gunsteren, A. Dinola, and J. R. Haak, Molecular-Dynamics with Coupling to an External Bath., Journal of Chemical Physics **81**, 3684–3690 (1984).

[62] D. J. Evans and B. L. Holian, The Nose-Hoover thermostat., Journal of Chemical Physics **83**, 4069–4074 (Oct. 1985).

[63] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Mol. Phys. **52**, 255–268 (1984).

[64] W. G. Hoover, Canonical dynamics: equilibrium phase-space distributions, Phys. Rev. A **31**, 1695–1697 (1985).

[65] M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method., J. Appl. Phys. **52**, 7182–7190 (1981).

[66] S. S. Nosé and M. L. Klein, Constant pressure molecular dynamics for molecular systems., Mol.Phys. **50**, 1055–1076 (1983).

[67] S. Iijima, Helical Microtubules of Graphitic Carbon, Nature **354,**, 56–58 (1991.).

[68] M. S. Arnold, A. A. Green, J. F. Hulvat, S. I. Stupp, and M. C. Hersam, Sorting carbon nanotubes by electronic structure using density differentiation, Nature Nanotechnology **1**, 60–65 (2006).

[69] A. G. Rinzler, Materials processing: Sorting out carbon nanotube electronics, Nature Nanotechnology **1**, 17 – 18 (2006).

[70] M.-F. Yu, Strength and Breaking Mechanism of Multiwalled Carbon Nanotubes Under Tensile Load, Science **287**, 637–640 (2000).

[71] P. G. Collins and A. Phaedon, Nanotubes for Electronics, Scientific American **67**, 62–69 (2000).

[72] M. S. Arnold, S. I. Stupp, and M. C. Hersam, Enrichment of Single-Walled Carbon Nanotubes by Diameter in Density Gradients, Nano Letters **5**, 713–718 (2005).

[73] S. Lebedkin, K. Arnold, F. Hennrich, R. Krupke, B. Renker, and M. M. Kappes, FTIR-luminescence mapping of dispersed single-walled carbon nanotubes, New J. Phys. **5**, 1–11 (2003).

[74] G. Hummer, J. C. Rasalh, and J. P. Noworyta, Water conduction through the hydrophobic channel of a carbon nanotube, nature **414**, 188–190 (2001).

[75] E. Dujardin, T. W. Ebbesen, A. Krishnan, and M. Treacy, Wetting of Singel Shell Carbon Nanotubes, Advanced Materials **10**, 1472–1475 (1998).

[76] M. P. Rossi, Y. Haihui, Y. Gogotsi, S. Babu, P. Ndungu, and J.-C. Bradley, Environmental Scanning Electron Microscopy Study of Water in Carbon Nanopipes, Nano Letters **4**, 989–993 (2004).

[77] O. Byl, J.-C. Liu, Y. Wang, W.-L. Yim, J. Johnson, and J. Yates, Unusual Hydrogen Bonding in Water-Filled Carbon Nanotubes, J. Am. Chem. Soc. **128(37)**, 12090–12097 (2006).

[78] K. Matsuda, T. Hibi, H. Kadowaki, H. Kataura, and Y. Maniwa, Water dynamics inside single-wall carbon nanotubes: NMR observations, Physical Review B. **74**, 073415 (2006).

[79] N. R. de Souza, A. I. Kolesnikov, C. J. Burnham, and C. K. Loong, Sturcture and dynamics of water confined in single-wall carbon nanotubes, Journal of Physics: condensed matter **18**, 2321–2334 (2006).

[80] M. Gordillo and J. Marti, Hydrogen bond structure of liquid water confined in nanotubes., Chemical Physics Letters **329**, 341–345 (2000).

[81] W. H. Noon, K. D. Ausman, R. E. Smalley, and J. Ma, Helical ice-sheets inside carbon nanotubes in the physiological condition, Chemical Physics Letters **355**, 445–448 (2002).

[82] I. Brovchenko, A. Geiger, A. Oleinikova, and D. Paschek, Phase coexistence and dynamic porperites of water in nanopores, Eur. Phys. J. E **12**, 69–76 (2003), hydrophobic hydrophylic tubes entscheidet LJ parameters.

[83] J. Wang, Y.Zhu, J. Zhou, and X.-H. Lu, Diameter and helicity effects on static properties of water molecules confined in carbon nanotubes., Phys. Chem. Chem. Phys. **6**, 829–835 (2004).

[84] B. Huang, Y. Xia, M. Zhao, F. Li, X. Liu, Y. Li, and C. Song, Distribution patterns and controllable transport of water inside and outside charge single-walled carbon nanotubes, J. Chem. Phys. **122**, 084708 (2005).

[85] T. Werder, J. Walther, R. Jaffe, T. Halicioglu, and P. Koumoutsakos, On the Water-Carbon Interaction for Use in Molecular Dynamics Simulations of Graphite and Carbon Nanotubes, Journal of Physical Chemistry B **107**(6), 1345–1352 (2003).

[86] K. Hinsen, The molecular modeling toolkit: A new approach to molecular simulations, J. Comp. Chem. **21**, 79–85 (2000).

[87] C. Simmerling, B. Strockbine, and A. Roitberg, All-atom structure prediction and folding simulations of a stable protein., J Am Chem Soc. **124**, 11258–9 (2002).

[88] J. Kubelka, J. Hofrichter, and W. Eaton, The protein folding ‘speed limit’., Current Opinion in Structural Biology **4**, 76–88 (2004).

[89] C. Levinthal, Are there pathways for protein folding?, Journal de Chimie Physique **65**, 44 (1968).

[90] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Theory of protein folding: The energy landscape perspective., Annu. Rev. Phys. Chem. **48**, 545–600 (1997).

[91] C. B. Anfinsen, Principles that govern the folding of protein chains., Science **181**, 223–230 (1973).

[92] T. Herges., *Entwicklung eine Kraftfelds zur Strukturvorhersage von Helixproteinen.*, PhD thesis, University of Dortmund, 2003.

[93] J. T. Pedersen and J. Moult, Protein folding simulations with genetic algorithms and a detailed molecular description, J. Mol. Biol. **269**, 240–259 (1997).

[94] A. Nayeem, J. Vila, and H. A. Scheraga, A comparative study of the simulated-annealing and monte carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [met]-enkephalin., J. Comput. Chem. **12**, 594–605 (1991).

[95] D. M. Leitner, , C. Chakravarty, R. J. Hinde, and D. J. Wales, Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms., Phys. Rev. E **56**, 363 (1997).

[96] T. Herges, H. Merlitz, and W. Wenzel, Stochastic optimizations methods for biomolecular structure prediction., J. Assoc. Lab. Autom. **7**, 98–104 (2002).

[97] T. Herges and W. Wenzel, In Silico Folding of a Three Helix Protein and Characterization of Its Free-Energy Landscape in an All-Atom Force Field, Physical Review Letters **94**(1), 018101 (2005).

[98] T. Herges and W. Wenzel, An All-Atom Force Field for Tertiary Structure Prediction of Helical Proteins, Biophys. J. **87**(5), 3100–3109 (2004).

[99] R. A. Abagyan and M. Totrov, Biased probability Monte Carlo confromation searches and electrostatic calculations for peptides and porteins., J. Mol. Biol. **235**, 983–1002 (1994).

[100] T. Herges, A. Schug, and W. Wenzel, Exploration of the free energy surface of a three helix peptide with stochastic optimization methods., J. Quantum. Chem. **99**, 854–893 (2004).

[101] F. Avbelj and J. Moult, Role of electrostatic screening in determining protein main chain conformational preferences, Biochemistry **34**, 755–764 (1995).

[102] D. Eisenberg and A. D. McLachlan, Solvation energy in protein folding and binding, Nature **319**, 199–203 (1986).

[103] K. A. Sharp, A. Nicholls, R. Friedman, and B. Honig, Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models., Biochemistry **30**, 9686–9697 (1991).

[104] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, The Protein Data Bank., Nucleic Acids Research **28**, 235–242 (2000).

[105] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing., Science **220**, 671–680 (1983).

[106] A. Verma, A. Schug, K. H. Lee, and W. Wenzel, Basin hopping simulations for all-atom protein folding., J. Chem. Phys. **124**, 044515 (2006).

[107] J. Schneider, I. Morgenstern, and J. M. Singer, Bouncing towards the optimum: Improving the results of monte carlo optimisation algorithms., Phys. Rev. E **58**, 5085–5095 (1998).

[108] W. Kabsch and C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features., Biopolymers **22**, 2577–2637 (1983).

[109] K. Hukushima and K. Nemoto, Exchange monte carlo method and application to spin glass simulations., J. Phys. Soc. Jpn. **65**, 1604–1608 (1996).

[110] E. Marinari and G.Parisi, Simulated tempering: a new Monte Carlo scheme., Eur. Phys. Lett. **19**, 451–458 (1992).

[111] A. E. Garcia and N. Onuchic, Folding a protein in a computer: An atomic description of the folding/unfolding of protein., A. Proc. Natl. Acad. Sci. **100**, 13898–13903 (2003).

[112] M. Born, Volumes and hydration warmth of ions., Z. Phys. **1**, 45–48 (1920).

[113] G. Hawkins, C. Kramer, and D. Truhlar, Pairwise solute descreening of solute charges from a dielectric medium, Chem. Phys. Lett. **246**, 122–129 (1995).

[114] G. Hawkins, C. Kramer, and D. Truhlar, Parametrized models of aqueous free energies of solvation based on pairwise solute descreening of solute charges from a dielectric medium, J. Phys. Chem. **100**, 19824–19839 (1996).

[115] J. Chen, W. Im, and C. L. Brooks, Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field., J. Am. Chem. Soc. **128**, 3728–3736 (2006).

[116] O. B. Ptitsyn, Sequential mechanism of protein folding., Dokl. Akad. Nauk SSSR **210**, 1213–1215 (1973).

[117] P. S. Kim and R. L. Baldwin, Intermediates in the folding reactions of small proteins., Annu. Rev. Biochem. **59**, 631–660 (1990).

[118] H. J. Dyson and P. E. Wright, Peptide conformation and protein folding., Curr. Opin. Struct. Biol. **3**, 60–65 (1993).

[119] M. Karplus and D. L. Weaver, Protein folding dynamics: The diffusion-collision model and experimental data., Protein Sci. **3**, 650–668 (1994).

[120] S. A. Islam, M. Karplus, and D. L. Weaver, Application of the diffusion-collision model to the folding of three-helix bundle proteins., J. Mol. Biol. **318**, 199–215 (2002).

[121] T. Yoda, Y. Sugita, and Y. Okamoto, Comparisons of force fields for proteins by generalized ensemble simulations., Chem. Phys. Lett. **386**, 460–467 (2004).

[122] T. Wang and R. Wade, Force field effects on a βsheet protein domain structure in thermal unfolding simulations., J. Chem. Theor. Comput. **2**, 140–148 (2005).

[123] T. Z. Lwin and R. Luo, Force field influences in beta-hairpin folding simulations., Protein Sci. **15**, 2642–2655 (2006).

[124] V. Hornak, R. Abel, B. Stockbrine, A. Roitberg, and C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters., Proteins: Struct., Funct., Bioinformatics **65**, 712–725 (2006).

[125] M. A. Sali, E. Shakhnovich, How does a protein fold ?, Nature **369**, 248–251 (1994).

[126] M. T. Lazaridis, 'new view' of protein folding reconciled with the old through multiple unfolding simulations., Science **278**, 1928–1931 (1997).

[127] M. Lee, G. Gippert, K. Soman, D. Case, and P. Wright., Threedimensional solution structure of a single zinc finger dna-binding domain., Science **245**, 635–637 (1989).

[128] M. W. M. Kochoyan, H. T. Keutmann, Architectural rules of the zinc-finger motif: Comparative two-dimensional nmr studies of native and "aromatic-swap" domains define a "weakly polar switch"., Proc. Nat. Acad. Sci. (USA) **88**, 8455–8459 (1991).

[129] Y. Choo and M. M. Isalan, Advances in zinc finger engineering., Curr. Opin. Struct. Biol. **10**, 411–416 (2000).

[130] J. H. Laity, B. M. Lee, and P. E. Wright, Zinc finger proteins: new insights into structural and functional diversity., Curr. Opin. Struct. Biol. **11**, 39–49 (2001).

[131] F. D. Urnov, J. C. Miller, Y. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory, and M. C. Holmes, Highly efficient endogenous human gene correction using designed zinc-finger nucleases., Nature **435**, 646–651 (2005).

[132] G. R. Desiraju, *Crystal Engineering. The Design of Organic Solids.*, Elsevier, Amsterdam, 1989.

[133] Desiraju, Crystal engineering: a holistic view., Angewandte Chemie Int. Ed. **48**, 2–14 (2007).

[134] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, Antechamber, an Accessory Software Package for Molecular Mechanical Calculations, Journal of Chemical Information and Computer Sciences (submitted).

[135] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, AM1: a new general purpose quantum mechanical molecular model, Journal of the American Chemical Society **107**(13), 3902–3909 (1985).

[136] A. Jakalian, B. L. Bush, K. B. Jack, and C. I. Bayly, ast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method, J. Comp. Chem. **21**, 132–146 (2000).

[137] A. Jakalian, B. L. Bush, and C. I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, J. Comp. Chem. **23**, 1623–1641 (2002).

Acknowledgments

I first want to thank Wolfgang Wenzel for giving me the opportunity to be a member of his group, for helping me and answering always with enthusiasm and patience, for introducing me to different research fields and for his dynamic and visionary supervision. I also want to thank my colleagues; Abhinav Verma for his immense help every time I had questions and for teaching me everything I know about scripting and computers at all, Bernhard Fischer for his valuable and funded comments, for very interesting discussions encoded in an understated sarcasm and for encouraging me, Konstantin Klenin for explaining me brownian dynamics and Holger Merlitz, Alexander Schug, Srinivasa Murthy Gopal, and Horacio Sanchez.

During this thesis, I had the opportunity to work and collaborate with several experimental groups. I thank Frank Hennrich for collaborating and working together with us, and for providing us with his experimental data on carbon nanotubes. I also thank Tobias Grab for further collaborations and synthesizing the molecules I worked on, and the organic chemistry group involved on a common project. Working at the INT was a great experience. I thank Matthias Hettler for his logistic and indispensable technical support. I thank Erika Schütze and Christine Batsch for being always understanding and helpful in the administrative tasks. I thank Benjamin Schmidt and Julia Setzler for their willingness to help and for a nice office atmosphere and many other people at the INT for a nice quotidianity, Sabine Blatt, Robert Maul, Ismael Tejero, Lucia Fernandez, Paloma Sevillano, Erika Schütze and Christine Batsch.

I want to sincerely thank my good friend Jasmin Aghassi, for her suport and help during all these years, for encouraging me and for the hard task of proofreading parts of this work. I also thank my parents Paloma and Enrique and my sister Maria for their unconditional support and for their humorfull and unconstrained objectivity. Finally, I thank Axel Thielmann, without whom I would not have been able to finish this work, again for his suport and for much more than I can write here.