Coarse-grained peptide models: conformational sampling, peptide association and dynamical properties for peptides.

Dissertation

zur Erlangung des Grades

"Doktor der Naturwissenschaften"

am Fachbereich Physik
der Johannes Gutenberg-Universität Mainz

Olga Bezkorovaynaya

Max-Planck-Institut für Polymerforschung

Abstract

A bottom-up coarse-graining (CG) procedure for peptides in aqueous solution is studied in this thesis. The coarse-graining procedure reduces the number of degrees of freedom of the system, enabling us to investigate larger systems and due to the smoother energy landscape one can get faster a better sampling of the system. The interaction potentials in our coarse-grained model are constructed in a such way, that the coarse-grained peptide reproduces conformations according to a high-resolution (atomistic) model.

In this work the influence of differently constructed bonded potentials on the reproduction of atomistic characteristics in structure-based CG simulation was investigated. In the coarse-graining procedure one loses by constuction microscopic structural details of the peptide. This can be for example correlations between degrees of freedom. In the thesis it is presented that those "lost" properties can be recovered in a backmapping procedure which reintroduces atomistic degrees of freedom into CG structures – as long as the overall conformational sampling of the molecule is correctly represented in the CG level of resolution. These correlations play an important role in secondary structure formation. Therefore they are crucial for a realistic conformational ensemble of the peptide. It is shown that for an exact agreement of the CG conformations with the atomistic reference additional bonded potentials are required such as 1-5 bond and 1,3,5-angle potentials.

It is shown that the intramolecular parameters (i.e. bonds, angles, torsions) determined for short oligopeptides are transferable to longer peptide sequences. But one has to be aware that bonded potentials should be used only in combination with those nonbonded interaction potentials, with which they were parametrized. So, they cannot necessarily be combined with other nonbonded interactions, for example a different water model.

Since the energy landscape is smoother in CG simulations, there is the acceleration in time and in principle the CG time does not corresponds one to one to the atomistic time any more. The dynamical properties of the peptide in water on the atomistic and CG levels were investigated in order to get an estimate for the speed-up of the dynamics in the CG model compared to the atomistic one. We found that these scaling factors are different for different dynamical properties and concluded that there is different "speed-up" for different types of motions (for example rotation and translation). This is an important observation for the kinetics of processes such as peptide aggregation.

Zusammenfassung

In dieser Arbeit wird ein vergröbertes (engl. coarse-grained, CG) Simulationsmodell für Peptide in wässriger Lösung entwickelt. In einem CG Verfahren reduziert man die Anzahl der Freiheitsgrade des Systems, so dass man grössere Systeme auf längeren Zeitskalen untersuchen kann. Die Wechselwirkungspotentiale des CG Modells sind so aufgebaut, dass die Peptid Konformationen eines höher aufgelösten (atomistischen) Modells reproduziert werden.

In dieser Arbeit wird der Einfluss unterschiedlicher bindender Wechselwirkungspotentiale in der CG Simulation untersucht, insbesondere daraufhin, in wie weit das Konformationsgleichgewicht der atomistischen Simulation reproduziert werden kann. Im CG Verfahren verliert man per Konstruktion mikroskopische strukturelle Details des Peptids, zum Beispiel, Korrelationen zwischen Freiheitsgraden entlang der Peptidkette. In der Dissertation wird gezeigt, dass diese "verlorenen" Eigenschaften in einem Rückabbildungsverfahren wiederhergestellt werden können, in dem die atomistischen Freiheitsgrade wieder in die CG-Strukturen eingefügt werden. Dies gelingt, solange die Konformationen des CG Modells grundsätzlich gut mit der atomistischen Ebene übereinstimmen. Die erwähnten Korrelationen spielen eine grosse Rolle bei der Bildung von Sekundärstrukturen und sind somit von entscheidender Bedeutung für ein realistisches Ensemble von Peptidkonformationen. Es wird gezeigt, dass für eine gute Übereinstimmung zwischen CG und atomistischen Kettenkonformationen spezielle bindende Wechselwirkungen wie zum Beispiel 1-5 Bindungs- und 1,3,5-Winkelpotentiale erforderlich sind. Die intramolekularen Parameter (d.h. Bindungen, Winkel, Torsionen), die für kurze Oligopeptide parametrisiert wurden, sind übertragbar auf längere Peptidsequenzen. Allerdings können diese gebundenen Wechselwirkungen nur in Kombination mit solchen nichtbindenden Wechselwirkungspotentialen kombiniert werden, die bei der Parametrisierung verwendet werden, sind also zum Beispiel nicht ohne weiteres mit einem andere Wasser-Modell kombinierbar.

Da die Energielandschaft in CG-Simulationen glatter ist als im atomistischen Modell, gibt es eine Beschleunigung in der Dynamik. Diese Beschleunigung ist unterschiedlich für verschiedene dynamische Prozesse, zum Beispiel für verschiedene Arten von Bewegungen (Rotation und Translation). Dies ist ein wichtiger Aspekt bei der Untersuchung der Kinetik von Strukturbildungsprozessen, zum Beispiel Peptid Aggregation.

Contents

Chapter 1

Introduction

Proteins are biopolymers constructed of 20 natural amino acid building blocks linked by peptide bonds [1, 2]. Proteins can be viewed as the driving forces of cellular machinery, as they are responsible for diverse functions ranging from molecular motors to signaling. They catalyze reactions, transport oxygen and lipids, ions and sugar, form the building blocks of viral capsids, traverse the membranes to yield regulated channels, and transmit information from the DNA to the RNA. They also synthesize new molecules and are responsible for their degradation. Proteins are the vehicles of the immune response and of viral entry into the cell [3].

One important group of proteins are structure proteins [1, 2]. A prominent example of a structure protein is collagen, which is present in animal tissues such as skin, muscle, tendon, etc. Another example of interesting structure proteins are silk proteins [4, 5], which exist in silkworm and spider silk [6, 7, 8]. They are composed of fibroin and have very interesting properties (spider silk is very strong and at the same time very thin), which one would like to understand and possibly copy for synthetic or biomimetic materials. The organic matrix of many biological composite materials, such as bone, teeth, nails, etc., is also formed by proteins [9]. Remarkable mechanical properties like the extensibility and strength of protein materials are caused by a protein's ability to form 3D structures [10, 11]. For example, the strength of interactions between the β-sheet regions of proteins determines the tensile strength of each kind of silk [12]. 3-D protein structures are hierarchically organized as can be seen in Fig. 1.1:

- The secondary structure is a regularly repeating local structure stabilized by hydrogen bonds. α-helices, β-sheets and turns are the most common examples [2].

- The tertiary structure is the overall shape of a single protein molecule.

It stems from the spatial interactions between the secondary structures. The tertiary structure is generally stabilized by non-local interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen bonds, disulfide bonds, and even post-translational modifications. The term "tertiary structure" is often used as synonym with the term "fold". The tertiary structure is what controls the basic function of the protein [2].

- The quaternary structure describes a structure formed by several independent protein molecules (separate polypeptide chains), usually called subunits, which function as a protein complex. [2].

Proteins can be very long, up to thousands of amino acids [2] and, in addition to having a large number of degrees of freedom (DOF), their architectures are complex and of great diversity [2]. Therefore, the study of their properties with computer simulations can be problematic because of the required computer power and due to difficulties in parametrizing all of their properties, e.g., creating a force field. But to start with, one does not need such a long polymer, at least at the stage of investigating the model. It is possible to study shorter sequences as model systems in order to get an idea about the behavior and properties of the protein. A molecule with sequences of up to 50-100 amino acids is called a peptide. Peptides represent the most favorable building blocks for the design and synthesis of nanostructures because they offer a great diversity of chemical and physical properties [15]. They can also be synthesized in large amounts, modified and decorated with functional elements, and can then be used in diverse applications. Using basic biological building blocks and a large number of diverse peptide structural motifs, it is possible to build new materials via a bottom-up approach [11, 16, 17, 18, 19, 20]. The development of new bioinspired materials frequently relies on peptide units which are able to self-assemble, organize and form nanostructures [21, 22]. Many different approaches to design peptide sequences for structure formation have been pursued in the past: peptides can be divided into sequences that lead to the formation of fibrillar structures, hydrogels, or sequences that result in amphiphilic peptides which function as surfactants [16, 23, 24, 25].

The ability of proteins to aggregate plays an important role in the formation of large protein complexes (e.g. virus shells) and protein-based materials (e.g. spider silk, bones). From the biomedical point of view, these aggregates appear to be of significant importance for the understanding of neurodegenerative diseases such as Alzheimer's [26, 27], Huntington's and Parkinson's [28] disease. These phenomena are caused by a drastic change in protein conformations in connection with the aggregation process. Examples for such con-

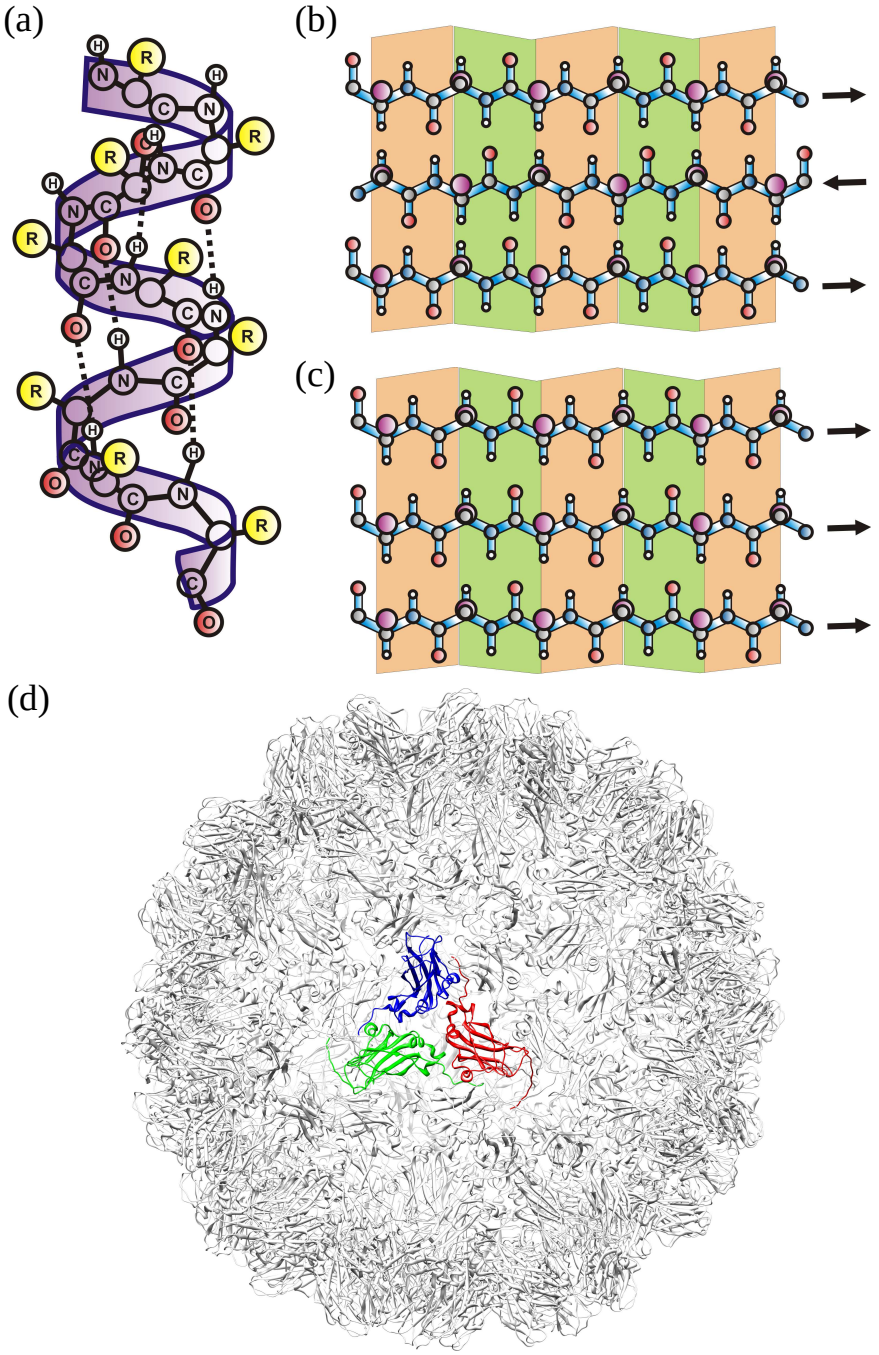

Figure 1.1: **3D-structures of proteins.** (a) Secondury structure, α-helix (adopted from [13]); (b) and (c) antiparallel and parallel β-sheets respectively (adopted from [14]); (d) colored part presents the tertiary strusture of the protein and whole picture present the virus, which is the example of quaternary structure.

formational changes or disorders are Kuru [16] and Creutzfeldt-Jakob [29] in humans, scrapie in sheep, BSE (Bovine spongiform encephalopathy) in cows [30], and certain types of cancer [16, 31, 32].

Partially driven by the need to understand neurodegenerative diseases and amyloid aggregation [33, 34, 35], the aggregation of peptides and proteins [36] has become a field of growing interest over the recent years. Another example of the importance of this development in medicine is the fact that some peptides can aggregate and serve as templates for tissue growth, in for example functional nerve networks, which can then even develop into active synapses [15, 37, 38, 39]. The ability of the nanostructures to provide a permissive environment for axonal regeneration of the tissue in the central nervous system after injury [15, 40] is used in tissue engineering and regeneration processes. Self-assembling materials in combination with a bioengineering platform can also be used to assist functional bone regeneration in cases of larger bone defects [21].

With these growing interests in protein- and peptide-based materials, understanding the relationship between sequences (structure and function) and mechanical properties needs more attention. Here, computer simulations are being used to gain a better molecular understanding of the systems and processes, as well as the driving forces and mechanisms behind them. Simulations can help to interpret and inspire experiments and to provide a picture of the underlying microscopic processes at a single molecule resolution frequently not accessible experimentally. The field of biomolecular simulation has been growing tremendously over the last 30 years. Atomistic biomolecular force fields have been developed which allow one to study peptides, proteins and other biomolecules on a microscopic level. Computer simulations are very powerful for refining structures of biomolecules in solution, in micelles or in membrane fragments, especially when the simulation refinement is based on experemental data like NMR or crystallography [41, 42]. Atomistic models provide important information on the nm scale (for example to study a protein in its native folded state), but it is not feasible to perform fully detailed atomistic simulations to obtain microscopic bulk properties of polymers or biological materials, or to observe the folding of a protein [43] on an atomistic level. This is because the huge number of degrees of freedom limits atomistic approaches for the investigation of phenomena at macroscopic length scales and the limited sampling also makes an assessment of possible force field problems difficult if the system gets trapped in the "wrong" configurations. So, one needs to reduce the number of DOF by coarsening the model and keeping only those DOF which are deemed relevant for the particular questions of interest. Thereby all potentials of the system will be smoother, all processes will run faster, and one can cover long time and

the length scales. Meso-scale or coarse-grained (CG) models contain enough information to retain the chemical identity of the parent polymer.

In a recent review, van Gunsteren et al. summerize several major challenges which biomolecular modeling is facing today [44]. They state that present day biomolecular modeling is limited by four main problems: 1) force field problems, 2) search (sampling) problems, 3) ensemble (sampling) problems, and 4) experimental problems. One example of a force field problem is that the free energy differences driving processes like folding are in the order of $1-10k_BT$, which are relatively small. However, these energies result from a summation over all atom pairs, so this number is actually very large for biological systems (10^6-10^8). Hence, it would be impossible to simulate folding or membrane formation. To reach the requested accuracy for the total nonbonded energy, the accuracy of the individual terms in the summation must be higher. There are sampling problems because such big systems have a large number of degrees of freedom and the energy hypersurface is very rugged, with energy basins and mountains of a wide range of depths or heights. This makes the search for a global energy minimum a daunting task. The ensemble problem consists of the fact that the state of a system is generally characterized not by one configuration or structure, but by a Boltzmann ensemble of configurations. Again, due to a large number of DOFs of the biological system, the ensemble is also large. Experimental data are necessary for the parametrization of force fields, but, for biomolecular systems these are scarce relative to the number of degrees of freedom involved. However, many different conformational ensembles may reproduce the same set of experimental data. Unfortunately, this makes the problem of deriving the conformational ensemble from experimental data for a biomolecular system under-determined. Different strategies have been pursued to overcome these challenges for biomolecular simulations, such as the development of enhanced sampling techniques, new force fields, etc. One important ingredient to overcome search and ensemble problems is the development of coarse-graining (CG) techniques.

CG methods to investigate proteins and peptides have already been developed for a long time and different ways of sampling such systems have been found. For example, there are one-bead [1] models, which reduce many degrees of freedom, but such models cannot easily describe secondary structure transitions [45]. Some people argue that this type of CG mapping scheme gives a 'natural' representation, because the amino acids are the 'building blocks' of proteins [45]. Models with a larger number of bead types in a back-

[1] A one-bead model is a model representing an amino acid with one single interacting center (bead) [45].

bone can describe the system better, but will need more time and computer resources [45]. In addition, intramolecular degrees of freedom of peptides within the backbone chain are usually correlated, since the underlying atomistic degrees of freedom are already correlated. Although the correlations in the atomistic simulation are usually reproduced, it is not so trivial to recover these correlations in CG simulation as well [46]. It is a rather difficult task, because one has to use pair interactions and it is better to avoid complex multi-parameter potentials (expensive in terms of parametrization, computational time and resources). Because of this correlation, one usualy keeps quite a high order of resolution for the backbones [47]. Also, the number of beads per amino acid side chain may vary in the CG mapping schemes. The choice of it depends on the properties which one wants to reproduce. Even one bead located on sidechains allows one to easily describe the non-bonded interactions [45] and an even higher resolution (almost atomistic) allows one to reproduce realistically both secondary structure properties as well as the hydrogen bonding behavior [48]. Different CG models are used in various areas such as self-assembly of peptide-based materials, peptide folding and in structure prediction algorithms [15, 48, 49, 50, 51, 52, 53]. Usually, these models are parameterized in such a way that they reproduce Ramachandran plots [2] [48, 55, 56, 57], which predict secondary structures [58, 59], known from experiments or an atomistic reference simulation [60, 61].

Often CG models are just one component of a multiscale simulation study where one investigates a system at several levels of resolution at the same time. In this case, the consistency of the different models at the individual levels of resolution is of particular importance. Suppose one wants to investigate a process on very large time and length scales for example, to study biomolecular aggregation or biological materials while at the same time being able to zoom in to a high resolution (to obtain local microscopic properties). Then it is important to combine both high and low resolution simulation models in such a way that it is possible to easily navigate in between them – either sequentially or by simultaneously simulating parts of the system at high and low resolution [62, 63]. In such a case, the switching/simultaneous treatment can only be done if the models at different levels of resolution are consistent both structurally and thermodynamically. Otherwise, each change of resolution would lead to the disruption of the structures, phase changes, etc. Several different coarse-graining approaches have been developed over the years which share the goal of being able to switch in between resolutions

[2] A graphical representation in which the dihedral angle of rotation about the alpha-carbon to carbonyl-carbon bond in polypeptides is plotted against the dihedral angle of rotation about the alpha-carbon to nitrogen bond [54].

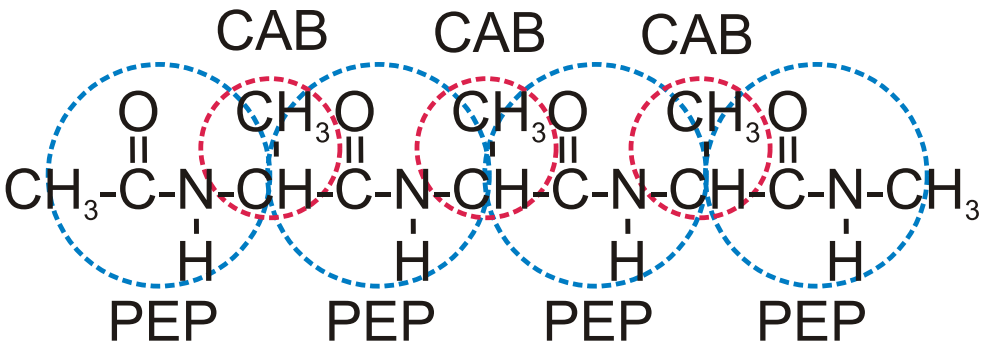

Figure 1.2: **Chemical structure of capped-alanine oligomer considered in this thesis, with a scheme of CG beads.**

and being transferable [64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76]. Usually in these approaches, a simulation at the higher resolution level (e.g. an atomistic simulation) is used as a reference to determine interaction functions for the lower resolution (CG) model.

Many of the CG methodologies and multiscale simulation methods (i.e. structure-based approaches [67], thermodynamics force field (MARTINI) [74], etc.) have been developed for or tested on model systems [66, 67, 77, 78] like homogeneous liquids [72, 79] and amorphous polymers [64, 65, 80]. Extending them to more complex systems such as biological materials, biomolecular systems with multiple components, etc. poses new and exciting challenges. One can address questions concerning aggregation and phase separation as well as complex conformational equilibria, which are in turn affected by the interactions with surfaces and interfaces.

The scope of this thesis is to investigate several aspects of coarse-graining with a special focus on extending the methodology to biomolecules (peptides) in aqueous solution. We have studied methods to develop CG interaction functions for multiscale simulations, conformational behavior on the CG level and a backmapped atomistic ensemble, and dynamic properties of the CG molecules. The central system of the thesis is a capped-alanine oligomer: $CH_3 - [CONH - CH_2CH_3]_3 - CH_3$ (Fig. 1.2). This choice is not accidental, as alanine oligomers are typical systems used to test simulation models for peptides (with structurally simple sidechains and no electrostatic charges). Also, for the creation of a CG model, the alanine oligomers are attractive since oligo-alanine by itself plays a significant role in the structure formation of silks (specifically spider silk). Thus, these simulations will also be useful for subsequent investigations of spider silk [6].

In this thesis, a CG model for oligo-alanine peptides is developed based on an atomistic (force field) description. We rely on a structure-based CG methodology to represent bonded and nonbonded interactions, such that they reproduce structural properties of the atomistic model. We assume that bonded and nonbonded interactions can be separated and parametrized independently. This assumption was originally made for amorphous polymer systems where bonded and nonbonded interactions can often be clearly separated. In this case, bonded interactions are determined based on sampling of single chains with long-range exclusions to avoid double counting (for details see Ref. [81]). Here, one assumes that the CG nonbonded "environment" has the same effect on the conformational sampling as the all-atoms nonbonded "environment" [60]. This assumption needs to be verified and in biomolecular systems, it is likely to be problematic since a hydrogen bonding and solvation effects significantly influence the conformational sampling. In Chapters 4 and 5 we will address the problems associated with interdependence of bonded and nonbonded interactions and between different bonded DOFs in details. For a short dipeptide (di-phenylalanine), a structure based CG model has previously been developed, which reproduces both conformations as well as the association behavior of the peptide in an atomistic description [60, 61]. In this thesis, the focus is on conformational sampling of a CG model for slightly longer peptides (capped tri- and tetra-Alanine, denoted as Ala_3 and Ala_4). It is investigated to which extend a (structure-based) CG model is capable of reproducing the conformational equilibrium of a corresponding atomistic model, with particular attention being paid on the problem of correlations between bond, angle and torsional degrees of freedom. These play a decisive role in the overall conformations adopted by biological molecules in aqueous solution.

The nonbonded interactions can be parametrized to reproduce thermodynamic properties (for example free energies) [69, 74, 82, 83] or structural properties [60, 61]. Reproducing one of the characteristics does not guarantee the reproduction of another [84]. Currently, some studies are in progress, which try to derive CG potentials, that can reproduce both kinds of properties (thermodynamic and structure) with reasonable agreement [85, 86].

In the present work, first calculations were done using the nonbonded potentials developed by Villa at al. [60, 61]. We also performed simulations to reparametrize the nonbonded potentials and to study the influences of pressure correction of the water model and of neighboring beads on the nonbonded interactions between peptides in solution.

Dynamical properties of the CG peptide are also considered. CG dynamics is generally faster than the dynamics of the underlying atomistic system as the averaging over detailed, local degrees of freedom leads to a

smoother potential energy surface and consequently faster dynamics of CG models [87]. To study the dynamics in a CG simulation, it is important to make a "bridge" between atomistic and CG systems. The "speed-up" of the dynamic processes depends on the CG model and it should be determined empirically for each model. It is not a priory clear whether different dynamical modes experience a different speed up in the coarse-graining process (for example translational, rotational and internal dynamics). This needs to be investigated since it influences the overall kinetics of the system and may lead to a different sampling of kinetically controlled processes. A better knowledge of the link between the time scales and their dependence on varying conditions can also be linked to the aspect of transferability and, in the end, to the predictive capability of the CG model.

The thesis is organized as follows: Chapter 2 gives a review of the theoretical background, which is the basis for molecular dynamics simulations. The general idea of CG methods, in particular the structure-based methods, are discussed in Chapter 3. We present our results and performance of the CG model for oligo-alanin in Chapter 4. Here, we will discuss what kind of properties we can reproduce with structure-based CG simulation. We can also obtain accurate results for characteristics which were not directly parametrized, and will present results for a backmapped system. Possible additional corrections to the nonbonded and bonded interactions are discused in Chapter 5. In this section we will present several variants of correction for CG potentials in order to obtain all kinds of properties with good agreement to the original atomistic system. In Chapter 6 the differences in non-bonded potentials are considered. We will describe the process of obtaining nonbonded potentials and possible ways of improving them. In Chapter 7, we focus on the hydrodynamical properties of the pure water and "peptide+water" system, like diffusion coefficients and rotational diffusion, and investigate the dependence of these properties on the system size and compare them for the atomistic and CG cases.

Chapter 2

Theoretical background: molecular dynamics (MD) simulation

In the present chapter I present briefly the theory, which is relevant to the simulations done in this thesis. First, I will consider the equations of motion, possible ways of their numerical determination [88, 89] and additional conditional (constraints) [90, 91]. Then I will discuss different ensembles [92, 93], periodic boundary conditions [94], the force fields [95, 96, 97] and the potentials of mean force (PMF) [98].

2.1 Molecular dynamics. Equations of motion

Molecular dynamics (MD) simulation [98] is a technique used to describe the motion of the system in a classical limit (with the laws of classical mechanics). The main justification of the MD method is that statistical ensemble averages are equal to time averages of a system, known as the ergodic hypothesis. The key idea in MD simulation is to describe how positions and velocities of each classical particle change with time use Newton's equations of motion. In other words, the behavior of an entire system can be determined if one has a set of initial conditions and forces of interactions for each of its components.

The potential energy surface on which the equations of motion are simulated is called force field. Though the descriptions of force fields include the covalent structure of the molecules, classical methods cannot be used in the case of chemical reactions which change the covalent structure (more details on force field can be found in Sec. 2.5). For investigation of such cases, the

reactive part of the molecular system should be treated differently, e.g. by quantum-chemical methods. Although classical MD simulations are limited by system size, the classical methods are faster than "*ab initio*" [99] methods.

As mentioned before, the MD simulation is based on the equations of motion, which should be solved for each particle of the system. The starting point is Hamilton's variational principle, which concisely summarizes most of classical mechanics into the statement that the phase space trajectory followed by a mechanical system is the one for which the time integral:

$$I = \int L \mathrm{d}t \tag{2.1}$$

is an extremum [88], where L is the Lagrangian:

$$L = L(q, \dot{q}) = T - U(\{q_i\}) = \frac{1}{2} \sum_i \dot{q}_i^2 m_i - U(\{q_i\}), \tag{2.2}$$

here T is a kinetic energy, U is a potential energy and m_i is the mass of one of the system's particle. Given set of N independent generalized coordinates[1] and velocities $\{q_i, \dot{q}_i\}$ that describe the state of a conservative system (one in which all forces derive from some potential energy function U), so that $L = L\left(\{q_i\}, \{\dot{q}_i\}, t\right)$, Lagrangian L can be shown to satisfy the Lagrange equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0, \tag{2.3}$$

where $i \in [1, N]$. If we take into account the Eq. 2.2, we can get the Newton's second law, which is a consequence of this result:

$$m_i \frac{\partial^2 \vec{q}_i}{\partial t^2} = -\frac{\partial U}{\partial q_i} = \vec{F}_i \tag{2.4}$$

There is an alternative formulation of the equations of motion. Replacement of the generalized velocities $\{\dot{q}_i\}$ in the Lagrange formulation by generalized momenta $p_i = \frac{\partial L}{\partial \{\dot{q}_i\}}$ (in case the coordinates are Cartesian and $p_i = m_i \dot{q}_i$) and consideration of the Hamiltonian as $H = H\left(\{q_i\}, \{p_i\}, t\right)$, where H is defined by:

$$H = \sum_i \dot{q}_i p_i - L, \tag{2.5}$$

give two first-order equations of motion associated with each coordinate:

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i} \end{aligned} \tag{2.6}$$

[1] The generalized coordinate can be replaced with Cartesian or polar coordinates.

In case, when H has no explicit time-dependence, there will be $\dot{H} = 0$, so H is a conserved quantity. Properly solving the equations of motion for a system without time- or velocity-dependent forces will therefore produce a microcanonical (NVE) ensemble, in practice, however, there will be errors that cause deviations from ideal behavior: errors in forces (because of truncation) will produce pseudorandom disturbances that cause energies to drift; the finite time step will cause integration errors and the total energy will not be exactly conserved. Therefore there are always modifications to the pure Newtonian equations of motion needed to generate long stable trajectories.

The important quality, which an integration algorithm should possess, is simplicity. A simple algorithm will involve the storage of only a few coordinates, velocities etc., and will be easy to program. There are three main criteria, which influence the choice of the algorithm:

- Time reversibility should be ensured, because it is inherent in the Newtonian equations of motion.
- The generated trajectories should conserve volume in phase space, i.e., the algorithm should be symplectic (the deviation from symplectic behavior produces time-dependent weight factors in phase space) [89].
- The computational effort is completely dominated by the force calculation, and therefore, it is better to prefer methods that use only one force evaluation per time step.

2.1.1 Verlet methods

Perhaps the most widely used method of integrating the equations of motion is that initially adopted by Verlet [100]. This method is a direct solution of the second-order equations Eq. 2.4. The method is based on positions $\vec{r}(t)$, accelerations $\vec{a}(t) = \ddot{\vec{r}}(t)$, and the positions $\vec{r}(t - \Delta t)$ from the previous step. The derivation of the Verlet formula follows from the Taylor expansion of the coordinate variable at times $t + \Delta t$ and $t - \Delta t$ [88]:

$$\begin{aligned} \vec{r}(t + \Delta t) &= \vec{r}(t) + \frac{\partial \vec{r}(t)}{\partial t}\Delta t + \frac{\vec{F}(t)}{2m}\Delta t^2 + \boldsymbol{O}(\Delta t^3) \\ \vec{r}(t - \Delta t) &= \vec{r}(t) - \frac{\partial \vec{r}(t)}{\partial t}\Delta t + \frac{\vec{F}(t)}{2m}\Delta t^2 + \boldsymbol{O}(\Delta t^3) \end{aligned} \tag{2.7}$$

Summing these two equations, we get:

$$\vec{r}(t + \Delta t) = 2\vec{r}(t) - \vec{r}(t - \Delta t) + \frac{\vec{F}(t)}{m}\Delta t^2 + \boldsymbol{O}(\Delta t^4). \tag{2.8}$$

The velocity $\vec{v}(t)$ is not directly involved in the solution, but if required it can be obtained from:

$$\vec{v}(t) = \frac{\partial \vec{r}(t)}{\partial t} = \frac{\vec{r}(t+\Delta t) - \vec{r}(t-\Delta t)}{2\Delta t} + \boldsymbol{O}(\Delta t^2). \tag{2.9}$$

The "leap-frog" algorithm [101, 102], which was used in this work, is equivalent to the Verlet scheme. It uses positions at integer time steps and velocities in between time steps (these velocities are used for calculation the new position):

$$\begin{aligned} \frac{\partial \vec{r}\left(t+\frac{1}{2}\Delta t\right)}{\partial t} &= \frac{\partial \vec{r}\left(t-\frac{1}{2}\Delta t\right)}{\partial t} + \frac{\vec{F}(t)}{m}\Delta t \\ \vec{r}\,(t+\Delta t) &= \vec{r}(t) + \frac{\partial \vec{r}\left(t+\frac{1}{2}\Delta t\right)}{\partial t}\Delta t \end{aligned} \tag{2.10}$$

As one can see after these calculation we will have position and velocity at different times, and one can not evaluate the kinetic and potential energy at the same time. Thus, the velocity at time t must be computed for example as an average:

$$v(t) = \frac{v\left(t+\frac{1}{2}\Delta t\right) + v\left(t-\frac{1}{2}\Delta t\right)}{2}. \tag{2.11}$$

Verlet and leap-frog schemes fulfill the criteria of being time-reversible and symplectic, and they require only one force evaluation per time step.

2.2 Constraints

Constraints are additional boundary conditions for the equation of motion, added in such a way that the system fulfills certain conditions, effectively reducing the number of degrees of freedom of the system. They can be described by a constraint equation $\sigma(\vec{r}) = 0$ depending only on the coordinates that should be satisfied all the times (holonomic constraints). For every constraint there is such an equation, for example:

- distance constraint between two particles: $\sigma = |\vec{r}_{ij}| - d_{ij} = 0$;
- angle constraint between two constrained bonds:
  $\sigma = \vec{r}_{ij}\vec{r}_{kj} - d_{ij}d_{kj}\cos\phi = 0$.

The way to introduce holonomic constraints into the equations of motion is by minimizing the action while preserving the constraints, using Lagrange multipliers. The new Lagrange equations are:

$$\frac{d}{dt}\left(\frac{\partial L'}{\partial \dot{r}_k}\right) - \frac{\partial L'}{\partial r_k} = 0, \tag{2.12}$$

where

$$L' = L + \sum_{s=1}^{m} \lambda_s \sigma_s(r), \tag{2.13}$$

while for all r along the path $\sigma_s(r) = 0$, $s = 1, ..., m$. Holonomic constraints do not modify the generalized momenta, because the constraints are not functions of $\dot{r}$, but the forces are:

$$m_i \ddot{\vec{r}}_i = \vec{F}_i + \sum_{s=1}^{m} \lambda_s \frac{\partial \sigma_s(r)}{\partial \vec{r}_i}. \tag{2.14}$$

The second term on the right side describes the constraint forces. For finding a proper set of Lagrange multipliers it is necessary to integrate these equations. There are two common methods: SHAKE [103] and LINCS [104]. Different modifications of these methods also exist: RATTLE [105] and P-SHAKE [106], SETTLE [107] (modifications of SHAKE) and P-LINCS [108] (modification of LINKS).

Constraints are used in the present thesis to constrain the bonds of the system and to calculate free energies along a reaction coordinates[2]. In many force field approaches all bonds in the system are constrained to their average bond length for the following reason: during the simulation the time step in the integration of the equations of motion needs to be chosen to be shorter than the shortest relevant time scale in a system. To describe fastest oscillations properly, one oscillation period should correspond to several time steps (usually of order 50 [90]). In typical atomistic simulations the intramolecular vibrations represent the motion with the highest frequency. If the high-frequency vibration of bonds is not of special interest, constraining these bonds allows for the use of a larger time step and therefore faster simulations.

2.2.1 SHAKE

One popular method of solving the constraint equations [103] can be used in conjunction with the Verlet algorithm:

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{(\Delta t)^2}{m_i}[\vec{F}_i^u(t) + \vec{F}_i^c(t)], \tag{2.15}$$

[2]A geometric parameter that changes during the conversion of one (or more) reactant molecular entities into one (or more) product molecular entities and whose value can be taken for a measure of the progress of an elementary reaction (for example, a bond length or bond angle or a combination of bond lengths and/or bond angles; it is sometimes approximated by a non-geometric parameter, such as the bond order of some specified bond [109]).

where $\vec{F}^u$ are the forces disregarding the constraints, and the constraint forces on particle i at time t are given by:

$$\vec{F}_i^c(t) = \sum_s \lambda_s(t) \frac{\partial \sigma_s}{\partial \vec{r}_i}. \tag{2.16}$$

The effect of the constraint force is to add a second contribution to the displacement of the particles. The algorithm first computes the new positions $\vec{r}_i'$ disregarding the constraints:

$$\vec{r}_i' = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{(\Delta t)^2}{m_i} \vec{F}_i^u(t), \tag{2.17}$$

and then corrects the positions with $\Delta\vec{r}_i$ such that

$$\sigma_s(\vec{r}' + (\Delta\vec{r})) = 0, s = 1, ..., m \tag{2.18}$$

where

$$\Delta\vec{r}_i = \frac{(\Delta t)^2}{m_i} \sum_s \lambda_s(t) \frac{\partial \sigma_s(\vec{r}(t))}{\partial \vec{r}_i}. \tag{2.19}$$

This set of m coupled equations for the m λ's is then solved sequentially and the procedure is iterated until convergence. This method resets the coordinates after an unconstrained time step to satisfy the constraints within a giving precision. SHAKE is simple and numerically stable because it resets all constraints within a prescribed tolerance but it is an iterative method. When the displacements are large no solutions may be found for this case, because each bond will be calculated one by one and an improvement one of them can lead to worsening previously made ones. Therefore, it is difficult to parallelize SHAKE [104]. At the same time it is not so difficult to parallelize another method for constraints, LINKS, which, at the same accuracy, is three to four times faster than the SHAKE algorithm [104].

2.2.2 LINCS (Linear Constraint Solver)

The LINCS method rewrites the equations of motion to include the constraints and does not reset the coordinates like SHAKE. This is done in the following way: Eq. 2.14 can be rewritten in the matrix notation:

$$\mathbf{M}\ddot{\mathbf{r}} = \mathbf{f} + \mathbf{C}^{\mathbf{T}}\lambda, \tag{2.20}$$

where $\mathbf{r}$ and $\mathbf{f}$ are $3N \times 1$ matrices containing coordinates and unconstrained forces correspondingly, $\mathbf{M}$ is the $3N \times 3N$ diagonal matrix of masses, and the constraint matrix $\mathbf{C}$ is defined by:

$$C_{si} = \frac{\partial \sigma_s}{\partial r_i}. \tag{2.21}$$

By taking the time derivative of :

$$\frac{\partial \sigma_s}{\partial t} = \sum_i \frac{\partial \sigma_s}{\partial r_i}\frac{\partial r_i}{\partial t} = (\mathbf{C}\dot{\mathbf{r}})s = 0, \tag{2.22}$$

the following relation is found:

$$\mathbf{C}\ddot{\mathbf{r}} = -\dot{\mathbf{C}}\dot{\mathbf{r}}. \tag{2.23}$$

By left-multiplying Eq. 2.20 by $\mathbf{CM^{-1}}$ and using relation Eq. 2.23 we obtain:

$$\lambda = -(\mathbf{CM^{-1}C^T})^{-1}(\mathbf{CM^{-1}f} + \dot{\mathbf{C}}\dot{\mathbf{r}}). \tag{2.24}$$

The matrix $\mathbf{CM^{-1}C^T}$ is non-singular and can be inverted, if the constraints are independent [110]. Substituting this expression for λ in Eq. 2.20 we obtain the equations of motion:

$$\ddot{\mathbf{r}} = (\mathbf{1} - \mathbf{TC})\mathbf{M^{-1}f} - \mathbf{T}\dot{\mathbf{C}}\dot{\mathbf{r}}, \tag{2.25}$$

where $\mathbf{T} = \mathbf{M^{-1}C^T}(\mathbf{CM^{-1}C^T})^{-1}$. The matrix $\mathbf{1} - \mathbf{TC}$ projects the accelerations due to the unconstrained forces onto the constraint hypersurface. The first term in Eq. 2.25 gives the constrained accelerations due to the systematic forces and the second term gives the constrained accelerations due to centripetal forces. In this work the LINKS algorithm was used.

2.3 Ensembles

As stated in Sec. 2.1, in case when $\dot{H} = 0$, H is a conserved quantity, and solving the equations of motion will result in microcanonical sampling (NVE). But there are always errors in practical simulations due to the finite time step and due to numerical accuracy. To avoid this it is necessary to make modifications to the Newtonian equations of motion Eq. 2.4. In most applications it is desirable to simulate at constant temperature T [94], i.e. to generate a canonical NVT ensemble, and in many applications one needs to simulate at constant pressure, i.e. in an isothermal-isobaric NpT ensemble. These ensembles are often favorable in simulations since they describe conditions, which come closer to experimental ones that the NVE ensemble.

There are several ways to control temperature and pressure in the system. Not all of them generate well-defined ensembles, but even such algorithms are still used, for example for equilibration of the system since they are very robust and numerically stable. So, the choice of the method depends on the purpose of the simulation.

In the following I will review the thermostats used in this thesis. Usually there are analogous barostats (i.e. on following the pressure to obtain an NpT ensemble) for each of these methods [111].

2.3.1 Stochastic methods

The Langevin thermostat (also called stochastic dynamics (SD) thermostat) is a stochastic method that involves the application of random forces together with friction forces, and produces a canonical ensemble. A friction force is added to the conservative force, which removes kinetic energy from the system, and a random force adds kinetic energy to the system. To generate a canonical ensemble these two contributions have to obey the fluctuation-dissipation theorem [93, 112, 113]. The equations of motion have the following form:

$$\dot{\vec{p}}_i = \vec{F}_i - \gamma_i \vec{p}_i + \vec{R}_i(t), \tag{2.26}$$

where $R_i(t)$ is a zero-average stationary random process without memory:

$$\langle \vec{R}_i(0) \otimes \vec{R}_i(t) \rangle = 2m_i\gamma_i k_B T \delta(t) \mathbb{1}. \tag{2.27}$$

Langevin dynamics can be seen as molecular dynamics with a stochastic thermostat, when $\frac{1}{\gamma_i}$ is large compared to the relevant time scales in the system. But, if $\frac{1}{\gamma_i}$ is small, the dynamics of the system will be different from molecular dynamics (see below), but sampled configurations correspond to the ones in the canonical ensemble. Due to the random and friction forces the implementation of this method with the leap-frog algorithm is not trivial and it is reviewed in details by W.F van Gunster and H.J.C. Berendsen in Ref. [114].

The Langevin thermostat acts locally, because it changes the momentum of each particle. Particles, which moves too slow, gain more energy by the noise term, while too fast particles are slowed down by the friction. This keeps numerical instabilities, which usually arise from inaccurate calculation of a local collisionlike process, effectively under control and prevents them from propagating. But there is also a disadvantage: it is not momentum conserving, which is one of the conditions to treat hydrodynamics correctly.

The method of dissipative particle dynamics (DPD) [115, 116] can circumvent this. As the SD thermostat DPD also uses friction and noise on a local scale, but those are applied to pairs of particles. The friction scales the relative velocities between pairs of nearby particles and the noise acts according to Newton's third law. Therefore this method reproduces hydrodynamic behavior on large length and time scales [117].

2.3.2 Weak-coupling methods

Weak-coupling methods (Berendsen thermostat and barostat [92]) are not stochastic in nature, and can be applied both for temperature and pressure

control. The idea is to rescale velocities per step in a such way that the total temperature of the system T satisfies:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}, \tag{2.28}$$

where T_0 is the desired temperature. The temperature T is given by the kinetic energy found after updating the velocities in a normal molecular dynamic step. This rate equation would cause a temperature deviation from T_0 to decay exponentially to zero with a time constant τ. The implementation in terms of a scaling factor λ for the velocities is given by equation:

$$\lambda^2 = 1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T} - 1\right). \tag{2.29}$$

The temperature is conserved and the scaling is complete for the smallest possible value of the time constant $\tau = \Delta t$, where Δt is the MD time step. This case corresponds to the Gaussian isokinetic thermostat [118, 119, 120] which produces a canonical ensemble in configuration space [94]. The kinetic energy does not fluctuate for small τ but the potential energy does. As τ increases, fluctuations in kinetic energy appear at the expense of potential energy fluctuations, to become equal and opposite at large τ. So, for τ much longer than the intrinsic correlation times for internal exchange of energy, the scaling has no effect and a microcanonical ensemble is obtained [94]. Strictly speaking, this thermostat does not generate a proper canonical ensemble. For very small systems the sampling will indeed be incorrect, because the error scales with $1/N$, but for larger systems most properties will not be affected significantly, except for the distribution of the kinetic energy itself, i.e. head capacity is not correct in the system [121]. A similar thermostat which does produce a correct ensemble is the stochastic velocity rescaling thermostat by Bussi et al, which will be described below.

The Berendsen thermostat is not acting locally but on a global scale, since only the kinetic energy of the whole system is monitored and all particle velocities are scaled by the same factor. Therefore artifacts can occur, for example for an inhomogeneous system, e.g. a big protein in water, the temperatures for the protein and the solvent can differ strongly while the "average" temperature of the system is correct. This problem can be solved by controlling separately protein and water by two independent thermostats.

Pressure control by weak coupling is possible by scaling coordinates, in similar way it is done for the temperature:

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p}. \tag{2.30}$$

The scaling of coordinates and volume is done the following way:

$$\vec{r}' = \chi\vec{r}, \tag{2.31}$$
$$V' = \chi^3 V, \tag{2.32}$$

at every time step with a scaling factor χ:

$$\chi^3 = 1 - \beta_T \frac{\Delta t}{\tau_p} \left(P_0 - P\right), \tag{2.33}$$

with β_T being the isothermal compressibility:

$$\beta_T = -\left(\frac{1}{V}\right)\frac{\partial V}{\partial P}. \tag{2.34}$$

The Berendsen barostat has advantage of smooth response, but the disadvantages are that it does not generate a known ensemble and fluctuations cannot be used. In order to generate true NpT ensemble one should use a barostat first introduced by Andersen [122], then generalized for anisotropic volume fluctuations by Parrinello and Rahman [123] and later the implementation of this barostat was cosidered in details by Kolb and Dünweg [124]. But if the pressure of the system is very far from equilibrium, the Parrinello-Rahman coupling may result in very large box oscillations that could even crash a run. In that case it is better to use the weak coupling scheme to reach the target pressure, and then switch to Parrinello-Rahman coupling once the system is in equilibrium [121]. Mainly we made simulations with constant pressure for equilibrating the system and getting the reference data for our CG model, therefore we used Berensen barostat.

2.3.3 Canonical sampling through stochastic velocity rescaling by Bussi et.al.

This kind of thermostat uses velocity rescaling to obtain a canonical ensemble. It was proposed by Bussi et al. [125]. The stochastic velocity rescaling method by Bussi et.al. consists in multiplying the velocities of all particles by the same factor α, calculated by enforcing the total kinetic energy K to be equal to the average kinetic energy at the target temperature, $\overline{K} = \frac{1}{2}N_f k_B T$, where N_f is the number of degrees of freedom. The rescaling factor α for the velocities is obtained as:

$$\alpha = \sqrt{\frac{\overline{K}}{K}}. \tag{2.35}$$

The same factor is used for all the particles, therefore there is neither an effect on constrained bond lengths nor on the center of mass motion [125]. It is proposed to modify the way the rescaling factor is calculated, so as to enforce a canonical distribution for the kinetic energy. Instead of forcing the kinetic energy to be exactly equal to $\overline{K}$, a target value K_t with a stochastic procedure aimed at obtaining the desired ensemble is selected. So, the rescaling factor with this modification is:

$$\alpha = \sqrt{\frac{K_t}{K}}, \tag{2.36}$$

where K_t is drawn from the canonical equilibrium distribution for the kinetic energy:

$$\overline{P}(K_t)dK_t \propto K_t^{(N_f/2-1)} \exp\left[-\frac{K_t}{k_B T}\right] dK_t \tag{2.37}$$

The system is evolved using Hamilton's equations (Eq. 2.6) between the rescalings. The number of integration time steps in between can be fixed or randomly varied. Both the Hamiltonian evolution and the stochastic velocity rescaling leave a canonical probability distribution unaltered. Under the condition that the Hamiltonian evolution is ergodic in the microcanonical ensemble, it follows that this method samples the canonical ensemble [126].

The stochastic velocity rescaling thermostat by Bussi et.al. is practically a Berendsen thermostat (see sec. 2.3.2), but with an additional stochastic term that ensures a correct kinetic energy distribution:

$$dK = (\overline{K} - K)\frac{dt}{\tau} + 2\sqrt{\frac{K\overline{K}}{N_f}}\frac{dW}{\sqrt{\tau}}, \tag{2.38}$$

where N_f the number of degrees of freedom, dW a Wiener process and τ is a coupling constant. Without the second summand this equation became the same as for Berendsen thermostat (see Eq. 2.28):

$$K = \frac{3}{2}k_B T, \tag{2.39}$$

$$dK = (\overline{K} - K)\frac{dt}{\tau} \Rightarrow d(\frac{3}{2}k_B T) = (\frac{3}{2}k_B\overline{T} - \frac{3}{2}k_B T)\frac{dt}{\tau} \Rightarrow \tag{2.40}$$

$$\Rightarrow \frac{dT}{dt} = \frac{\overline{T} - T}{\tau}$$

2.4 Periodic boundary conditions

Typically molecular dynamics is applied to systems containing several hundred to a million atoms. Such systems are dominated by surface effects.

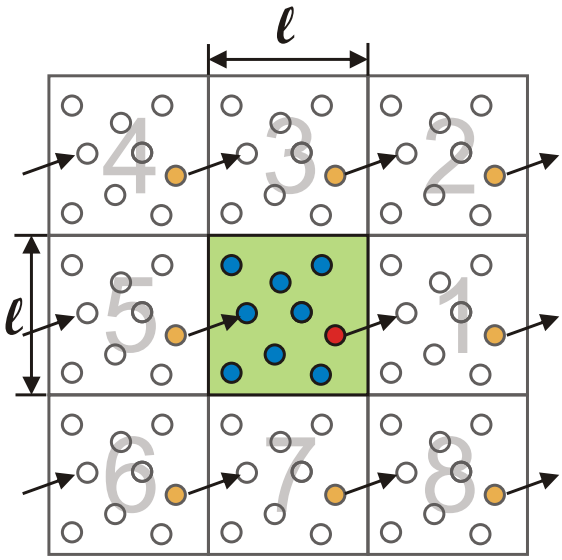

Figure 2.1: **All replicated in three dimensions.**

Using periodic boundary conditions is the way to overcome this problem. The system is exactly replicated in three dimensions, see Fig. 2.1. These replicas are called image cells. The simulated system is thus a unit cell of a periodic lattice. Each unit cell can have an arbitrary triclinic shape, defined by three basis vectors a, b, c with arbitrary angles α, β, γ between the basis vectors.

For simulations of proteins and peptides, the optimal unit cell has a minimal volume such that there is a prescribed minimum distance between any atom of the molecule and any atom of any neighboring image. This condition ensures that the interaction between images is small, while the volume minimizes the computational time spent on the less interesting solvent. A rhombic dodecahedron is the best choice for approximately spherical molecules [94]. In this work, a cubic box was used for simplicity.

Use of periodic boundary conditions removes unwanted surface effects at the expense of introducing artificial periodicity into the simulated system. Each atom in the unit cell interacts in principle with all $N - 1$ other atoms and all their images, including its own periodic image. It is implied that with periodic boundary conditions the potential functions are also periodic. Concerning the long ranged Coulombic interactions, this periodicity is taken into account by the use of lattice sums, which will be discussed later. Artifacts of periodicity can be avoided by modification of the interaction potentials such,

that these potentials vanish for distances larger than half of the smallest length of the unit cell. Thus only nearest image interactions can occur. For sufficiently short-ranged forces, like the Lennard-Jones model, it is necessary to consider only those image cells that adjoin the primary cell (in 2D system there are eight such image cells). The modification of potentials by adding cut-offs brings its own artifacts. There is no good solution to avoid periodicity artifacts completely: sudden cut-offs cause additional noise and erroneous behavior whereas smooth cut-offs strongly modify the interaction. The best strategy is to use consistent forces and potentials by inclusion of complete lattice sums, in combination with studying the behavior of the system as a function of the box size.

Effects of periodicity are more pronounced on dynamic properties, particularly time correlation functions. The imposed periodicity in space causes a suppressed but finite periodicity in time. Thus unrealistic correlations occur when the delay time for sampling the correlation function exceeds the time needed for spatial translations to become periodic, τ_{pbc}. This periodic correlation time τ_{pbc} is equivalent to the time required for a longitudinal wave to traverse the simulation cell of side a. The sound propagates by longitudinal wave as well and because of it, we can use the sonic velocity ω to obtain an expression for τ_{pbc}:

$$\tau_{pbc} = \frac{a}{\omega} = \frac{1}{\omega}\left(\frac{N}{\rho}\right)^{1/3}, \tag{2.41}$$

where the sonic velocity is denoted as $\omega = \frac{1}{\sqrt{\rho m \kappa_s}}$, m is the atomic mass, κ_s is an adiabatic compressibility, and ρ is the density of the system. An additional finite-size effect in hydrodynamic interactions between periodic images are discussed in more details in Chapter 7.

2.5 Force fields

Classical atomistic force fields can be based on different parametrization principles, use different reference data, are typically specialized for different applications and thus yield different results. Force field functions and parameter sets are derived from both experimental data and high-level quantum mechanical calculations. Ideally, a force field should be transferable between different molecules and valid for a broad range of environments and conditions. But usually important contributions are omitted, when a system is described by effective interactions between classical particles, which makes force fields typically non-transferable. High level (*ab initio*) quantum calculations should, ideally, provide a proper potential energy surface for molecules, but

they are severely limited in system size and cannot be applied to or produce reference data for bulk systems. For larger systems one has to make an approximation to the full quantum potential energy surface. Even high-quality *ab initio* calculations are not accurate enough to produce overall accuracies better than k_BT, required to obtain accurate thermodynamic properties. For this reason, many classical force fields are parametrized according to both experimental reference data of bulk systems such as density, dielectric properties, solvation thermodynamic data etc. and according to QM reference calculations. The force field used in this work is the Gromos53A6 [127] force field, a biomolecular force field particularly well suited for the simulation of proteins and peptides. Therefore it was chosen for atomistic reference calculations for the CG model for oligoalanine peptides. The aim of this work was to investigate how one can create a method for obtaining CG models with the all atom force field as a reference and we did not evaluate the agreement of the Gromos53A6 force field with experimental data, but assumed that its accuracy is reasonably good. Of course later we may revisit the atomistic force field and extend our method to other atomistic reference force fields. Here the CG model may even be useful to explore time and length scales that are not accessible to the atomistic model alone, which opens new possibilities to evaluate the quality of the atomistic system. The CG model is supposed to reproduce the conformational properties of the atomistic model, employing the atomistic force field as a starting point for the creation of the CG potentials.

There are two type of atomistic force fields: "All-atom" force fields, which provide parameters for every atom in a system (including hydrogen) and "united-atom" ones, which treat hydrogen and carbon atoms contained in methyl and methylene groups as a single interaction center [128]. The idea to include several atoms in one "super atom" as in the "united-atom" force field, can also be used for the creation of a coarse-grained model and its respective force field.

Atoms are the mass points which move according to the force field's laws and serve as the source points for the different terms in its description. The basic functional form of a force field encapsulates both bonded terms, relating atoms that are linked by covalent bonds and nonbonded (also called "noncovalent") terms, describing the long-range electrostatic and van der Waals forces. These atoms can belong to one molecule or to different ones. The specific decomposition of the terms depends on the force field used. For example, in the MM3 force field (Molecular Mechanics force field) the total energy of the system includes the stretching energy, bending energy, torsion energy, nonbonded term and also cross-terms like: stretch-bend and torsion-stretch terms [97]. A general form for the total energy in an additive force

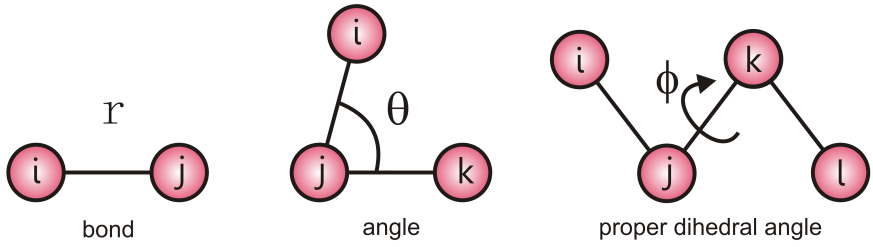

Figure 2.2: **Bonded interactions**

field can be written as $E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}}$, where the components of the covalent and noncovalent contributions are given by the following summations:

$$\begin{aligned} E_{\text{bonded}} &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} \\ E_{\text{nonbonded}} &= E_{\text{electrostatic}} + E_{\text{van der Waals}} \end{aligned} \tag{2.42}$$

These two types are computationally different: bonded interactions concern atoms which are read from a fixed list, whereas atoms involved in non-bonded interactions fluctuate and must be updated regularly. In the following we will mostly focus on the force field terms relevant for the Gromos force field used in this thesis.

2.5.1 Bonded interactions

There are several types of bonded potentials, which depend on involved atoms and differ from each other by the number of neighboring atoms involved in the interaction (Fig. 2.2).

- A covalent bond between two atoms is usually modeled as a harmonic oscillator in force fields which do not allow bond breaking:

$$V_b(r_{ij}) = \frac{1}{2} k_{\text{b}} (r_{ij} - b_0)^2, \tag{2.43}$$

  where $r_{ij} = |\vec{r}_i - \vec{r}_j|$ is the distance between two atoms, b_0 and k_{b} are the parameters, which differ for each type of bond. A more realistic description of a covalent bond at higher stretching is provided by the computationally more expensive Morse potential:

$$V_{\text{morse}}(r_{ij}) = D_{ij}[1 - \exp(-\beta_{ij}(r_{ij} - b_0))]^2. \tag{2.44}$$

  As mentioned in Sec. 2.2 in many atomistic simulations bonds are constrained to the average bond length.

In coarse-grained polymer simulations the beads are often connected by a finitely extensible nonlinear elastic (FENE) bonded potential [129]:

$$V_{FENE}(r_{ij}) = -\frac{1}{2}k_{ij}^{\mathrm{b}}b_{ij}^2 \log\left(1 - \frac{r_{ij}^2}{b_{ij}^2}\right). \tag{2.45}$$

At short distances the potential asymptotically goes to a harmonic potential with force constant k_{b} (Eq. 2.43), while it diverges at distance b.

- A covalent bond angle is described by a harmonic angular potential of the form:

$$V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^{\theta}(\theta_{ijk} - \theta_{ijk}^0)^2, \tag{2.46}$$

where $\theta = \arccos\frac{\vec{r}_{ij}\vec{r}_{kj}}{r_{ij}r_{kj}}$ is the angle between atoms i, j and k. Also for this kind a potential a simplified function can be used:

$$V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^{\theta}(\cos(\theta_{ijk}) - \cos(\theta_{ijk}^0))^2, \tag{2.47}$$

- The dihedral angle ϕ is defined by the positions of four atoms i, j, k, l, which are connected by three consecutive bonds. ϕ is an angle between the normals $\vec{n}$ and $\vec{m}$ to the two planes i, j, k and j, k, l:

$$\phi = \arccos\frac{\vec{n}\vec{m}}{|\vec{n}||\vec{m}|}, \tag{2.48}$$

where $\vec{n} = \vec{r}_{ij} \times \vec{r}_{kj}$ and $\vec{m} = \vec{r}_{jk} \times \vec{r}_{lk}$. In the gromos force field the potential for the proper dihedral angle is represented as:

$$V_d(\phi_{ijkl}) = k_\phi(1 + \cos(n\phi - \phi_0)). \tag{2.49}$$

In this periodic potential all minima are equivalent (in other words, the trans and the two gauche states for the threefold periodic dihedral, as between two sp^3 carbon atoms are equivalent). The actual energetic difference between the minima is caused by the introduction of an extra nonbonded interaction between the first and fourth atom, called the 1-4 interaction.

- There is another type of dihedral angle, which keeps groups planar and prevents molecules from flipping over to their mirror images for tetrahedral sites with only three bonds between 4 atoms, which span two planes. This type of angle is named improper dihedral and is defined by a harmonic potential:

$$V_{\mathrm{improper}}(\xi) = \frac{1}{2}k_\xi(\xi - \xi_0)^2, \tag{2.50}$$

where ξ is improper dihedral angle.

2.5.2 Nonbonded interactions

The nonbonded terms are most computationally involved, as they include many more interactions per atom. They are typically pairpotentials and functions of the distance r_{ij} between the two atoms. Atom pairs, which are already involved in bonded interaction like the 1-2 and 1-3 along a covalently-bonded chain, are usually excluded from the nonbonded ones. The 1-4 interaction can also be excluded from the nonbonded potentials, or used in a modified form. The van der Waals term is usually computed with a Lennard-Jones potential and the electrostatic term via Coulomb's law. The Lennard-Jones potentials are truncated with a cut-off. For calculation of the electrostatics, either also cut-off truncation or in case of simulations with periodic boundary condition full lattice sums over a periodic lattice (see Sec. 2.5.4) can be used.

Lennard-Jones interactions have the characteristics of short-range repulsion and long-range dispersion interactions:

$$V_{LJ}(r_{ij}) = \frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6}, \tag{2.51}$$

where the coefficients C^{12} and C^6 are constants whose values depend on the depth of the energy well and the equilibrium separation of the two atoms' nucleiare and they are determined as:

$$\begin{aligned} C_6^{ij} &= (C_6^{ii} C_6^{jj})^{\frac{1}{2}}, \\ C_{12}^{ij} &= (C_{12}^{ii} C_{12}^{jj})^{\frac{1}{2}}, \end{aligned} \tag{2.52}$$

where C^{ij} is the interaction coefficient for the two types of particles, and C^{ii} is the coefficient for the interaction between the same particles. This is so called geometric mixing rule. This potential also can be written as:

$$V_{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \tag{2.53}$$

There is the connection between these two types of coefficients:

$$\begin{aligned} \sigma &= \sqrt[6]{\frac{C_{12}}{C_6}}, \\ \varepsilon &= \frac{C_6^2}{4C_{12}} \end{aligned} \tag{2.54}$$

and the other way around:

$$\begin{aligned} C_{12} &= 4\varepsilon\sigma^{12}, \\ C_6 &= 4\varepsilon\sigma^6 \end{aligned} \tag{2.55}$$

Coulomb interaction between charges or partial charges on the atoms is expressed as:

$$V_C(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_1 r_{ij}}, \tag{2.56}$$

where ε_0 is a electric constant, ε_1 is the relative dielectric constant of a medium.

2.5.3 Long-range dispersion interaction

Since the numerical effort to calculate non-bonded interactions increases with N^2, where N being the number of atoms in the system, we usually truncates nonbonded interactions beyond a cut-off distance. The influence of the particles beyond the cut-off is either neglected, or treated in a different way that is less computationally demanding that N^2.

The easiest way to apply a cut-off is an abrupt force and potential truncation. In this case the force is no longer the derivative of the potential and the potential itself is no longer conservative, and its derivative contains a delta function at the cut-off distance (this leads to artifacts). The use of a truncated force without delta function shifts the potential to zero at the cut-off distance and at the same time this sudden jump in the force leads to extra noise, heating and artifacts in the density distribution. These artifacts can be avoided, if the force is shifted to zero at the cut-off radius, but this has an even more severe influence on the effective potential, which now deviates from the exact potential over a wide range. In practical MD algorithms, several kinds of switching functions are used, switching the force smoothly off at the cut-off distance.

The error due to neglect of long-range interactions beyond the cut-off radius r_c can be reduced by increasing r_c, but this leads to an expensive long pair list and consequently increased computational effort. The error can be taken into account in the computation of energy and pressure of the system for three possible functions for cut-off: truncated potential, truncated force and shifted force. The average number density is ρ and the radial distribution function is $g(r)$. Then the correction to the potential energy and therefore to the internal energy u per particle can be expressed as:

$$\Delta u = \frac{1}{2}\rho \int_0^\infty \Delta V(r) 4\pi r^2 g(r) dr, \tag{2.57}$$

and the pressure correction as:

$$\Delta P = -\frac{2\pi}{3}\rho^2 \int_0^\infty r^3 g(r) \frac{d\Delta V(r)}{dr} dr. \tag{2.58}$$

For a simple case of truncated potential with an interatomic dispersion interaction $\Delta V = V^{disp} = -C_6 r^{-6}$, assuming that $g(r) = 1$ for $r \geq r_c$ and the number of particles within r_c is bigger than 1, the corrections will be [94]:

$$\begin{aligned} \Delta u &= -\tfrac{2\pi}{3}\rho C_6 r_c^{-3}, \\ \Delta P &= -2\pi\rho^2 C_6 r_c^{-3}. \end{aligned} \tag{2.59}$$

2.5.4 Long-range Coulomb interaction

As was shown before, the contribution of a truncated tail of the potential energy can be assessed with Eq. 2.57, from which it is clear, that the result converges only for cases when the energy function decays faster than r^{-3}. But the charge-charge interaction between ions ($V_{cc} \sim r^{-1}$) and the dipole-dipole interaction between molecules ($V_{dd} \sim r^{-3}$) do not satisfy this requirement. If the potential is truncated with the spherical surface, the resulting sphere around a given ion could be charged, since the number of anions and cations need not balance at any instant. The tendency of ions to migrate back and forth across the spherical surface would create artificial effects at $r = r_c$. This can be compensated by distributing a new charge over the surface of the sphere, equal in magnitude and opposite in sign to the net charge of the sphere, so as to guarantee local electroneutrality. This is rather like shifting the potential. However, some undesirable structural effects are inevitable. In contrast, the cut-off can be done with the cubic surface, but it is also not a solution of all problems, because similarly charged ions will tend to occupy positions in opposite corners of the cube: the periodic image structure will be imposed directly on what should be an isotropic liquid, and this results in a distortion of the liquid structure [95]. There are two methods which can be used to tackle the problem of long-range forces. The lattice methods, such as the Ewald sum, include the interaction of an ion or molecule with all its periodic images. These methods will tend to overemphasize the periodic nature of the model fluid. Another approach is the reaction field methods, which assumes that the interaction from molecules beyond a cut-off distance can be handled in an average way, using macroscopic electrostatics.

The Ewald sum is a technique for efficiently summing the interaction between a charged point and all its periodic images [130]. The potential energy can be written as:

$$V_{qq} = \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{\vec{n}}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j |\vec{r}_{ij} + \vec{n}|^{-1}\right), \tag{2.60}$$

where q_i, q_j are the charges, the sum over $\vec{n}$ is the sum over all simple cubic lattice points, $\vec{n} = (n_x L, n_y L, n_z L)$ with n_x, n_y, n_z integers. In the Ewald

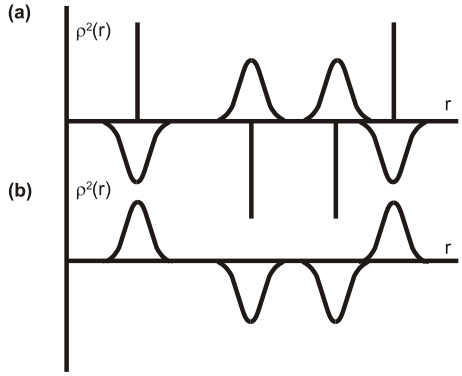

Figure 2.3: **Charge distribution in the Ewald sum**: (a) Original point charges plus screening distribution.(b) Cancelling distribution.

method, each point charge is surrounded by a charge distribution of equal magnitude and opposite sign, which spreads out radially from the charge. This distribution is conveniently taken to be Gaussian:

$$\rho_i^q(\vec{n}) = q_i \kappa^3 \frac{\exp(-\kappa^2 r^2)}{\pi^{\frac{3}{2}}}, \tag{2.61}$$

where κ determines the width of the distribution, and $\vec{r}$ is the position relative to the center of the distribution. This extra distribution acts like an ionic atmosphere, to screen the interaction between neighboring charges. The screened interactions are now short-ranged, and the total screened potential is calculated by summing over all the molecules in the central cube and all their images in the real space lattice of image boxes (Fig. 2.3a). A charge distribution of the same sign as the original charge, and the same shape as the distribution $\rho_i^q(\vec{r})$ is also added (Fig. 2.3b). This cancelling distribution reduces the overall potential to that due to the original set of charges [95]. The cancelling distribution is summed in reciprocal space. So the Fourier transforms of the cancelling distributions are added and the total is transformed back into real space. Thus, the final potential energy will contain a real space sum plus a reciprocal space sum minus a self-term plus the surface

term:

$$V_{qq} = \frac{1}{2}\frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{|\vec{n}|=0}^{\infty} q_i q_j \frac{\mathrm{erfc}(\kappa|\vec{r}_{ij}+\vec{n}|)}{|\vec{r}_{ij}+\vec{n}|} + \left(\frac{1}{\pi L^3}\right)\sum_{\vec{k}\neq 0} q_i q_j \frac{4\pi^2}{k^2}\exp\left(\frac{-k^2}{4\kappa^2}\right)\cos(\vec{k}\vec{r}_{ij})\right) - \frac{1}{4\pi\varepsilon_0}\frac{\kappa}{\pi^{\frac{1}{2}}}\sum_{i=1}^{N} q_i^2 + \frac{1}{4\pi\varepsilon_0}\frac{2\pi}{3L^3}\left|\sum_{i=1}^{N} q_i\vec{r}_i\right|^2 . \tag{2.62}$$

Here $\mathrm{erfc}(x)$ is the complementary error function:

$$\mathrm{erfc}(x) = (2/\pi^{\frac{1}{2}})\int_0^{\infty}\exp(-t^2)dt, \tag{2.63}$$

which falls to zero with increasing x. Particle-mesh Ewald (PME) is a method proposed by Darden [131, 132] to improve the performance of the reciprocal sum. Instead of directly summing wave vectors, the charges are assigned to a grid using cardinal B-spline interpolation. This grid is then transformed with a fast Fourier transformation (FFT) algorithm and the reciprocal energy term obtained by a single sum over the grid in k-space. The potential at the grid points is calculated by inverse transformation, and by using the interpolation factors we get the forces on each atom. The PME algorithm scales as $N\log(N)$, and is substantially faster than ordinary Ewald summation on medium to large systems (it scales as N^2). On very small systems it might still be better to use Ewald to avoid the overhead in setting up grids and transforms. The system studied in this thesis is big enough and we use PME.

The Particle-Particle Particle-Mesh (P^3M) methods of Hockney & Eastwood can also be used for the treatment of long range electrostatic interactions [133]. With this algorithm the charges of all particles are spread over a grid of dimensions (n_x, n_y, n_z) using a weighting function called the triangle-shaped charged distribution:

$$W(\vec{r}) = W(x)W(y)W(z) \tag{2.64}$$

$$W(\xi) = \begin{cases} \frac{3}{4} - \left(\frac{\xi}{h}\right)^2, & |\xi| \leq \frac{h}{2} \\ \frac{1}{2}\left(\frac{3}{2} - \frac{|\xi|}{h}\right)^2, & \frac{h}{2} < |\xi| < \frac{3h}{2} \\ 0, & \frac{3h}{2} \leq |\xi| \end{cases} \tag{2.65}$$

where ξ(x, y or z) is the distance to a grid point in the corresponding dimension. It is not easy to calculate the full long-range virial tensor with P^3M, but it is possible to obtain the trace. This means that the sum of the pressure components is correct (and therefore the isotropic pressure) but not necessarily the individual pressure components.

2.5.5 Virtual interaction sites

Some force fields use interaction sites that do not coincide with atomic positions. For example, if one wants to place a partial charge at the position of a "bond" midway between two atoms i and j: $\vec{r_{vs}} = \frac{1}{2}(\vec{r}_i + \vec{r}_j)$. Generally these sites are functions of **n** atomic positions:

$$\vec{r}_{vs} = f(\vec{r}_1, \vec{r}_2, ..., \vec{r}_n). \tag{2.66}$$

Virtual sites do not participate directly in the equations of motion and they have no mass. Also, they are reconstructed after every dynamic step. The force $\vec{F}$, acting on a virtual site, is distributed among the atoms with mass on which the site depends in a consistent way. The potential energy can be written as:

$$V = V(\vec{r}_{vs}, \vec{r}_1, \vec{r}_2, ..., \vec{r}_n) = V^*(\vec{r}_1, \vec{r}_1, ..., \vec{r}_n), \tag{2.67}$$

and the force on the particle i in this case is equal:

$$\vec{F}_i = -\frac{\partial V^*}{\partial \vec{r}_i} = -\frac{\partial V}{\partial \vec{r}_i} - \frac{\partial \vec{r}_{vs}}{\partial \vec{r}_i}\frac{\partial V}{\partial \vec{r}_{vs}} = \vec{F}_i^{direct} + \vec{F}_i', \tag{2.68}$$

where $\frac{\partial V^*}{\partial \vec{r}_i}$ is a matrix product, $\vec{F}_i^{direct}$ is the normal force acting on particle i due to the other particles and $\vec{F}_i'$ is the part of the force acting on the virtual site, which is distributed to the particle i [96]. The simplest way of construction of virtual sites is a linear combination of $\vec{r}_i$:

$$\vec{r}_{vs} = \sum_{i=1}^{N} w_i \vec{r}_i. \tag{2.69}$$

Then the force is redistributed using the same weight:

$$\vec{F}_i' = w_i \vec{F}_{vs}. \tag{2.70}$$

The weight w_i for atom i is expressed via the weights a_i, which usually corresponds to the masses of the atoms involved in the virtual sites:

$$w_i = a_i \left(\sum_{j=1}^{N} a_j \right)^{-1} \tag{2.71}$$

There are three commonly used options for setting the weight a_i:

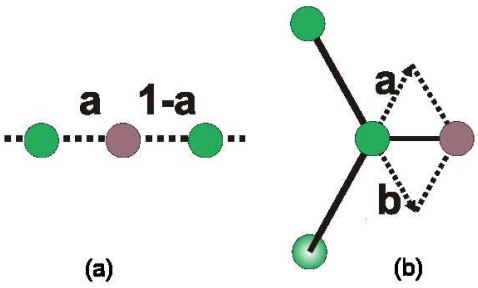

Figure 2.4: **Examples of virtual site geometric construction.** Green beads represent positions of atoms, brown bead represents the position of virtual site. Parameters a and b are used for obtaining the position of virtual site. In case shown on panel (a) the position of virtual site is calculated as $\vec{r}_{vs} = \vec{r}_i + a\vec{r}_j$, where $\vec{r}_i$ and $\vec{r}_j$ are radius-vectors of two atoms. In case shown on panel (b) the position of virtual site is calculated as $\vec{r}_{vs} = \vec{r}_i + a\vec{r}_{ij} + b\vec{r}_{ik}$, where $\vec{r}_{ij}$ and $\vec{r}_{ik}$ are vector between central atom i and two atoms on the sides.

- center of geometry (in case of equal weights);
- center of mass (a_i is the mass of atom i);
- center of weights: a_i is defined by the user.

In order to compare the conformational sampling of the CG model with the atomistic simulation, in this work we (re)introduced atomistic details into the CG structures, using virtual sites [134] corresponding to our CG units (backmapping process). In this work two types of virtual site construction were used (Fig. 2.4(a,b)). More details about backmapping process used in this work can be found in A.4.

2.6 Potential of Mean Force (PMF)

The potential of mean force (PMF) is the free energy of the system projected on a reaction coordinates. The simplest way to obtain it is by inverting the probability distribution of interest given that the system has been sampled extensively. But it is often impossible to measure a full distribution of degrees of freedom, due to sampling problems. In that case methods such as umbrella sampling or thermodynamic integration [135] can be used. In the present thesis, thermodynamic integration using distance constraints was used to compute the PMF between two molecules.

In that method one constrains the distance between two molecules and measures the average constrained force $< \vec{f}_c >$ required to keep this distance [136]. The mean force is then the negative of the constrained force. So integration of this force will give the free energy difference:

$$V(r_2) - V(r_1) = \int_{r_1}^{r_2} dr \left\langle -\vec{f}(r) \right\rangle = \int_{r_1}^{r_2} dr \left\langle \vec{f}_c(r) \right\rangle . \tag{2.72}$$

But it is necessary also take into account the entropic part of the force, which pulls apart noninteracting particles. This force is presented as:

$$-\frac{d}{dr}\left[-k_B T \log 4\pi \left(\frac{r}{r_0}\right)^2\right] = \frac{2k_B T}{r}. \tag{2.73}$$

Here $k_B \log 4\pi (r/r_0)^2$ is just the entropy of the phase space volume $4\pi r^2$. Putting at all together, the expression for potential of mean force reads as:

$$V_{PMF}(r) = \int_{r_m}^{r} dr' \left[\left\langle \vec{f}_c \right\rangle (r') + \frac{2k_B T}{r'}\right] + C, \tag{2.74}$$

where r_m is the maximum distance between the centers of mass of the molecules.

The method helps to find the potential of mean force for molecules, which appear in a very low concentration.

Chapter 3

Theoretical background: development of coarse-grained models

3.1 Introduction

Coarse-grained (CG) simulation models reduce the number of degrees of freedom compared to an atomistic representation, thus, decreasing the computational time for a given system and allowing investigations of larger systems and longer timescales. Various levels of coarse-graining can be employed, keeping only those degrees of freedom, which are deemed relevant for the particular problem of interest. In multiscale simulation approaches where atomistic and CG simulations are used in combination (either by switching barrier and force between the two levels or by using them simultaneously) special attention is being paid to compatibility of the two resolution levels. There are several approaches to develop CG models based on an atomistic reference simulation:

- structure-based;
- force matching;
- thermodynamic-property based (do not necessarily directly link to the atomistic data).

There are multiple strategies to develop CG models for biomolecules, peptides and proteins, and the methodology and the level of coarse-graining very much depend on the properties and processes which are to be investigated with the model. For biomolecules, such as peptides and proteins, a variety of

CG models have been developed in the past [48, 49, 55, 137, 138, 139, 140, 141], vastly differing depending on the investigated processes (i.e. protein folding or protein aggregation).

Early CG models were parametrized on generic properties of aminoacids such as hydrophobicity and hydrophilicity [139, 140, 142, 143, 144]. These models can be used to study protein folding mechanisms, in the simplest form as a coil-globule transition, under varying conditions such as temperature or pH, with different composition and distribution of the beads in the chain with molecular dynamics [143] or Monte Carlo [142, 144] simulations. Another early CG approach to protein folding are Go-like models [145, 146]. In this model, the significant attraction between beads corresponds to the native contacts and all non-native contacts are just repulsive. A different coarse-graining method for proteins is the Elastic Network Model (ENM) in which the peptide is represented as a network of beads connected by elastic springs. Such a model usually has one bead per amino acid [147]. A physics-based united-residue force field (UNRES model) also exists [148, 149]. Here, each amino acid residue is represented by only two interaction sites and a force field has been derived as a potential of mean force of the UNRES chain.

CG peptide models are frequently devised based on secondary structure arguments, leading to coarse-grained models with which one can successfully study folding of secondary structures. Those types of coarse-grained peptide models are determined independently from atomistic simulation models with parameters based on experimental data such as knowledge of secondary structure propensities of specific amino acids, hydrophobicity arguments, etc. [56, 52, 150, 151, 152, 153, 154].

One of the widely used CG approaches is a structure-based one, where CG potentials are refined until the CG model reproduces a set of structural properties of the reference system, like a set of reference distribution functions. Structure based CG methods are, for example, iterative Boltzmann inversion [67] and inverse Monte Carlo [66]. Force-matching (FM-CG) is another possibility to determine effective potentials in a CG model by minimizing the difference between the forces in the CG and in an atomistic reference simulation [68, 70, 73]. A third approach, which is used in the MARTINI force field [69, 74], determines CG interaction functions in such a way that thermodynamic properties, e.g. solvation free energies or partitioning data, are reproduced. Irrespective of the parametrization approach, be it based on structural or on thermodynamic arguments, the resulting CG models should preserve both structural and thermodynamic properties of the atomistic level in a multiscale simulation.

The above mentioned approaches using thermodynamic properties, structure or forces to parametrize the CG level have been applied to peptide sys-

tems as well [60, 61, 70, 74, 155, 156, 157].

In the present study we rely on the structure-based CG methodology. The procedure of finding potentials and choosing mapping scheme will be explained in Chapters 4 and 6.

3.2 Structure-based coarse-graining

The main idea of the structure-based coarse-graining method is to obtain an accurate reproduction of typical structural target functions, which are the intramolecular bond, angle and torsion distributions or pair-correlation functions. CG models can be obtained with the Boltzmann inversion or iterative Boltzmann inversion procedure [67, 158] (which was used in this thesis) or with inverse Monte Carlo [66]. In these approaches CG models are parametrized in such a way that structural properties reproduce those of atomistic simulations or known experimental data. The interaction potentials for the CG model are chosen according to this requirement. As opposed to CG force matching (FM-CG) method (where the aim is to reproduce the mean forces on the groups of atoms representing a CG bead by the CG simulation as closely as possible), one has a correspondence between the target DOF and the DOF as parameter in interaction in structure base methods potential.

3.2.1 Boltzmann inversion

The possibility to obtain the structure-based potentials relies on the assumption that particles interact with each other via central forces. A one-to-one correspondence exists between the potential and the distribution function of interest, such as bond length, bond angles and torsions. One should also note that according to the Henderson theorem [159, 160], the pairwise coarse-grained potential U(r) is unique up to an additive constant and exists [161, 162]. This means that all structure-based iterative methods must converge to the same coarse-grained potential, provided that their aim is to exactly reproduce pair correlation functions of the reference system. However, in practice, there can be several rather different potentials which reproduce a given structure within small errors. In order to accurately obtain a potential from these distribution they have to be weighted by the corresponding Jacobian which stems from transformation from spherical to Cartesian coordinates (note that at this point we do not discuss interdependencies between different degrees of freedom, these aspects will be discussed later):

- bond distributions:
$$g(r,T) = \frac{P(r,T)}{4\pi r^2}, \tag{3.1}$$
- angle distributions:
$$g(\theta,T) = \frac{P(\theta,T)}{\sin\theta}, \tag{3.2}$$
- torsion distributions:
$$g(\varphi,T) = P(\varphi,T). \tag{3.3}$$
- structure functions, e.g. radial distributional function of one particle with respect to another one:
$$g_{12}(r) = \frac{V}{4\pi r^2}\sum_{i}^{N_1}\sum_{j}^{N_2} P(r) \tag{3.4}$$

These weighted distributions should be then Boltzmann-inverted to obtain intramolecular potentials between the CG beads:

$$U(\xi) = -k_B T \ln g(\xi), \tag{3.5}$$

where ξ is any parameter of interest, k_B is the Boltzmann constant, and T is the temperature. Note that the weighted distributions depend on the temperature, i.e. the potentials (PMF) are temperature dependent.

In principle, this potential of mean force could be used directly to perform coarse-grained simulation. However, because the correction terms for higher density (in case of radial distribution function) are not included in this approach and often potentials for different pairs of particle types are needed that mutually influence each other, and because the atomistic system may be inhomogeneous, this potential of mean force can serve only as initial guess and can be improved upon in an iterative scheme.

Iterative Boltzmann inversion

Iterative Boltzmann inversion (IBI) [67] is a natural extension of the Boltzmann inversion method (Eq. 3.5). Since the goal of the coarse-grained model is to reproduce the distribution functions of the reference system as accurately as possible, one can iteratively refine the coarse-grained potentials using a numerical scheme. The method starts from an initial guess for the nonbonded potential Eq. 3.5. This potential will not usually reproduce the original distribution correctly due to correlations between DOFs or coupling to the environment. The first guessed potential is used for a CG simulation

which yields a distribution $g_0(\xi)$ that is different from target one $g_{target}(\xi)$. Thus, the potential is then modified by a correction term ΔU:

$$U^{CG}_{NB,i+1} = U^{CG}_{NB,i} + \Delta U = U^{CG}_{NB,i} + k_B T \ln \frac{g_i(\xi)}{g_{target}(\xi)}. \tag{3.6}$$

This corrected potential is again used for a CG simulation and the process of potential modification is repeated as many times as necessary until $g_{target}(r)$ is reproduced. Some times this formula is used with a prefactor $0 < \lambda \leq 1$ for the correction term ΔU to avoid overshooting in the numerical scheme.

3.2.2 Bonded interactions

In order to determine CG interaction potentials, one commonly distinguishes between bonded/covalent intramolecular interactions and nonbonded ones. These are determined separately based on the assumption that the total potential energy can be separated into bonded/covalent and nonbonded contributions [163]:

$$U^{CG} = U^{CG}_B + U^{CG}_{NB} \tag{3.7}$$

This separation, which is not used in all structure-based CG approaches [164], has the advantage of a certain modularity. The (local) intramolecular interaction potentials obtained in the present study could in principle be adapted to different intermolecular interaction functions, either from structure-based CG approaches or from other thermodynamics-based models.

In structure-based CG approaches, which are based on atomistic simulations, one first determines the CG beads positions based on the atomistic sampling. With these, one defines the (CG) bond, angle and dihedral distributions which correspond to the atomistic simulation. Via Boltzmann inversion, one then computes the corresponding potentials:

$$\begin{aligned} U^{CG}(r,T) &= -k_B T \ln(P(r,T)/4\pi r^2) + \text{const}_r \\ U^{CG}(\theta,T) &= -k_B T \ln(P(\theta,T)/\sin\theta) + \text{const}_\theta \\ U^{CG}(\varphi,T) &= -k_B T \ln(P(\varphi,T)) + \text{const}_\varphi \end{aligned} \tag{3.8}$$

This relies on the assumption that the probability distribution describing the conformations factorize into bond, angle and torsion contributions:

$$P^{CG}(r,\theta,\varphi,T) = P^{CG}(r,T)P^{CG}(\theta,T)P^{CG}(\varphi,T) \tag{3.9}$$

This assumption is problematic for systems where degrees of freedom are correlated, for example in biological systems with distinct secondary structures, intramolecular hydrogen bonding, etc. It is tested for our model in the Chapter 4.

As will be shown in Chapter 4, the procedure outlined above does not yield a CG model for tri-alanine which satisfactorily reproduces the atomisticlly sampled conformational equilibrium. For this reason, two corrections were added: a pressure correction [165, 166] was applied to the water-water (nonbonded) interaction potential (sec. 3.2.3) and an iterative correction was applied to some of the bonded interaction potentials which is detailed in the following.

Refining bonded potentials

Eq. 3.8 shows how CG bond, angle and dihedral potentials are obtained by Boltzmann inverting distributions from an atomistic simulation. In the present case, it was found that the resulting CG model (i.e. the combination of all covalent potentials and nonbonded interactions, also with the solvent) does not exactly reproduce the distributions from the atomistic simulation for all degrees of freedom. For this reason, an additional refining step, which is completely analogous to the iterative procedure for nonbonded interactions, was introduced (see Eq. 3.6). In case of the distribution of an angular DOF θ, $g_{target}(\xi) = P^{ref}(\theta, T)$ is the reference angular distribution from atomistic simulation and $g_i(\xi) = P_i(\theta, T)$ is the distribution after the ith iteration. For example P_n.

3.2.3 Nonbonded interactions

Nonbonded interaction potentials can be obtained in different ways. In case of structure-based coarse-graining, the aim is to find potentials which reproduce the structure of the investigated system. Usually, the radial distribution functions (RDF) are measured. Only pair potentials are used to describe nonbonded interactions to avoid computationally expensive multi-body interactions.

Pressure correction of nonbonded potentials

In Ref. [60] the iterative Boltzmann inversion procedure has been used to determine the nonbonded interaction potential for the water-water interaction. In iterative Boltzmann inversion Eq. 3.6, a CG potential is self-consistently refined until a structural property, in this case the radial distribution function obtained from atomistic simulations, is reproduced in the CG simulation. Here, $g^{target}(\xi)$ is the radial distribution function (RDF) of the atomistic simulation, and $g_i(\xi)$ is the RDF at iteration number i. The resulting structure based on the CG water model exhibits an artificially high pressure. As shown

in Chapter 4 below, it was investigated if this high solvent pressure causes artifacts in the conformational sampling of the CG peptide model. To this end, a pressure correction was applied to the CG water-water interaction as described in Refs. [165] and [166]. Originally, pressure correction implements with a linear correction method [67] at the cost of losing some of the accuracy in the RDF:

$$\Delta V(r) = A\left(1 - \frac{r}{r_{cut}}\right), \tag{3.10}$$

where $V(r)$ is a potential, A is a constant equals to $-0.1k_BT$. But Wang et.al. [166] determine this constant in a different way:

$$-\left[\frac{2\pi N\rho}{3r_{cut}}\int_0^{r_{cut}} r^3 g_i(r)\mathrm{d}r\right] A_i \approx (P_i - P_{target})\, V, \tag{3.11}$$

where pressure P_i is given via the virial expression through potential $V_i(r)$ and RDF $g_i(r)$ of the ith step the iteration:

$$P_iV = Nk_BT - \frac{2}{3}\pi N\rho\int_0^{+\infty} r^3\frac{\mathrm{d}V_i(r)}{\mathrm{d}r}g_i(r)\mathrm{d}r, \tag{3.12}$$

and the target pressure P_{target}, which is correct and to which we want to match the pressure in the system, is determined as:

$$P_{target}V \approx Nk_BT - \frac{2}{3}\pi N\rho\int_0^{+\infty} r^3\frac{\mathrm{d}}{\mathrm{d}r}\left[V_i(r) + \Delta V_i(r)\right] g_i(r)\mathrm{d}r. \tag{3.13}$$

But it should be mentioned, that the pressure correction changes a potential in a such way, that the isothermal compressibility may deviate significantly from one of the all atom model [166].

The iterative procedure for this thesis was performed using the VOTCA package [75].

3.3 Nonbonded interactions in low concentrated solute solvent systems

When solvent degrees of freedom are not explicitly present in the CG system, potentials inverted from atomistic distributions can be used directly as effective bead-bead nonbonded interactions (in this thesis solute refers to the peptide beads). But when water explicitly exists in the CG system, a new effective bead-bead nonbonded interactions are needed. For solvent-solvent and solute-solvent interactions the normal iterative Boltzmann inversion procedure can be used. But CG solute-solute nonbonded interactions are affected

by solvent, however, they should ideally reproduce the thermodynamic association behavior of the atomistic system. For this, the influence from solvent should be removed from solute-solute nonbonded potentials.

First, the potential of mean forces V_{PMF}^{atom} for atomistic simulation should be calculated. Since in dilute systems this cannot be easily obtained from inverting of a distribution function due to sampling, enhanced sampling methods such as thermodynamic integration or umbrella sampling can be used (see Chapter 6). Next step is to remove the solvent influence from it. For this constraint simulations for peptide-peptide interaction in water (or another solvent) between CG beads are carred out, where direct interaction between peptide beads is excluded [60]. In this case potential $V_{PMF,excl}^{CG}$ contains only influence from solvent. The effective bead-bead nonbonded potential, can then be obtained:

$$V_{eff}(r) = V_{PMF}^{atom}(r) - V_{PMF,excl}^{CG}(r), \tag{3.14}$$

where each potential $V_{PMF}(r)$ is calculated with Eq. 2.74.

By construction the resulting V_{eff} reproduces the target PMF in the CG model. This procedure of determination of nonbonded interaction will be discussed in details in Sec. 6. For concentration transferability issues concerning these potentials see Villa et.al. [167].

3.4 Back mapping

A method to reintroduce atomistic details into a CG structure is called "back mapping", or "inverse mapping". Every CG structure corresponds to an ensemble of atomistic configurations. Therefore, there is no unique way for finding a set of atomistic coordinates, which will correspond to the target CG structure. Thus, the aim is to find a structure from the ensemble of all possible ones, which displays the correct statistical weight of those degrees of freedom not resolved in the CG description.

As was already mentioned, the limitation of computer resources does not allow for long time-scale and large length-scale simulations, but a combination of CG and back mapping can solve this problem as well as help to obtain well-equilibrated atomistic structures and trajectories. These results can then be directly compared to experimental data, for example, data from NMR or X-ray diffraction measurements [155, 168].

For polymeric melts, it is possible to obtain backmapped atomistic structures by taking rigid all-atom chain fragments from a correctly sampled atomistic distribution (these chain fragments can correspond to a single chain or a small set of CG beads). These fragments should be fitted onto the CG

structure and the resulting all-atom structure should be relaxed by energy minimization and a short equilibration [155, 157, 168, 169].

When polymer chain is too flexible to use pre-sampled atomistic structures or when molecules have a too low molecular weight such that the atomistic structure already diffuses away from the CG target coordinates during short equilibration runs, the same strategy can not be applied. (For example, the azobenzene liquid crystal [72]). For such system it is better to restrain the atomistic coordinates in a way that they always satisfy the CG mapping scheme. Thus, not single atoms, but groups of atoms which define a CG mapping point together, are restrained [170]. This method also allows the insertion of flexible chain units at random initial atomistic positions. The resulting structures are relaxed while restrained to the CG mapping points and can be easily equilibrated.

In this work, back mapping was used to see whether it is possible to get back all atomistic properties, which were lost during CG simulations, see section 4.5.

3.5 Dynamics of coarse-grained peptide simulations. Time scales.

As was mentioned above this thesis considers a coarse-grained model which is parametrized in order to reproduce structural properties of atomistic peptides. The length scaling between atomistic and CG models is automatically given by our mapping scheme. But for dynamics it is not so easy to make a link between time scales of the two systems directly from the mapping scheme. CG simulations are faster than corresponding atomistic systems, which is caused by several reasons:

- The CG model contains a reduced number of degrees of freedom (DOF) compared to the atomistic one;
- The interaction potentials are smoother in CG models;
- The effective bead friction is reduced due to smoother energy landscape or/and smaller energy barriers [171].

The first item influences only on a “speed-up” in computational costs. But this point does not directly affect dynamical properties. The second item has two consequences. First, the potentials are often softer, leading to the possibility to use larger time steps. The smoother potentials cause the reduced friction, which is very important for the investigation of the dynamical

properties of the system. But it is not immediately clear if all energy barriers are lower in the same way for all components of the CG model and whether the average friction coefficient will describe motion of the system correctly for all components, or not.

For the polymer models such as Rouse and Zimm the connection between friction coefficient and diffusion coefficient are $D \propto N^{-1}\zeta^{-1}$ and $D \propto \zeta^{-1} \propto N^{-1/2}$ [172] (the Rouse model describes the simple polymer behavior, while the Zimm model takes into account the hydrodynamic interactions [172]). As was said the friction coefficient of the CG systems ζ^{CG} is reduced compared to the atomistic one ζ^{atom} due to a fact that effective CG potentials are softer than atomistic potentials. Thereby the atomistic value for the friction and thus the dynamic properties are closer to the experimental results [173] than the same data in CG simulation.

In order to get the quantitative prediction of polymer dynamics for CG model, the time scale has to be found in such a way, that CG results can be fitted to atomistic or experimental one. Possible methods to do it are already widely discussed and used in polymer research [64, 65, 69, 168, 174, 175, 176].

Usually the time scaling factor is determined as:

$$\tau_D = \frac{D^{CG}}{D^{atom}} = \frac{\zeta^{atom}}{\zeta^{CG}}. \tag{3.15}$$

It provides the link between dynamical properties in atomistic and CG simulations. In practice one should calculate the mean square displacement (MSD) from both simulations. It will give the results for diffusion coefficient according to Einstein relation:

$$D = \lim_{t\to\infty} \frac{\left\langle (R(t) - R(0))^2 \right\rangle}{6t}, \tag{3.16}$$

or

$$\left\langle (R(t) - R(0))^2 \right\rangle \propto 6Dt, \tag{3.17}$$

which means, that one can make a plot of MSD vs time t (the slope will correspond to the diffusion coefficient) for both simulations and then shift CG curve along the time axis in order to fit it to the atomistic one. The necessary shift will correspond to the scaling factor τ_D.

An analytical prediction of time scales in CG simulation is usually not possible. In principle, one can determine a time scale through the average mass m_{CG} of the CG beads, the known length scale σ_{CG} and strength of the interaction parameter in the nonbonded excluded volume interaction (in case, when a Leonnard-Jones interactions used) ε_{CG}:

$$1\tau_{CG} = 1\left(\frac{m_{CG}\sigma_{CG}^2}{\varepsilon_{CG}}\right)^{\frac{1}{2}}, \tag{3.18}$$

but his value is not equal to the time scale obtained directly from MSD measure τ_D quantity. The main reason for the difference between two results is the fact, that Eq. 3.18 does not take into account a change of friction coefficient in CG simulation [64, 177, 171].

Nevertheless, Depa and Maranas suggested a way to obtain a value for the time scaling factor in CG simulation [175]. They suggested a method on the basis of an assumption that an acceleration in CG system is caused by a reduction of the nearest neighbors. A surrounding area of the particle is named "cage" and they suppose that it is more difficult to escape from this cage for atomistic particles then for CG ones, because at the second case the CG nonbonded potentials are less attractive. Depa and Maranas decided to apply an argument from hyperdynamics method [178] in order to describe a transition of the particle from one local cage to another. Originally the hyperdynamics method allows to study transition between two minima of the potential energy surface, which happened rarely and is not easy to investigate with pure MD simulations. In case of using this method the simulations are accelerated by modifying the potential energy surface with a "bias" potential. Provided that all potential are agreeable to the certain standards, the method gives a computational speed up with help of time-averaged "boost factor":

$$\frac{t_{hyper}}{t_{MD}} = \left\langle \exp\left(\frac{\Delta V(r_N)}{k_B T}\right)\right\rangle_{\Delta t}, \tag{3.19}$$

where $\Delta V(r_N)$ is a bias potential and r_N is the N-particle configuration. For the calculation of CG speed up in compare to atomistic simulation they used the same equation, just now the CG boost is time invariant, and the time averaging is not required:

$$\frac{t_{CG}}{t_{atom}} = \exp\left(\frac{\langle \Delta V\rangle}{k_B T}\right), \tag{3.20}$$

where $\langle \Delta V\rangle$ is an ensemble averaged bias potential per particle:

$$\langle \Delta V\rangle = \int_0^\infty \langle N_{CG}(r)\rangle\, U_{CG}(r) dr - \int_0^\infty \langle N_{atom}(r)\rangle\, U_{atom}(r) dr \tag{3.21}$$

Here, the atomistic and CG potential are weighted with the ensemble averaged number density of neighbors for each particle, $\langle N_x(r)\rangle$:

$$\langle N_x(r)\rangle = \frac{\langle P_x(r)\rangle}{n_x}, \tag{3.22}$$

where P_x is the ensemble averaged number density of intermolecular Lennard-Jones pairs, and n_x is the total number of particles. As shown by Depa and

Maranas [175] the agreement of the time scaling factors obtained from fitting of two MSD and from this method is quite good with deviation in results about 7% for polyethylene.

In this thesis the time scaling factor is considered for the peptide with two types of beads in water. However, the method described above is not investigated for such systems and also it is not clear if it can be used for polymers with extended side groups, such as proteins. And as far as we want to make an universal CG model, which later one can use for proteins as well, we did not use the prediction method and in order to obtain the time scales we just made a comparison of the diffusion coefficients for atomistic and CG systems. In addition, different dynamics processes might have different time scaling factors, what we also will discuss in Chapter 7. As far as we wanted to consider system of peptide and solvent (water), we used the Zimm theory for our model, but we did not check characteristic scaling factors for the system, because this asymptotic scaling does not apply to our small peptides.

Chapter 4

A first coarse-grained model for oligoalanine

In this chapter we will present the steps needed to set up a coarse-grained model based on an atomistic one. In particular it will be shown which additional refinements were necessary in the particular case of oligo-alanine peptides in aqueous solution compared to previous, synthetic polymer systems. It will be explained how in an inverse mapping procedure atomistic coordinates can be obtained for structures from a coarse-grained (CG) simulation. Presently, local aspects are considered, i.e., intramolecular CG potentials are determined in such a way that the local chain conformations are in good agreement with the sampling of an atomistic model (analogous to atomistic studies when potentials for backbone torsion angles are determined [179], i.e., non-local intramolecular interactions such as hydrogen bonds that lead to helix formation are added at a later stage). This is important to ensure that the CG model samples conformations that are fully compatible with an original atomistic sampling after reinserting atomistic coordinates. To test the sampling one can, for example, consider the Ramachandran angles, which indicate correlations between adjacent torsional degrees of freedom. It will be shown to which extent the resulting CG potentials for bonds, angles and torsions are transferable to longer sequences as well as to which extent the interplay of intramolecular bonded interactions and intermolecular interactions with a surrounding medium needs to be accounted for in the parametrization process.

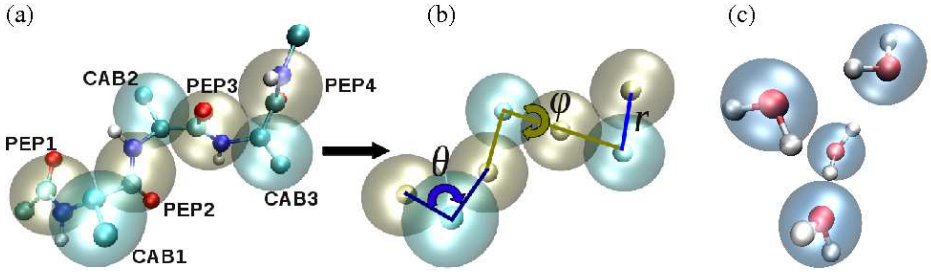

Figure 4.1: **CG representation of the system.**(a): Relation between atomistic (united atom, small beads) and CG model (large transparent beads) of Ala_3 with their respective CG bead names used throughout the manuscript. (b): CG internal degrees of freedom (small spheres indicate the centers of the CG beads); (c) Representation of CG model for water.

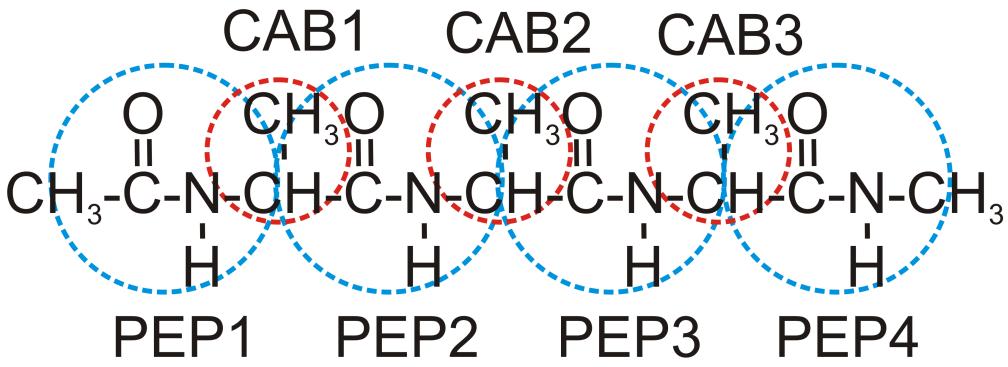

Figure 4.2: **Chemical structure of capped-alanine oligomer considered in this thesis, with a scheme of CG beads.**

4.1 Mapping scheme

The mapping scheme is the relation between the atomistic and the CG degrees of freedom as was discussed in Chapter 3. The one of the capped oligo-alanine peptides studied here is illustrated in Fig. 4.1(a). The CG peptide consists of a linear chain with two types of spherical beads, representing the α and β carbon united atoms "CAB" and the peptide group "PEP" (see Fig. 4.2). This mapping scheme is an analogue of the one in Ref. [60] for di-phenylalanine, with the difference that alanine residues require no side chain beads, since the C_β united atom is already represented in the "CAB" bead. The CG Ala_3 peptide consists of 7 beads in total. They are uncharged, the positions are given by the centers of mass of the corresponding

constituting atoms which also determine the bead masses.

While the atomistic simulations were performed with an explicit water model, two different types of water representations were considered for the CG system: an implicit water representation, where the effect of the water molecules on the peptide is accounted for via the CG interactions of the peptide beads [60], and an explicit CG water model [61]. In the latter case each water molecule was represented by one CG bead (Fig.4.1(c)).

4.2 Nonbonded interactions

In the present chapter the focus is on conformational sampling of a single peptide in atomistic and CG representations, not on the association behavior of several peptides, however, in Sec. 6 we also consider a system with two ALA_2 peptides in the water in order to investigate a intermolecular interactions. For this reason, at first nonbonded peptide-peptide and peptide-water interactions were not newly parametrized, but taken from Ref. [60, 61], where A. Villa et al. suggested a structure-based CG model for diphenylalanine, with bonded potentials according to a Boltzmann inversion procedure and nonbonded potentials with calculation of potentials of mean force. The C_{alpha} - C_{beta} ("CAB") bead's parameters correspond to the "C$\alpha\beta$" bead's parameters in Ref. [60], which had been parametrized based on methane, while the peptide ("PEP") bead's parameters had been divided based on N-Methyl-Acetamide, Fig. 6.4(a), (note that this bead can therefore also be used for the capped ends since it incorporates contributions from the capping). CG non-bonded interactions within the peptide chain were only applied between beads further than four bonds apart, all interactions between beads closer to each other along the peptide chain were handled by bond, angle and torsion potentials. As discussed in detail in Ref. [61], no additional exclusions for long range nonbonded interactions along the peptide chain in the atomistic sampling of the reference distributions were applied during parametrization of the CG bonded interactions. In Sec. 6 the parametrization of non-bonded potentials will be discussed in details.

4.3 Bonded interactions

As mentioned above, Ala_3 is represented by 7 CG beads, see Fig. 4.1(b), connected by 6 bonds, 5 angles and 4 dihedral angles. Each of them was modeled by a separate potential obtained through Boltzmann inversion according to Eq. 3.8. It should be noted, however, that the respective equiva-

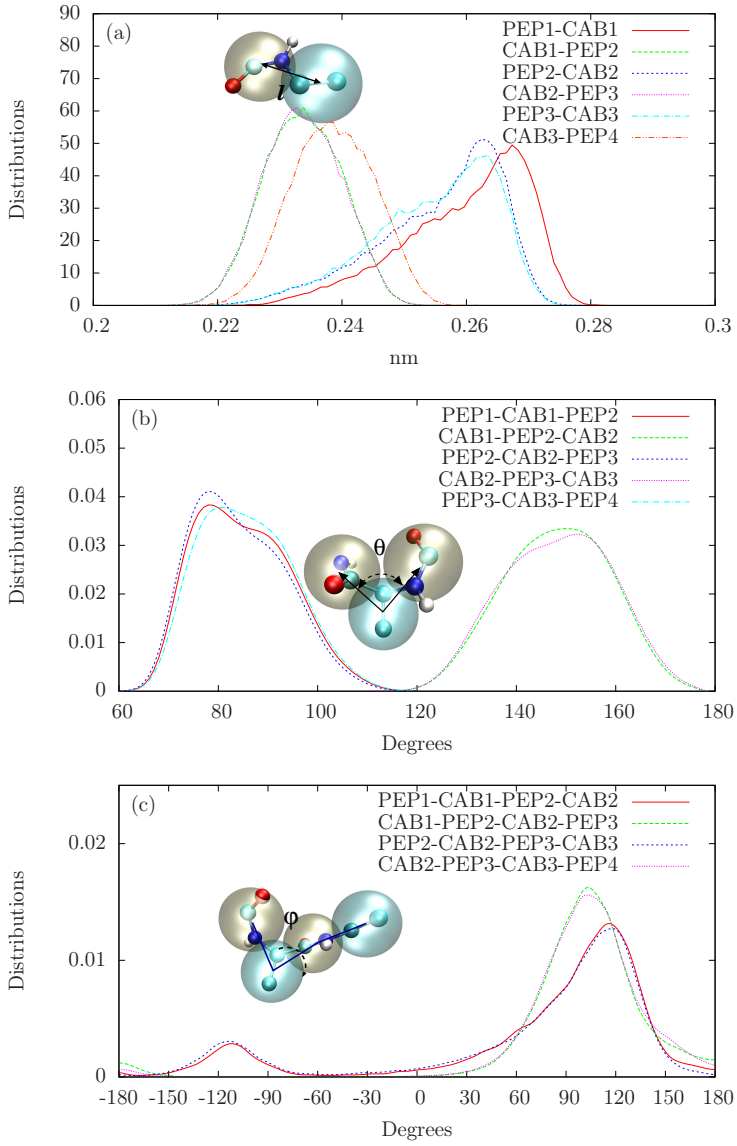

Figure 4.3: **Intramolecular distributions obtained from atomistic simulations and presented in CG degrees of freedom.** (a) Bond distributions; (b) Angle distributions; (c) Dihedral distributions. The names of the beads in the legend correspond to the names on the scheme 4.1a. Pictures illustrate only one of the two types of each DOF: (a) PEP-CAB; (b) PEP-CAB-PEP; (c) PEP-CAB-PEP-CAB.

lent bonds, angles and torsions at different positions along the peptide chain exhibit very similar distributions, with only minor differences when a chain end is involved. This can be seen in Fig. 4.3. Fig. 4.3(a) represents the bond distributions between the neighbor beads. Here are two main groups of distributions corresponding to the PEP-CAB and CAB-PEP bonds, where only

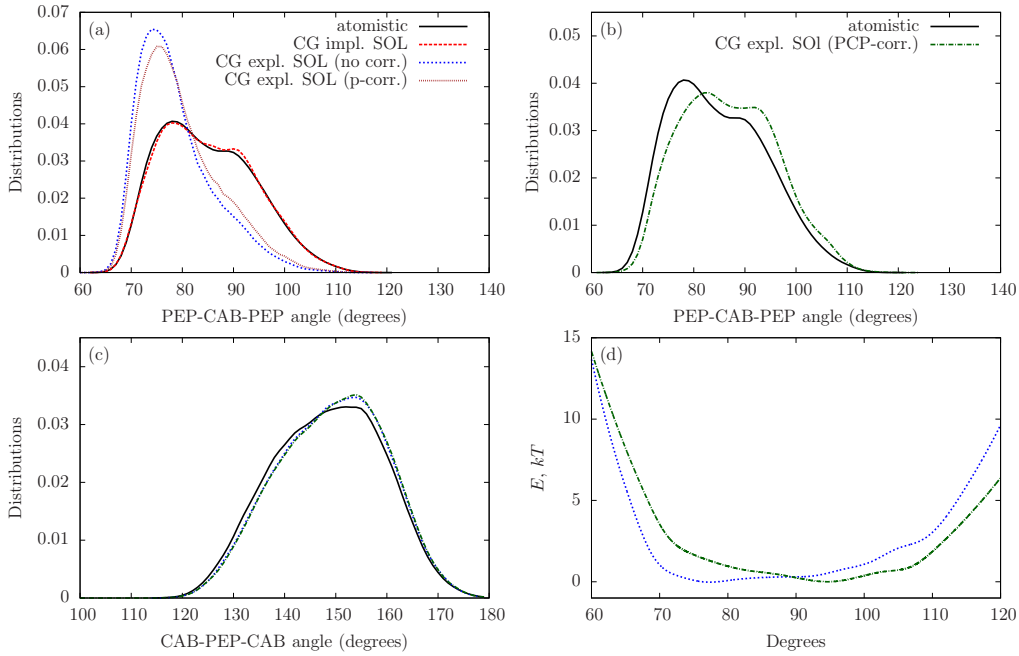

Figure 4.4: **Distributions.** Panels a,b: Comparison of PEP-CAB-PEP angle distributions from simulations with different tabulated potentials (atomistic: black line; CG model with implicit solvent: red line; CG model with explicit solvent without correction: blue line; CG with pressure-corrected solvent: brown line; CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent: green line); panel c: distribution for CAB-PEP-CAB angle; panel d: tabulated potentials for PEP-CAB-PEP angle (uncorrected potential: blue line; potential after 4 iterations: green line).

distributions for first PEP1-CAB1 and last CAB3-PEP4 bonds are shifted a little bit towards bigger distances. This difference for the end beads results from their position in the chain with only one neighbor. Panels (b) and (c) show angle and torsion distributions correspondingly. In these cases all curves for each type of angles are approximately the same.

At first, we set out to confirm that the atomistic distributions used for parametrization were reproduced by the CG peptide model after adding up all interactions (bonded and nonbonded ones). Indeed, our results showed that in both implicit solvent and explicit solvent models all bonds and the two angles between the CAB-PEP-CAB groups, see Fig. 4.4(c), were very well represented in the sampling of the CG peptide. Thus, these degrees of freedom (DOF) are not too sensitive to the environment. However, for the three PEP-CAB-PEP angles, the initial agreement was not as good. Fig. 4.4(a) shows that for a PEP-CAB-PEP angle, the model with the potentials ob-

tained from non-iterative Boltzmann inversion in an implicit solvent representation (red line) reproduces the atomistic distribution (black line) very well. The corresponding explicit solvent model, however, (with a CG water representation, blue line) fails to reproduce the atomistic reference distribution. This observation had not been made previously for di-phenylalanine, possibly because the peptide has an explicit side chain bead, which – due to the additional angular potentials and the excluded volume interactions with other chain segments – prevents distortions at the CAB bead which more easily occur in the linear oligo-Ala segment. In order to get a better agreement for the PEP-CAB-PEP angle, two types of refinements were applied.

Since the original iterative Boltzmann inverted water-water interaction exhibits a comparatively high pressure, we suspected that this high pressure causes a distortion of the peptide conformations. Thus, we applied a pressure correction to the water-water interaction potential, such that it reproduces atmospheric pressure conditions. As can be seen from Fig. 4.4(a), this had only a small impact on the conformational sampling of the peptide (brown line).

Furthermore, we iteratively refined the corresponding potential for the PEP-CAB-PEP angle (see Sec. 3.2.2). As a target of this iterative Boltzmann inversion (Eq. 3.8), we used the distribution of the PEP1-CAB1-PEP2 angle obtained from atomistic simulation of Ala_3 (note that as mentioned above the atomistic distributions of all PEP-CAB-PEP angles are almost identical). For the short simulations in the iterative procedure, we did not use the entire CG Ala_3 peptide, but a shorter PEP-CAB-PEP fragment in CG solvent. This was done not only to save computational time (by not having to equilibrate the entire peptide for each iteration step) but also to separate the respective angle from other degrees of freedom in the peptide (of course one then has to test, if the resulting potential gives the correct distributions of the respective angle in the peptide chain with all other interactions present). After 4 iteration steps, we obtained a new CG PEP-CAB-PEP potential shown on Fig. 4.4(d), which was then applied to the three respective angles in Ala_3. The new CG model (with pressure corrected solvent and refined PEP-CAB-PEP angle) reproduces the angular distribution of the atomistic peptide as shown in Fig. 4.4(a). The distributions of the bonds and the CAB-PEP-CAB angles are not affected by the change in the PEP-CAB-PEP potential, see Fig. 4.4(c).

It is important to mention that this new tabulated potential also slightly improved the sampling of the PEP-CAB-PEP-CAB dihedral angles (Fig. 4.5). As one can see, the black (atomistic case) with the green (CG case with pressure corrected water and refined PEP-CAB-PEP angle) curves show better agreement than the other two corresponding to the simulation cases without

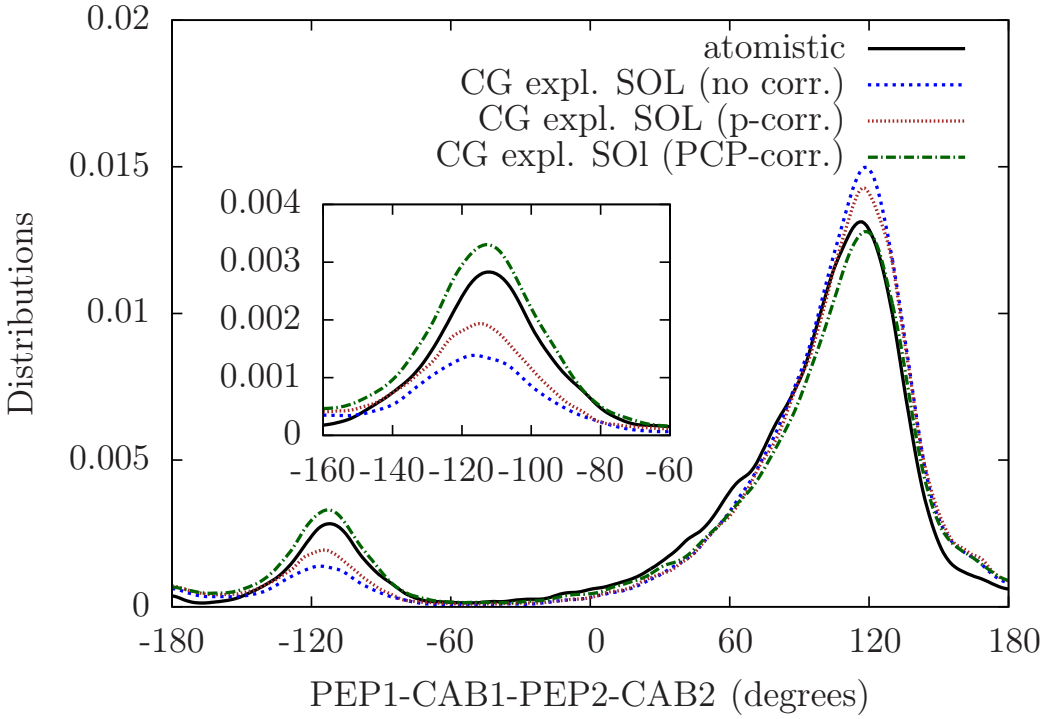

Figure 4.5: **Distributions of PEP1-CAB1-PEP2-CAB2 dihedral angle from simulations with different tabulated potentials** (atomistic: black line; CG with explicit solvent without correction: blue line; CG with pressure-corrected solvent: brown line; CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent: green line).

any corrections (blue line) and to the calculation with only pressure correction (brown line). Two analogous dihedral angles PEP1-CAB1-PEP2-CAB2 and PEP2-CAB2-PEP3-CAB3 show very similar distributions, and in both cases the corrected CG model and the atomistic simulation agree very good, as well as the agreement between two CAB-PEP-CAB-PEP dihedral angles.

The fact that the correction to the PEP-CAB-PEP angle also removes a discrepancy in the sampling of a "neighboring" dihedral angle shows that it is potentially problematic to "blindly" iterate all potentials at once, since the degrees of freedom are coupled.

4.4 Analysis of other structural properties

As an independent quantity which had not directly entered into the parametrization process, the end-to-end distance (between PEP1 and PEP4) was chosen to assess the agreement of the overall conformational sampling of the CG peptide with the atomistic one.

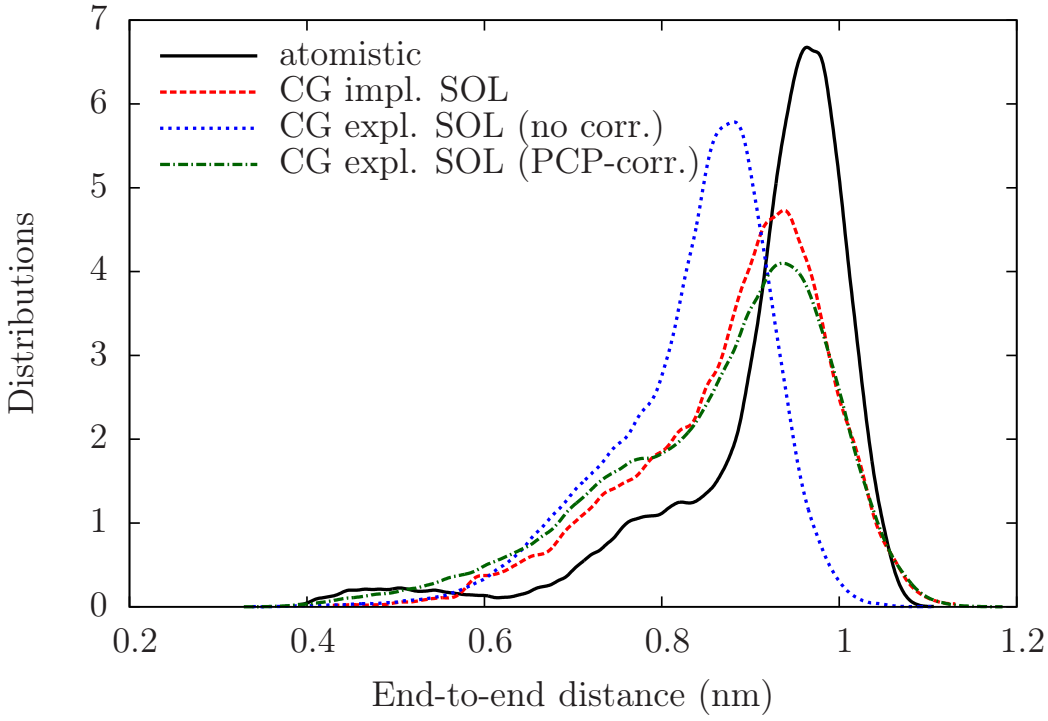

Figure 4.6: **End-to-end distance distribution of Ala_3** (i.e distance between CG centers PEP1 and PEP4); atomistic: black line; implicit solvent CG model: red line; CG with explicit solvent without correction: blue line; CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent: green line.

Fig. 4.6 shows the end-to-end distance in the atomistic sampling (black line), the implicit solvent CG model (red line), the uncorrected CG model with CG solvent (blue line), and the CG model with pressure and angle corrections (green line). One can see that (just as for the PEP-CAB-PEP angular distributions, which is also the main cause of the shift) the uncorrected explicit solvent model does not agree well with the atomistic reference, the peptide appears to be more compressed. The implicit solvent model and the explicit solvent model with corrections show a better (but still not optimal) agreement with the end-to-end distance distributions in the atomistic reference. There is still a shift towards smaller distances, possibly due to the nonbonded interactions within the peptide chain, that will be further investigated in Chapter 5.

The values of the mean end-to-end distances for atomistic and CG cases are very similar: 0.91 nm and 0.86 nm (see Tab. 5.1), the difference is only about 5%, but each area of the distribution corresponds to some physical state which is important to reproduce in CG model as well. In addition, for longer chains the discrepancy in the distributions and mean values is

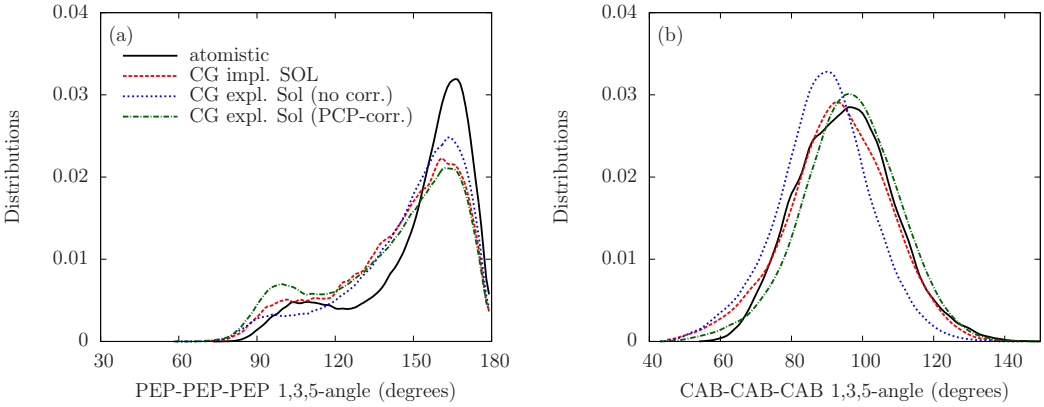

Figure 4.7: **Distributions.** Panel a: Distribution of the average of PEP-PEP-PEP 1,3,5-angle (atomistic: black line; implicit solvent CG model: red line; CG with explicit solvent without correction: green line; CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent: blue line). Panel b: Distribution of the average of CAB-CAB-CAB 2,4,6-angle (atomistic: black line; implicit solvent CG model: red line; CG with explicit solvent without correction: green line; CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent: blue line)

likely to become bigger. Thus, it is important to solve the problem with the compressed conformation of CG peptide.

Another sign of the compressed conformations is the distribution of the angle between adjacent peptide groups along the chain (e.g. the PEP-PEP-PEP 1,3,5-angle, see Fig. 4.7(a)) which is somewhat overemphasizing smaller angles. At the same time, the angle between adjacent CAB groups is nicely reproduced as presented on Fig. 4.7(b). Here we observe also an improvement in the CG model with pressure corrected solvent and corrections to the PEP-CAB-PEP angle.

Thereby, we found out that the direct use of inverted potentials for bonds, angles and torsions does not produce a CG model which can reproduce the atomistic conformational sampling sufficiently close. In order to obtain a better agreement, corrections of the model by iteration of specific DOF can be added. However, local improvement of the bonded potentials does not necessarily solve the problem on a global level, as it cannot fit the end-to-end distance distribution of CG system to the atomistic one. There are three possible interpretation of the above results for our model: 1) assumption that CG degrees of freedom are uncorrelated is not correct; 2) the nonbonded interactions are not correct; 3) the assumption that noncorrected bonded potentials (bonds, angles and torsions) are sufficient to describe the peptide's behavior and reproduce atomistic conformations is not correct. The possible ways to settle it will be discussed in Sec. 4.5, Sec. 5 and 6.

4.5 Analyzing the backmapped structures

Despite the previosly discissed problems, it is interesting to investigate if it is possible to get back atomistic properties from our CG model. As mentioned above, our CG model does not reproduce all atomistic distributions exactly. We check, if we can reconstruct all atomistic characteristics well with this CG model. The details of the backmapping procedure are described in Sec. A.4.

The parametrization procedure outlined in the section 3 assumes that the CG degrees of freedom (DOF) are independent (Eq. 3.9). For biomolecules, this assumption is most likely problematic. Therefore, one should investigate how important the correlations between the degrees of freedom are for the conformational sampling of the molecule and to which extent they are reproduced by the CG model.

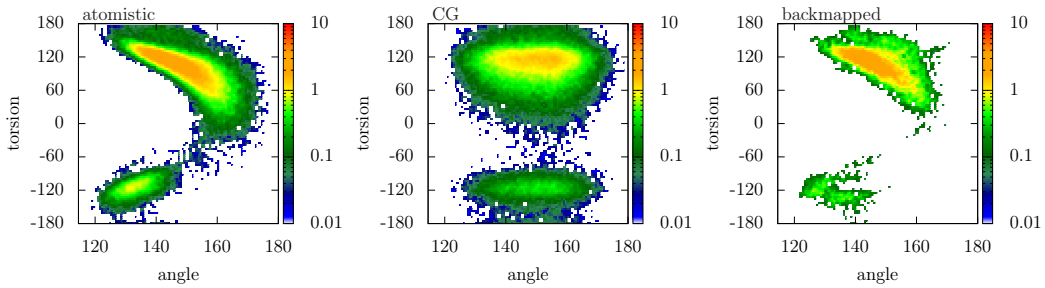

Figure 4.8: **2-D correlation between CAB1-PEP2-CAB2 angle and PEP1-CAB1-PEP2-CAB2 dihedral angle in logarithmic representation.** Left panel: atomistic simulation; middle panel: CG simulation; right panel: distribution from backmapped ensemble based on 100 structures (analyzed in CG coordinates).

Fig. 4.8 represents the correlation between a CG angle (CAB1-PEP2-CAB2) and a CG torsion (PEP1-CAB1-PEP2-CAB2) degrees of freedom along the peptide chain. The left panel shows that these two CG degrees of freedom are highly correlated in the atomistic sampling. However, while the coarse-grained sampling reproduces the distributions of each individual degree of freedom very well, it fails to capture the correlations (middle panel). But we found that when we reinvestigated the correlation in the "backmapped" ensemble, i.e. the atomistic ensemble corresponding to the CG trajectory which was sampled with position restraints to keep the atomistic structure very close to the CG reference, we obtained the correlation between the CG degrees of freedomagain (right panel). The same observation could be made for all other correlated sets of degrees of freedom along the peptide chain. This indicates that the "lost" correlations for this peptide are connected to the fact that for alanine, we have a linear CG model without an explicit side chain bead. Thus, locally the steric effect of the

side chain was lost, even though the overall conformational properties of the chain segments were well enough covered by angle and dihedral potentials, so that the correlations could be restored by backmapping.

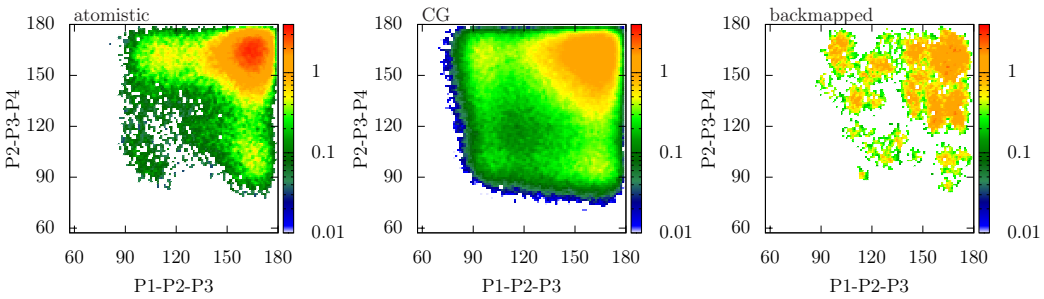

Figure 4.9: **2-D correlation between PEP1-PEP2-PEP3 angle and PEP2-PEP3-PEP4 angle in logarithmic representation.** Left panel: atomistic simulation; middle panel: CG simulation; right panel: distribution from backmapped ensemble (analyzed in CG coordinates).

Fig. 4.9 represents another type of correlations in the peptide. Here the correlations between PEP1-PEP2-PEP3 and PEP2-PEP3-PEP4 1,3,5-angles are shown. One can see, that there is a strong correlation between these 1,3,5-angles in the atomistic case. In the CG case we observe the same pattern, but the energy landscape is more full than in the atomistic case. It is so, because CG simulations are faster and can cover more possible conformations than atomistic simulations. The plot for the backmapping simulation also shows a possible correlation, but this sampling was done only for 100 conformations, thus, it is not as full as the other two.

The main task, however, is to reproduce the atomistic properties from the CG simulation. For checking this possibility we analyzed the atomistic properties of the ensemble of backmapped structures further and compared them to those of the original ensemble from (free) atomistic simulations.

Since the (atomistic) backbone torsional angles at the C_α atoms (Ramachandran angles) characterize the peptide chain conformations and secondary structure formation, we analyzed their distributions. Fig.4.10(a) and (b) show the dihedral angle distributions for two of the Φ and Ψ angles along the backbone. Given that we analyze only 100 random conformations from the CG trajectory, the agreement with the original atomistic sampling is reasonable (the agreement of the other torsions not shown is qualitatively the same). Fig. 4.10(c) and (d) show the Ramachandran plot, i.e. the correlation of the Φ and Ψ angles averaged over all C_α atoms for the original atomistic and the backmapped ensembles. The agreement shows that the local chain properties of the atomistic peptide chain are reasonably conserved (or can be well recovered) in the coarse-graining and subsequent backmapping process.

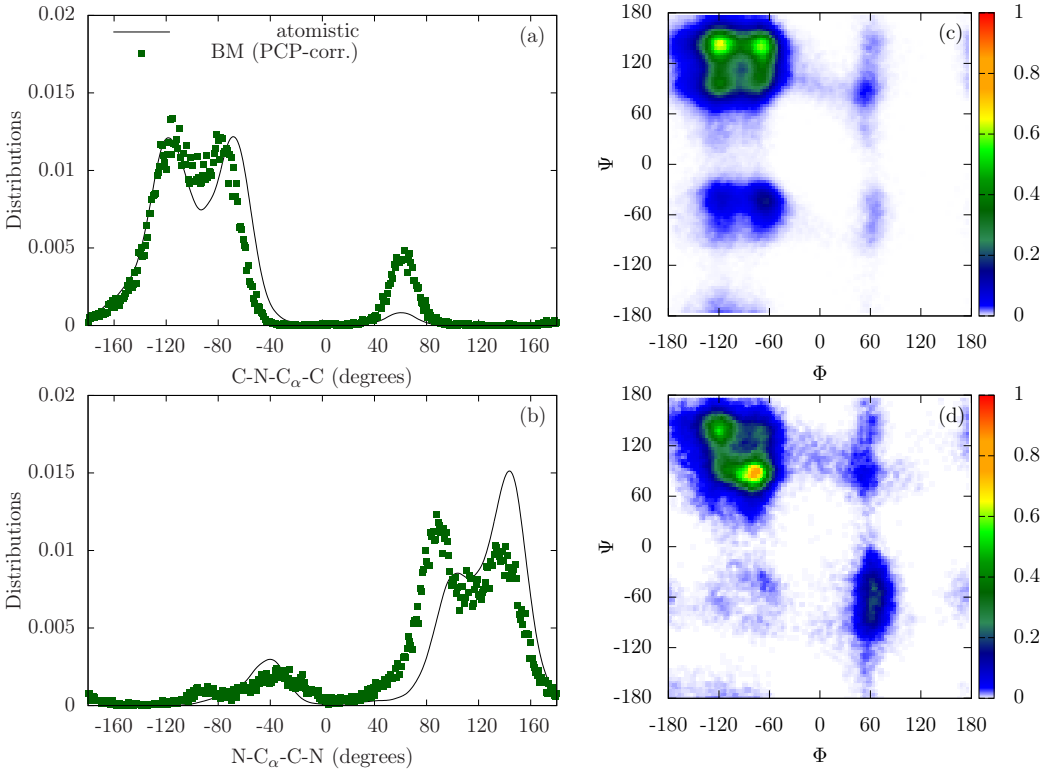

Figure 4.10: **Distributions for atomistic dihedral angles in original atomistic (black) and backmapped (green) ensemble:** (a) distribution for Φ ($C - N - C_\alpha - C$) angle; (b) distribution for Ψ ($N - C_\alpha - C - N$) angle; (c) Ramachandran plot for original atomistic simulation; (d) Ramachandran plot for back mapped ensemble.

Thereby, even though the CG model is not able to reproduce exactly the behavior of the atomistic peptide, the "backmapped" properties of the chain are rather close to the original ones. In the next Chapter we will discuss possibilities to improve the CG model, in order to better reproduce all atomistic distributions including the end-to-end distance. There we will also consider the question, if the improved CG model delivers better results through "backmapping" process.

Chapter 5

Refinement of the CG model for oligoalanine

5.1 Iterative bonded potentials

As shown in Fig. 4.6, the end-to-end distance of Ala_3 is reproduced by the first CG model not well enough, but there is still a shift in the peak. The CG peptide is more compressed than the atomistic one which can also be seen from 1,3,5-angle between PEP beads, see Fig. 4.7(a). In addition, we noticed, that the distance between these two types of beads is also slightly too short. It can be seen in Fig. 5.1(a), where the black line corresponds to the atomistic 1-5 distance and the green line corresponds to the CG simulation. Because of all these data we initially suspected that there is a too strong attraction between PEP beads. To test this the existing nonbonded potential between PEP beads in order to weaken the attraction between them. Several empirical modifications to the PEP-PEP interaction were made which are shown in Fig. 5.2 panel (b).

Fig. 5.2 shows the different variations of potentials, panel (b), and the corresponding end-to-end distance distributions, panel (a). The green line corresponds to the original nonbonded potential, the brown line corresponds to the original potential multiplied by $\exp(-4x^4)$, in order to reduce the long-range attractive interactions, and the blue line shows the case, when the potential was shifted up to the zero and then attractive part was cut. As it can be seen from the Fig. 5.2(a), the end-to-end distributions have approximately the same form with all these potentials. Thereby, we can conclude that the problem is not in the too attractive nonbonded potentials between PEP beads. But, as was mentioned before, the problems are always observed in PEP-PEP interactions.

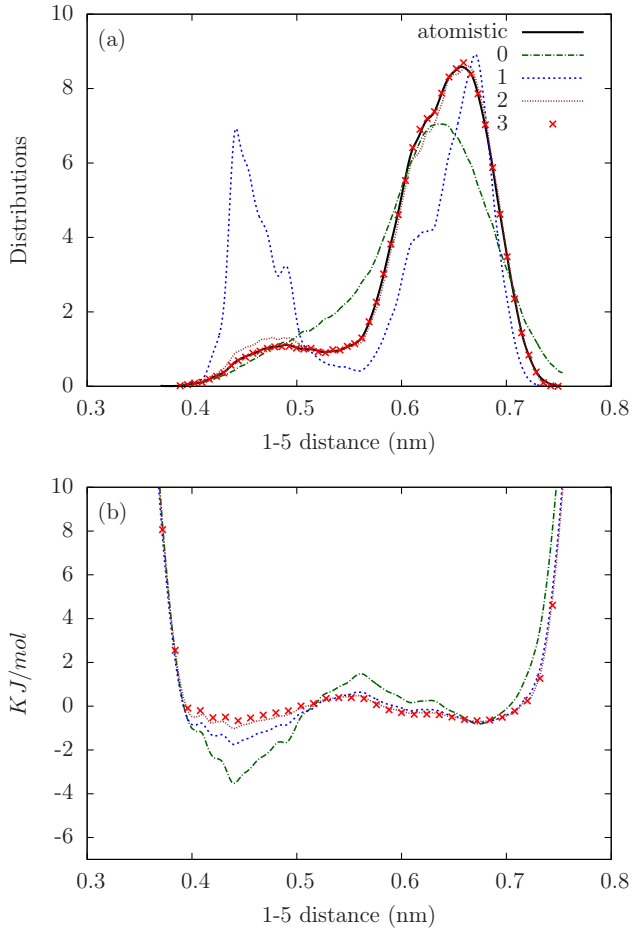

Figure 5.1: **Iterative process for getting bonded potential for interaction between 1-5 beads of PEP type.** Panel (a): changes of 1-5 distance distribution and comparison with the aim atomistic distribution (black line); panel (b): changes of potential during iterative process.

Thus, following the example of improving a CG model for polystyrene [80] an additional "bonded" potential was introduced, which is supposed to fix this problem. In this case, instead of using the standard PEP-PEP non-bonded potential for the 1-5 beads (i.e. for the PEP beads, which have three units between), an additional 1-5 pair potential was applied, in order to get the correct sampling of the 1-5 distance distribution. This tactic, as it will be shown in the following, improves not only the 1-5 distance, but also the end-to-end distance distribution. However, result still corresponds to the aim data not exactly (sec. 5.1.1). Thus, we tried another method and added another type of potential, 1,3,5-angle potential, in order to get correct sampling of the 1,3,5-angle distribution. The results of this approach will be discussed

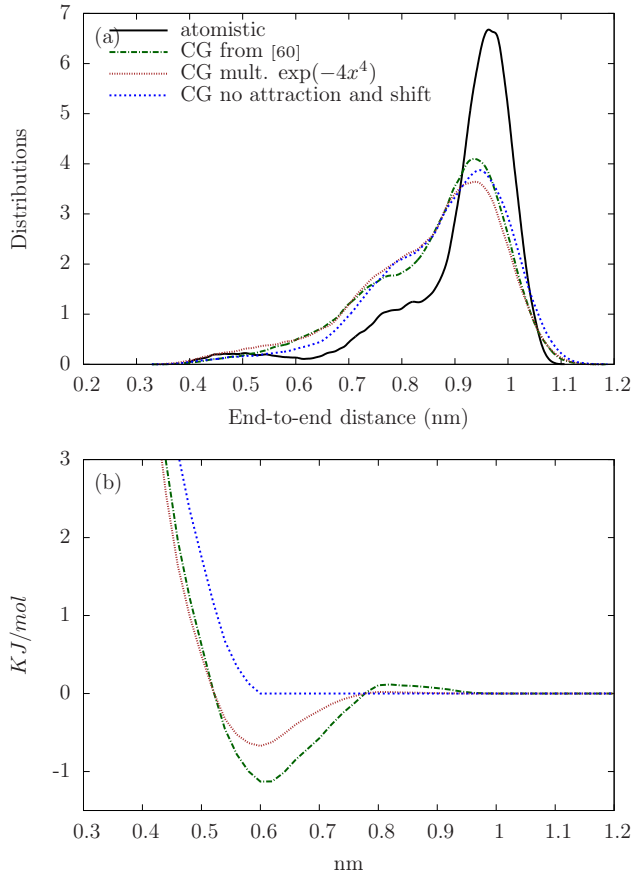

Figure 5.2: **Change of the nonbonded potential between PEP-PEP beads.** (a) Different potentials; (b) corresponding end-to-end distances.

in sec. 5.1.2.

5.1.1 1-5 pair potential

In order to get new "bonded" potentials the iterative Boltzmann inversion was used (Eq. 3.6). Now $g_{target}(r)$ is the atomistic distribution for the 1-5 distance between PEP beads. Fig. 5.1 shows the process of getting such potential, which helps to fit 1-5 distance distribution. After already 3 iterative steps the atomistic distribution was reproduced exactly with CG simulation.

In Fig. 5.3 other properties of the peptide conformation are presented. Panel (a) shows the end-to-end distance distribution. It can be seen, that with an additional 1-5 potential the distribution (red points) is closer to the atomistic one (black line) in comparison with a simple CG simulation without this potential (green line). Panels (b) and (c) show 1,3,5-angle distri-

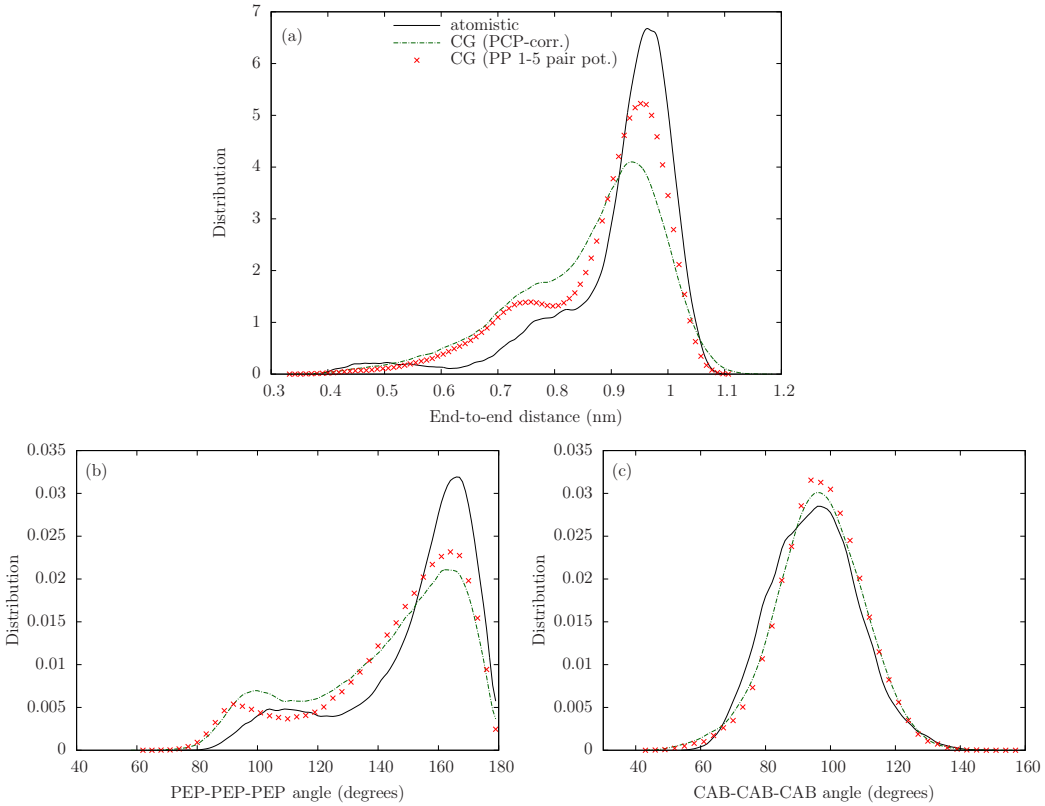

Figure 5.3: **Comparison of distributions for atomistic case (black line), CG case without additional 1-5 potential (green line) and CG with additional 1-5 pair potential (red dots).** Panel (a): End-to-end distance; panel (b): PEP1-PEP2-PEP3 angle; panel (c): CAB1-CAB2-CAB3 angle.

butions (the angle between PEP1-PEP2-PEP3 beads and the CAB1-CAB2-CAB3 beads correspondingly). The PEP-PEP-PEP angle became better, and CAB-CAB-CAB angle almost did not change, what is good, as far as it was reproduced pretty good already before.

Thereby, 1-5 distance distribution can be fixed with an additional corrective 1-5 pair potential and it positively influences on the PEP-PEP-PEP 1,3,5-angle. However, the resulting distributions still have differences from atomistic one. It means, that the CG peptide can have such conformations in space, that some of it structure characteristics correspond to atomistic ones, when other characteristics are different. 1-5 pair potential works only in 2 dimensional space while in third dimension peptide is not fixed. In order to get better correspondence to the atomistic results, we decided to add instead another corrective 1,3,5-angle potential between PEP beads, which works in 3 dimensions and can make all structure properties of CG peptide closer to

the atomistic characteristics.

5.1.2 1,3,5-angle potential

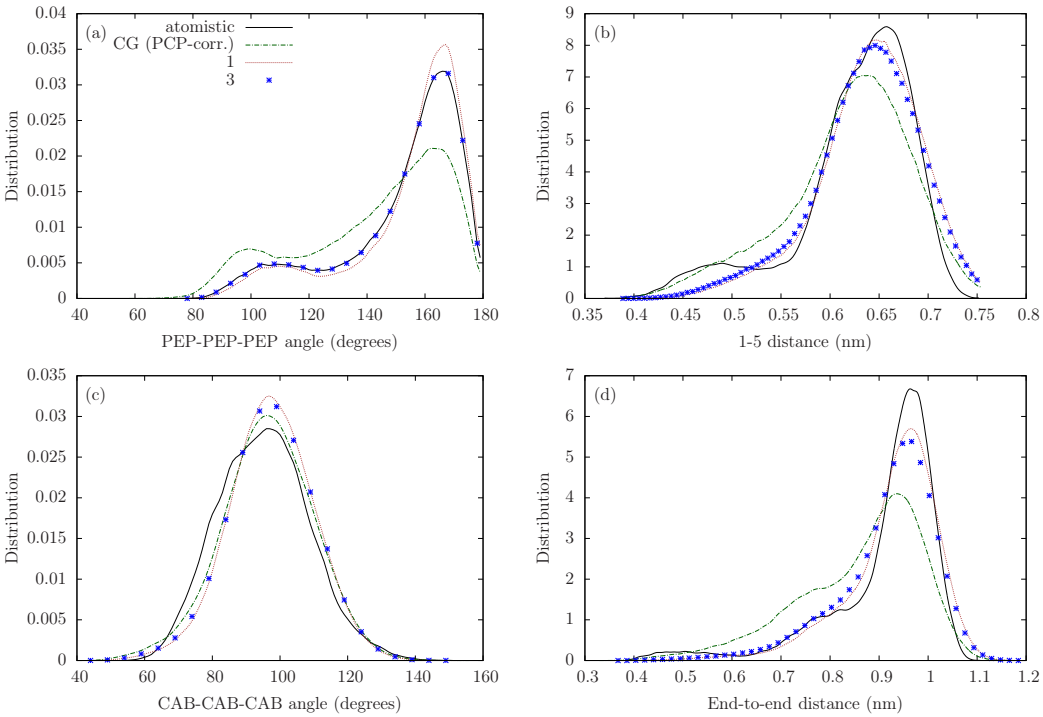

Figure 5.4: **Comparison of distributions for atomistic case (black line), CG case without additional 1,3,5-angle potential (green line) and CG with additional 1,3,5-angle potential after first iterative step (brown line) and after third/last iterative step (blue line).** Panel (a): PEP-PEP-PEP angle; panel (b): 1-5 distance between PEP beads; panel (c): CAB-CAB-CAB angle; panel (d): end-to-end distributions.

For creating the 1,3,5-angle potential the iterative Boltzmann inversion procedure was used again (see Eq. 3.6) as in the case of 1-5 pair potential. The target distribution $g_{target}(r)$ now is the average of atomistic distribution of PEP-PEP-PEP 1,3,5-angle. Fig. 5.4 shows how all characteristics are changed during the iterative process. After three iterative steps the 1,3,5-angle distribution for the PEP-PEP-PEP angle fitted the target one exactly (panel (a)). At the same time, distribution for the another type of 1,3,5-angle did not change and still corresponded well to the atomistic distribution, panel (c). Distributions for the 1-5 distance between PEP beads (panel (b)) and end-to-end distance (panel (d)) became closer to the atomistic distributions.

Thus, 1,3,5-angle potential corrected the CG model better than 1-5 pair potential, because with 1,3,5-angle potential all structure characteristics fit-

Table 5.1: **Mean value of end-to-end distance**

simulation case	end-to-end distance (nm)	
	ALA_3	ALA_4
atomistic	0.9115	1.1623
CG	0.8624	0.9574
CG 100 str. for BM	0.8495	
BM for CG	0.8867	
CG with 1-5-pair pot.	0.8828	0.9996
CG with 1,3,5-angle pot.	0.9172	1.0819
CG with 1,3,5-angle pot. (100 str. for BM)	0.9284	
BM for CG with 1,3,5-angle pot. (100 str.)	0.9659	
CG with 1,3,5-angle pot. (200 str. for BM)	0.9220	
BM for CG with 1,3,5-angle pot. (200 str.)	0.9620	
CG with 1,3,5-angle pot. (400 str. for BM)	0.9691	
BM for CG with 1,3,5-angle pot. (400 str.)	0.9596	
CG with 1,3,5-angle pot. and 1-5 pair pot.	0.8987	1.15389
CG with 1,3,5-angle pot. and 1-5 pair pot.(200 str. for BM)	0.9378	
BM for CG with 1,3,5-angle pot. and 1-5 pair pot. (200 str.)	0.9253	

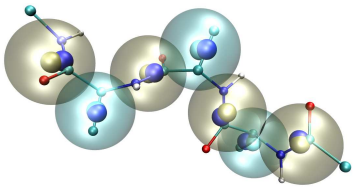

Figure 5.5: **Correspondence between centers of CG beads (the same color as the CG beads have) and the centers of virtual sites (dark blue).**

ted atomistic distributions with good agreement. CG end-to-end distance distribution reproduces atomistic one almost exact. From the Tab. 5.1 it can be seen, that mean value for the end-to-end distance for all our simulations cases is similar, but the CG simulation with additional 1,3,5-angle potential gives the closest to the atomistic simulation result, what also confirm that this CG simulation run gives at this point best results. It is important to mention, that for both CG simulations with additional potential distributions for all degrees of freedom (bonds, angles and torsions) are reproduced well.

From the Tab. 5.1 it also can be seen, that after backmapping procedure the mean value of the end-to-end distance is bigger than the atomistic one. This is because of the backmapping model. As described in section 2.5.5, a backmapping model with virtual sites was used. For the backmapping representation of "CAB" beads the virtual site scheme presented on the Fig. 2.4(a) was used. This scheme correspond pretty good to the CG mapping scheme. For the "PEP" beads another virtual site scheme was used, the Fig. 2.4(b). It does not correspond exactly to the original CG scheme and therefore the backmapped peptide does not fit the CG conformation exactly, as it can be seen from picture Fig. 5.5.

As was shown before, Fig. 4.8 (middle panel), during CG simulation the correlations between degrees of freedom are lost. For cases, when there is additional "bonded" potential, these correlations are also gone (see Fig. 5.6).

Thus, we can conclude, that the additional 1,3,5-angle potential improves the behavior of the peptide Ala_3. As can be seen from the Fig. 5.7 the 1,3,5-angle potential allows to improve the result of CG simulation not only for PEP-PEP-PEP angle distribution, panel (a), based on which it was parametrized, but this set of potentials also gives good agreement for 1-5 distance distribution, panel (b). And also the end-to-end distance for the CG simulation with 1,3,5-angle potential fits the atomistic one in the closest way, panel (d). For another type of 1,3,5 angle – CAB-CAB-CAB-angle – the

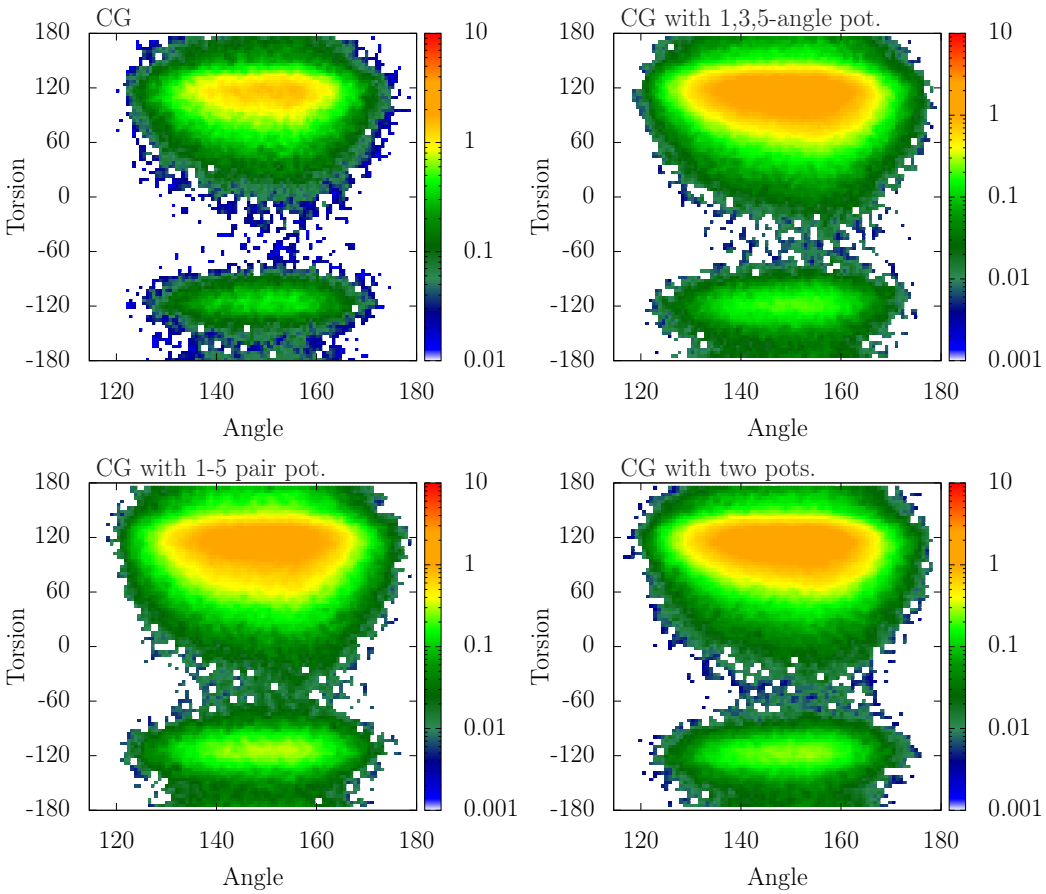

Figure 5.6: **2-D correlation between CAB1-PEP2-CAB2 angle and PEP1-CAB1-PEP2-CAB2 dihedral angles in logarithmic representation.** Left top panel: CG simulation; right top panel: CG simulation with additional bonded 1-5 pair potential; left bottom panel: CG simulation with additional bonded 1,3,5-angle potential; right bottom panel: CG simulation with additional bonded 1,3,5-angle and 1-5 pair potentials.

distributions is approximately the same for all sets of potentials and these results are in the satisfactory agreement with atomistic distribution, panel (c).

After this conclusion we can say, that it is good to reproduce all atomistic characteristics, but are all of them so important? It would be interesting to identify the correlations causing 1,3,5 angle and to find out whether the different regions in its distribution can be linked to typical secondary structure elements sampled by the polypeptide chain. For this reason two idealized peptide chains were investigated, a model oligo-Ala α-helix and a β-strand, which were built using pymol [180] and mapped to CG coordinates. The 1,3,5 PEP-PEP-PEP angles found for these model structures are indicated

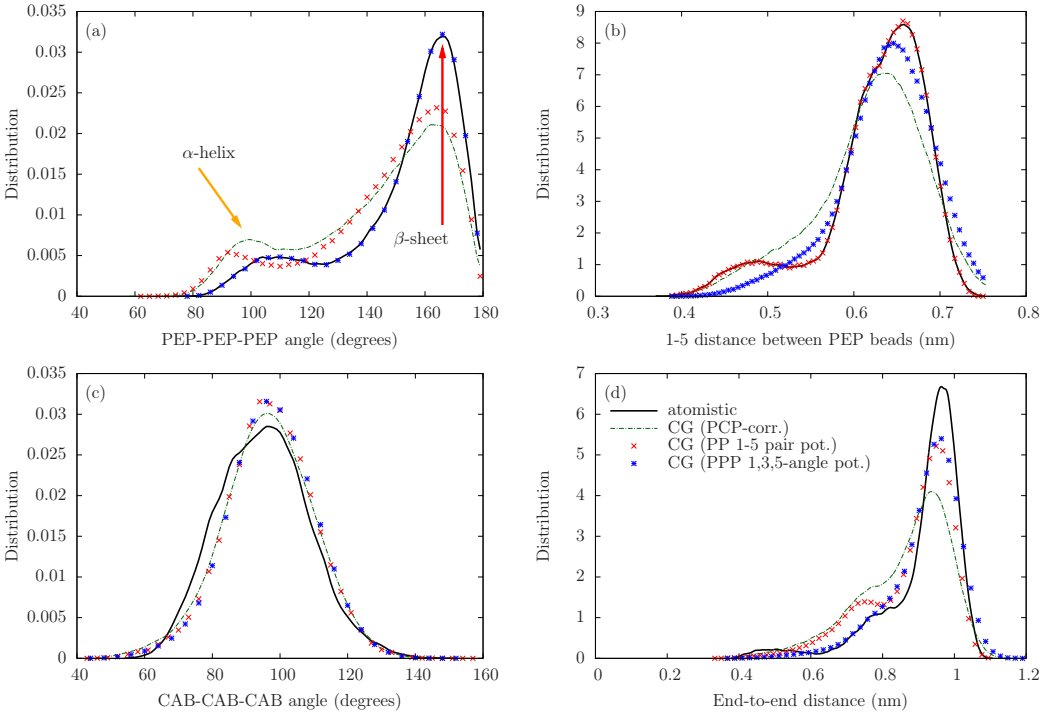

Figure 5.7: **Comparison of the different DOFs for different sets of potentials** (CG with iterative improvement of angle - green line; CG with additional 1-5 pair potential - red dots; CG with additional 1,3,5-angle potential - blue dots) with atomistic distribution (black line). Panel (a): PEP-PEP-PEP angle, the arrows indicate the 1,3,5 PEP-PEP-PEP angle for a α-helix and a β-strand model structure; panel (b): 1-5 distance; panel (c): CAB-CAB-CAB angle; panel (d): end-to-end distance.

by arrows in Fig. 5.7(a). The value of approximately 100 degrees found for the α-helix clearly corresponds to the shoulder of the distribution, while the β-strand with a value of approximately 170 degrees clearly represents the major peak. Thus, the 1,3,5 PEP-PEP-PEP angle is directly linked to the secondary structure adopted by the respective chain segment. This is not unexpected since the 1,3,5 PEP-PEP-PEP angle covers two CAB groups, i.e. adjacent sets of Ramachandran angles, which should be correlated in secondary structure elements. For this reason we would argue that imposing an additional 1,3,5 PEP-PEP-PEP angle potential is a well justified addition in order to reproduce the atomistic conformational sampling by the CG model. This also nicely illustrates that upon constructing a CG model it is important to chose a CG mapping/CG degrees of freedom where the important secondary structure elements can be well represented [55].

As far as on the CG level all results from simulation with additional 1,3,5-angle potential show better agreement with atomistic data than results from

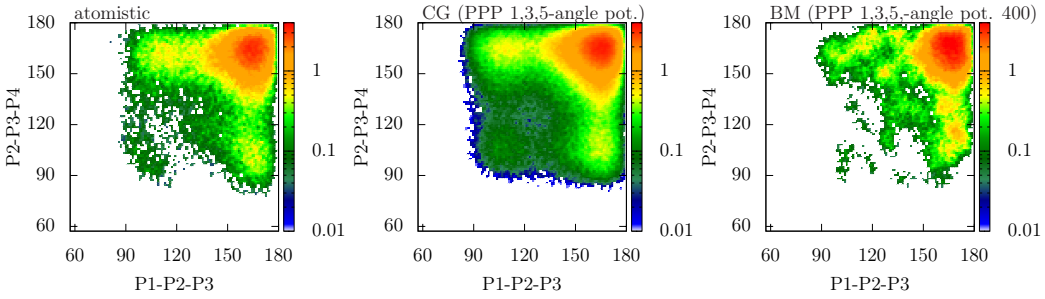

Figure 5.8: **2-D correlation between PEP1-PEP2-PEP3 angle and PEP2-PEP3-PEP4 angle in logarithmic representation.** Left panel: atomistic simulation; middle panel: CG simulation with PEP-PEP-PEP 1,3,5-angle potential; right panel: distribution from backmapped ensemble from 400 structures (analyzed in CG coordinates).

CG simulation without any additional bonded potential, it is interesting to check, if backmapped results of such model also will give better agreement on the atomistic level.

In section 4.5 we already checked the correlation between PEP-PEP-PEP 1,3,5-angles. It would be interesting to see if the 1,3,5-angle potential makes this correlation better. Fig. 5.8 represents this type of correlations for atomistic, CG and backmapped cases for sets with additional PEP-PEP-PEP 1,3,5-angle potentials. In principle this new results bring not much improvement compared to a set without additional potential (Fig. 4.9). So, it is not necessary to use this additional potential if one is interested particularly in this characteristic and in nothing else.

From the new CG simulation also only 100 structures were randomly chosen. In Fig. 5.9 the distributions of the atomistic dihedral angles for atomistic simulation, old backmapped ensemble (see 4.5) and for new backmapped ensemble are presented. One can see, that new system allows to get rather good correspondence to the atomistic distribution compared to the old model. The panel (a) for Φ angle shows that the new model gives approximately the same result as the old one, there is just improvement on the second peak of plot (of course, it can be just a result of "correct" chosen peptide conformations from CG trajectory). But the distribution for the Ψ angle presented on the panel (b) is definitely better for the new CG model than for old one, because before it was shifted to the left and there was a "inversed" balance in the main peaks. The new distributions have already a nice agreement with the atomistic data.

The Ramachandran plot looks also better for the new CG model than for the old one as it can be seen from the Fig. 5.11. Thereby, one additional 1,3,5-angle potential allows to improve results not only on the CG level but also on the atomistic level as well.

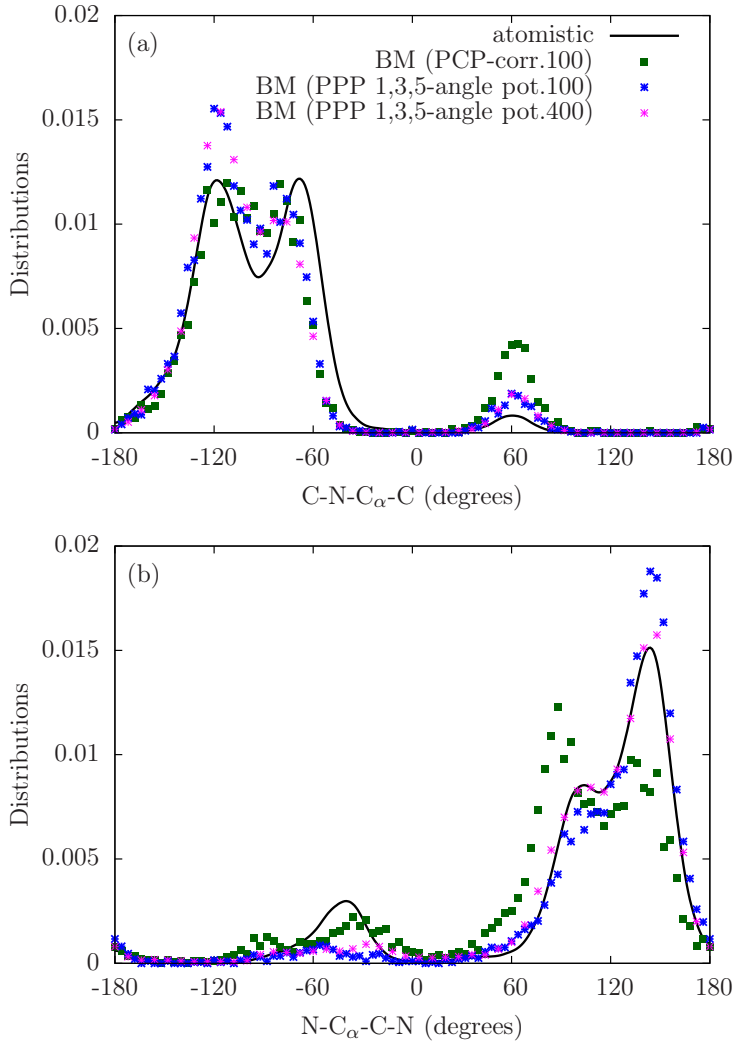

Figure 5.9: **Distributions for atomistic dihedral angles** in original atomistic (black), backmapped ensemble after CG simulation without additional bonded potential (green), and backmapped ensemble after CG simulation with 1,3,5-angle potential in case of 100 structures (blue), backmapped ensemble after CG simulation with 1,3,5-angle potential in case of 400 structures (violet): (a) distribution for Φ ($C-N-C_\alpha-C$) angle; (b) distribution for Ψ ($N-C_\alpha-C-N$) angle.

5.1.3 1-5 pair potential and 1,3,5-angle potential together

As a logical continuation of our investigation we did another set of simulations with two additional 1-5 pair and 1,3,5-angle potentials together. In Fig. 5.10 we show the result of this set. On panel (a) the PEP-PEP-PEP 1,3,5-angle is presented. It can be seen that the new set (violet points) reproduces atomistic

distribution almost exactly. At the same time the 1-5 distance distribution, panel (b), is reproduced also well. Before, with only one type of additional potentials we could fit only one of these distributions exactly and the another one was only approximately as the atomistic one. The main result of this set is the end-to-end distance. Now the CG distribution is in good agreement with the atomistic one and almost reproduces it. Of course, these results are so good, because now we have more potentials and fix our molecule stronger. It can be not the best way, because we parametrized our two additional potentials separately and each of them partly contains interactions, which are included in another one, therefore we might add them twice. As a result we might reduce the possible sampling space.

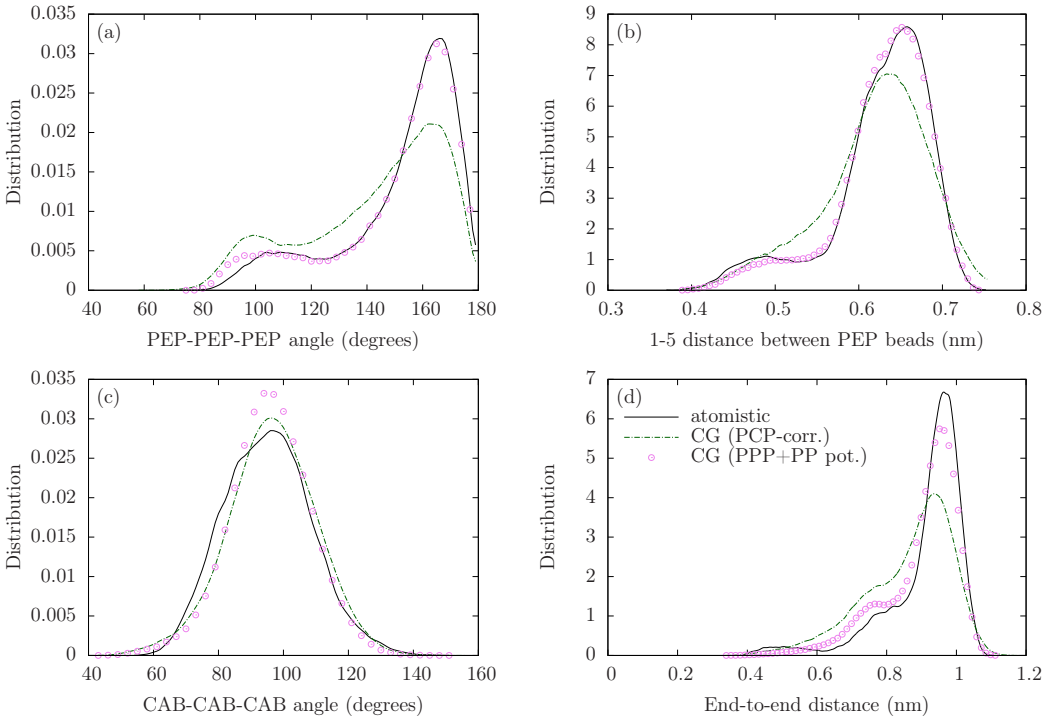

Figure 5.10: **Comparison of the different DOFs for different sets of potentials** (CG with iterative improvement of angle - green line; CG with two additional bonded potentials: 1-5 pair potential and 1,3,5-angle potential - violet line) with atomistic distribution (black line). Panel (a): PEP-PEP-PEP angle; panel (b): 1-5 distance; panel (c): CAB-CAB-CAB angle; panel (d): end-to-end distance.

5.1.4 Backmapping process with the bigger number of conformations

As was shown above, after the backmapping procedure we can get back satisfying atomistic properties, but they are not identical to the original one. There are two main reasons for this. First, we already talked about the shortcoming of our backmapped model model (the center of virtual sites for the "PEP" beads do not correspond exactly to the center of CG bead), and this can give some discrepancy to our results.

Second, before we always used only 100 randomly chosen CG conformations for getting atomistic properties back. And we assumed that a larger number of conformations would give better results. In order to check this assumption we did another set of backmapping procedure but now with 400 CG conformations. The whole procedure was the same as before.

The results for the new backmapping set are presented on the figures 5.9 and 5.11. One can see that the analyzing of 400 CG structures helps to improve resulting data. First of all, in Fig. 5.9(a) it is shown, that the smallest maximum has now better agreement with atomistic one, than it was with 100 structures. From the panel (b) it is clear that the main peak is now better reproduced than it was before. There is still some discrepancy for the small peak, but we can still say that we take into account only 400 CG structures and still can lose some properties.

Any way, from Fig. 5.11 one can see, that now we have better agreement with atomistic Ramachandran plot than it was for 100 structures, because now we have the dots at the top-right area, where they were missing before.

Thus, now we can say with bigger confidence, that increasing the number of CG conformations for the backmapping procedure will improve the results and will make the agreement between the backmapped data and the original atomistic one better.

In order to have a full impression about different sets, we did also the Ramachandran plot for the CG simulations with additional 1-5 pair potential between "PEP" beads and for the case of using two types of additional potentials together. These backmapped processes were done for 200 CG structures each, in order to have good sampling but do not spend to much time and computer resources. It can be seen, that in comparisson to the simple CG simulation without additional bonded potentials the new set with 1-5 pair potential improves the results a little bit, but there is still a shift in the maximum at the point (-60;90), while in the atomistic case the maximum is placed approximately at (-120;140). At the same time it shows better the area corresponding to the α region, while, for example, the CG model with 1,3,5 PEP-PEP-PEP angle potentials cover this area but does not have the

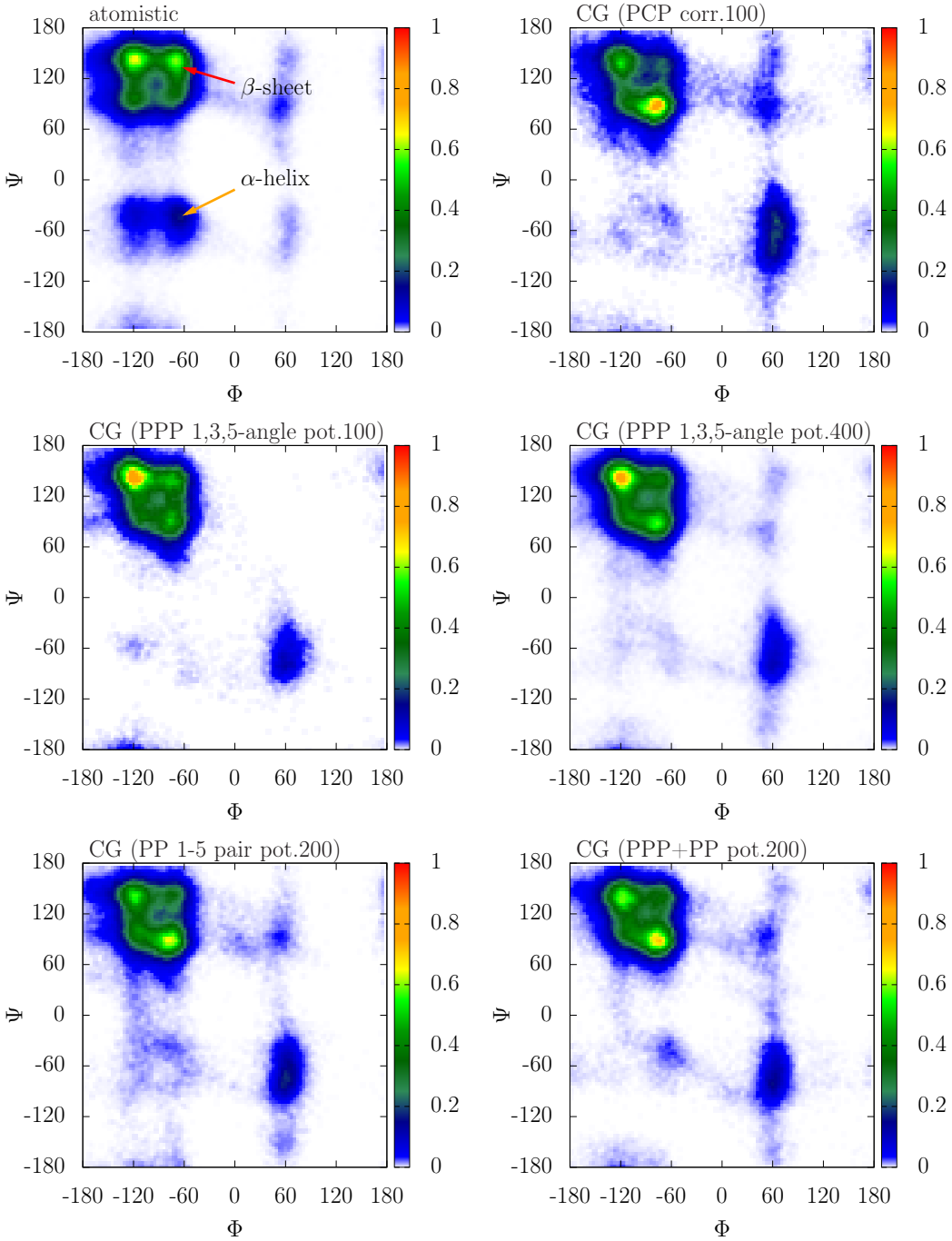

Figure 5.11: **Ramachandran plots:** atomistic case, CG with iterative improved PEP-CAB-PEP angle case, CG with new additional 1,3,5-angle potential with 100 structures case, CG with 1,3,5-angle potential with 400 structures case, CG with new additional 1-5 pair potential with 200 structures case and CG with 1,3,5-angle and 1-5 pair potentials together with 200 structures case. On the upper left panel we indicate the areas that correspond to backbone segments with α-helical or β-sheet local conformations

.

stress on it.

5.2 Extension of the model to Ala$_4$

In order to test if the bonded potentials obtained for alanine in Ala$_3$ can be transferred to other oligomers, we applied them without reparametrization to Ala$_4$, i.e. a peptide with 9 CG beads. For this we made three sets of calculations with different potentials. First set was done in pressure-corrected water with all nonbonded and bonded potentials, where PEP-CAB-PEP potential was improved with iterative Boltzmann procedure. In the second set the additional 1-5-bond potential was used. And in the third set the 1,3,5-angle potential was used instead of 1-5 bond potential. In Fig. 5.12 one can see the results for some degrees of freedom.

Fig.5.12(a) presents the distribution for the dihedral angle PEP3-CAB3-PEP4-CAB4. One can see that two sets with only one of the additional bonded potential give approximately the same results as a CG set without any additional potentials and quite good correspondence to the atomistic distribution, but any way there are some small differences: set without additional bonded potential gives distribution with a lower main peak than the atomistic distribution has, but the second set with 1-5 bonded potential improves this a little bit, while the set with 1,3,5-angle potential almost exactly reproduce atomistic distribution. The case with two additional bonded potentials gives a little bit worse result, as far as the small peak is even lower than the atomistic one, and the main peak is bigger. So, we have a disbalance in distribution, but the areas are still represented correctly.

Almost the same results are presented in Fig. 5.12(b) where the CAB1-PEP2-CAB2-PEP3 dihedral angle is shown. As opposed to the PEP3-CAB3-PEP4-CAB4 angle, where the set with two additional potentials gives a worse result compared to the two other sets, in case of this angle the two additional potentials improve its distribution in a way that it corresponds to the atomistic curve perfectly.

Fig.5.12(c) shows the distribution for another dihedral angle PEP2-CAB2-PEP3-CAB3, which is exactly the centeral dihedral angle. One can see, that two sets without any additional bonded potential and with 1-5 bonded potential give almost the same results, which have not fitted the atomistic distribution. They cover the correct area and have the peaks at the right positions, but there is the disbalance in the peaks: one is too small and the another one is too high. But the set with additional 1,3,5-angle potential solves this problem and the balance of the peaks is correct. The set with two additional potentials is better than CG set without any additional potentials

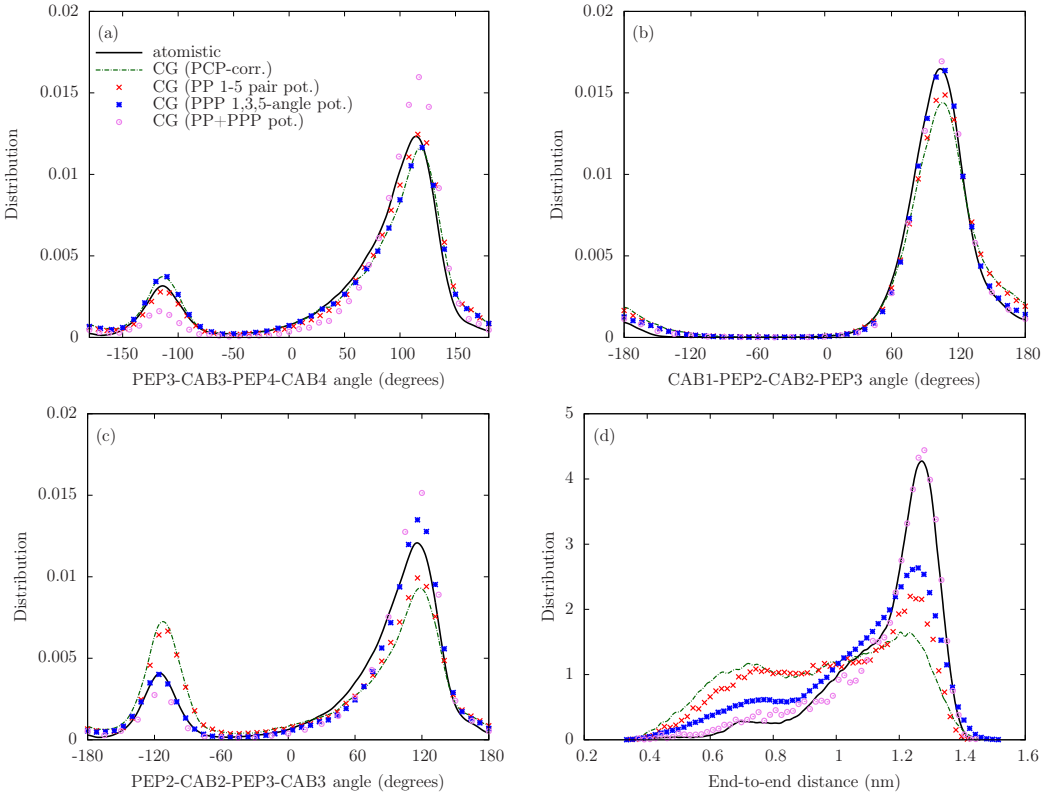

Figure 5.12: **Distributions for the case with 9 beads.** Black line corresponds to the target atomistic simulation; green line corresponds to the CG simulation with potentials, which were parametrized for the 7 bead model; red dots corresponds to the CG simulation with additional 1-5 potential; blue dots corresponds to the CG simulation with additional 1,3,5-angle potential; violet dots corresponds to the CG simulation with two additional potentials: 1-5 pair potential and 1,3,5-angle potential. (a) distribution for PEP3-CAB3-PEP4-CAB4 angle; (b) distribution for CAB1-PEP2-CAB2-PEP3 angle; (c) distribution for PEP2-CAB2-PEP3-CAB3 angle; (d) End-to-end distance.

and CG set with 1-5 pair potential, but it is not so good as CG simulation with only 1,3,5-angle potential.

Fig. 5.12(d) shows the end-to-end distribution. The range of end-to-end distances is reasonably well reproduced by the CG model without additional bonded potentials, but there is a tendency to more compact peptide chains which is probably due to the nonbonded interactions within the peptide chain. Since the chain is longer than in Ala_3, one would expect to see a stronger effect of the nonbonded interactions on such overall conformational properties of the peptide, and exactly this we can observe here, because now the end-to-end distribution has completely different form and does not have one high peak and two shoulders (in case of Ala_3 the CG distribution keeps

the correct form).The 1-5 bonded potential helps a little bit to improve the distribution and in the simulation with this additional potential we can observe a global peak at the correct position, but distribution has still another form. But the set with 1,3,5-angle potential gives better result and the form of distribution is closer to the atomistic one than in case of two other CG simulations. We have a biggest problems with this kind of distribution and it is very important to mention, that the simulation with two additional bonded potentials gives distribution, which almost exactly corresponds to the atomistic result.

Of course, one can say, that reproducing the same range of distances is already good and the average end-to-end distance will be approximately the same for each set. In the table 5.1 the average values of the end-to-end distance for ALA_4 are presented. One can see, that the data are close to the atomistic result in all sets, but any way, the simulation with two additional 1,3,5-angle and 1-5 pair potentials has the best agreement with atomistic one. And I have to add, that for biological systems to get the same average is not enough because each point of distribution corresponds to a specific conformation and it is important to reproduce all possible conformations with the same statistical weight. With two additional potentials now we can reproduce it correctly.

Chapter 6

Coarse-graining nonbonded interactions determined from PMF calculations

In Chapters 4 and 5 the focus had been done on intramolecular (bonded) interactions of the peptide molecule in water, and nonbonded potentials had been used from ref. [60, 61]. Those nonbonded potentials had been parametrized based on small molecules representing peptide fragments. The small molecules relevant for our ALA_n system were methylacetamide and methane, representing the peptide group (PEP) and $C_\alpha - C_\beta$ (CAB) beads respectively. For pair-interactions the potentials of mean force (PMFs) between pairs of these molecules in aqueous solution had been determined from atomistic simulation and used as target functions for the coarse-grained (CG) model. The description of the approach to determine CG potentials based PMFs is presented in Chapter 2 Sec. 6. In the case of implicit solvent we found that the PMFs from atomistic simulation of solute pairs in explicit water can be used directly as effective bead-bead nonbonded interactions for the CG system, there was no need for additional iteration steps to remove correlations or multibody effects. When the CG system has an explicit representation of the solvent, the atomistic PMFs cannot be used directly, because the mediation of the surrounding water is incorporated into this solute-solute interaction potential of mean force. In order to get potentials with "pure" interaction between the beads $V_{bead-bead}(r)$ without water contribution, the method (described in details in Section 2.6) can be used [60, 61]:

$$V_{bead-bead}(r) = V_{PMF}^{atom}(r) - V_{PMF,excl}^{CG}(r). \tag{6.1}$$

Here $V_{PMF}^{atom}(r)$ is the target potential of mean force calculated in atomistic simulations of solute pairs in explicit water, $V_{PMF,excl}^{CG}(r)$ is the potential of

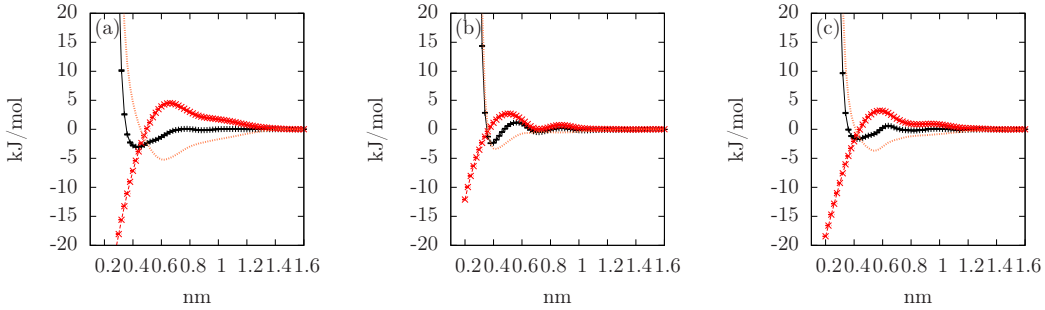

Figure 6.1: **Procedure to generate CG nonbonded potentials between peptide beads:** panel (a) – PEP-PEP; panel (b) – CAB-CAB; panel (c) – PEP-CAB. Black lines correspond to the V_{PMF}^{atom}, red lines: $V_{PMF,excl}^{CG}$; orange lines: effective nonbonded potential $V_{PMF}^{atom} - V_{PMF,excl}^{CG}$.

mean force calculated in CG simulation with explicit CG water and excluded direct bead-bead interaction, this means that all other CG interactions (CG water-water and CG bead-water) are present. By subtracting this from the target PMF we remove the contribution of the environment from the potential of mean force obtained in atomistic simulation. In this procedure one assumes a previously determined set of CG water-water and solute-water interaction potentials. To determine these, iterative Boltzmann inversion had been used, with the radial distribution functions (RDF) from atomistic simulation as a target. This method had first been introduced and tested for small dipeptides by Villa et al [60, 61].

This procedure of parametrization based on fragment molecules is possible because of the assumptions that the fragments are a good representative for the beads as part of a larger (chain) molecule, and that the interactions for the peptide representation on the CG scale are additive. In the present chapter we will closely investigate these assumptions. Here we proceed in two steps:

- first, we reparametrize all non-bonded interactions (as in ref. [60, 61]) since now we use a pressure-corrected CG water model. In principle, the pressure correction is important mostly for the mixed systems, where one consider processes on the two levels at the same time: atomistic and CG ones,- and pressure of both parts should be the same, or the CG part would put higher pressure on the atomistic part. In our case this correction makes no difference, but as far as we already did it like a first order of correction for our angle distribution (see sect. 4.3) we decided to continue to work with this model. This also gives us an impression how sensitive these nonbonded interactions depend on each other, i.e. how drastically the solute-solute interactions change if we

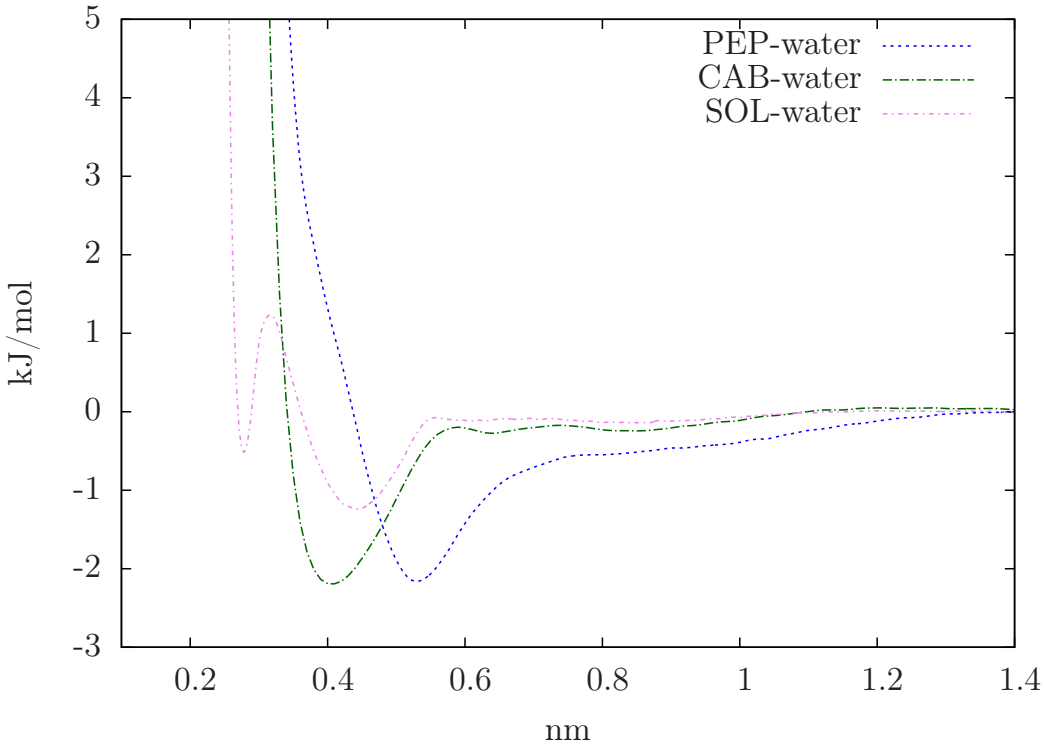

Figure 6.2: **CG nonbonded potentials from iterative Boltzmann inversion for PEP-water (blue), CAB-water (green), water-water (violet).**

apply changes to the solvent-solvent potentials.

- second, these new potentials will be used in order to test the assumption, that the CG potentials, which are parametrized on molecules fragment, can be transferred to the situation where the bead is a part of a chain.

As shown before, the CG peptide contained two types of beads, one represents the C_α and C_β atoms (CAB bead) and another one represents the peptide group $[CONH]$ (PEP bead). In total we have three solute-solute pair interactions (CAB-CAB, PEP-PEP and PEP-CAB) for which we have to determine the PMFs between two representative fragments molecules. Fig. 6.1 shows for all three pairs the two potentials of mean force, $V_{PMF}^{atom}(r)$ and $V_{PMF,excl}^{CG}(r)$, and the resulting effective potential $V_{bead-bead}(r)$ (according to Eq. 6.1).

The potentials for the interactions of peptide beads with solvent beads and between solvent beads, that is water-water and water-peptide, were generated to reproduce the radial distribution functions of liquid water and dilute peptide-water solutions (as it was done in [60, 61]). In the case of

solute-solvent interactions the CG solvent-solvent potential which had been determined previously is applied and the CG solute-solvent potential is refined iteratively (Eq. 3.6) until the target RDF is reproduced. Fig. 6.2 shows the resulting potentials obtained with iterative Boltzmann inversion procedure.

As mentioned before, new nonbonded potentials were obtained for the system with a pressure-corrected water model. The pressure correction resuls in an additional attraction between the water beads (Eq. 3.10). Fig. 6.3(a) shows the pressure-corrected water-water interaction (black line) in comparison to the non pressure-corrected one. Fig. 6.3(b) and (c) shows the CAB-water and PEP-water interaction potentials, parametrized by iterative Boltzmann inversion, using the CG water potentials with (black line) and without pressure correction. We see that both the solvent-solvent and solute-solvent potentials from iterative Boltzmann inversion with pressure correction (black line) are more attractive than without (red line). With these new pressure-corrected solvent-solvent and solute-solvent potentials we now determine new solute-solute interactions according to Eq. 6.1. The resulting potentials are shown in Fig. 6.3(d,e,f) - black lines. Again we observe that the potentials are more attractive compared to the parametrization without pressure correction. In all cases the weak repulsion for distances more than 6-7 $\mathring{A}$ is removed.

First, as was explained above, the PEP-PEP potential was reparametrized for our model based on the two methylacetamide molecules, see Fig. 6.4(a). In Fig. 6.3(f) it can be seen that the potential with pressure correction (black line) has a more steep attraction part compared to the old model (with water made without pressure correction). One also sees that the potential is longer ranged, which has implications on the maximum distance up to which the PMF needs to be computed (see Eq. 2.72; since the PMF has to be computed up to distances where the constraint force is zero). It also implies that the CG calculations of the peptides with the water model with pressure correction need to be performed with a cut-off of appropriate length (in our case 1.4 nm). It means, that the PEP-beads interact now with each other on slightly longer distances.

In the approach to determine PEP-PEP interactions based on the Methylacetamide PMF we assume the interaction to be isotropic, however in the peptide the PEP beads are not anisotropic in their nonbonded interaction. It means that interaction potential depends on the fact with which "sides" two peptides are oriented to each other. The example of possible orientations are presented on Fig. 6.5. Panels (a,b,c) show some possible interaction ways between PEP beads. And when we make parametrization of nonbonded potential based only on two PEP beads we take into account all such inter-

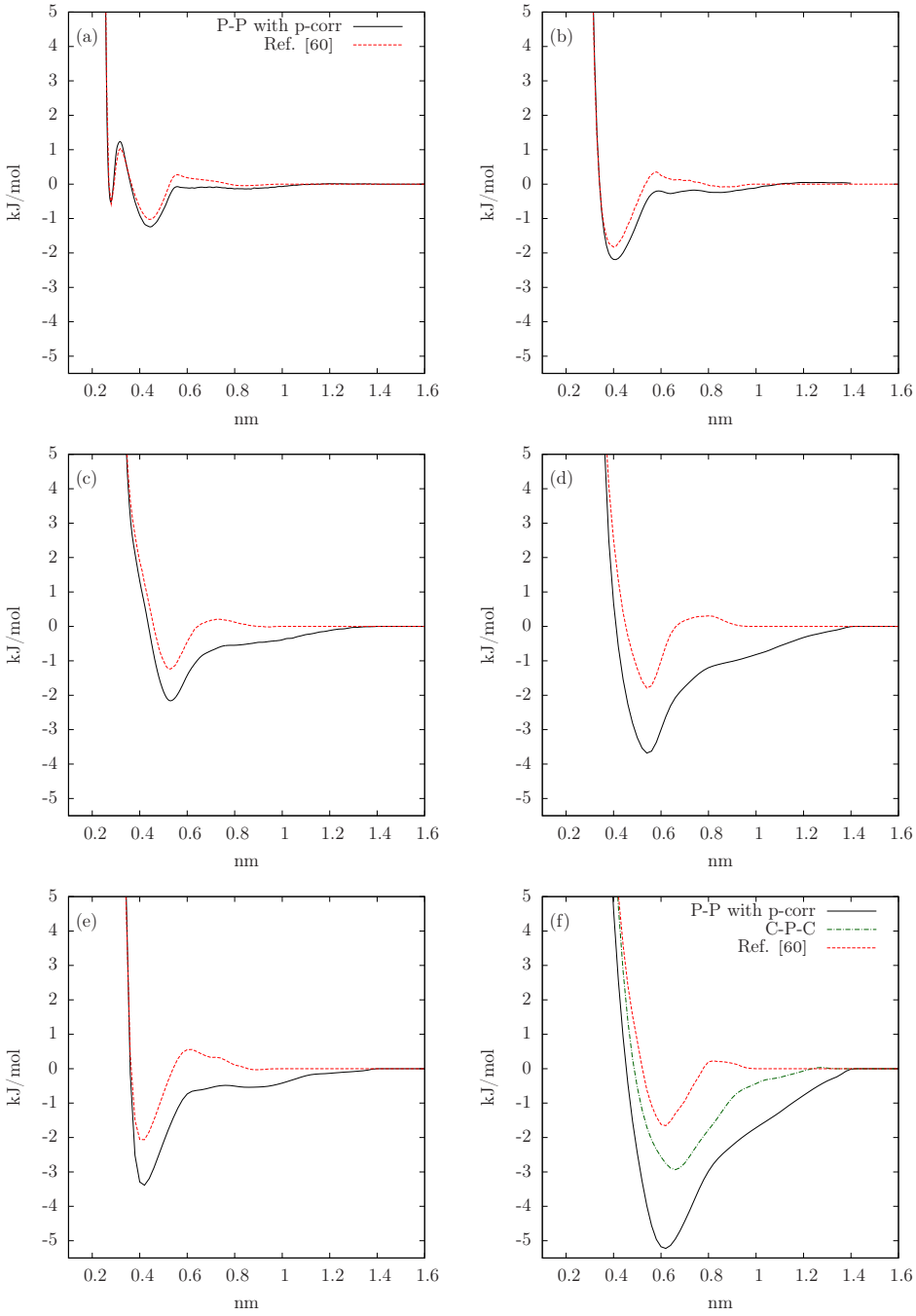

Figure 6.3: **Comparison of old nonbonded potentials from Ref. [60] (red line) with new ones (black line), which were obtained with pressure corrected water.** (a) Nonbonded potentials between water-water beads; (b) Nonbonded potentials between CAB-water beads; (c) Nonbonded potentials between PEP-water beads; (d) Nonbonded potentials between PEP-CAB beads; (e) Nonbonded potentials between CAB-CAB beads; (f) Nonbonded potentials between PEP-PEP beads (here the green line corresponds to the potential, which was obtained based on the CG peptide fragment - CAB-PEP-CAB, it is discussed later).

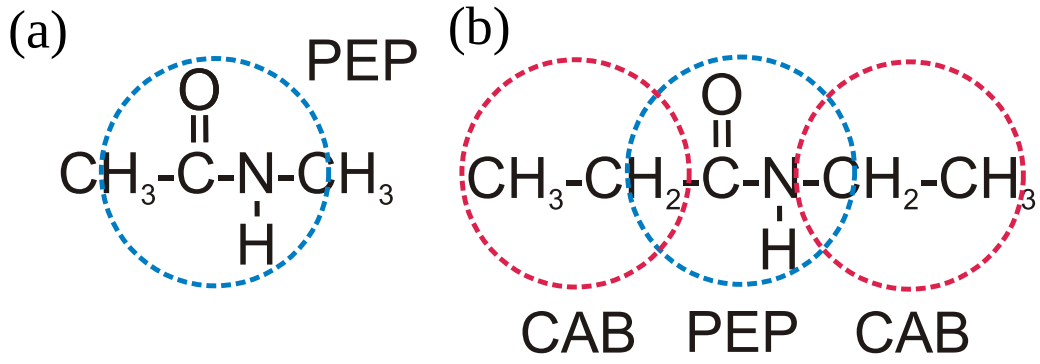

Figure 6.4: **Panel (a) - chemical structure of methylacetamide; panel (b) - chemical structure of ethylpropionamide corresponded to the CAB-PEP-CAB peptide fragment.**

actions and assume that in the chain they will work in the same way. But as one can see on the panels (d,e) not all interactions between PEP beads are realized in the chain: on the panel (e) interaction is approximately the same as on the panel (b), but panel (d) shows that in such orientation we have already different interaction in compare to panel (a) caused by neighbor CAB beads. And as long as we consider the whole chain of beads, where each unit is fixed relatively to its neighbors, beads have no longer the possibility to interact with each other with any "sides", but just with some of them and also there is the influence from their neighbors. In order to take into account all these factors, we decided to reinvestigate carefully the PEP-PEP interactions.

In order to consider the influence from the chain as well, the PEP-PEP potential was again parametrized, but in this case the parametrization was based on another molecule (ethylpropionamide) shown in Fig. 6.4(b), which would on the CG level correspond to the CAB-PEP-CAB fragment of the peptide. In Fig. 6.3(f) it can be seen that this new potential (green line) is placed "between" the old and new potentials parametrized on two methylacetamides. Fig. 6.6 shows the different potentials of mean force that contribute to the new PEP-PEP interaction based on the CAB-PEP-CAB fragments. The new CG interaction is again computed according to Eq. 6.1. The black line in Fig. 6.6(a) shows the atomistic target PMF from Eq. 6.1 between two ethylpropionamide (structure Fig. 6.4(b)) molecules in explicit solvent. For comparison this new atomistic PMF (of ethylpropionamide) is shown in Fig. 6.6(b), blue line, next to the old target PMF (green line) between two methylacetamide molecules (structure Fig. 6.4(a)). We can see that the influence of the two additional CH_2 elements on the atomistic targets is comparatively small.

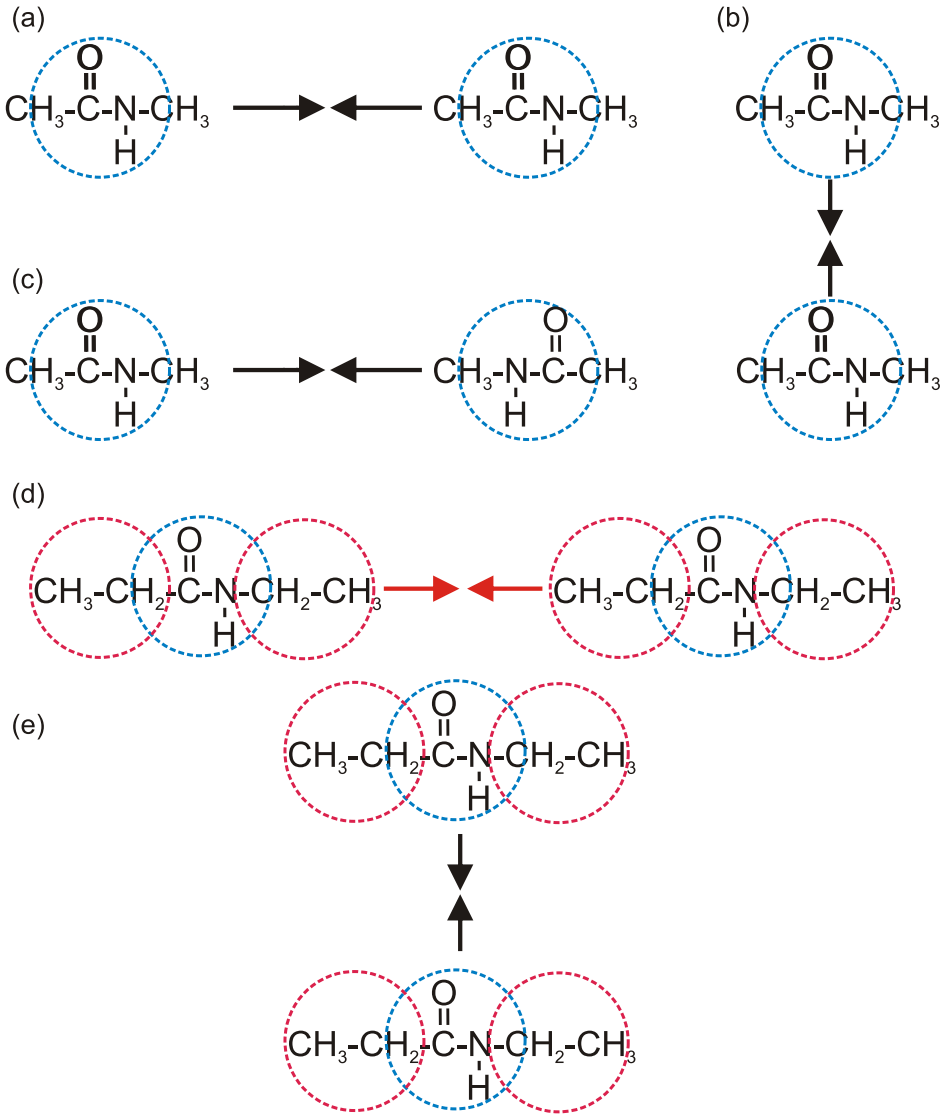

Figure 6.5: **(a,b,c) Examples of possible orientations of PEP beads and "interactions" sides. On the panel (c) the chemical formula is written wrong, it is done in such way just for illustration of possible way of interaction between PEP beads. (d,e) Examples of possible interactions between PEP beads, when they are in the chain.**

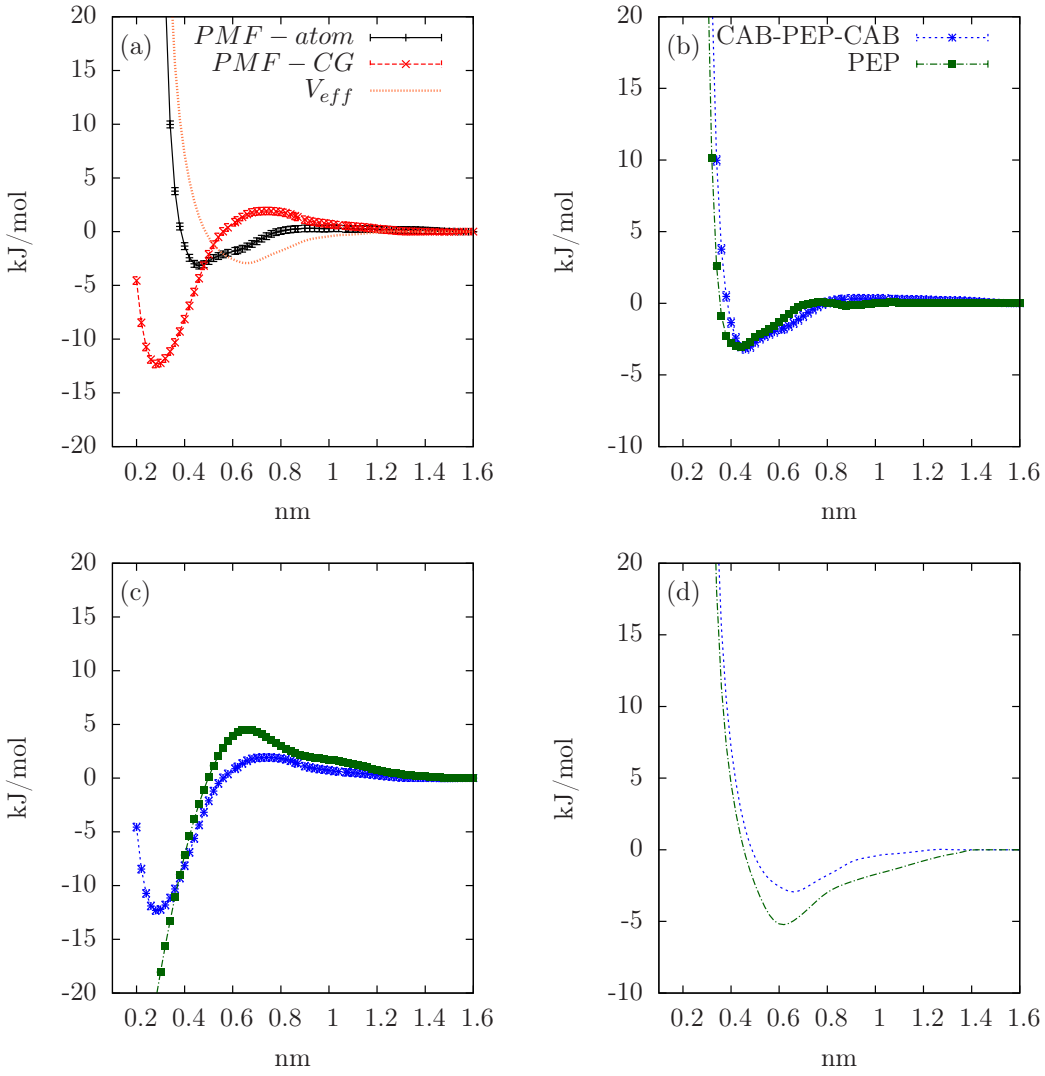

Figure 6.6: **(a) Procedure to generate CG nonbonded potentials between PEP-PEP beads with CAB-PEP-CAB pieces according to Eq. 6.1; (b) comparison V_{PMF}^{atom} in case of parametrization with two PEP beads (green line) and two pieces CAB-PEP-CAB (blue line); (c) comparison $V_{PMF,excl}^{CG}$ in case of parametrization with two PEP beads (green line) and two peaces CAB-PEP-CAB (blue line); (d) comparison effective potentials $V_{PMF}^{atom} - V_{PMF,excl}^{CG}$ in case of parametrization with two PEP beads (green line) and two peaces CAB-PEP-CAB (blue line).**

The red line in Fig. 6.6(a) shows the CG PMF between two CAB-PEP-CAB molecules (in the CG water model with pressure correction) where the direct interaction between the central PEP beads is turned off (but all other interactions: PEP-CAB, PEP-water, CAB-water, water-water and bonded ones,- were in). This potential corresponds to the $V^{CG}_{PMF,excl}$ term in Eq. 6.1. Again this $V^{CG}_{PMF,excl}$ is presented in Fig. 6.6(c) in order to get the direct comparison to the old parametrization procedure where we used the PMF between two CG PEP beads (not in chain) with direct interaction turned off. We see that the new $V^{CG}_{PMF,excl}$ (blue line) exhibits an repulsive contribution at close distances which is due to the interaction between the CAB beads of the two molecules.

The orange dotted line in Fig. 6.6(a) shows the new effective PEP-PEP interaction, resulting from $V^{atom}_{PMF} - V^{CG}_{PMF,excl}$, based on CAB-PEP-CAB fragments. It is also directly compared in Fig. 6.6(d) to the parametrization based on PEP fragments (with otherwise identical interactions). Wee see that by accounting for the presence of the neighboring beads in the parametrization, the effective PEP-PEP interaction has a less deep attraction compared to the parametrization based on the two PEP beads. Now we have three different sets of nonbonded potentials: 1) potential from Ref. [60] combined with the new water model, with pressure correction (Ref. [60] with p-corr); 2) our new potential parametrized on two PEP beads accounting for the new CG water model with pressure correction (P-P with p-corr); 3) the potential parametrized on CAB-PEP-CAB pieces (C-P-C).

In order to check the ability of the three effective nonbonded potentials between PEP beads to reproduce the atomistic PMF of ethylpropionamide $CH_3 - CH_2 - CONH - CH_2 - CH_3$, the PMFs between two CAB-PEP-CAB molecules were computed on the CG level. During this simulation all interactions were taken into account. We performed simulations with all three different PEP-PEP potentials. The result of the test is presented on the Fig. 6.7. The potential based on CAB-PEP-CAB fragments fits the atomistic potential exactly, while the two other simulations show severe deviations. Of course it is not surprising that the new PEP-PEP potential better reproduces the atomistic ethylpropionamide PMF, since it was constructed this way. By extending the molecule further we can test if our accounting for the presence of the chain in the PEP-PEP interaction gives also an improvement for the interactions between longer peptides. Therefore we simulated ALA_2 (corresponding to our CG peptide chain with 5 beads, i.e. the sequence PEP-CAB-PEP-CAB-PEP).

Three independent PMF calculation for the two CG ALA_2 peptides were done, where all interactions were included. In each simulation a different PEP-PEP potentials was used (all other parameters were the same). The

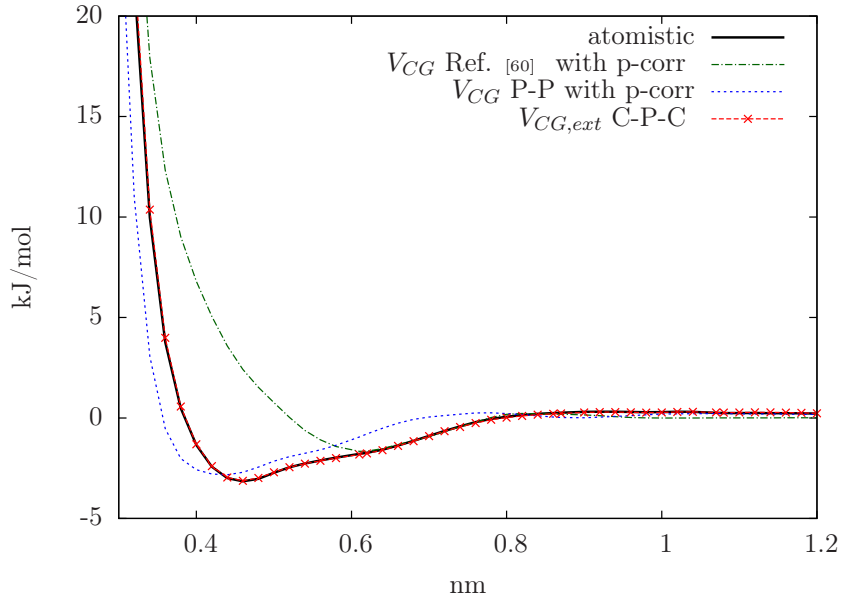

Figure 6.7: **PMF calculation for CAB-PEP-CAB fragment of peptide.** Black line corresponds to the atomistic simulation; all other lines correspond to the CG PMF simulation without excluded interactions between PEP-beads: blue – simulation with first potential from [60, 61], red – simulation with "extended" PEP-PEP potential, parametrized on two CAB-PEP-CAB pieces (ethylpropionamides); green – simulation with PEP-PEP potentials, parametrized on two PEP beads.

result is presented in Fig. 6.8, where the black line is the atomistic PMF, the green line corresponds to the CG simulation with old potential (ref.[60]), the blue line shows the PMF for the simulation with our new potential based on the two PEP beads with all pressure corrections, and the red line corresponds to the PMFs with the potential parametrized on CAB-PEP-CAB fragments. It can be seen that the simulation with the fragment-based potential reproduces the atomistic PMF almost exactly, while all other potentials give completely different results. Thus, the conclusion can be made, that it is important to take into account the influence of neighbor beads (in our case CAB beads).

Nonbonded potentials play a significant role in such processes as aggregation, where the interaction between two systems (two molecules or molecule and surface) is important. In order to understand the influence of our differently parametrized nonbonded potentials on the interaction between our peptides and to find out if there is a difference between them, we decided to study the orientation of ALA_2 peptides relative to each other. Therefore, we investigate the angle between the "end-to-end" vectors θ of the two peptides during the PMF calculation. For the distributions we used the PMF calcu-

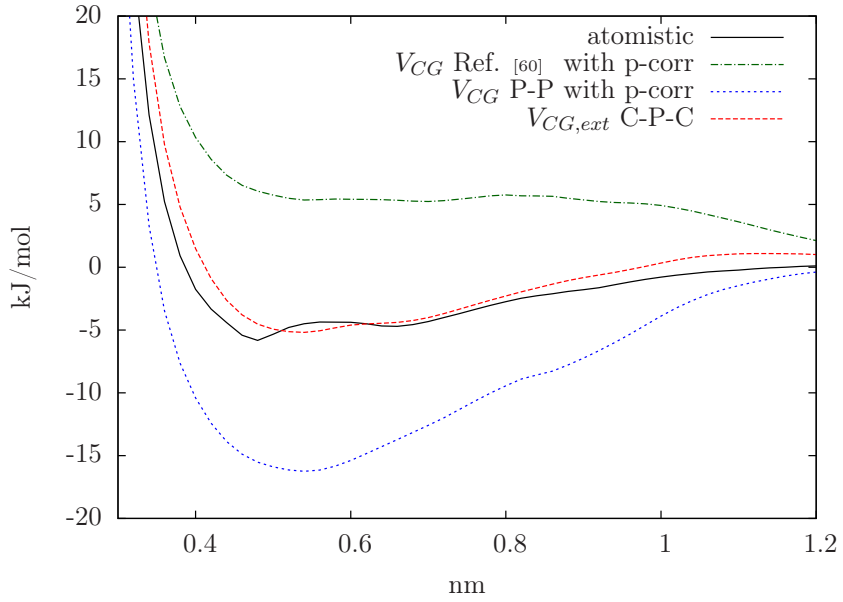

Figure 6.8: **PMF calculation for ALA2 peptide.** Black line corresponds to the atomistic simulation; blue line corresponds to the CG simulation with old potential; red line corresponds to the CG simulation with "extended" PEP-PEP potential, parametrized on two CAB-PEP-CAB pieces; green line corresponds to the CG simulation with normal PEP-PEP potentials, parametrized on two PEP beads.

lation results of the simulation with all potentials in the system. We chose three distances where the peptides are close to each other (0.4, 0.5 and 0.6 nm distance between the two central PEP beads) and three distances where they are far apart, i.e. where the two peptides should not "feel" each other (1.4, 1.5, 1.6 nm between middle beads). Fig. 6.9 presents the distributions of $\cos\theta$ for all 6 cases. One can see that closely located peptides (panels a, b, c) have a preferance for certain relative orientations (where $\cos = \pm 1$). It means that peptides prefer to be parallel to each other. On large distances there are no interactions between the peptides and their behaviors are independent, which is confirmed by the flat distributions shown on panels (d, e, f). In the atomistic cases (black lines) it seems that for the case of the distance between central PEP beads equaled to 0.4 nm the $\cos\theta$ distributions have peaks not at ± 1 but around ± 0.6 (corresponding to approximately to 60 and 120 degrees). The reason for this could be sterical effects or hydrogen bonds between the end beads (the typical distance between two hydrogen-bonds in β-sheet is $\sim$0.5 nm) which might favor special orientations and which are very specific and not captured by the CG model.

To analyze this, we have more closely investigated two cases: one, where the distance between the peptides is 0.4 nm and one, where the distance

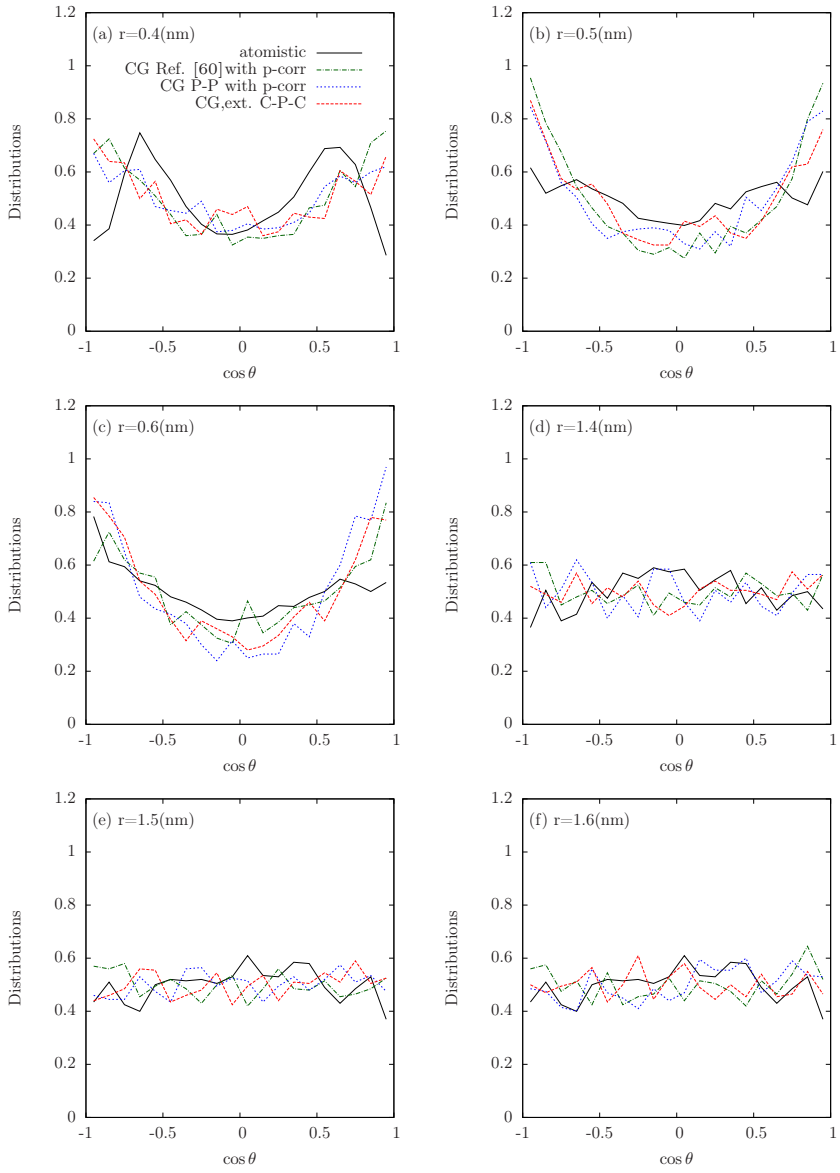

Figure 6.9: **Distribution of a cosine of an angle between two end-to-end vectors of the ALA_2 peptides.** (a): distance between ALA_2 is equal to 0.4 nm; (b): distance between ALA_2 is equal to 0.5 nm; (c): distance between ALA_2 is equal to 0.6 nm; (d): distance between ALA_2 is equal to 1.4 nm; (e): distance between ALA_2 is equal to 1.5 nm; (f): distance between ALA_2 is equal to 1.6 nm. Black line corresponds to the atomistic simulation; blue line corresponds to the CG simulation with old potential; red line corresponds to the CG simulation with "extended" PEP-PEP potential, parametrized on two CAB-PEP-CAB pieces; green line corresponds to the CG simulation with normal PEP-PEP potentials, parametrized on two PEP beads.

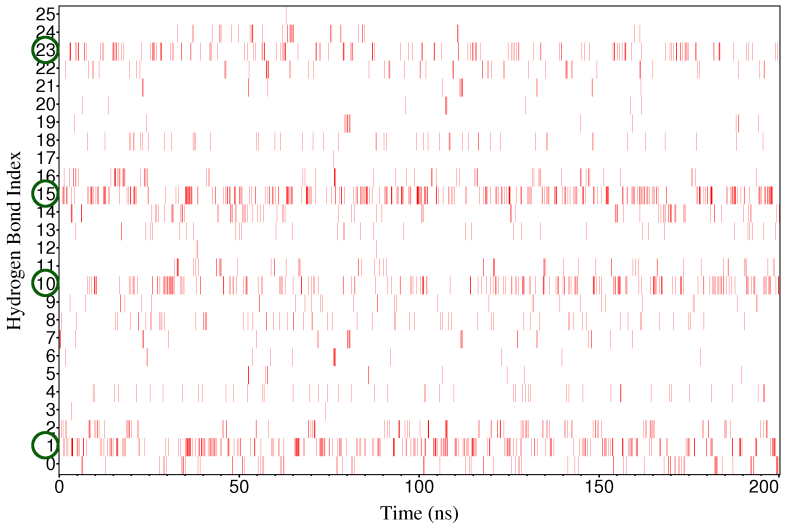

Figure 6.10: **Hydrogen bond existence map for two peptides with the distance between them equal to 0.4 nm. Green circles point two remarkable "donor-hydrogen-acceptor" triplets, which exhibit hydrogen bonds particularly frequently.** These cases correspond to the situation, when hydrogen bond exist between the center of one of peptide and a terminal peptide group of the other peptide, a representative snapshot is shown on Fig. 6.12(a). Corespondent names of these triplets can be found in A.5.4, in Tab. A.1 and Tab. A.2.

between the peptides is 0.5 nm. In Fig. 6.10 and Fig. 6.11, one can see the existence maps for the hydrogen bonds (which demonstrate the presence of hydrogen bonds) for these two cases. A list of hydrogen bonds is given in the Appendix A.5.4. They show points in time when a "donor-hydrogen-acceptor" triplet contains a hydrogen bond (red strokes). The four lines with more frequent strokes (indicated by a green circle around the hydrogen bond index) correspond to configurations where hydrogen bonds between the central peptide group of one molecule and a terminal peptide group of the other molecule are formed. Here the acceptor is located in the middle of one peptide and the donor and its respective hydrogen are at the end of another peptide (representative snapshot see in Fig. 6.12(a)). At the same time, peptides with a 0.5 nm distance between central groups (Fig. 6.11) show more frequent hydrogen bonding as well as two more pronounced lines, which correspond to the "donor-hydrogen-acceptor" triplets when the acceptor is located in the center of one peptides and the donor and hydrogen at the center of the other (representative snapshot see in Fig. 6.12(b)). Thus, the different orientation between the closest peptides (0.4 nm) is caused not only by hydrogen bonds (they exist for larger distance as well) but also by the steric effect. This can be confirmed when looking back at the potential

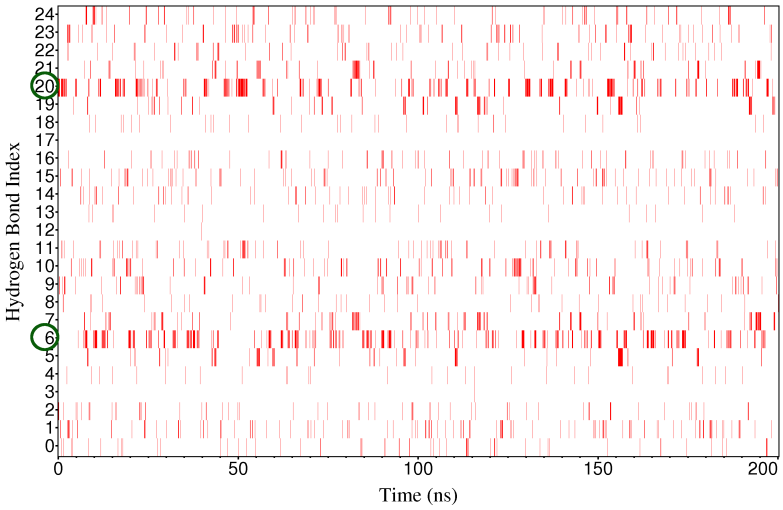

Figure 6.11: **Hydrogen bond existence map for two peptides with the distance between them equal to 0.5 nm. Green circles point two remarkable "donor-hydrogen-acceptor" triplets, which exhibit hydrogen bonds particularly frequently.** These cases correspond to the situation, when hydrogen bond exist between two centers of the peptides, a representative snapshot is shown on Fig. 6.12(b). Corespondent names of these triplets can be found in A.5.4, in Tab. A.1 and Tab. A.2.

between the central groups of the peptides for the ALA_2 molecules Fig. 6.8. There, one can see that at a distance of 0.4 nm,the potential is repulsive and hence the peptides try to move away from each other. For details see A.5.4.

Fig. 6.9 shows that in terms of the relative peptide-peptide orientation all the CG models give the same results. This means that the different nonbonded interactions (with and without pressure correction and with or without accounting for the chain in the PMF calculations) do not influence the way parallel and antiparallel peptide orientations are favored over perpendicular ones. Combining this with the significant difference in the peptide-peptide association (Fig. 6.8) we can conclude that the best model for nonbonded interactions is the one where the chain had been accounted for.

We also checked the end-to-end distances of the peptides for each of the positions in order to clarify if the interactions between two molecules will affect their conformation and if yes, if the atomistic and the CG models agree on this. The distributions are the same for all distances, this means that there is no significant influence from intermolecular interactions on the conformations of the peptides (data not shown). The dynamics behavior of two peptides will be considered in more details in Chapter 7.

We would like to conclude this chapter about nonbonded CG interac-

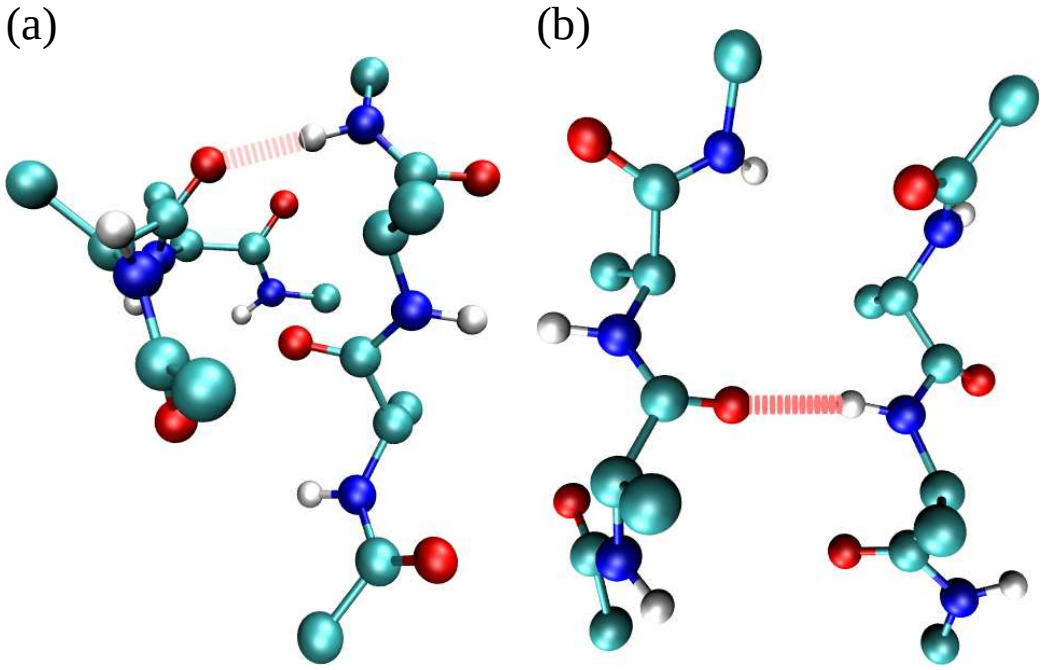

Figure 6.12: **Orientation of peptides when they are close to each other and have hydrogen bond.** (a) Distance between peptides is 0.4 nm, they are almost perpendicular; (b) distance between peptide is 0.5 nm, they are parallel.

tions by revisiting the model for ALA_3 developed in Chapter 4 and 5. Now it would be interesting to see, what kind of results we get for this system with our new nonbonded potentials and if we can directly use the bonded potentials, which were early obtained with the nonbonded interactions from Ref. [60]. So, we performed CG simulations for the ALA_3 peptide for following sets of interactions: one was done with bonded potentials directly obtained from Boltzmann inversion without correction. Another simulation was done with bonded interactions, where PEP-CAB-PEP-angle potential was corrected with iterative procedure for the set of nonbonded potentials from Ref. [60] (see Sect. 4.3). Last two simulations were done with our new nonbonded potentials based on two PEP fragments and based on two CAB-PEP-CAB fragments. First, they were calculated with the previously obtained bonded PEP-CAB-PEP-angle potential (with and without iterative correction) and at the end with the bonded potentials specially parametrized for them. Fig. 6.13 presents the result for all these sets. Panel (a) shows the PEP-CAB-PEP-angle distribution for the case of using the bonded potentials from direct Boltzmann inversion of the atomistic distribution without iteration correction. One can see, that all sets of nonbonded potentials give a very similar distribution of the PEP-CAB-PEP-angle towards smaller val-

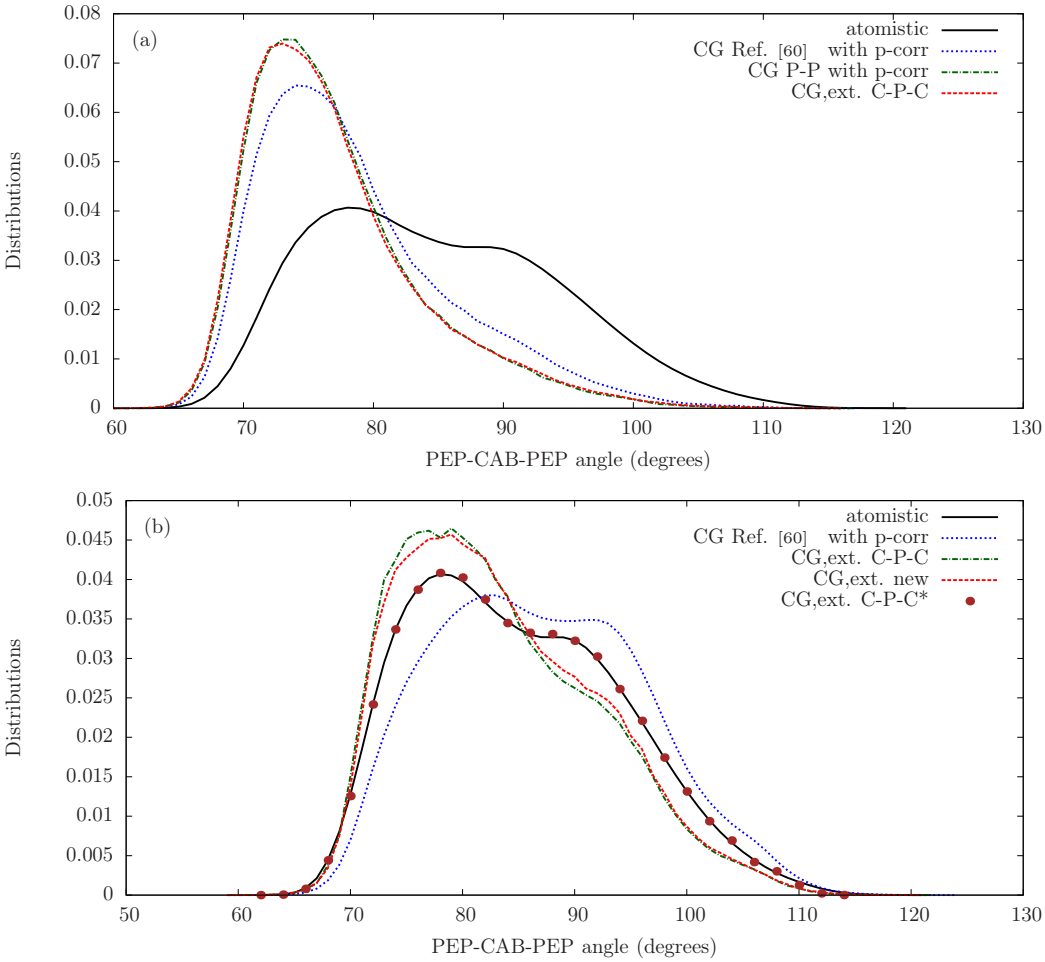

Figure 6.13: **PEP-CAB-PEP-angle distribution for ALA_3 peptide with different nonbonded potentials.** (In all panels black line corresponds to the atomistic simulation; blue line corresponds to the CG simulation with nonbonded potentials from Ref. [60]; red line corresponds to the CG simulation with PEP-PEP potential, parametrized on two CAB-PEP-CAB pieces, including pressure correction for all nonbonded interactions; green line corresponds to the CG simulation with PEP-PEP potentials, parametrized on two PEP fragments): (a) Comparison of simulations with the three sets of nonbonded potentials with bonded interactions without iterative angle correction; (b) Comparison simulations with the three sets of nonbonded potentials with iteratively corrected angle potential (iterations were done in the system with nonbonded interactions from Ref. [60]); "CG, ext. new*" (red dots) corresponds to a simulation with nonbonded potentials based on the CAB-PEP-CAB fragments and iteratively corrected angle potential, parametrized especially for this set of nonbonded potentials.

ues. Panel (b) presents that the set, with potentials from Ref. [60] and with iteratively corrected bonded PEP-CAB-PEP-angle potential, which was parametrized for this sets of nonbonded interactions based on the PEP-CAB-PEP fragments, gives better agreement with atomistic curve than with a "first guess" for bonded potentials (panel (a)). Also this result is better compared to the another two CG sets with the new nonbonded potentials but with the same corrected PEP-CAB-PEP-angle potential (panel(b)). It can mean that corrected potentials are not independent on nonbonded potentials which were used for their parametrization and in order to get nice correspondence between CG and atomistic result, one have to obtain new bonded potentials for each set of nonbonded ones. Therefore, we made new iteratively corrected bonded PEP-CAB-PEP-angle potentials for each new set of nonbonded interactions (parametrized on methylacetamide and ethylpropionamide). These new iteratively corrected bonded potentials are the same for the both new CG sets. Thus, we present only result for the CG simulation with nonbonded interaction between PEP beads based on CAB-PEP-CAB fragments - red dots on the panel (b). One can see, that with a new parametrized PEP-CAB-PEP-angle potential one obtains by construction exact agreement for this type of angle. It means that indeed one has to parametrize all bonded potentials specially for the each set of nonbonded interactions in order to obtain good results and cannot transfer potentials, which were made with another nonbonded parameters.

Next step would be to reparametrize PEP-PEP-PEP 1,3,5-angle potential for the new nonbonded sets as well, because it depends not only on nonbonded interactions but also on bonded ones. But at first it makes sense to remodel another bead, CAB, in order to take into account the influence of the chain, as far as we showed above that this aspect is important for the peptide-peptide association.

Chapter 7

Dynamics in coarse-grained simulations of peptides

It is known that dynamical properties obtained in molecular dynamics (MD) simulations are finite-size dependent due to hydrodynamic interactions between periodic copies, as was shown by Dünweg and Kremer [181, 182] for the diffusion coefficient of polymers. It is also known, that there is a "speed-up" of the dynamics of a coarse-grained (CG) system beyond the simple reduction of number of degrees of freedom (DOF), which stems from the fact that the system evolves on a "smoother" potential energy landscape, where the "finest" degrees of freedom have been integrated out [174]. The magnitude of this additional "speed-up" needs to be determined empirically by mapping dynamic properties on the atomistic and CG scales.

In simulations of peptides and proteins, hydrodynamic effects play a significant role, for example to study aggregation phenomena or reactions in biological cell. In Ref. [183, 184] it was shown that simulations of small peptides in water with and without hydrodynamic interactions (HI) give different results on the same level of resolution (in this case CG). For example, the diffusion coefficients obtained from the simulations with and without HIs are different by a factor of 3 [183]. However, if the factor for another type of dynamics (rotation) is not 3 but different, it would be already noticeable for investigation of aggregation phenomena and we would get two unlike results. As mentioned in Ref. [185] the translational and the rotational diffusion coefficients and the timescales of the internal dynamics can only be reproduced correctly with HI. In order to include HI to the system, one has to use a thermostat which preserves HI, for example dissipative particle dynamics [115, 116] and the stochastic velocity-rescale thermostat (vr) by Bussi et.al. [125].

However, including the HI does not solve all problems. Since one usually

use periodic boundary conditions for simulations, it is important to remember that system will interact not only with the solvent around it, but also with its periodic images. These HI lead to finite size effects [186] (the result will depend on the simulation box size).

The present chapter investigates various aspects of dynamic properties of the CG peptide in the aqueous CG solution developed in previous sections 4-6:

- Comparison of finite size effects on atomistic and coarse-grained levels of resolution for pure water and mixed peptide-water systems.
- Comparison of time mapping for different dynamic properties in the mixed peptide-water system.

7.1 Theory

7.1.1 Classical polymer theory: two polymer models

In order to determine any dynamical characteristics, one first can consider classical polymer physics. Some statistical properties of the polymers are:

1. End-to-end distance:

$$< R^2 >=< (\vec{r}_N - \vec{r}_1)^2 >; \tag{7.1}$$

2. Radius of gyration:

$$< R_G^2 >= \frac{1}{2N^2} \sum_{ij} < r_{ij}^2 >, \tag{7.2}$$

   which represents the mean square length between all the pairs of the segments in a chain [187];

3. Hydrodynamic radius:

$$\left\langle \frac{1}{R_H} \right\rangle = \frac{1}{N^2} \sum_{i \neq j} \left\langle \frac{1}{r_{ij}} \right\rangle. \tag{7.3}$$

In classical polymer theory are two models for describing polymer chains [172]. One is called the Rouse model. It assumes that the chain is ideal, phantom

(self-crossing) and the solvent is immobile. It considers the equation of motion for each monomer of a polymer chain in a solvent:

$$m\frac{\partial^2 \vec{x}_n}{\partial t^2} = \vec{f}_n^{ch} + \vec{f}_n^{fr} + \vec{f}_n^{r}, \tag{7.4}$$

where $\vec{f}_n^{ch}$ is the force between neighboring monomers in the chain, $\vec{f}_n^{fr}$ is the force from friction against the solvent and $\vec{f}_n^{r}$ is a random force, which arises when the given monomer collides with solvent molecules. For conventional motion of the monomer in a dense solvent, the left part of Eq. 7.4 (the inertial term) is not significant and thus the equation can be written as:

$$\vec{f}_n^{ch} + \vec{f}_n^{fr} + \vec{f}_n^{r} = 0. \tag{7.5}$$

The force from friction of the monomer against the solvent is proportional to the velocity, because the monomer (according to the assumption of this model), moves with the thermal velocity in a viscous immobile solvent:

$$\vec{f}_n^{fr} = -\xi\frac{\partial x}{\partial t}, \tag{7.6}$$

where ξ is the friction coefficient. The random force contains a non-regular interaction term of the monomer with the solvent and it mean value is equal to zero:

$$\langle \vec{f}_n^{r}(t)\rangle = 0, \tag{7.7}$$

Its second momentum is:

$$\langle \vec{f}_{ni}^{r}(t)\vec{f}_{mj}^{r}(t')\rangle = 2\xi T\delta_{nm}\delta_{ij}\delta(t-t'), \tag{7.8}$$

where δ_{ij} is Kronecker's symbol. In this case, Eq. 7.4 for one monomer will be:

$$\xi\frac{\partial x}{\partial t} = \vec{f}_n^{r}. \tag{7.9}$$

From this equation, one can find the position of the monomer at each moment in time:

$$\vec{x}(t) = \frac{1}{\xi}\int_0^t \vec{f}^{r}(t')dt'. \tag{7.10}$$

Therefore the mean square displacement of the monomer is:

$$\langle \vec{x}^2(t)\rangle = \frac{1}{\xi^2}\langle\left(\int_0^t \vec{f}^{r}(t')dt'\right)\left(\int_0^t \vec{f}^{r}(t'')dt''\right)\rangle = \frac{6T}{\xi}t = 6D_0 t. \tag{7.11}$$

where D_0 is the monomer diffusion coefficient. For the whole polymer chain of N monomers the diffusion coefficient is [172]:

$$D = \frac{D_0}{N}. \tag{7.12}$$

Another model for polymer chains is called Zimm model. It already takes into account the fact that the solvent is entrained by monomers (i.e. the hydrodynamic interaction). In this model, the force of friction also contains the pertrubation of the solvent velocity field ($\vec{v}(x)$):

$$\vec{f}_n^{fr} = -\xi \left(\frac{\partial \vec{x}}{\partial t} - \vec{v}(x_n) \right). \tag{7.13}$$

In this case, all processes in the systems are already different. For example, if we suppose that the entrainment is strong enough to carry away all of the solvent inside the polymer chain during motion, then the coefficient of diffusion of the whole polymer coincides with the Stokes coefficient of diffusion of a solid sphere of the same size as the polymer coil:

$$D = \frac{T}{\xi_{coil}} \tag{7.14}$$

The Zimm model is closer to reality than the Rouse model [172].

The fact whether there is hydrodynamic interaction in the model or not is important because the results are not the same. In the Rouse model, the dynamic scaling for the diffusion coefficient is (see Eq. 7.12):

$$D \propto N^{-1} \propto R_G^{-1/\nu}, \tag{7.15}$$

there, ν is the critical exponent which describes the solvent (good, bad or θ-solvent), while with hydrodynamic interaction, the scaling is:

$$D \propto R_H^{-1} \propto R_G^{-1}. \tag{7.16}$$

7.1.2 Diffusion coefficient with finite-size effect

In fact the description of the behavior of a polymer chain in a solvent is complex, especially when one considers a periodic boundary box where chains interact with all periodic images. The diffusion coefficient of the center of mass of a polymer can be better described using the Kirkwood-Zimm theory [188, 189, 190] which takes into accountthe hydrodynamic interaction

(the correlation of the stochastic displacements mediated by fast diffusive momentum transport through the solvent) via the Oseen tensor:

$$D_{ij} = D_0 \delta_{ij} \mathbb{1} + (1 - \delta_{ij}) \frac{k_B T}{8\pi\eta r_{ij}} \left(\mathbb{1} + \hat{\vec{r}}_{ij} \otimes \hat{\vec{r}}_{ij} \right). \tag{7.17}$$

Using this result, the diffusion coefficient can be found through the Kirkwood formula which uses the hydrodynamic radius R_H:

$$D = \frac{D_0}{N} + \frac{k_B T}{6\pi\eta} \left\langle \frac{1}{R_H} \right\rangle. \tag{7.18}$$

Here, D_0 is the monomer diffusion coefficient.

In the Rouse model [191], the diffusion coefficient is simpler (see Eq. 7.12) because it assumes that every monomer is coupled to a viscous background with uncorrelated stochastic displacements and the diffusion tensor in this case is diagonal:

$$D_{ij} = D_0 \delta_{ij} \mathbb{1}. \tag{7.19}$$

Due to the long-range nature of the hydrodynamic interactions, one can expect strong effects on the dynamical properties from the finiteness of the MD box, even if the chain fits nicely into the box and static properties are not affected. Simulations are performed with periodic boundary conditions and the chain has an infinite number of periodic images with which it interacts via hydrodynamic interactions. Use of a hard wall would remove periodic images and the polymer would not interact with itself and there would be unwanted surface effects. The finite size problem can only be overcome by using a very large simulation box, which is computationally unfeasible.

Quantitatively, the finite-size effect can be taken into account by modifying the diffusion tensor appropriately [183]. Denoting the linear size of cubic MD box as L, one has off-diagonal elements, $i \neq j$:

$$\begin{aligned} &D_{ij} = D(\vec{r}_{ij}), \\ &D(\vec{r}) = \tfrac{k_B T}{\eta L^3} \textstyle\sum_{\vec{k} \neq 0} \tfrac{\mathbb{1} - \hat{k} \otimes \hat{k}}{k^2} \exp(i\vec{k}\vec{r}), \end{aligned} \tag{7.20}$$

where $\vec{k} = (2\pi\vec{n})/L$ runs over the reciprocal lattice vectors of the MD box and $\vec{n}$ is a vector of integers. This form comes from the superposition of hydrodynamic modes. Note that, $\vec{k} = 0$ is excluded because this would correspond to an overall translational motion of the system and we study the diffusion of the chain relative to the fluid which is globally at rest. Also this exclusion is necessary not only for physical reasons, but also mathematically in order to keep the expression finite. An attempt to write down the hydrodynamic

interaction with periodic images by just adding up the infinite number of Oseen tensors in real space would lead to a mathematically ill-defined formula, because this series contains a $\vec{k} = 0$ component and hence it diverges. For this reason, the Fourier representation (Eq. 7.20) is more useful.

The diagonal elements also need to be modified due to the hydrodynamic interactions of beads with their own periodic images:

$$D_{ij} = D_0 \mathbb{1} + \lim_{r \to 0} \left(D(\vec{r}) - \frac{k_B T}{8\pi\eta r}(\mathbb{1} + \hat{r} \otimes \hat{r}) \right), \tag{7.21}$$

where $\hat{r} \otimes \hat{r}$ is the tensor product of the unit vector in the $\vec{r_{ij}}$ direction with itself.

Then, for cubic simulation box of length L, the diffusion coefficient corrected for system-size effects is equal to:

$$D_{cor} = D_{PBC} + \frac{2.837297 k_B T}{6\pi\eta L}, \tag{7.22}$$

where D_{PBC} is the diffusion coefficient of the simulation and the constant 2.837297 was obtained firstly by Dünweg and Kremer [182, 186].

7.2 Results

In the following section, we will first focus on pure water systems to investigate the effect of finite system sizes on HI in atomistic and CG models. Then we will consider these aspects in a mixed system, namely the ALA_3 peptide in aqueous solution that had been introduced in the previous Chapters.

7.2.1 Pure water systems

As was shown in ref. [184] the diffusion coefficient D in molecular dynamics simulations with periodic boundary conditions is dependent on the box length, due to hydrodynamic interactions of the system with its periodic copies. To obtain the atomistic reference data for the pure water system, molecular dynamics simulations were performed for several box lengths with 1000, 2000, 3000, 5000, 10000 and 50000 water molecules at equal density. As a reference density, ρ_{1000}, we used the smallest system with 1000 water molecules. For details see

To determine the diffusion coefficient, first, the mean square displacement (MSD) was calculated which is linked to D via the Einstein relation [192]:

$$\lim_{t \to \infty} \left\langle ||r_i(t) - r_i(0)||^2 \right\rangle = 6Dt \tag{7.23}$$

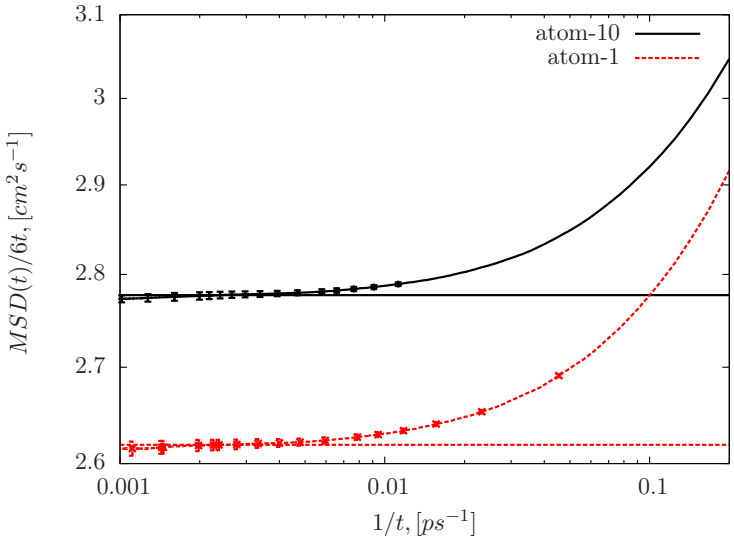

Figure 7.1: **Atomistic simulation of pure water, showing how to accurately determine the diffusion coefficient from $MSD(t)/6t$ vs $1/t$ in a double logarithmic scale.** For 1 thousand (red line, atom-1) and 10 thousands (black line, atom-10) atomistic water molecules, with 3.10633 nm and 6.69259 nm box sizes correspondingly.

To determine diffusion coefficients accurately the MSD/t vs. $1/t$ were plotted in the double logarithmic scale. This representation allows us to determine the exact plateau of the curve, which corresponds to the diffusive regime with sufficient statistics (two examples for 10^3 and 10^4 water molecules are presented in Fig. 7.1). Typical error bars are also shown in the Figure.

Since we want to investigate the dynamic properties of the systems, we have to use an appropriate thermostat. The stochastic velocity-rescaling thermostat by Bussi et.al.(Sec. 2.3.2) was used in our work, a global thermostat that produces a canonical ensemble and preserves HI. To demonstrate the contrast, the same simulations were performed with stochastic dynamics (sd, Langevin thermostat), see Fig. 7.2. In the case of the sd simulation, there is no dependence of the diffusion coefficient on the box size L. As can be seen from the figure, the expected linear dependence between the diffusion coefficient of the system and the inverse box length is present when the stochastic velocity-rescale thermostat by Bussi et.al. was used. From this plot, we can then find the finite-size corrected diffusion constant D_{cor}^{atom} for pure water in an atomistic simulation, which is $(2.93 \pm 0.01) \times 10^{-5} (cm^2/s)$ (Eq. 7.22).

In order to be certain that our results are correct, we performed one more round of simulations to check that the time coupling constant of the stochastic velocity-rescaling thermostat does not influence the hydrodynamic

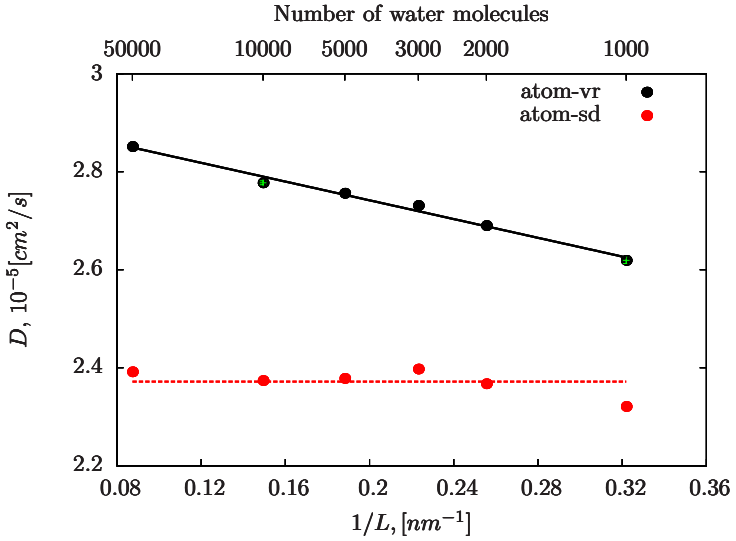

Figure 7.2: **Comparison of finite-size effect in simulations with stochastic velocity-rescale thermostat (vr) by Bussi et.al. and stochastic dynamics (sd).**

Table 7.1: **Diffusion coefficients and viscosity of water for the pure water system and mixed peptide-water system in atomistic and CG cases.**

system		$D_{cor} \times 10^{-5}(\frac{cm^2}{s})$	$\eta \times 10^{-4}$ Pa×s
pure water	atomistic	2.93±0.01	6.50±0.16
	CG	15.46±0.06	2.28±0.27
water with peptide	atomistic	2.93±0.06	5.82±0.82
	CG	15.44±0.04	2.27±0.19

finite-size effect. Fig. 7.3 shows data for three box sizes in three sets of simulations: with τ (see Sec. 2.3.3) equal to 0.05 (ps), 0.1 (ps), and 1 (ps). From this plot, one can conclude that the diffusion constants are independent of the thermostat settings. This means, that our results are independent of the thermostat settings.

The main idea of this chapter however is to find out the relation between dynamic properties in atomistic and CG simulations. Therefore, the same simulations have been performed for CG system, where the pressure-corrected CG water potential was determined with iterative Boltzmann inversion and used as described in Introduction and Chapter 4. In Fig. 7.4(a) it is shown that there is also finite-size effect for the case of coarse-grained water. It is important to mention that the time scale in this Figure does not yet account for the "speed-up" that is inherent to CG models because of the smoother energy landscape in the CG representation. In order to get the cor-

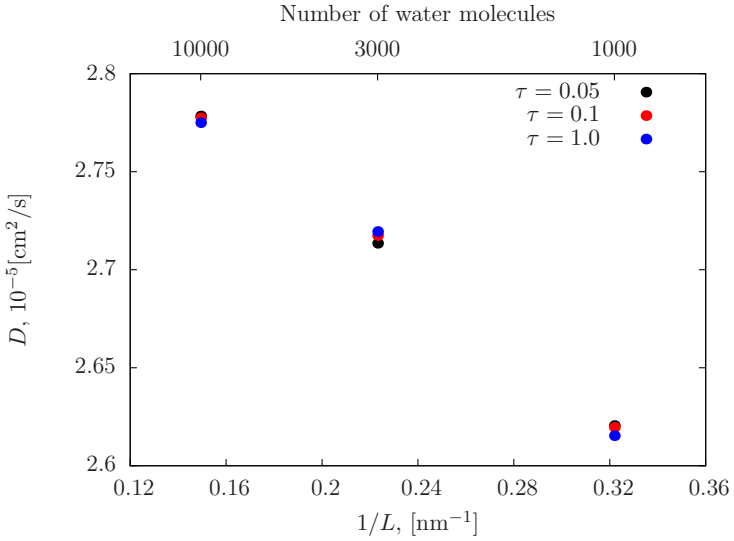

Figure 7.3: **Atomistic simulations of pure water with different time coupling constants.**

respondence between these time units and the atomistic ones, a *timemapping* has to be applied. For this the finite-size corrected diffusion coefficient D_{cor}^{CG} for the CG case was found to be $((15.46 \pm 0.06) \times 10^{-5}(cm^2/s))$ (in order to compare results for diffusion coefficients and viscosities in atomistic and CG cases, see Tab. 7.1)[1]. From the two diffusion coefficients, D_{cor}, at the atomistic and coarse-grained levels, one can now determine a time scaling factor, τ_D, which characterizes the speed-up of the dynamics in the CG compared with the atomistic system:

$$\tau_D = \frac{D_{cor}^{CG}}{D_{cor}^{atom}} = 5.28 \pm 0.04. \tag{7.24}$$

This time scaling factor cannot only be determined based on the finite-size corrected diffusion constants, but also based on the individual system size. The dependence of τ_D on the system size is shown in Fig. 7.4(b). Let us take Eq. 7.22 for the atomistic and coarse-grained cases and find the ratio of D_{PBC}^{CG} to the D_{PBC}^{atom}:

$$\frac{D_{PBC}^{CG}}{D_{PBC}^{atom}} = \frac{D_{cor}^{CG} - A\frac{1}{L}}{D_{cor}^{atom} - B\frac{1}{L}}, \tag{7.25}$$

[1]It should be noted that we used a KJ/mol as a unit for energy and nm as a unit for the distance. From this data, we obtain ps for time. We do not use any reduced units (such as for Leonard-Jones parameters σ and ϵ).

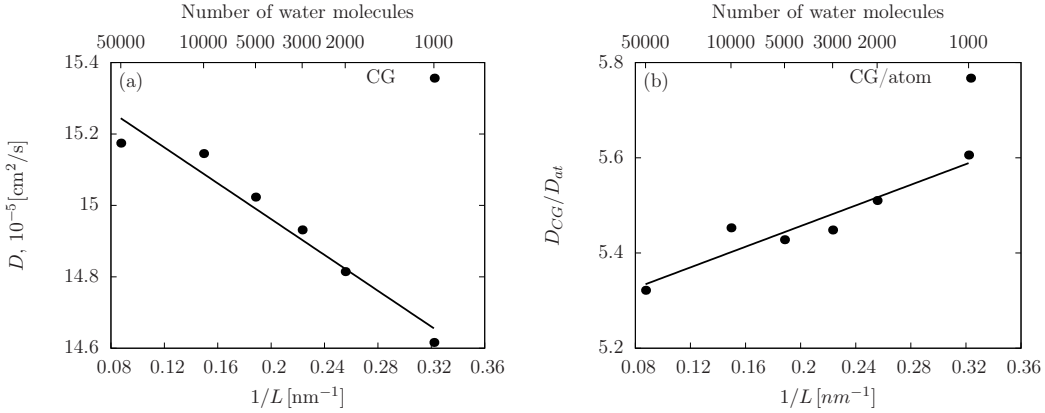

Figure 7.4: **Panel (a): dependence of diffusion coefficient on box length in CG simulation; panel (b): connection between diffusion coefficient of CG to atomistic simulation.**

where

$$A = \frac{2.837297 k_B T}{6\pi\eta_{CG}}, \tag{7.26}$$

and

$$B = \frac{2.837297 k_B T}{6\pi\eta_{atom}}. \tag{7.27}$$

The viscosity in the atomistic model is larger than that in the CG model (see Tab. 7.1). Therefore, coefficient B is smaller than A. Thus, Eq. 7.25 can be approximated as (for details see Appendix A.5.2):

$$\frac{D_{PBC}^{CG}}{D_{PBC}^{atom}} \approx \frac{D_{cor}^{CG}}{D_{cor}^{atom}} - \frac{A}{D_{cor}^{atom}} \frac{1}{L}. \tag{7.28}$$

From this equation, it already can be seen that there is a dependence between the ratio of diffusion coefficients for the CG and atomistic simulations and the box size. The scaling factor is now equal to the slope of the plot, 5.24 ± 0.04.

7.2.2 Mixed "peptide-water" system

Next, we compare the diffusion in atomistic and CG mixed "peptide-water" systems. When the peptide is added to the box, the concentration of the water changes. However, we would still like to keep the same water density as that without the peptide for all boxes. Therefore, an additional calculation of the number of water molecules for the new systems needs to be done. We have used the density of the smallest box (V_1) as reference. Each other box was obtained from the smallest one plus an additional box (ΔV), in which

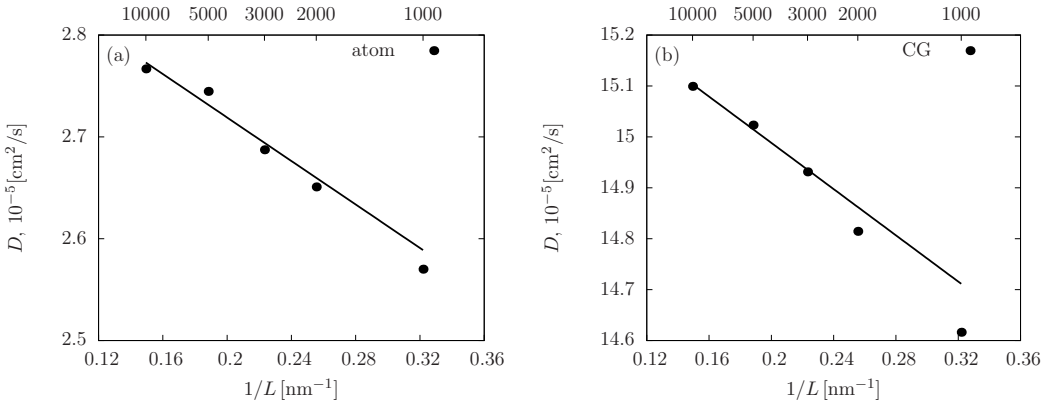

Figure 7.5: **Simulation of the mixed system "peptide-water", dependence of diffusion coefficients of the water on the box size.** (a) Atomistic simulations; (b) CG simulations.

the density of the water should be the same as in the system of pure water without a peptide. For details see A.5.3.

Fig. 7.5 shows that with this setup, the dependence of the diffusion coefficient of water in CG simulations (as well as in the atomistic cases) is the same in pure water and in mixed systems. For the mixed "peptide-water" system the diffusion coefficient of water was found to be $D_{mixed} = (15.44 \pm 0.04) \times 10^{-5}(cm^2/s)$ and the viscosity of water $\eta_{mixed} = (2.27 \pm 0.19) \times 10^{-4}(Pa \times s)$, which lies within the error bars of the same result for the pure water system, see Tab. 7.1. It is important to mention that for the mixed system, we used only data for the three biggest boxes (3, 5 and 10 thousands water molecules). The end-to-end distance of our peptide, as was shown previosly, is about 1 nm and the box sizes of the smallest systems (with 1 and 2 thousands water molecules) are approximately 3.1 and 4 nm. In this case, our solution is not exatly dilute but close to a semi-dilute regime, which can effect the result for the diffusion coefficient.

It is natural that we obtained the same dependence for the mixed system, because it is caused by interactions of particles with their periodic images through water and the additional peptide molecule would not change it.

Then, we investigated the finite size effect and the difference between atomistic and CG dynamics for the diffusion of the peptide molecule. First, we studied translational diffusion.

Long simulation times are necessary to obtain good statistics for a single peptide in water. Unfortunately, even 100 ns is not enough to obtain a flat plateau (which corresponds to the normal diffusion regime) on the MSD/t vs $1/t$ plot (Fig. 7.6(a)). However, we can take rough estimates of the diffusion coefficient from the atomistic and CG simulation (Fig. 7.6(b)), and find ap-

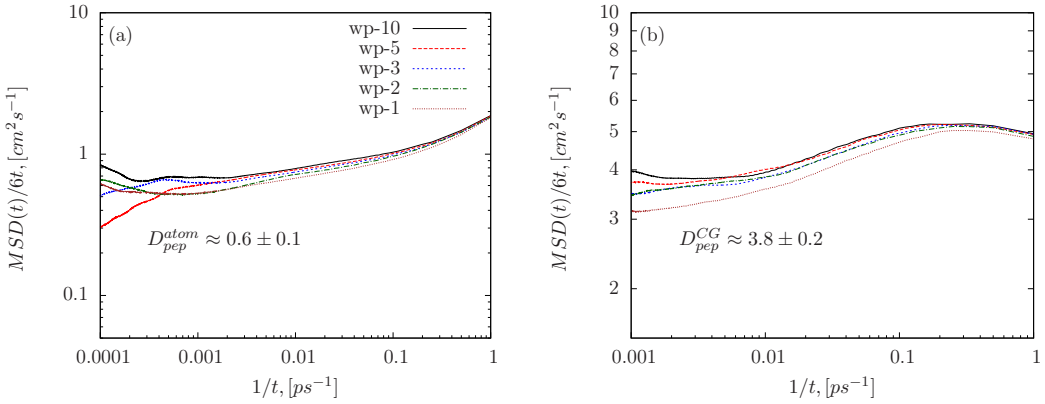

Figure 7.6: **Distribution of $MSD(t)/6t$ vs $1/t$ for the peptide in different boxes in the double logarithmic scale.** Panel (a): atomistic simulations; panel (b): coarse-graining simulations.

proximately a scaling factor, which is in the order of $\tau_D^{pep} 3.8/0.6 \approx 6.3 \pm 1.1$. This scaling factor (within the error bars) is similar to the one for pure water, $\tau_D = 5.24$. Thus, the speed-up for water dynamics in the pure water system and the "water+peptide" mixed system are similar. This indicates that the processes that determine peptide diffusion resemble the ones that influence water diffusion, which is logical since it is due to friction and random forces of the water molecules.

In addition to the translational diffusion coefficient, we investigated another dynamic property in the peptide-water system, namely, rotational diffusion. For this, we can compute the rotation auto-correlation function:

$$C = \int_0^\infty \cos \angle(\vec{n}(\xi)\vec{n}(t+\xi))d\xi, \tag{7.29}$$

where $\vec{n}$ is end-to-end vector. Two correlation functions for atomistic and CG cases are presented in Fig. 7.7. The curves are fitted with the law $y = \exp(-t/\tau)$, where τ is the correlation time. From Table 7.2, it is seen that τ is approximately the same for each box size for the atomistic and CG cases. However, there is a difference between τ_{atom} and τ_{CG} and the scaling factor for this characteristic is:

$$\tau^{RAF} = \frac{\tau_{atom}}{\tau_{CG}} \simeq 13.3 \tag{7.30}$$

This value means, that the speed-up of different processes of peptide is different in CG simulations, because the scaling factor for rotational diffusion is larger than for translational diffusion (13.3 and 5.24). This can be a problem in simulations of protein, which have translational and rotational

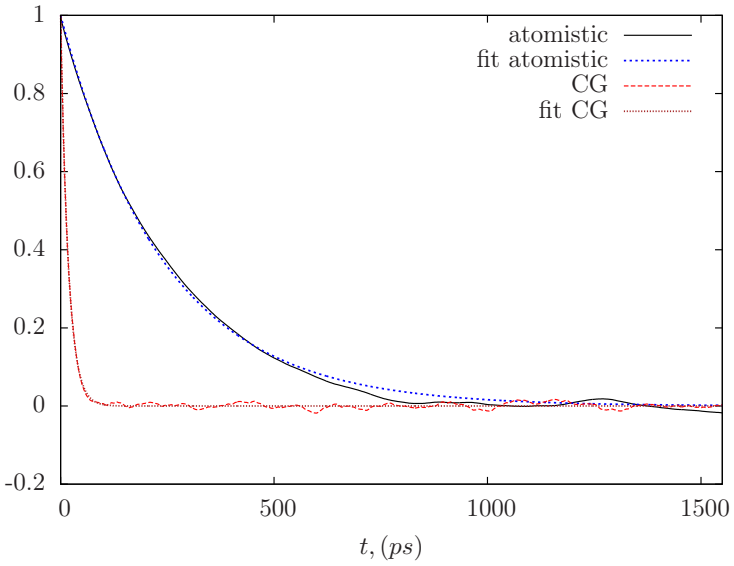

Figure 7.7: **Rotation correlation function for the case of approximately 5000 water molecules in atomistic (black line and blue line - fit) and CG (red line and brown line - fit) cases.**

motion. These processes are important for aggregation of proteins, but since the "speed-up" of dynamics of two movements is different, it is difficult to predict what kind of result will be obtained on the CG level.

To simulate not one but a system of several peptides, it is necessary to investigate real biological processes like aggregation phenomena. In Chapter 6, we have already considered such a system where we calculated the PMF between two peptides. Here, we want to investigate a dynamic property that depends on the peptide-peptide interaction in order to see how such a dynamic property is affected by coarse-graining. Therefore, we have studied a

Table 7.2: **Correlation times τ for the case of atomistic and CG simulations (CG simulations were done with the same potentials as in Chap. 4 and one additional run was done for a box with 3000 water molecules with additional 1,3,5-angle and 1-5 bond potentials (see Sec. 5.1.3)).**

number of water mol.	box length $L(nm)$	$\tau_{atom}(ps)$	$\tau_{CG}(ps)$	$\tau_{CG}^{2pot}(ps)$
1000	3.10633	229.75	18.32	
2000	3.91398	242.05	17.64	
3000	4.48010	250.76	17.70	22.43
5000	5.31182	242.54	17.88	
10000	6.69259	239.54	17.77	

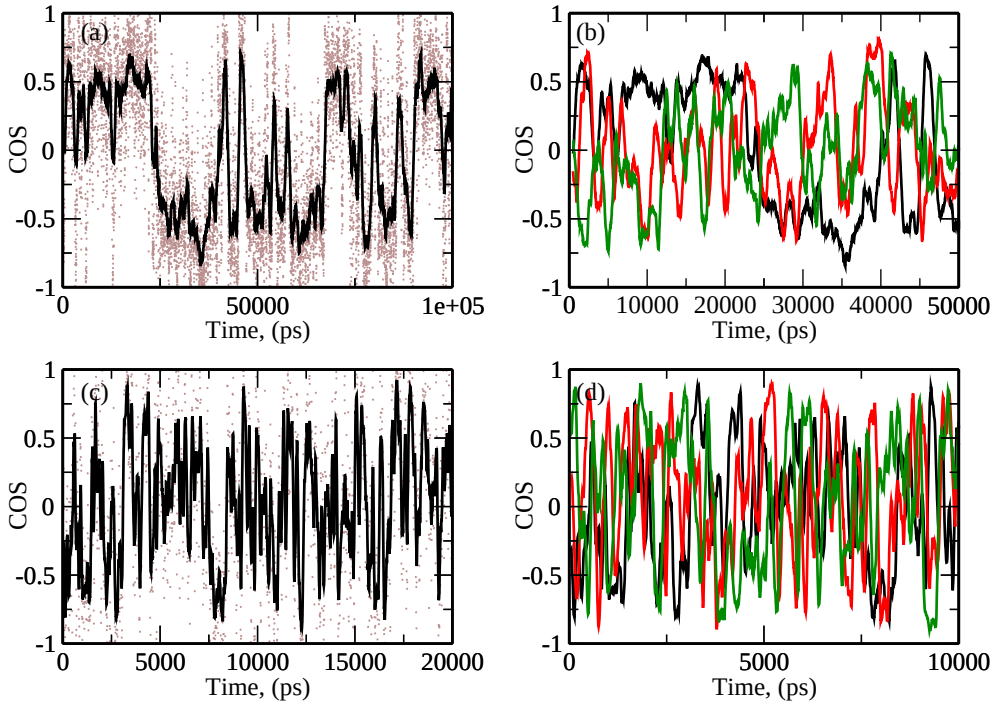

Figure 7.8: **Time dependence of the cosine of the angle between "end-to-end" vectors of two peptides ALA_2 in atomistic simulations.** (a) Distance between central PEP beads is equal to 0.4 nm, brown dots correspond to the time dependence of cosine, black line corresponds to the running average of this plot; (b) running averages for the cases when distance between central PEP beads are equal to 0.4 nm (black line), 0.5 nm (red line), 0.6 nm (green line); (c) distance between central PEP beads is equal to 1.4 nm, brown dots correspond to the time dependence of cosine, black line corresponds to the running average of this plot; (d) running averages for the cases when distance between central PEP beads are equal to 1.4 nm (black line), 1.5 nm (red line), 1.6 nm (green line). Running averages were done over each 100 ps for peptide at close distance and over each 10 ps for peptides far apart.

parameter which can characterize the orientation of peptides relative to each other: namely the angle between the two end-to-end vectors of the peptides. From the simulations performed to compute the peptide-peptide PMF using distance constraints between the two peptides centers (Chapter 6) we need to consider two cases: (1) Simulations at constraint distances where the peptides are close to each other (0.4, 0.5 and 0.6 nm) and (2) simulations where they are far apart and the distance between them is larger than the cut-off of the non-bonded potentials (1.4 nm). For these two sets of simulations we can compute the angle, θ, between the two end-to-end vectors. (The distributions of $\cos\theta$ have already been shown in Chapter 6, Fig. 6.9). In order to study the dynamics of this peptide-peptide reorientation process, we plan

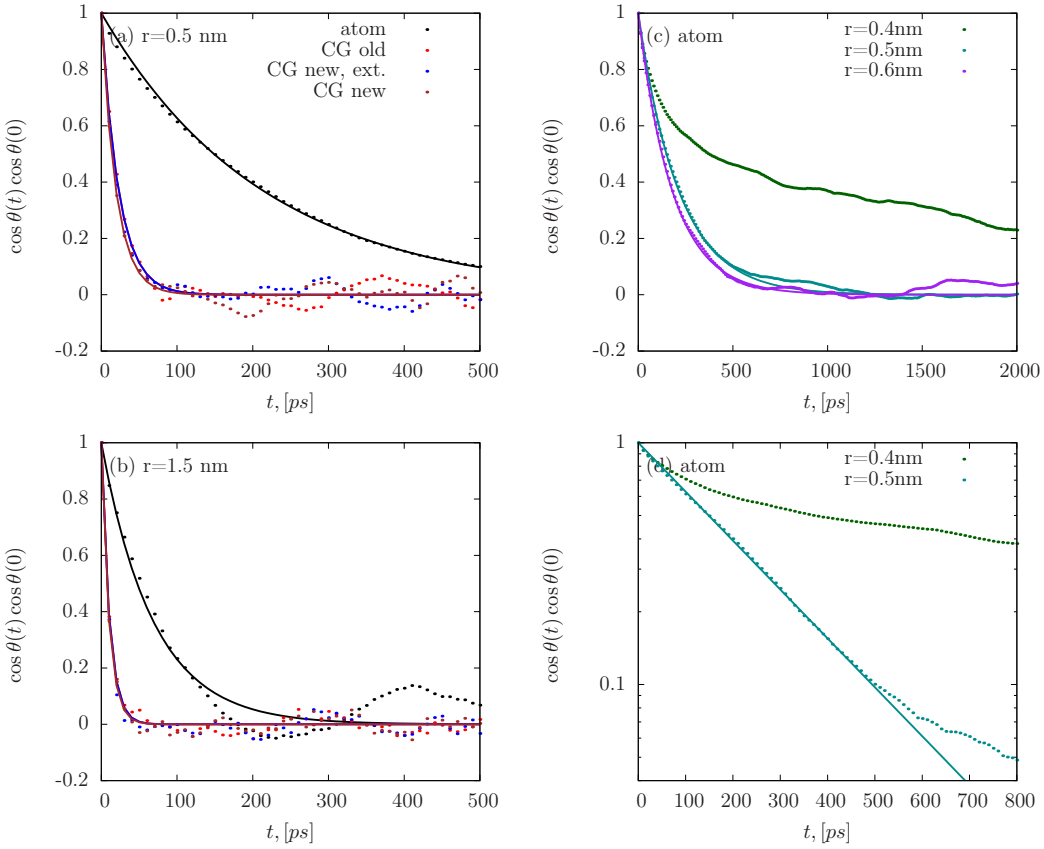

Figure 7.9: **Autocorrelation function of $\cos\theta$ between "end-to-end" vectors of $\mathbf{ALA_2}$.** (a) Autocorrelation function for atomistic and CG simulations for the distance between peptides central equal to 0.5 nm; (b) autocorrelation function for atomistic and CG simulations for the distance between peptides central equal to 1.5 nm; (c) autocorrelation functions for three atomistic sets with distances between centers of peptides equal to 0.4 nm, 0.5 nm and 0.6; (d) autocorrelation functions for atomistic cases with distance 0.4 and 0.5 nm, in higher resolution for x-axis.

to compute the correlation time from the autocorrelation function of $\cos\theta$. But before this is done, firstly, we need to check whether the sampling of this reorientation process has been sufficient. Especially in atomistic simulations at close peptide distances this is critical, since the reorientation is expected to be slow due to specific interactions, e.g. hydrogen bonds. To this end, we have monitored $\cos\theta$ as a function of time (Fig. 7.8). In order to make the transitions more visible, we perform a running average. Non averaged data are indicated as dotted lines in panels (a) and (c); a sensible time window for the running average has to be significantly shorter than the relevant correlation time and was chosen to be 100 ps for peptides at close distances and 10 ps for peptides far apart. One can see that closely arranged peptides

have two preferred positions corresponding to approximately parallel peptide orientations (0 and π angles), panels (a, b). Due to the method of preparation of the PMF calculations, all systems have approximately the same conformation for each distance at first. One can see in Fig. 7.8(b and d) that the memory of the starting conformation is short and the pairs of peptides placed at different distances from each other have different orientations at the same moments.

Almost the same conclusions can be made from Fig. 7.9, where the autocorrelation functions for $\cos\theta$ for different distances are presented. On panels (a) and (b), the autocorrelation functions for the distances, between central PEP beads, equal to 0.5 and 1.5 nm correspondingly are shown. The curves for the atomistic simulation (black line) and for the CG simulations with three sets of potentials (see Sec. 6) demonstrate that the relaxation times in CG cases are smaller than in those atomistic case for both types of distances. A special case is at distance the 0.4 nm (panel c). For such a close position of peptides, the relaxation time is difficult to measure (panel d). It seems as if the peptides get stuck. As was discussed in Chapter 6, it can be two possible reasons:

- A steric effect.
- The "end" beads (first and last) can come closer than 0.4 nm and between them hydrogen bonds may be formed.

And as we concluded before, the steric effect plays significant role in this behaviour.

At the same time, we can see that the dynamics of two peptides are faster when they are far apart and do not feel each other any more compared to the close placement (Fig. 7.9(a,b)). The same behavior occurs in the CG cases (when peptides are far apart they have smaller relaxation times than in close position). We can see that the dynamics of peptides on the close and far distances for atomistic simulations are different by approximately factor of 3:

$$\frac{\tau_{vect}^{r=0.5nm}}{\tau_{vect}^{r=1.5nm}} \simeq \frac{208.68}{68.58} \simeq 3, \tag{7.31}$$

and for CG simulation we have a factor of 2:

$$\frac{\tau_{vect}^{r=0.5nm}}{\tau_{vect}^{r=1.5nm}} \simeq \frac{22.34}{10.65} \simeq 2. \tag{7.32}$$

The "speed-up" factors (s) between atomistic and CG cases for the close and far distances are also slightly different:

- $r = 0.5nm$

$$s_{r=0.5} = \frac{\tau_{vect}^{atom}}{\tau_{vect}^{CG}} \simeq \frac{208.68}{20} \simeq 10 \tag{7.33}$$

- $r = 1.5nm$

$$s_{r=1.5} = \frac{\tau_{vect}^{atom}}{\tau_{vect}^{CG}} \simeq \frac{68.58}{10} \simeq 6 - 7 \tag{7.34}$$

What is also remarkable is that the atomistic movement of the peptides at the close distance in 0.4 nm between them is very slow in comparison to larger distances (0.5 nm and 0.6 nm), but in CG case the correlation times for all three close distances are approximately the same (see Tab. 7.3). This fact can point out that we miss in our CG model some essential atom-atom interactions which appear at the close distances (like hydrogen bonds) which "lock" the peptides in one position for quite a long time (compared to placement at lager distances).

All the relaxation times for all our sets have been presented in the Table 7.3. A delicate finite size effect influences the dynamics of two (or more) peptides as well. If one considers the interaction between them in water, the volume exclusion as well the drag of water should be taken into account, because when one peptide moves it pushes the water away and this water will also carry along another peptide. Furthermore, if the simulation box is small, the influence of such movement would be more pronounced for periodic images.

At the end, we checked the CPU time which one needs for atomistic and CG simulations. For a test, we performed a simulation with 1000 water molecules for 100 ps on one processor for atomistic and CG levels. The time step was 0.001 ps. It was determined that real simulation times for the atomistic case was 627.627 seconds and for CG simulations it was 358.911 seconds. So, the "speed-up" is approximately 1.7. But due to the softer potential landscape one can also increase the time step, which will lead to the acceleration of other process as well. We tried two bigger time steps: 0.002 and 0.01, which are in 2 and 10 times larger than the atomistic time step. Real simulation times for these two cases are approximately the same: 336.691 and 329.141 seconds correspondingly, but this time is smaller than for the CG simulation with the same time step as the atomistic simulation. The "speed-ups" concerning the atomistic case are 1.86 and 1.91 correspondingly. However, if one looks at the performance, we find out that the speed of the atomistic simulation is 13.766 ns/day and the CG simulation with time step equal to atomistic one is 24.073 ns/day. For two other CG simulations with bigger time steps the speed of the simulation is 51.323 ns/day for time steps equal to 0.002 and 262.504 ns/day when the time step 10 times bigger than

Table 7.3: **Correlation times, τ_{vect}, for the orientation dynamics of two peptides.**

r, nm	case	$\tau_{vect}(ps)$	$\tau_{vect}^{atom}/\tau_{vect}^{CG}$
0.4	atomistic	906.62	
	CG old	27.86	32.54
	CG new	18.02	50.31
	CG new, ext.	18.29	49.57
0.5	atomistic	208.68	
	CG old	22.34	9.34
	CG new	19.46	10.72
	CG new, ext.	20.59	10.13
0.6	atomistic	177.04	
	CG old	19.29	9.18
	CG new	21.50	8.23
	CG new, ext.	22.37	7.91
1.4	atomistic	59.61	
	CG old	11.03	5.40
	CG new	11.55	5.16
	CG new, ext.	10.76	5.54
1.5	atomistic	68.58	
	CG old	10.65	6.44
	CG new	9.96	6.89
	CG new, ext.	9.57	7.17
1.6	atomistic	65.06	
	CG old	10.48	6.21
	CG new	9.51	6.84
	CG new, ext.	9.08	7.16

atomistic one. Therefore, these data demonstrate that CG simulations allow one to simulate systems faster than atomistic simulations do.

In conclusion, we have considered the dynamics of peptides in water in the atomistic and CG cases and found out that different dynamics of peptides have different scaling factors between atomistic and CG simulation cases, which can be a problem when one wants to investigate such processes as aggregation of peptides at the CG level, where different types of behavior play significant role. Also, we have found out that the dynamics of two molecules depends on their position (if they are far away or close to each other) and that in the CG simulations, we can repeat such tendencies although the factor is slightly different, which also should be taken into account during CG investigations of the system.

Chapter 8

Conclusion

In this thesis, a coarse-grained model for alanine oligomers, ALA_n, in aqueous solution was developed with the main aim to reproduce the conformational distributions of the respective peptides in all atom reference simulations. Several aspects of coarse-grained modeling were investigated starting from conformational properties, via nonbonded interactions to questions concerning the dynamics in coarse-grained simulations.

To parametrize the CG interactions, we extended a methodology previously developed by Villa et al. ([60, 61]) for amphiphilic dipeptides to ALA_3 which has in the CG description the form of a linear chain of two types of beads. First, we encountered the fact, that our initial model with CG potentials directly obtained from Boltzmann inversion did not reproduce the atomistic structures. This was different from the observations made for diphenylalanine where Boltzmann inverted bond, angle and dihedral distributions resulted in a CG model that reproduce the atomistic conformations with good agreement. The reason for this discrepancy most likely lies in the fact that for the linear CG model for oligoalanines without any side chains the influence of the solvent environment on the angle distributions is stronger than in the case of diphenylalanine where the backbone geometry is stiffer due to several angle and improper torsion potentials at the branching points of the sidechains. The observed discrepancy in angle distributions between CG and atomistic ALA_3 could be corrected by a new set of CG angle potentials obtained from iterative Boltzmann inversion. It was observed that the iterative correction to the angle potentials did not have any negative effect on the sampling of other CG degrees of freedom such as bonds or other torsions. On the contrary, the sampling of one of the backbone dihedral distributions was also improved with the corrected angle potential. This interdependence led us to conclude that these intramolecular distributions are in fact not entirely independent and that therefore a certain sequence (starting with the stiffer

degrees of freedom) of parametrizing CG potentials is advisable instead of iterating several potentials simultaneously.

A CG model for ALA_3 with CG bonds, angles, and dihedrals is well capable of reproducing local conformational properties, but already properties not used for parametrization such as the distance distribution between two peptide groups separated by five CG bonds or the angle distribution between three consecutive peptide groups are not reproduced anymore. For longer chains this results in a non negligible discrepancy in the end to end distribution, making the initial model rather unsatisfactory. In order to reproduce also these properties, additional CG potentials along the peptide backbone were needed, for example 1,5 PEP-PEP distance or 1,3,5 PEP-PEP-PEP angle potentials. The need for these potentials can be explained by the fact that these 1,3,5 PEP-PEP-PEP angles and 1,5 PEP-PEP distances are characteristic for certain secondary structure propensities of the peptide chain (for α helical and β strand chain segments). Since secondary structure formation is intimately linked with intrachain correlations, most importantly between different backbone dihedral angles, the factorization into independent bond angle and torsions which is assumed for the Boltzmann inversion procedure does not hold for these biomolecular systems. For the mapping scheme chosen in this thesis correct backbone conformational sampling could be achieved with the above mentioned additional interaction functions between consecutive peptide units along the backbone.

By construction we lose some microscopic properties of the peptide in the coarse-graining procedure. As a consequence of the assumption that the degrees of freedom are independent so that one can obtain potentials via the Boltzmann inversion procedure we lose for example some correlations between CG backbone angles and torsions that had been present in the original atomistic sampling. After a backmapping procedure that reintroduces atomistic degrees of freedom into the CG conformations in such a way that the atomistic structures lie on top of the CG reference conformations we investigated the structural properties of the resulting backmapped ensemble. We could show that this backmapped ensemble again reproduces all these correlations with satisfactory agreement with the original atomistic data. This is an important observation that shows that the CG sampling is sufficiently close to the original atomistic one that some properties which are lost in the coarse-graining process can be regained after backmapping, which is important for the subsequent analysis of the backmapepd ensemble in terms of secondary structures and for comparison to experiment. We also extended the ALA_3 model to ALA_4 and showed that the parameters determined for the shorter oligopeptide (ALA_3) can be transferred to the longer peptide sequence.

In a next step we revisited the nonbonded interaction functions in the CG model, which had initially been taken from previous work by Villa et al ([60, 61]). We investigated the influence of a pressure correction on the nonbonded interactions and more importantly the influence of the nature of the fragment molecules which were used in the parametrization procedure based on calculations of potentials of mean force. We found that in order to obtain the correct peptide-peptide association behavior it is of importance to include the influence of the surrounding chain in the parametrization of the nonbonded interaction between two peptide beads. Thus, one should base the nonbonded interactions not on potentials of mean force between two fragments corresponding to individual coarse-grained beads but rather on small pieces of the chain molecule that also include the neighboring beads.

We found that the bonded interactions are heavily coupled with the nonbonded interaction functions, in particular in a case where bonded interactions need to be refined by iterative Boltzmann inversion in presence of the nonbonded interactions, as is for example the case in some of the peptide angles. Here the iteratively refined potentials obtained for one set of nonbonded interactions could not be directly used with the newer set developed later in the course of the projects. Therefore, we recommend the following order for the coarse-graining process: first, nonbonded potentials for the surrounding medium (CG water) and the interactions of the peptide beads with this medium should be obtained, then nonbonded potentials between different peptide beads should be obtained based on molecule fragments accounting for the neighboring beads; then, a first set of bonded (bond, angle, torsion) potentials should be obtained from straightforward Boltzmann inversion of the respective distributions (assuming the degrees of freedom are independent), then one might refine some of these bonded potentials (iteratively, here one should follow the order to start with the stiffest degrees of freedom since these interactions are potentially interdependent), and last, if it is necessary one can add some additional special (bonded or nonbonded) interaction functions in order to reproduce overall (less local) conformational properties of the chain such as secondary structure elements etc.

In the present thesis, we have developed a CG model for oligoalanines that reproduces the peptide-peptide association behavior obtained form atomistic simulations and a procedure to determine bonded interaction functions such that the correct chain conformations are obtained. In the future we plan to apply this peptide model to aggregates of several oligo alanine peptide strands, for example small crystalline units of oligo alanine beta sheets, again comparing to atomistic reference simulations in order to assess the quality of the model for such a situation. Such oligoalanine aggregates are found for example in spider silk fibers and are assumed to be important for the re-

markable mechanical properties of these fibers. Here, a coarse-grained model is an important step towards studying such fibers at mesoscopic time and length scales that are important for the overall properties of the materials.

In the last part of the present thesis we investigated dynamic properties of the peptide in water on two levels: atomistic and coarse-grained. It is well known that in CG simulations due to smoother potentials and less number of degrees of freedom processes are faster than in an atomistic model. We investigated the dynamics in the CG model for different types of dynamic processes, including the influence of hydrodynamic interaction of translational and rotational diffusion of the peptides. We found that translational diffusion experiences a different "speed-up" upon coarse-graining than rotational reorientation, and – to make matters even more complicated – that the speed up of the dynamics of relative reorientation of two peptides depends on the peptide distance, probably due to hydrodynamic interactions between the two peptides. It is expected that these phenomena exhibit in addition strong finite-size effects of the periodic system and further dependence on the peptide concentration. Since all these aspects are of importance if one wants to use such models to study aggregation phenomena further research along those lines is planned for the future.

Appendix A

For all simulations we used the GROMACS program suite [104, 193].

A.1 Abbreviations used in the thesis

Abbreviations	Meaning
CG	coarse-graining
BM	backmapped
CG impl. SOL	coarse-graining simulation with implicit solvent
CG expl. SOL (no corr.)	coarse-graining simulation with explicit solvent, without any corrections
CG expl. SOL (p-corr.)	coarse-graining simulation with explicit solvent and with pressure corrected water
CG expl. SOL (PCP corr.) & CG (PCP corr.)	coarse-graining simulation with corrected PEP-CAB-PEP angle and pressure corrected water
CG (PP 1-5 pair pot.)	coarse-graining simulation with explicit solvent, with pressure corrected water and additional 1-5 pair potential between "PEP" beads
CG (PPP 1,3,5-angle pot.)	coarse-graining simulation with explicit solvent, with pressure corrected water and additional 1,3,5-angle potential for PEP-PEP-PEP angles
CG (PP+PPP pot.)	coarse-graining simulation with explicit solvent, with pressure corrected water and additional 1,3,5-angle and 1-5 pair potentials together

A.2 Simulation details of atomistic system

MD simulations were carried out using the leap-frog algorithm with a time step of 0.002 ps. Constant temperature and pressure were maintained via the weak coupling algorithm [92], where the temperature was set to 300K with a coupling constant of 0.1 ps and the pressure was set to 1 bar with a coupling constant of 0.5 ps and a compressibility of 4.5e-5 bar^{-1}. The LINCS algorithm was applied to constrain all bonds [104]. Electrostatic interactions were computed with the PME method [194] with a real-space cut-off of 1.0 nm, and a cut-off of 1.4 nm was applied for Lennard-Jones interactions with long range dispersion corrections for energy and pressure. We used the GROMOS 53a6 force field [127] with the SPC/E water model [195, 196]. Atomistic simulations of Ala_3 were carried out for 90 ns, while the atomistic simulations of Ala_4 were carried out for 200 ns.

A.3 Simulation details of coarse-grained system

For the CG systems, a leap-frog stochastic dynamics integrator was used with a time step of 0.002 ps (note that in principle the time step could be increased in the CG model but we did not do so here) and an inverse friction constant of 0.1 ps. The temperature was set to 300 K, and simulations were carried out at constant volume conditions (at the average volume of the corresponding atomistic system at 300 K and 1 bar). Bonded and non-bonded interactions were computed using tabulated potentials. The bonded potentials were obtained using Boltzmann inversion procedure, but instead of directly employing the numerically inverted distributions as potentials in the simulation, the potentials are first smoothed using cubic splines (now this can be done with VOTCA package [75]). All CG simulations of Ala_3 were carried out for 100 ns and CG simulations for Ala_4 - for 100 ns.

Only the simulations for PMF calculations were done with NpT ensemble with a reference pressure equal to 1 bar with Berendsen barostat but also with the leap-frog stochastic dynamics integrator and with the time step of 0.002 ps.

A.4 Simulation details of back-mapped system

For backmapping, the energy minimization was done via a steepest descent algorithm. In subsequent simulations, the temperature was maintained at 300 K using a Langevin thermostat with a coupling constant of 1.0 ps. Individual simulations were run for 200 ps each.

In order to compare the conformational sampling of the CG model with the atomistic simulation, we (re)introduced atomistic details into the CG structures. This can be done using virtual sites (previously called "dummies") [134] corresponding to our CG units. Molecular dynamics simulations with position restraints on these virtual sites "force" (like in [197]) and equilibrating an initial atomistic structure on top of the CG one – initially dihedral restraints are applied to avoid unphysical trans-cis flips of the peptide group. The backmapping process was broken down into several steps. The CG structures to be backmapped were selected from the CG trajectory. For each of these structures, an initial atomistic conformation was selected from a database of atomistic structures (based on a similar inertia tensor and end to end distance) and then fitted on to the CG structure. The superimposed structures were then processed according to the following steps:

1. Energy minimization with harmonic position restraints force constants (100 $\mathrm{kJmol^{-1}nm^{2}}$) on the virtual sites and dihedral restraints ($180° \pm 70°$), which help to keep dihedral angle in peptide group in trans conformation.

2. Stochastic dynamics simulation with the same restraints.

3. Stochastic dynamics simulation with stronger position restraints (10000 $\mathrm{kJmol^{-1}nm^{-2}}$).

4. Water was added to the box.

5. Energy minimization.

6. Molecular dynamics with position restraints (10000 $\mathrm{kJmol^{-1}nm^{-2}}$) and dihedral restraints.

7. Molecular dynamics with position restraints (10000 $\mathrm{kJmol^{-1}nm^{-2}}$) without dihedral restraints.

We backmapped 100 randomly chosen structures from the CG trajectory and sampled (after initial preparation as described in Section II. E) 200 ps of

end atomistic simulation with an explicit solvent and virtual atom restraints to maintain the atomistic coordinates on top of the CG reference.

A.5 Simulation details for investigation of system's dynamics

A.5.1 Computational details

For investigation of dynamic properties of the systems "pure water" and "water+peptide" there are several boxes were considered. At first boxes with 1000, 2000, 3000, 5000, 10000 and 50000 water molecules were created. Later the peptide Ala_3 were added there. For the investigation of system's dynamics in atomistic and CG cases, a leap-frog algorithm with a time step of 0.001 ps in atomistic simulation and 0.002 ps in CG simulation were used with an inverse friction constant of 0.1 ps. The Constant temperature of 300 K were maintained via velocity rescaling [125] coupling with a stochastic term. The simulations were carried out at constant volume conditions (in CG case at the average volume of the corresponding atomistic system at 300 K and 1 bar). Bonded and nonbonded interactions in CG simulation were computed using tabulated potentials. All atomistic and CG simulations of pure water and water with Ala_3 were carried out for 100 ns.

For pure water, were done additional simulation with a leap-frog stochastic dynamics integrator with a time step of 0.001 ps in atomistic simulation and 0.002 ps and a coupling constant of 0.1 ps.

A.5.2 Inspection of CG and atomistic ration dependence on box size

Correspondingly on the plot 7.4(b) there is the dependence of ratio of diffusion coefficients in CG and atomistic simulation from box size. The dots indicating diffusion coefficients for each simulation box were fitted with equation $F(x) = ax + b$ for atomistic and CG simulations (plots 7.2 and 7.4(a)), and the value of these coefficients are $a = 1.08 \pm 0.18$(nm) and $b = 5.24 \pm 0.04$, x corresponds to the $1/L$. This equation we can compare with Eq. 7.28. Then, we get such correspondences:

$$\begin{aligned} a &\cong \tfrac{A}{D_{cor}^{atom}}, \\ b &\cong \tfrac{D_{cor}^{CG}}{D_{cor}^{atom}} \end{aligned} \tag{A.1}$$

From the extrapolation of plots Fig. 7.1(a) and Fig. 7.4(a) we got the meaning of "real" diffusion coefficient for the atomistic and coarse-graining simulations: $D_{cor}^{atom} = (2.93 \pm 0.01) \times 10^{-5}(\mathrm{cm}^2/\mathrm{s})$ and $D_{cor}^{CG} = (15.46 \pm 0.06) \times 10^{-5}(\mathrm{cm}^2/\mathrm{s})$; and also the meaning of constant $A = (2.51 \pm 0.29) \times 10^{-12}(\mathrm{cm}^3/\mathrm{s})$ (in principle, the value of A is negative, because of the slope, as it can be seen from the plot Fig. 7.4(a), but there is the sign minus before the constant A in the Eq. 7.28, so, lets use just the plus, as in the equation for the fitting). Constant A was obtained from plot's slope and the order was calculated as:

$$\begin{array}{l} A = \frac{2.837297 k_B T}{6\pi\eta_{CG}} = \frac{2.837297 * 1.38 * 10^{-23} * 300}{6 * \pi * 10 * -5} \frac{\mathrm{J*K*m^3}}{\mathrm{K*J*s}} \sim \\ \sim 10^{-18} \frac{\mathrm{m}^3}{\mathrm{s}} \sim 10^{-12} \frac{\mathrm{cm}^3}{\mathrm{s}} \end{array} \tag{A.2}$$

Here, we took into account that the value of water viscosity is in the order of $10^{-5} \frac{\mathrm{J*s}}{\mathrm{m}^3}$. With all these data we can calculate, what kind of values should correspond to the constants a and b in the fitting equation. Thus, $A/D_{cor}^{atom} = 0.86 \pm 0.10(\mathrm{nm})$ and $D_{cor}^{CG}/D_{cor}^{atom} = 5.28 \pm 0.04$, which is in good agreement with data obtained directly from the extrapolation of the plot 7.4(b):

$$\begin{array}{l} a \cong \frac{A}{D_{cor}^{atom}} \rightarrow 1.08 \pm 0.18 \cong 0.86 \pm 0.10 \\ b \cong \frac{D_{cor}^{CG}}{D_{cor}^{atom}} \rightarrow 5.24 \pm 0.04 \cong 5.28 \pm 0.04 \end{array} \tag{A.3}$$

A.5.3 Process of keeping the same density for all boxes

We decided to use in our simulations the density, which was found for the smallest box, in our case it is the box with 1000 water molecules. The density was calculated within usual equation:

$$\rho = \frac{Nm}{L^3} = const. \tag{A.4}$$

For all bigger boxes we devided the volume for the volume equal to the smallest box (V_1) and suround volume (ΔV) which we call "additional" box. We assume that the density at the V_1 is already the same as at the reference box (with 1000 water molecules). Then, we should find the number of water molecules, which we already have in "additional" box:

$$\Delta N = N_{wp}^x - N_{wp}^1, \tag{A.5}$$

where N_{wp}^x is the number of water molecules in the system (of x water molecules) with peptide, N_{wp}^1 is the number of water molecules in the smallest box after adding the peptide. Then the volume of "additional" box was calculated like a difference of volumes between box of system with x molecules and the smallest box:

$$\Delta V = V_x - V_1. \tag{A.6}$$

Then we used the same value for the density, as before (Eq. A.4):

$$\frac{N_1}{V_1} = \frac{\Delta N_{aim}}{\Delta V}, \tag{A.7}$$

where N_1 and V_1 are the number of molecules (1000) and box volume for the smallest box, ΔN_{aim} is the number of water molecules, which suppose to be in "additional" box for getting the same density as without peptide, and ΔV is the volume of "additional" box. From this equation:

$$\Delta N_{aim} = N_1 \left(\frac{\Delta V}{V_1} \right). \tag{A.8}$$

After all this steps it is possible to find out, if some water molecules should be added to or removed from the box:

$$\delta N = \Delta N_{aim} - \Delta N^x_{wp} \tag{A.9}$$

A.5.4 Hydrogen bond existence map

Hydrogen bond existence map is done with GROMACS program - *g_hbond*. It shows the presence of hydgrogen bond at each moment of simulation time. Also it shows not just abstract hydrogen bonds in the system, but for the concrete group of "donor-hydrogen-acceptor". On the y-axis numbers corresponds to the number of line in file with the list of such groups (hbond.log). This list is presented in Tab. A.1. Names of the atoms in this list correspond to the names in conformational file (conf.gro), see Tab. A.2. In this case we have two similar peptides. In order to distinguish between them we named one ALA (from alanine) and another one PEP (from peptide). So, the numbers from y-axis of Fig. 6.10 and Fig. 6.11 correspond to the ones from first column of Tab. A.1.

Table A.1: List of "donor-hydrogen-acceptor" groups in the system.

Number	Donor	Hydrogen	Acceptor
0	PEP4N	PEP4H	ALA1O
1	PEP4N	PEP4H	ALA2O
2	PEP4N	PEP4H	ALA3O
3	PEP4N	PEP4H	PEP1O
4	PEP4N	PEP4H	PEP2O
5	PEP3N	PEP3H	ALA1O
6	PEP3N	PEP3H	ALA2O
7	PEP3N	PEP3H	ALA3O
8	PEP3N	PEP3H	PEP1O
9	PEP2N	PEP2H	ALA1O
10	PEP2N	PEP2H	ALA2O
11	PEP2N	PEP2H	ALA3O
12	ALA4N	ALA4H	ALA1O
13	ALA4N	ALA4H	ALA2O
14	ALA4N	ALA4H	PEP1O
15	ALA4N	ALA4H	PEP2O
16	ALA4N	ALA4H	PEP3O
17	ALA4N	ALA4H	PEP4N
18	ALA3N	ALA3H	ALA1O
19	ALA3N	ALA3H	PEP1O
20	ALA3N	ALA3H	PEP2O
21	ALA3N	ALA3H	PEP3O
22	ALA2N	ALA2H	PEP1O
23	ALA2N	ALA2H	PEP2O
24	ALA2N	ALA2H	PEP3O
25	ALA2N	ALA2H	PEP4N

Table A.2: List of "donor-hydrogen-acceptor" groups in the system.

Molecule	Atomtype	Index	x	y	z	V_x	V_y	V_z
1ALA	CA	1	0.868	1.708	2.904	-0.1079	0.5120	0.1229
1ALA	C	2	0.850	1.564	2.952	0.0957	0.3648	-0.2416
1ALA	O	3	0.816	1.543	3.068	-0.4433	0.2596	-0.4179
2ALA	N	4	0.865	1.472	2.857	-0.3681	0.0992	-0.0600
2ALA	H	5	0.876	1.503	2.762	-1.9325	-0.8982	-0.6058
2ALA	CA	6	0.862	1.328	2.886	0.1408	0.1388	0.1929
2ALA	CB	7	0.773	1.257	2.784	0.1194	0.4129	0.0212
2ALA	C	8	0.996	1.253	2.885	0.2147	0.2717	0.1498
2ALA	O	9	1.100	1.298	2.838	0.7828	-0.3241	0.8166
3ALA	N	10	1.001	1.157	2.977	-0.3107	0.0581	-0.0387
3ALA	H	11	0.933	1.141	3.049	-0.8430	1.5990	-0.1627
3ALA	CA	12	1.114	1.065	2.992	-0.1896	0.1171	-0.5571
3ALA	CB	13	1.168	1.073	3.135	-0.0725	-0.5135	-0.5627
3ALA	C	14	1.091	0.918	2.952	-0.1439	-0.1179	0.2664
3ALA	O	15	0.987	0.855	2.968	-0.1070	-0.2202	0.1067
4ALA	N	16	1.204	0.852	2.927	0.3268	0.9006	-0.3870
4ALA	H	17	1.278	0.870	2.991	-1.2776	2.6393	1.1097
5ALA	CA	18	1.204	0.719	2.863	0.0702	0.7551	-0.0882
1PEP	CA	19	1.514	0.989	2.277	-0.5805	0.4301	-0.2078
1PEP	C	20	1.457	0.956	2.416	0.3173	-0.2820	-0.0014
1PEP	O	21	1.456	0.840	2.459	-0.2323	-0.4117	-0.3504
2PEP	N	22	1.378	1.054	2.458	-0.1694	-0.4685	-0.4725
2PEP	H	23	1.346	1.134	2.408	-1.9855	-0.1260	1.1294
2PEP	CA	24	1.310	1.031	2.586	-0.0242	0.0971	-0.2912
2PEP	CB	25	1.360	1.121	2.699	0.2167	-0.8364	0.3560
2PEP	C	26	1.161	1.059	2.564	-0.0944	-0.0070	0.0554
2PEP	O	27	1.115	1.173	2.567	0.0774	0.0648	-0.0569
3PEP	N	28	1.091	0.952	2.527	0.0001	-0.4345	1.0854
3PEP	H	29	1.135	0.875	2.481	-1.8280	-0.5570	-0.5904
3PEP	CA	30	0.944	0.966	2.530	-0.0234	-0.2545	-0.4050
3PEP	CB	31	0.903	0.955	2.383	0.1843	-0.0136	-0.4821
3PEP	C	32	0.883	0.866	2.628	-0.3236	-0.1496	-0.4845
3PEP	O	33	0.909	0.745	2.633	-0.2903	-0.1132	0.3312
4PEP	N	34	0.762	0.903	2.668	-0.1878	-0.1833	-0.0429
4PEP	H	35	0.723	0.988	2.632	-0.1426	-0.8721	-1.7985
5PEP	CA	36	0.678	0.837	2.769	0.1512	-0.5025	0.0327

Bibliography

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular biology of the Cell*. Garland Science, 4 edition, 2002.

[2] A. Finkelstein and O. Ptitsyn. *Protein physics: a course of lectures*. Academic Press, Amsterdam; Boston, 2002.

[3] R. Nussinov and G. Schreiber, editors. *Computational Protein-Protein Interactions*. CRC Press Taylor&Francis Group edition, 2009.

[4] J. G. Hardy and T. R. Scheibel. Composite materials based on silk proteins. *Progress in Polymer Science*, 35(9):1093 – 1115, 2010. Topical Issue on Biomaterials.

[5] C. Vepari and D. L. Kaplan. Silk as a biomaterial. *Progress in Polymer Science*, 32(8-9):991 – 1007, 2007.

[6] L. Brunetta. *Spider silk : evolution and 400 million years of spinning, waiting, snagging, and mating*. Yale University Press, New Haven, 2010.

[7] R. V. Lewis. Spider silk: ancient ideas for new biomaterials. *Chemical Reviews*, 106(9):3762–3774, 2006. PMID: 16967919.

[8] Spider silk: from soluble protein to extraordinary fiber. *Angewandte Chemie (International Ed. in English)*, 48(20).

[9] M. J. Buehler, S. Keten, and T. Ackbarow. Theoretical and computational hierarchical nanomechanics of protein materials: Deformation and fracture. *Progress in Materials Science*, 53(8):1101 – 1241, 2008.

[10] D. Woolfson and M. Ryadnov. Peptide-based fibrous biomaterials: some things old, new and borrowed. *Current Opinion in Chemical Biology*, 10(6):559 – 567, 2006. Model systems / Biopolymers.

[11] T. Scheibel. Protein fibers as performance proteins: new technologies and applications. *Current Opinion in Structural Biology*, 16:427–433, 2005.

[12] C. Hayashi. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International Journal of Biological Macromolecules*, 24(2-3):271–275, 1999.

[13] http://designmatrix.wordpress.com/2009/01/28/the-rational-essence-of-proteins-and-dna/.

[14] http://xray.bmc.uu.se/Courses/bioinformatik/Intro/Introduction.html.

[15] G. Colombo, P. Soto, and G. E. Peptide self-assembly at the nanoscale: a challenging target for computational and experimental biotechnology. *TRENDS in Biotechnology*, 25(5):211–218, 2007.

[16] X. Zhao and S. Zhang. Self-Assembling nanopeptides become a new type of biomaterial. volume 203 of *Advances in Polymer Science*, pages 145–170. Springer Berlin / Heidelberg, 2006. 10.1007/12_088.

[17] G. Payne. Biopolymer-based materials: the nanoscale components and their hierarchical assembly. *Current Opinion in Chemical Biology*, 11(2):214 – 219, 2007.

[18] G. Petsko and D. Ringe. *Protein Structure and Function*. New Science Press Ltd, London, UK, 2003.

[19] J. Clarke and L. Regan. Protein engineering and design: from first principles to new technologies. *Current Opinion in Structural Biology*, 20(4):480 – 481, 2010. Membranes / Engineering and design.

[20] P. Bryan and J. Orban. Proteins that switch folds. *Current Opinion in Structural Biology*, 20(4):482 – 488, 2010. Membranes / Engineering and design.

[21] C. E. Semino. Self-assembling peptides: From bio-inspired materials to bone regeneration. *Journal of Dental Research*, 87(7):606 –616, 2008.

[22] K. Sanford and M. Kumar. New proteins in a materials world. *Current Opinion in Biotechnology*, 16(4):416–421, 2005.

[23] R. Davies, A. Aggeli, A. Beevers, N. Boden, L. Carrick, C. Fishwick, T. McLeish, I. Nyrkova, and A. Semenov. Self-assembling beta-sheet tape forming peptides. *Supramol. Chem.*, 18(5):435–443, 2006.

[24] P. Tamamis, L. Adler-Abramovich, M. Reches, K. Marshall, P. Sikorski, L. Serpell, E. Gazit, and G. Archontis. Self-assembly of phenylalanine oligopeptides: Insights from experiments and simulations. *Biophysical Journal*, 96(12):5020 – 5029, 2009.

[25] E. Gazit. Self-Assembled peptide nanostructures: The design of molecular building blocks and their technological utilization. *Chem. Soc. Rev.*, 36(8):1263–1269, 2007.

[26] P. Lansbury and H. Lashuel. A century-old debate on protein aggregation and neurodegeneration enters the clinic. *Nature*, 443:774–779, 2006.

[27] E. Hone, I. Martins, M. Jeoung, T. Ji, S. Gandy, and R. Martins. Alzheimer's disease amyloid-beta peptide modulates apolipoprotein e isoform specific receptor binding. *J. Alzheimer's Disease*, 7(4):303–314, 2005.

[28] E. Wanker. Protein aggregation in huntington's and parkinson's disease: implications for therapy. *Mol Med Today*, 6:387–391, 2000.

[29] S. Mead, S. Joiner, M. Desbruslais, J. A. Beck, M. O'Donoghue, P. Lantos, J. D. F. Wadsworth, and J. Collinge. Creutzfeldt-Jakob disease, prion protein gene codon 129VV, and a novel PrPSc type in a young british woman. *Archives of Neurology*, 64(12):1780–1784, Dec. 2007. PMID: 18071044.

[30] http://www.cdc.gov/ncidod/dvrd/bse/.

[31] J. Amiel, D. Trochet, M. Clement-Ziza, A. Munnich, and S. Lyonnet. Polyalanine expansions in human. *Human Molecular Genetics*, 13:235–243, 2004.

[32] S. Lindquist and S. Henikoff. Self-perpetuating structural states in biology, disease, and genetics. *Proc. Natl. Acad. Sci. U.S.A.*, 99(4):16377, 2002.

[33] J. E. Straub and D. Thirumalai. Principles governing oligomer formation in amyloidogenic peptides. *Current Opinion in Structural Biology*, 20(2):187–195, 2010.

[34] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75(1):333–366, 2006.

[35] Amyloids: not only pathological agents but also ordered nanomaterials. *Angewandte Chemie (International Ed. in English)*, 47(22).

[36] C. Wu and J. Shea. Coarse-grained models for protein aggregation. *Current Opinion in Structural Biology*, 21(2):209–220, 2011. PMID: 21371882.

[37] S. Yoshi, M. Oka, M. Shima, A. Taniguchi, and M. Akagi. Bridging a 30-mm nerve defect using collagen filaments. *J. Biomed. Mater. Res. A*, 67:467–474, 2003.

[38] T. Holmes, S. de Lacalle, X. Su, G. Liu, A. Rich, and S. Zhang. Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds. *Proc. Natl. Acad. Sci. U.S.A.*, 97(12):6728–6733, 2000.

[39] G. Silva, C. Catherine Czeisler, K. Niece, E. Beniash, D. Harrington, J. Kessler, and I. Samuel. Selective differentiation of neural progenitor cells by high-epitope density nanofibers. *Science*, 303(5662):1352–1355, 2004.

[40] R. Ellis-Behnke, Y. Liang, S. You, D. Tay, S. Zhang, K. So, and S. G.E. Nano neuro knitting: Peptide nanofiber scaffold for brain repair and axon regeneration with functional return of vision. *Proc. Natl. Acad. Sci. U.S.A.*, 103(13):5054–5059, 2006.

[41] W. Gunsteren. The role of computer simulation techniques in protein engineering. *Protein Engineering*, 2(1):5–13, 1988.

[42] S. Schmidler, J. Liu, and D. Brutlag. Bayesian segmentation of protein secondary structure. *Journal of Computational Biology*, 7:233–248, 2000.

[43] C. M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, 2003. PMID: 14685248.

[44] W. van Gunsteren and et al. et al. Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.*, 45:4064–4092, 2006.

[45] V. Tozzini. Minimalist models for proteins: a comparative analysis. *Quarterly Reviews of Biophysics*, 43:333–371, 2010.

[46] O. Engin, A. Villa, C. Peter, and M. Sayar. A challenge for peptide coarse graining: Transferability of fragment-based models. *Macromol. Theory Simul.*, in press.

[47] W. Han and Y.-D. Wu. Coarse-grained protein model coupled with a coarse-grained water model: molecular dynamics study of polyalanine-based peptides. *Journal of Chemical Theory and Computation*, 3(6):2146–2161, 2007.

[48] J. Maupetit, P. Tuffery, and P. Derreumaux. A coarse-grained protein force field for folding and structure prediction. *Proteins*, 69:394–408, 2007.

[49] C. Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Curr.Opin.Struct.Biol.*, 18:10–15, 2008.

[50] R. Hills, L. Lanyuan Lu, and G. Voth. Multiscale coarse-graining of the protein energy landscape. *PLoS Comput Biol*, 6:1, 2010.

[51] G. A. Voth, editor. *Coarse-Graining of Condensed Phase and Biomolecular Systems*. Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.

[52] G. Bellesia and J. Shea. Self-assembly of beta-sheet forming peptides into chiral fibrillar aggregates. *J. Chem. Phys.*, 126:245104, 2007.

[53] M. McCullagh, T. Prytkova, S. Tonzani, N. D. Winter, and G. C. Schatz. Modeling self assembly processes driven by nonbonded interactions in soft materials. *J. Phys. Chem. B*, 112:10388, 2008.

[54] http://www.biology-online.org/dictionary/Ramachandran_plot.

[55] V. Tozzini, W. Rocchia, and J. McCammon. Mapping all-atom models onto one-bead coarse-grained models: general properties and applications to a minimal polypeptide model. *J. Chem. Theory Comput*, 2:667–673, 2006.

[56] T. Bereau and M. Deserno. Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.*, 130:235106, DOI:10.1063/1.3152842, 2009.

[57] P. Derreumaux and N. Mousseau. Coarse-grained protein molecular dynamics simulations. *J. Chem. Phys.*, 126:025101, 2007.

[58] P. Das, S. Matysiak, and C. Clementi. Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *PNAS*, 102(29):10141–10146, 2005.

[59] A. Liwo, S. Oldziej, Pincus, M.R., R. Wawak, S. Rackovsky, and H. Scheraga. A united-residue force field for off-lattice protein-structure simulations, i: Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.*, 18:849–873, 1997.

[60] A. Villa, C. Peter, and N. van der Vegt. Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Phys. Chem. Chem. Phys.*, 11:2077–2086, DOI: 10.1039/b818144f, 2009.

[61] A. Villa, N. van der Vegt, and C. Peter. Self-assembling dipeptides: including solvent degrees of freedom in a caorse-grained model. *Phys. Chem. Chem. Phys.*, 11:2068–2076, DOI: 10.1039/b818146m, 2009.

[62] M. Praprotnik, L. Delle Site, and K. Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annu. Rev. Phys. Chem.*, 59:545–571, 2008.

[63] S. Poblete, M. Praprotnik, K. Kremer, and L. Delle Site. Coupling different levels of resolution in molecular simulations. *J. Chem. Phys.*, 132:114101–114107, 2010.

[64] W. Tschöp, K. Kremer, J. Batoulis, T. Burger, and O. Hahn. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.*, 49:61–74, 1998.

[65] W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, and T. Burger. Simulation of polymer melts. II. From coarse-grained models back to atomistic description. *Acta Polym.*, 49:75–79, 1998.

[66] A. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Phys. Rev. E*, 52:3730–3737, 1995.

[67] D. Reith, M. Puetz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.*, 24:1624–1636, 2003.

[68] S. Izvekov, M. Parrinello, C. Burnham, and G. Voth. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.*, 120:10896–10913, 2004.

[69] S. Marrink, A. de Vries, and A. Mark. Coarse grained model for semi-quantative lipid simulations. *J. Phys. Chem. B*, 108:750–760, 2004.

[70] J. Zhou, I. Thorpe, S. Izvekov, and G. A. Voth. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophysical J.*, 92:4289–4303, 2007.

[71] G. Ayton, W. Noid, and G. Voth. Multiscale modeling of biomolecular systems: in serial and in parallel. *Current Opinion in Structural Biologyy*, 17:192–198, 2007.

[72] C. Peter, L. Delle Site, and K. Kremer. Classical simulations from the atomistic to the mesoscale: coarse graining an azobenzene liquid crystal. *Soft Matter*, 4:859–869, 2008.

[73] W. Noid, J. Chu, G. Ayton, V. Krishna, S. Izvekov, G. Voth, A. Das, and H. Andersen. The multiscale coarse-graining method. i.a rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.*, 128:244114, DOI:10.1063/1.2938860, 2008.

[74] L. Monticelli, S. Kandasamy, X. Periole, R. Larson, D. Tieleman, and S. Marrink. The martini coarse-grained force field: Extension to proteins. *J. Chem. Theory and Comput.*, 4:819–834, 2008.

[75] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory and Comput.*, 5:3211–3223, 2009.

[76] A. Shehu, L. Kavraki, and C. Clementi, C. Multiscale characterization of protein conformational ensembles. *Proteins: Structure, Function, and Bioinformatics*, 76(4):837–851, 2009.

[77] Q. Sun and R. Faller. Systematic coarse-graining of atomistic models for simulation of polymeric systems. *Comput. Chem. Eng*, 29:2380–2385, 2005.

[78] A. Lyubartsev, A. Mirzoev, L. Chen, and A. Laaksonen. Systematic coarse-graining of molecular models by the newton inversion method. *Faraday Discuss.*, 144:43–56, 2010.

[79] C. F. Abrams and K. Kremer. Combined coarse-grained and atomistic simulation of liquid bisphenol a-polycarbonate: Liquid packing and intramolecular structure. *Macromolecules*, 36:260–267, 2003.

[80] D. Fritz, V. A. Harmandaris, K. Kremer, and N. F. A. van der Vegt. Coarse-grained polymer melts based on isolated atomistic chains: Simulation of polystyrene of different tacticities. *Macromolecules*, 42(19):7579–7588, 2009.

[81] O. Engin, A. Villa, C. Peter, and M. Sayar. A challenge for peptide coarse graining: Transferability of fragment-based models. *Macromolecular Theory and Simulations*, page in press, 2011.

[82] S. Marrink, H. Risselada, S. Yefimov, D. Tieleman, and A. de Vries. The martini forcefield: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111:7812–7824, 2007.

[83] E. Gazit. Mechanisms of amyloid fibril self-assembly and inhibition: Model short peptides as a key research tool. *FEBS J.*, 272:5971–5978, 2005.

[84] R. Baron, D. Trzesniak, A. de Vries, A. Elsener, S. Marrink, and W. van Gunsteren. Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models. *Chem. Phys. Chem.*, 8:452–461, 2007.

[85] J. W. Mullinax and W. G. Noid. Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.*, 131(10):104110, 2009.

[86] M. Johnson, T. Head-Gordon, and A. Louis. Representability problems for coarse-grained water potentials. *J. Chem. Phys.*, 126(14):144509, 2007.

[87] D. Fritz, C. R. Herbers, K. Kremer, and N. F. A. van der Vegt. Hierarchical modeling of polymer permeation. *Soft Matter*, 5(22):4556, 2009.

[88] D. Rapaport. *The art of molecular dynamics simulation.* Cambridge University Press, 1995.

[89] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics.* Cambridge University Press, 2004.

[90] K. Binder. *Monte Carlo Simulations of Liquids.* Oxford Science Publications, Oxford, 1987.

[91] R. Feynman, R. Leighton, and M. Sands. *The Feynman lectures of Physics*, volume 1. Addison-Wesley, Reading, 1963.

[92] H. Berendsen, J. Postma, W. van Gunsteren, A. DiNola, and J. Haak. Molecular dynamics with coupling to an external bath1. *J. Chem. Phys.*, 81:3684, DOI: 10.1063/1.448118, 1984.

[93] D. Chandler. *Introduction to Modern Statistical Mechanics.* Oxford University Press., 1987.

[94] H. Berendsen. *Simulating the Physical World.* Cambridge University Press, New York, 1st edition, 2007.

[95] M. Allen and D. Tildesley. *Computer Simulation of Liquids.* Clarendon press. Oxford, 1991.

[96] H. Berendsen and W. F. van Gunsteren. *Molecular Liquids - Dynamics and Interactions.* Reidel Dordrecht, The Netherlands, nato asi c 135 edition, 1984.

[97] P. Todebush and J. Bowen. *Free energy calculations in rational drug design.* Kluwer Academic/Plenum Publischers, New York, 2001.

[98] W. van Gunsteren and H. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew.Chem.Int.Ed.Engl.*, 29(9):992–1023, 1990.

[99] D. Marx and J. Hutter. Ab initio molecular dynamics: Theory and implementation. *Modern Methods and Algorithms of Quantum Chemistry, NIC Series*, 3:329–477, 2000.

[100] L. Verlet. Computer "Experiments" on classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159(1):98–103, 1967.

[101] M. Griebel, S. Knapek, and G. W. Zumbusch. *Numerical simulation in molecular dynamics: numerics, algorithms, parallelization, applications.* Dpringer-Verlag Berlin Heidelberg, 2007.

[102] R. Hockney and J. Eastwood. *Computer Simulation using Particles.* Taylor & Francis Group, New York, 1988.

[103] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23:327–341, 1977.

[104] H. Hess, H. Bekker, H. Berendsen, and J. Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18:1463–1472, 1997.

[105] H. Andersen. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52(1):24 – 34, 1983.

[106] P. Gonnet. P-shake: A quadratically convergent shake in. *J. Comput. Phys.*, 220(2):740 – 750, 2007.

[107] S. Miyamoto and P. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, 1992.

[108] B. Hess. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory and Comput.*, 4(1):116–122, 2008.

[109] M. Nic, J. Jirat, B. Kosata, A. Jenkins, and A. McNaught, editors. *IUPAC Compendium of Chemical Terminology.* Blackwell Scientific Publications, Oxford, Research Triagle Park, NC, 2.1.0 edition, 1997.

[110] S. de Leeuw, J. Perram, and H. Petersen. Hamilton's equations for constrained dynamical systems. *J. Stat. Phys.*, 61(5):1203–1222, 1990.

[111] P. Hünenberger. Thermostat algorithms for molecular dynamics simulations. *Adv. Polym. Sci.*, 173:105–149, 2005.

[112] H. Nyquist. Thermal agitation of electric charge in conductors. *Phys. Rev.*, 32:110–113, 1928.

[113] H. Callen and T. Welton. Irreversibility and generalized noise. *Phys. Rev.*, 83:34–40, 1951.

[114] W. F. V. Gunsterena and H. J. C. Berendsena. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1(3):173–185, 1988.

[115] P. Hoogerbrugge and J. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhys. Lett.*, 19(3):155–160, 1992.

[116] T. Soddemann, B. Dünweg, and K. Kremer. Dissipative particle dynamics: A useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations. *Phys. Rev. E*, 68(4):046702, 2003.

[117] P. Español. Hydrodynamics from dissipative particle dynamics. *Phys. Rev. E*, 52(2):1734–1742, 1995.

[118] W. Hoover, A. Ladd, and B. Moran. High-strain-rate plastic flow via non-equilibrium molecular dynamics. *Phys. Rev. Lett.*, 48(26):1818–1820, 1982.

[119] D. Evans. Computer "experiment" for nonlinear thermodynamics of couette flow. *J. Chem. Phys.*, 78:3297–3302, 1983.

[120] D. Evans, W. Hoover, B. Failor, B. Moran, and A. Ladd. Nonequilibrium molecular dynamics via gauss's principle of least constraint. *Phys. Rev. A*, 28(2):1016–1021, 1983.

[121] B. Hess, D. van der Spoel, and E. Lindahl. GROMACS USER MANUAL.

[122] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384, 1980.

[123] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A Molecular-Dynamics study. *Physical Review Letters*, 45(14):1196–1199, 1980.

[124] A. Kolb and B. Dunweg. Optimized constant pressure stochastic dynamics. *The Journal of Chemical Physics*, 111(10):4453, 1999.

[125] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007.

[126] I. Manousiouthakis and M. Deem. Strict detailed balance is unnecessary in monte carlo simulation. *J. Chem. Phys.*, 110:2753, 1999.

[127] C. Oostenbrink, A. Villa, A. Mark, and W. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *J. Comput. Chem.*, 25:1656–1676, 2004.

[128] C. D. Wick, M. G. Martin, and J. I. Siepmann. Transferable potentials for phase equilibria. 4. united-atom description of linear and branched alkenes and alkylbenzenes. *J. Phys. Chem. B*, 104(33):8008–8016, 2000.

[129] H. R. Warner. Kinetic theory and rheology of dilute suspensions of finitely extendible dumbbells. *Industrial & Engineering Chemistry Fundamentals*, 11(3):379–387, 1972.

[130] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

[131] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n-log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[132] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald potential. *J. Chem. Phys.*, 103(19):8577–8592, 1995.

[133] B. A. Luty, I. G. Tironi, and W. F. van Gunsteren. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.*, 103:3014–3021, 1995.

[134] K. Feenstra, B. Hess, and H. Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.*, 20:786–798, 1999.

[135] D. Frenkel and B. Smit. *Understanding molecular Simulation*. Academic press, 2002.

[136] B. Hess, C. Holm, and N. van der Vegt. Osmotic coefficients of atomistic NaCl (aq) force fields. *The Journal of Chemical Physics*, 124(16):164509, 2006.

[137] V. Tozzini. Multiscale modeling of proteins. *Accounts Chem. Res.*, 43:220–230, 2010.

[138] J. Pfaendtner and G. Voth. Molecular dynamics simulation and coarse-grained analysis of the arp2/3 complex. *Biophysical Journal*, 95(11):5324 – 5333, 2008.

[139] T. Head-Gordon and S. Brown. Minimalist models for protein folding and design. *Current Opinion in Structural Biology*, 13:160–167, 2003.

[140] D. Klimov, M. Betancourt, and D. Thirumalai. Virtual atom representation of hydrogen bonds in minimal off-lattice models of alpha helices: effect on stability, cooperativity and kinetics. *Fold Des*, 3:481–496, 1998.

[141] N. Buchete, J. Straub, and D. Thirumalai. Development of novel statistical potentials for protein fold recognition. *Current Opinion in Structural Biology*, 14:225–232, 2004.

[142] A. Khokhlov and P. Khalatur. Protein-like copolymers: computer simulation. *Physica A*, 249:253–261, 1998.

[143] V. Vaselevskaya, A. Klochkov, A. Lazutin, P. Khalatur, and A. Khokhlov. HA (HydrophobicAmphiphilic) Copolymer Model: Coil-Globule Transition versus Aggregation. *Macromolecules*, 37:5444–5460, 2004.

[144] J. van den Oever, F. Leermakers, G. Fleer, V. Ivanov, N. Shushatina, A. Khokhlov, and P. Khalatur. Coil-globule transition for regular, random, and specially designed copolymers: Monte Carlo simulation and self-consistent field theory. *Phys. Rev. E*, 65:041708, 2002.

[145] N. Go and H. Abe. Noninteracting local-structure model of folding and unfolding transition in globular proteins. i. formulation. *Biopolymers*, 20:991–1011, 1981.

[146] T. Hoang and M. Cieplak. Molecular dynamics of folding of seconary structures in go-type models of proteins. *J. Chem. Phys.*, 112:6851, DOI:10.1063/1.481261, 2000.

[147] Z. Zhang, J. Pfaendtner, A. Grafmueller, and G. Voth. Defining coarse-grained representations of large biomolecules and biomolecular complex from elastic network models. *Biophysical Journal*, 97:2327–2337, 2009.

[148] A. Liwo and C. Czaplewski. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.*, 115:2323, DOI:10.1063/1.1383989, 2001.

[149] A. Liwo, M. Khalili, and H. Scheraga. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc. Natl Acad. Sci. USA*, 102:2362–2367, 2005.

[150] S. Brown, N. Fawzi, and T. Head-Gordon. Coarse-grained sequences for protein folding and design. *Proceedings of the National Academy of Sciences of the United States of America*, 100:10712–10717, 2003.

[151] S. Gopal, S. Mukherjee, Y. Cheng, and M. Feig. Primo/primona: A coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins*, 78:1266–1281, 2010.

[152] H. Nguyen and C. Hall. Phase diagrams describing fibrillization by polyalanine peptides. *Biophysical Journal*, 87:4122–4134, 2004.

[153] M. Schor, B. Ensing, and P. Bolhuis. A simple coarse-grained model for self-assembling silk-like protein fibers. *Faraday Discuss*, 144:127–141, 2009.

[154] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Oldziej, K. Wachucik, and H. Scheraga. Modification and optimization of the united-residue (unres) potential energy function for canonical simulations. i. temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B*, 111:260–285, 2007.

[155] B. Hess, S. Leon, N. van der Vegt, and K. Kremer. Long time atomistic polymer trajectories from coarse grained simulations: bisphenol-A polycarbonate. *Soft Matter*, 2:409–414, 2006.

[156] A. Rzepiela, L. Schafer, N. Goga, H. Risselada, A. De Vries, and S. Marrink. Software news and update reconstruction of atomistic details from coarse-grained structures. *J. Comput. Chem.*, 31:1333–1343, 2010.

[157] I. F. Thorpe, J. Zhou, and G. A. Voth. Peptide folding using multiscale coarse-grained models. *J. Phys. Chem. B*, 112:13079–13090, 2008.

[158] R. Falleri. Automatic coarse graining of polymers. *Polymer*, 45(11):3869 – 3876, 2004.

[159] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen. Multiscale modeling of emergent materials: biological and soft matter. *Phys. Chem. Chem. Phys.*, 11:1869–1892, 2009.

[160] R. L. Henderson. A uniqueness theorem for fluid pair correlation functions. *Physics Letters A*, 49(3):197 – 198, 1974.

[161] J. T. Chayes and L. Chayes. On the validity of the inverse conjecture in classical density functional theory. *Journal of Statistical Physics*, 36:471–488, 1984.

[162] J. T. Chayes, L. Chayes, and E. H. Lieb. The inverse problem in classical statistical mechanics. *Communications in Mathematical Physics*, 93:57–121, 1984.

[163] C. Peter and K. Kremer. Multiscale simulation of soft matter systems-from the atomistic to the coarse-grained level and back. *Soft Matter*, 5:4357–4366, 2009.

[164] F. Müller-Plathe. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *ChemPhysChem.*, 3:754–769, 2002.

[165] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. *Phys. Rev. E*, 73:066701, 2006.

[166] H. Wang, C. Junghans, and K. Kremer. Comparative atomistic and coarse-grained study of water:what do we lose by coarse-graining? *Eur. Phys. J. E*, 28:221–229, DOI: 10.1140/epje/i2008–10413–5, 2009.

[167] A. Villa, C. Peter, and N. F. A. van der Vegt. Transferability of non-bonded interaction potentials for Coarse-Grained simulations: Benzene in water. *Journal of Chemical Theory and Computation*, 6(8):2434–2444, 2010.

[168] V. Harmandaris, N. Adhikari, N. van der Vegt, and K. Kremer. Hierarchical modeling of polystyrene: From atomistic to coarse grained simulations. *Macromolecules*, (39):6708, 2006.

[169] W. W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, and T. Bürger. Simulation of polymer melts. II. from coarse-grained models back to atomistic description. *Acta Polymerica*, 49(2-3):75–79, 1998.

[170] X. Xiaoyu Chen, P. Carbone, G. Santangelo, A. Di Matteo, G. Milano, and F. Müller-Plathe. Backmapping coarse-grained polymer models under sheared nonequilibrium conditions. *Phys. Chem. Chem. Phys.*, 11:1977–1988, 2009.

[171] D. Fritz, K. Koschke, V. A. Harmandaris, N. F. A. van der Vegt, and K. Kremer. Multiscale modeling of soft matter: scaling of dynamics. *Physical Chemistry Chemical Physics*, 2011.

[172] A. Grosberg and A. Khokhlov. *Statistical Physics of Macromolecules*. American Institute of Physics, 1994.

[173] V. A. Harmandaris and K. Kremer. Predicting polymer dynamics at multiple length and time scales. *Soft Matter*, 5(20):3920, 2009.

[174] V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, K. Kremer, B. A. Mann, R. Voelkel, H. Weiss, and C. Liew. Ethylbenzene diffusion in polystyrene: Atomisticcoarse grained simulations and experiments. *Macromolecules*, (40):7026, 2007.

[175] P. K. Depa and J. K. Maranas. Speed up of dynamic observables in coarse-grained molecular-dynamics simulations of unentangled polymers. *The Journal of Chemical Physics*, 123(9):094901, 2005.

[176] P. K. Depa and J. K. Maranas. Dynamic evolution in coarse-grained molecular dynamics simulations of polyethylene melts. *The Journal of Chemical Physics*, 126(5):054903, 2007.

[177] V. A. Harmandaris and K. Kremer. Dynamics of polystyrene melts through hierarchical multiscale simulations. *Macromolecules*, 42(3):791–802, 2009.

[178] A. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78(20):3908–3911, 1997.

[179] R. Best, N. Buchete, and G. Hummer. Are current molecular dynamics force fields too helical? *Biophysical Journal*, 95:L07–L09, 2008.

[180] http://en.wikipedia.org/wiki/PyMOL.

[181] B. Dünweg. Molecular dynamics algorithms and hydrodynamic screening. *J. Chem. Phys.*, 99(9):6977–6982, 1993.

[182] B. Dünweg and K. Kremer. Molecular dynamics simulation of a polymer chain in solution. *J. Chem. Phys.*, 99(9):6983–6997, 1993.

[183] U. Winter and T. Geyer. Coarse grained simulations of a small peptide: Effects of finite damping and hydrodynamic interactions. *J. Chem. Phys.*, 131(10):104102, 2009.

[184] T. Frembgen-Kesner and A. H. Elcock. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory and Comput.*, 5(2):242–256, 2009.

[185] T. Geyer. Many-particle brownian and langevin dynamics simulations with the brownmove package. *BMC Biophys*, 4(7), 2011.

[186] I.-C. Yeh and G. Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B*, 108(40):15873–15879, 2004.

[187] M. Doi and S. F. Edwards. *The theory of polymer dynamics*. Oxford University Press, New York, 1986.

[188] J. G. Kirkwood and J. Riseman. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *J. Chem. Phys.*, 16(6):565–573, 1948.

[189] J. J. Erpenbeck and J. G. Kirkwood. Mestatisticalchanics of irreversible processes in polymer solutions. *J. Chem Phys.*, 29(4):909–913, 1958.

[190] B. H. Zimm. Dynamics of polymer molecules in dilute solution: Viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.*, 24(2):269–278, 1956.

[191] J. Prince E. Rouse. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.*, 21(7):1272–1280, 1953.

[192] M. Allen and D. Tildesley. *Computer Simulation of Liquids*. Clarendon press. Oxford, 1991.

[193] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.

[194] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An w log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[195] H. Berendsen, J. Grigera, and T. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

[196] P. Mark and L. Nilsson. Structure and Dynamics of the TIP3P, SPC and SPC/E Water Models at 298 K. *J. Phys. Chem. A*, 105:9954–9960, 2001.

[197] A. J. Rzepiela, L. V. Schfer, N. Goga, H. J. Risselada, A. H. D. Vries, and S. J. Marrink. Reconstruction of atomistic details from coarse-grained structures. *Journal of Computational Chemistry*, 31(6):1333–1343, 2010.

BIBLIOGRAPHY