## Multiscale Simulation of Small Peptides: Consistent Conformational Sampling in Atomistic and Coarse-Grained Models

Olga Bezkorovaynaya,<sup>[a]</sup> Alexander Lukyanov,<sup>[a]</sup> Kurt Kremer,<sup>[a]</sup> and Christine Peter<sup>[a]</sup>\*

A bottom-up coarse-graining procedure for peptides in aqueous solution is presented, where the interactions in the coarse-grained (CG) model are determined such that the CG peptide samples conformations according to a high-resolution (atomistic) model. It is shown that important aspects of conformational sampling, such as correlated degrees of freedom (DOF) which play an important role in secondary structure formation, can be reproduced in the CG description. In some cases, microscopic structural/conformational details are lost in the coarse-graining process. We show that these "lost" properties can be recovered in a backmapping procedure

which reintroduces atomistic DOF into CG structures — as long as the overall conformational sampling of the molecule is correctly represented in the CG level of resolution. Thus, it is possible to link an existing all-atom model of a biomolecular system with a CG description such that after inverse mapping one can recover structures at high resolution with the correctly sampled (according to the atomistic model) conformational properties.

## Introduction

To investigate biological systems covering a wide range of relevant length scales or processes happening on long timescales, simulation approaches at various levels of resolution are being developed. Coarse-grained (CG) simulation models, which reduce the number of degrees of freedom (DOFs) compared to an atomistic representation, have been applied to many types of systems from synthetic polymers to biological systems such as biomembranes or proteins.<sup>[1–10]</sup> For proteins and peptides a variety of CG models have been used,<sup>[8, 10–12]</sup> vastly differing depending on the investigated problem, for example, protein folding,<sup>[12, 13]</sup> association of large protein complexes,<sup>[14–16]</sup> or mechanical properties of structure proteins in biological materials.<sup>[17]</sup> To investigate general principles behind processes such as protein folding or amyloid aggregation, generic CG models which do not target at quantitative predictions of specific molecular systems have been successfully used.<sup>[18-21]</sup> Parameters in such generic CG models are determined independently from atomistic simulation models, based on experimental data such as secondary structure propensities of specific amino acids, hydrophobicity, and so on,<sup>[17-22]</sup> structurally accurate CG models are frequently devised from data sets of structures.<sup>[9, 23-27]</sup>

A different type of CG models is used as an element of a multiscale simulation framework, where one investigates a system at the same time at several levels of resolution—going back and forth between the levels to have access to both mesoscopic time and length scales as well as the corresponding microscopic picture. In this case, the thermodynamic and structural consistency of the models at the different resolution levels is of particular importance. Several coarse-graining approaches have been developed over the years, which share the goal of being able to switch resolution and systematically connect the scales.<sup>[28–35]</sup> Frequently in these approaches, a simulation at the higher resolution (e.g., atomistic) level is used as a reference to determine interaction functions for the lower resolution (CG) model. Parameters for the CG interactions can be determined based on different data obtained from the atomistic reference, most importantly structural information,<sup>[28, 29]</sup> mean forces<sup>[30, 31, 33]</sup> or thermodynamic data.<sup>[34, 36]</sup> These multiscale simulation methodologies have been successfully applied to model systems, homogeneous liquids, amorphous polymers, and so on.<sup>[2, 28, 29, 32, 37-40]</sup> Extending them to more complex systems such as multiple component systems, biomolecules in aqueous solution, biological materials, and so on poses new and exciting challenges. One needs to address guestions concerning aggregation and phase separation as well as complex conformational equilibria which are in turn affected by the interaction with surfaces and interfaces. For peptides, several CG models for multiscale simulation purposes have been already developed using the aforementioned approaches.<sup>[31, 34, 41–46]</sup> Questions related to the conformations sampled by a CG peptide model are whether it is possible to reproduce the folding equilibria found at the atomistic level. Correlations between DOFs are characteristic for peptides and proteins, for example, manifested in the Ramachandran plot, and the question arises whether they can be reproduced by the CG model, especially if one tries to avoid complex, computationally expensive multiparameter

 <sup>[</sup>a] O. Bezkorovaynaya, A. Lukyanov, K. Kremer, C. Peter Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany E-mail: peter@mpip-mainz.mpg.de

Contract/grant sponsor: Eurosim Early Stage Training Project; Contract/grant sponsor: German Science Foundation; Contract/grant number: PE 1625/1-1

potentials. If a suitable CG model can be found, one needs to investigate to which extent the interaction functions are transferable.<sup>[11,27]</sup> Regarding intramolecular interactions, a major question is whether they correctly predict the conformations, if the molecule is transferred to a different environment or to a situation where the conformational equilibrium changes, for example, when molecules aggregate.

In this study, we address some of these questions while we develop a CG model for oligoalanine peptides based on an atomistic (force field) description. For a short dipeptide (di-phenylalanine), a structure-based CG model had been previously developed that reproduces both the conformations and the association behavior of the peptide observed in atomistic simulations.<sup>[41,42]</sup> Here, we focus on the conformational sampling of a CG model for slightly longer peptides (capped Ala3 and Ala<sub>4</sub>). Oligoalanine was chosen, as it provides an ideal test system to concentrate on the bonded/intramolecular aspects of CG modeling, while leaving challenges concerning side-chain specific nonbonded interactions for later. In addition, oligoalanine is an important component of many silk proteins, and aggregates of oligoalanine-rich peptides are intensively studied to understand the remarkable mechanical properties of silk fibers.<sup>[47, 48]</sup> To obtain further insight into these biological materials, a multiscale model that links CG and atomistic simulation level and thus provides access to both the necessary length and timescales and local microscopic properties would be highly desirable.

We investigate to which extent the CG model reproduces the conformational equilibrium of the atomistic reference with particular attention to the problem of correlations between intramolecular DOFs. Thus, we want to assure that the CG model samples conformations which are fully compatible with the original atomistic sampling after reinserting atomistic coordinates. Note, that this study explicitly focuses on local aspects of conformational sampling, that is, nonlocal (nonbonded) intramolecular interactions such as hydrogen bonds that lead to helix formation will be added at a later stage. We will show to which extent the resulting intramolecular CG potentials are transferable to longer sequences and we will also show to which extent the interplay of bonded and nonbonded interactions, also with a surrounding medium, needs to be accounted for in the parameterization process.

## Methods

#### CG model: mapping scheme

The mapping scheme, that is, the relation between atomistic and CG DOFs, for the capped oligoalanine peptides is shown in Figure 1a. The CG peptide consists of a linear chain with two types of spherical beads, one representing the  $\alpha$  and  $\beta$  carbon atoms (denoted as "CAB") and one representing the peptide groups (denoted as "PEP"). The positions of the CG centers are defined by the center of mass of the constituting atoms. This mapping scheme is an analog of the one applied in Ref. [41] for di-phenylalanine, with the difference that alanine residues require no CG side chain beads. Although the atomistic simulations were performed with an explicit water model, two different types of water representations were considered for the CG system: an implicitwater representation, where the effect of the water molecules on the peptide is accounted for via the CG interactions between the peptide beads,<sup>[41]</sup> and an explicit CG water model.<sup>[42]</sup> In the latter case, each water molecule was represented by one CG bead, see Figure 1c.

#### CG bonded and nonbonded interactions

To determine CG interaction potentials, we distinguish between bonded intramolecular interactions and nonbonded ones. These are determined separately based on the assumption that the total potential energy can be separated into bonded and nonbonded contributions:

$$U^{\rm CG} = U^{\rm CG}_{\rm B} + U^{\rm CG}_{\rm NB} \tag{1}$$

Note that this separation, which is not done in all structure-based CG approaches,<sup>[4]</sup> has the advantage of a certain modularity. The (local) intramolecular interaction potentials obtained in this study could in principle be adapted for use with different nonbonded interaction functions from various coarse-graining approaches. Bond, angle, and torsion potentials obtained via Boltzmann inversion (see below) can, for example, also be used in combination with nonbonded interaction functions determined via force matching.<sup>[49,50]</sup> This also implies that the results of this study, regarding correlations between DOF and regarding the necessity for special intramolecular interaction functions (see below) are not only limited to structure-based coarse-graining approaches but also hold for approaches where nonbonded interactions are parameterized differently, for example, based on force matching or thermodynamic data.

In CG approaches, which are parameterized based on atomistic reference simulations one first maps the atomistic to CG coordinates. From the latter, one obtains distributions in CG DOF corresponding to the atomistic sampling which serve as reference distributions. For intramolecular interactions, these are bond, angle, and dihedral distributions. In structure-based approaches, these distributions are Boltzmann inverted to obtain the corresponding potentials of mean force:

$$U^{CG}(r,T) = -k_{B}T\ln(P(r,T)/r^{2}) + const_{r}$$
  

$$U^{CG}(\theta,T) = -k_{B}T\ln(P(\theta,T)/\sin\theta) + const_{\theta}$$
  

$$U^{CG}(\varphi,T) = -k_{B}T\ln(P(\varphi,T)) + const_{\varphi}$$
(2)

Whether these potentials of mean force can be directly used as (tabulated) CG interaction potentials depends on several conditions.

First, it requires that the probability distribution describing the conformations factorizes into bond, angle, and torsion contributions, that is the intramolecular DOF are uncorrelated:

$$P^{CG}(r,\theta,\varphi,T) = P^{CG}(r,T)P^{CG}(\theta,T)P^{CG}(\varphi,T)$$
(3)

How well this assumption holds for a given molecule may depend on the choice of the CG mapping scheme, as was, for example, nicely shown for the case of two CG polystyrene models.<sup>[51]</sup> The assumption of uncorrelated DOFs is particularly problematic for biological systems with distinct secondary structures. Here, certain correlations between intramolecular DOF are characteristic for the conformations adopted by the molecule and need to be accounted for in the CG model.<sup>[10, 23, 43, 52–56]</sup> In this manuscript, we will address questions regarding the independence of DOF and correlations along the peptide backbone for our small model peptide.

A second assumption is that bonded and nonbonded interactions can be separated according to eq. (1). In polymeric systems, a clean separation is achieved by obtaining bond, angle, and torsion distributions from sampling an isolated polymer chain with exclusions of long-range nonbonded interactions.<sup>[32, 39]</sup> This approach is problematic for biomolecules in water, where the solvent has an important influence on the conformational equilibrium sampled by the molecule. For a small dipeptide, the distributions from sampling by atomistic explicit-solvent simulations (without applying additional exclusions for long-range nonbonded interactions along the peptide chain) could be inverted and directly used as potentials.<sup>[42]</sup> The resulting CG model reproduced the conformational sampling from atomistic simulations very well. Whether this approach can be applied also for longer peptide chains will be tested in this manuscript.

For the above reasons, potentials obtained from Boltzmann inversion according to eqs. (2) may not succeed at producing the correct conformational equilibrium of the peptide in the CG model (i.e., after combining all covalent potentials and nonbonded interactions including the solvent). In that case, one can introduce an additional refining step, which is completely analogous to the iterative procedure commonly used for nonbonded interactions.<sup>[7,29]</sup> For example for an angular DOF,  $\theta$ , the iterative refinement is done as follows:

$$V_{i+1}(\theta) = V_i(\theta) + k_{\rm B}T \ln\left[\frac{P_i(\theta, T)}{P^{\rm ref}(\theta, T)}\right]$$
(4)

Here,  $P^{\text{ref}}(\theta, T)$  is the reference angular distribution from atomistic simulation and  $P_i(\theta, T)$  is the current distribution after the *i*th iteration.

In this study, we primarily focus on the conformational sampling of a single peptide in atomistic and CG representations, not on the association behavior of several peptides. For this reason, nonbonded peptide-peptide and peptide-water interactions were not specifically parameterized but taken from Ref. [42]. Please note, already at this point, that this is not meant to imply that CG bonded and nonbonded interactions are uncoupled, it just means that the bonded interaction functions are parameterized under the condition of a given set of nonbonded potentials (we will further address this issue in the course of the manuscript). The peptide-solvent and solventsolvent interactions had been determined by iterative Boltzmann inversion.<sup>[7,29]</sup> In iterative Boltzmann inversion, a CG potential (tabulated interaction function) is self-consistently refined until a structural property, in this case, the radial distribution function that had been previously obtained by atomistic simulations, is reproduced by the CG model. Thus, the resulting CG water model reproduces the water-water (center-of-mass) radial distribution function which is almost identical to the oxygen-oxygen pair correlation function.<sup>[57]</sup> One feature of the CG peptide model in explicit-CG water is that by construction of the peptide-water interaction functions the peptide retains a structurally intact (CG) solvation shell, as was shown in Ref. [42]. For interactions between solute beads at low concentration (here between peptide beads), an analogous method had been devised based on pair potentials of mean force at infinite dilution. Thus, the tabulated interactions are parameterized to reproduce the different association strengths of solute pairs (corresponding to the different peptide beads) in water (for details we refer to Ref. [42]). CG nonbonded 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. All CG interaction functions for oligoalanines obtained in this study are presented in supporting Information. Note, however, thatas will be discussed below-the primary result of this work is not so much a CG forcefield but rather a recommendation which types of CG interactions are needed and how they can be parameterized to achieve conformational sampling that is consistent with an atomistic representation. Therefore, the precise nature of the intramolecular CG interactions will depend on the underlying atomistic forcefield as well as on the chosen type of CG nonbonded interactions.

#### Backmapping

To compare the conformational sampling of atomistic and CG simulations on the atomistic level of resolution, one can (re)introduce atomistic coordinates corresponding to the CG structures. This can be done using virtual sites<sup>[58]</sup> at the positions of the CG beads.<sup>[32]</sup> Molecular dynamics simulations with position restraints on these virtual sites "force" an initial atomistic structure on top of the CG one and equilibrate it around these virtual site positions. The backmapping process is broken down into the following steps. CG structures are randomly selected from the CG trajectory. For each of these structures, an initial atomistic conformation is selected from a database of atomistic structures (based on a similar inertia tensor and end to end distance) and fitted on to the CG structure (please note that the use of presampled peptide conformations was primarily chosen for convenience, it is equally well possible to generate atomistic initial structures of the backmapping process from scratch, see, for example Ref. [41]). Several steps of energy minimization and stochastic dynamics simulations are carried out with harmonic position restraints on the virtual sites and dihedral angle restraints  $(180^{\circ} \pm 70^{\circ})$  on the peptide group to avoid unphysical trans-cis flips before and after addition of water to the system follow. Data for the analysis of the backmapped ensemble were obtained by 200 ps of molecular dynamics (MD) simulations (with position restraints,  $k_p = 10,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ ; without dihedral restraints) for each structure.



Figure 1. a) Relation between atomistic (united atom, small beads) and CG model (large transparent beads) of Ala<sub>3</sub> with their respective CG bead names used throughout the manuscript.b) CG internal DOFs (small spheres indicate centers of the CG beads) and c) CG water representation.

#### **Computational details**

For all simulations, we used the GROMACS program suite.<sup>[59,60]</sup> Atomistic MD simulations were carried out using the leap-frog algorithm with a time step of 2 fs. Constant temperature and pressure were maintained via the weak coupling algorithm,<sup>[61]</sup> where the temperature was set to 300 K 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  $4.5 \times 10^{-5}$  bar<sup>-1</sup>. The LINCS algorithm was applied to constrain all bonds.<sup>[59]</sup> Electrostatic interactions were computed with the Particle Mesh Ewald method<sup>[62]</sup> with a real-space cutoff of 1.0 nm, and a cutoff 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<sup>[63]</sup> with the SPC/E water model.<sup>[64]</sup> Atomistic simulations of Ala<sub>3</sub> were carried out for 90 ns, whereas the atomistic simulations of Ala<sub>4</sub> were carried out for 200 ns.

For the CG simulations, a leap-frog stochastic dynamics integrator (Langevin thermostat) was used with an inverse friction constant of 0.1 ps. Simulations were carried out at 300 K and constant volume conditions (at the average volume of the corresponding atomistic system). Bonded and nonbonded interactions were computed using tabulated potentials that had been smoothed using cubic splines. All CG simulations of Ala<sub>3</sub> and Ala<sub>4</sub> were carried out for 100 ns.

## **Results and Discussion**

## A first CG model for Ala<sub>3</sub>

Ala<sub>3</sub> is represented by seven CG beads, see Figure 1b, connected by six bonds, five angles, and four dihedral angles, each of which was modeled by a separate potential table obtained through



**Figure 2.** Distributions sampled by Ala<sub>3</sub> with different CG potentials (atomistic reference: solid black line; CG with implicit solvent: small red circles; CG with explicit solvent without correction: dashed blue line; CG with pressure-corrected solvent: gray squares; CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent: green dotted line). Panel a: PEP-CAB-PEP angle; panel b: CAB-PEP-CAB angle; panel c: PEP1-CAB1-PEP2-CAB2 dihedral angle; panel d: end-to-end distance, that is, distance between CG centers PEP1 and PEP4.

Boltzmann inversion according to eq. (2). It should be noted that the respective equivalent 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.

First, we tested whether the atomistic distributions used for parameterization are reproduced by the CG peptide model after adding up all interactions (bonded and nonbonded). Indeed, all bonds (data not shown) and the two angles between the CAB-PEP-CAB groups, see Figure 2b, are very well represented in the sampling of the CG peptide. However, for the three PEP-CAB-PEP angles the agreement is not as good. Figure 2a shows for one of the PEP-CAB-PEP angles that the model with the potentials obtained from Boltzmann inversion (noniterative) in an implicit-solvent representation reproduces the atomistic distribution very well, whereas the corresponding explicit-solvent model (CG water representation; potentials from Ref. [42]) fails to reproduce the atomistic reference distribution. A similar observation had not been made previously for diphenylalanine, possibly because the peptide has an explicit-sidechain 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. Typically, CG models obtained with iterative Boltzmann inversion exhibit a comparatively high pressure,<sup>[57]</sup> and we suspected that the high (solvent) pressure in the system may cause the distortion of the peptide conformations. Thus, we applied a pressure correction to the water-water interaction potential (see supporting Information and Ref. [57]), such that it reproduces atmospheric pressure conditions. Figure 2a shows that this had only a small impact on the conformational sampling of the peptide. Note that in all subsequent simulations, we from now on use this pressure corrected CG solvent model. Next, we iteratively refined the corresponding potential for the PEP-CAB-PEP angle, see eq. (4). Target of this iterative Boltzmann inversion was the distribution of the PEP-CAB-PEP angle obtained from atomistic simulation of Ala<sub>3</sub>. For the short simulations in the iterative procedure, we did not use the entire CG Ala<sub>3</sub> peptide, but a shorter three bead 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 DOFs in the peptide (of course one then has to test, if the resulting potential yields the correct distributions of the respective angle in the peptide chain with all other interactions present). The potential obtained after four iteration steps was then applied as new CG PEP-CAB-PEP potential to the three respective angles in Ala<sub>3</sub>. The new CG model with pressure corrected solvent and refined PEP-CAB-PEP angle reproduces the angular distribution of the atomistic peptide quite well as shown in Figure 2a. Other bond or angle distributions (see e.g., Fig. 2b), are not affected by the change in the PEP-CAB-PEP potential.

Ala<sub>3</sub> has four CG torsional DOFs. For the two equivalent CAB-PEP-CAB-PEP dihedral angles, the agreement between atomistic and CG sampling is very good, and the sampling remains unaffected by all changes applied to other interaction functions (data not shown). For the two PEP-CAB-PEP-CAB dihedral angles, the situation is slightly different. Figure 2c shows the distribution of the PEP1-CAB1-PEP2-CAB2 dihedral angle (the PEP2-CAB2-PEP3-CAB3 torsion behaves identically) obtained from atomistic sampling and the various CG models with explicit solvent. The model with iteratively refined PEP-CAB-PEP angle shows significantly better agreement with the atomistic reference data than the unrefined CG model. This is an important observation: the correction applied to the PEP-CAB-PEP angle also removes a discrepancy in the sampling of a "neighboring" dihedral angle, that is, discrepancies in two distribution functions result from a single physical source. This shows that it is potentially problematic to "blindly" iterate all potential tables at once. One should rather refine intramolecular potentials on a case-by-case basis as the DOFs are coupled. If one needs to correct several distributions it appears to be advisable to start with the stiffer DOFs that involve fewer beads (i.e., bonds or angles) before turning to torsions or even more complex DOFs (see below). It should be noted though that for mutually interdependent potentials of similar stiffness, for example, two correlated torsion angles, it might be necessary to optimize them simultaneously while keeping track of the correlations between them.

As an independent quantity, which had not directly entered the parameterization process, the end-to-end distance (i.e., the distance between the PEP1 and PEP4 beads) is analyzed as one possible measure for the agreement of the overall conformational sampling of the CG peptide with the atomistic one. Figure 2d shows the end-to-end distance in the atomistic sampling, the implicit-solvent CG model, the uncorrected explicit-solvent CG model, and the CG model with pressure and angle corrections. One sees that the uncorrected explicitsolvent model does not agree well with the atomistic reference, the peptide appears to be more compressed. The major cause of that shift is the "distorted" PEP-CAB-PEP angle distribution in the explicit solvent CG model. The implicit-solvent model and the explicit-solvent model with corrected PEP-CAB-PEP angle potential show a better agreement with the atomistic reference. However, there is still a small shift toward smaller distances in the end-to-end distance distribution, which appears in spite of the very good agreement of all local conformational distributions up to 1,4 bead separations along the peptide chain. Even though for Ala<sub>3</sub> this shift is not large, it is not very satisfactory as for longer peptides it might easily add up to a large discrepancy between atomistic and CG model. There are two possible explanations for the observed effect. One possibility is that the intramolecular nonbonded interactions between beads separated by more than three bonds need to be reparameterized. Another possibility is that there are conformational effects beyond bonds, angles, and torsions that influence the overall conformational sampling of the molecule, for example, due to correlated DOFs.

#### Refining the CG model for Ala<sub>3</sub>

In the following, we investigate the cause of the above discrepancy in the end-to-end distance and suggest a method to improve the CG model. As a basis, we use the CG model



**Figure 3.** Distributions sampled by Ala<sub>3</sub> with different CG potentials (atomistic reference: black solid line; CG with iterated PEP-CAB-PEP potential and pressurecorrected solvent: green dotted line; CG with additional 1-5 PEP-PEP distance potential: red crosses; CG with additional 1,3,5 PEP-PEP-PEP angle potential: blue squares) Panel a: 1,3,5 PEP-PEP-PEP angle, the arrows indicate the values for an  $\alpha$ -helical and a  $\beta$ -strand model structure; panel b: 1-5 PEP-PEP distance; panel c: 1,3,5 CAB-CAB-CAB angle; panel d: end-to-end distance (pink circles: CG model with additional 1-5 PEP-PEP distance potential and 1,3,5 PEP-PEP angle potential simultaneously). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

with pressure corrected solvent and iterative refinement of the PEP-CAB-PEP angle, the other models are not further discussed. First, we analyze other internal distances and angles in the system. Figure 3b shows that-completely analogous to the end-to-end distance-the CG 1,5 distance between two PEP groups which are separated by four CG bonds is too short in the sampling of the CG model compared to the atomistic simulations. Similarly, the distribution of the angles between adjacent PEP groups along the chain (i.e., the angle between PEP1-PEP2-PEP3 or PEP2-PEP3-PEP4 beads, from now on denoted as 1,3,5 PEP-PEP angle) is somewhat overemphasizing smaller values, see Figure 3a. In contrast, the angle between adjacent CAB groups along the chain, that is, the 1,3,5 CAB-CAB-CAB angle (Fig. 3c), and the corresponding 1,5 distance between CAB groups are nicely reproduced. The question is now whether the cause of the observed discrepancy in the 1,5 PEP-PEP distance and the 1,3,5 PEP-PEP-PEP angle distributions lies in the nonbonded PEP-PEP interactions (which might be too attractive) or whether the explanation should be rather found in the conformational behavior of the chain segment between the 1,5 PEP PEP beads which might be not properly described by the set of (uncoupled) bond, angle, and torsion potentials. Modifying the nonbonded interactions between PEP beads and reducing the attraction rather drastically had only very little effect on the 1,5 PEP-PEP distance or the end-to-end distance distribution (data not shown). From this, we concluded that the cause of the observed discrepancy does probably not lie primarily in the nonbonded

complex conformational behavior of the chain segments. Next, we used iterative Boltzmann inversion to determine a special "pseudobond" interaction potential between the 1,5 PEP-PEP beads which is tailor made to reproduce the corresponding distance distribution (in analogy to Fritz et al.<sup>[39]</sup> for polystyrene). However, if the real cause of the discrepancy between the CG and atomistic 1,5 PEP-PEP distance distribution lies in the conformations sampled by the chain segment between the two PEP beads, a different type of additional potential might be more physical. Instead of applying a 1,5 (pseudobond) interaction one could also impose a 1,3,5 angular potential (again determined by iterative Boltzmann inversion). Figure 3 shows that for both types of additional potentials one of the distributions (1,5 PEP-PEP distance or 1,3,5 PEP-PEP-PEP angle) is by construction exactly reproduced while the respective other distribution is improved moderately compared to the CG model without additional potential. In both cases, the end-to-end distance distribution covers the correct total range, and the height of the major peak around 0.95 nm has significantly improved compared to the CG model without additional potential. In the case of the 1,5 PEP-PEP distance potential, one still observes a larger deviation in the shoulder of the distribution around 0.7 nm. For this reason, we conclude that the 1,3,5 PEP-PEP-PEP angle potential produces a slightly better result than the 1,5 PEP-PEP distance potential. (Note, that Fig. 3d also contains data from a CG simulation where both a 1,5 PEP-PEP distance

interactions along the chain but is rather connected to more



Figure 4. End-to-end distance distributions sampled by Ala3 with different CG models. Atomistic reference: black solid line; explicit-solvent CG model with bond, angle, torsion potentials from Boltzmann inversion without iterative refinement: black dotted line; implicit-solvent CG model with bond, angle, torsion potentials from Boltzmann inversion: grav dot-dashed line: explicit-solvent CG model with iterated PEP-CAB-PEP potential and pressure-corrected solvent: black dot-dashed line; implicit-solvent CG model with additional 1.3.5 PEP-PEP-PEP angle potential (Boltzmann inversion, no iterative refinement): grav crosses: explicit-solvent CG model with additional 1,3,5 PEP-PEP-PEP angle potential (iterative Boltzmann inversion): black crosses. The arrows indicate the effect of the different refinement steps: (1) model with chain conformations where coupling between bonded and nonbonded (solvent) interactions causes deviations in local properties (PEP-CAB-PEP angle); (2) models with correct bond, angle, and torsion distributions, but where conformations involving a 1,3,5 PEP-PEP-PEP segment deviate; (3) models where local chain conformations up to 1.3.5 PEP-PEP-PEP segments are correct. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and a 1,3,5 PEP-PEP-PEP angle potential were applied simultaneously, we will not comment on this model here but at a later stage in the manuscript.) Neither of the additional potentials changes the regular bond, angle, and torsion distributions in the system (data not shown) and also the CAB-CAB-CAB angle distribution remains unaffected, see Figure 3c. A similar refinement of the 1,3,5 PEP-PEP-PEP segment conformations can be obtained for the implicit-solvent CG model. In this case, we have applied a 1,3,5 PEP-PEP-PEP angle potential obtained from Boltzmann inversion of the corresponding distribution (without iteration). Here, the end-to-end distance distribution is significantly improved compared to the implicit-solvent model with only "regular" bond, angle and torsion potentials (see Fig. 4).

If the need for such an additional potential stems from the fact that the conformations of the chain segment between two 1,5 PEP beads cannot be well represented by the original uncoupled bond, angle, and torsion potentials, it would be interesting to identify the correlations causing this effect and to find out whether the different regions in the 1,3,5 angle distribution can be linked to typical secondary structure elements sampled by the polypeptide chain. We investigated two idealized peptide chains, a model oligo-Ala  $\alpha$ -helix, and a  $\beta$ -strand, which were mapped to CG coordinates. The 1,3,5 PEP-PEP-PEP angles found for these model structures are indicated by arrows in Figure 3a. The value of ~100° found for the  $\alpha$ -helix clearly corresponds to the shoulder of the distribution, whereas the  $\beta$ -strand with a

value of  $\sim$ 170° clearly represents the major peak. Thus, the 1,3,5 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, that is, adjacent sets of Ramachandran angles which should be correlated in secondary structure elements. For this reason, we would argue that imposing a 1,3,5 PEP-PEP-PEP angle potential is a well-justified addition to the CG model to obtain the correct conformational sampling. Similar potential terms that account for correlations along the backbone and propensities for the formation of secondary structure elements have been discussed for other CG models, an excellent review of different approaches can be found in Ref. [10]. Our results also nicely illustrate the importance to choose a CG mapping where the typical secondary structure elements can be well represented<sup>[10, 54]</sup> and separated in the potentials. Figure 4 summarizes the effects of the different refinement steps on the conformations of the Ala<sub>3</sub> chain. Here, we show the end-to-end distance distribution for both the explicit-solvent and the implicit-solvent CG models in comparison with the atomistic reference. In the case of the CG model with explicit solvent and bond, angle, and torsion potentials from direct Boltzmann inversion, we observe coupling between bonded and nonbonded (solvent) interactions which causes deviations in local properties, in this case, the PEP-CAB-PEP angle. This leads to a large shift of the end-toend distribution toward more compact structures (red arrow). This artifact can be overcome by iterative refinement of the PEP-CAB-PEP angle. The resulting explicit-solvent model shows an end-to-end distance distribution which is very similar to the implicit-solvent CG model obtained from Boltzmann inversion without iteration (where by construction coupling between solvent and angle cannot occur). These are distributions (blue arrow) characteristic for models with correct individual bond, angle, and torsion distributions, but distortions in conformations involving 1,3,5 PEP-PEP-PEP segments causing the observed shift toward shorter end-to-end distances compared to the atomistic reference. Finally, the green arrow indicates end-to-end distance distributions characteristic for models where local chain conformations up to 1,3,5 PEP-PEP-PEP segments are correct (both for the explicit- and the implicit-solvent case). These distributions agree very well with the atomistic reference, which points out that the local properties of 1,3,5 PEP-PEP-PEP segments are crucial for a CG model to be able to describe and distinguish secondary structure propensities of a peptide chain. Note that these are still very local properties which are not related to hydrogen bonding effects that drive, for example, the formation of an  $\alpha$ -helix.

The assumption of independent CG DOFs in the coarsegraining procedure [eq. (3)] is most likely problematic for biomolecules. We already showed in the previous paragraphs that a simple set of bond, angle, and torsion potentials obtained from (iterative) Boltzmann inversion is not entirely sufficient to reproduce the atomistic conformational sampling of the Ala<sub>3</sub> peptide. Therefore, we investigated which correlations between CG DOFs occur in the atomistic sampling of Ala<sub>3</sub> and whether they are reproduced by the CG model. We found correlations along the peptide chain, namely between



Figure 5. Two-dimensional distribution of correlated CAB1-PEP2-CAB2 angle and PEP1-CAB1-PEP2-CAB2 dihedral angle. Upper left: atomistic reference (analyzed in CG coordinates); upper right: CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent; lower left: CG with additional 1,3,5-angle potential; lower right: backmapped (BM) ensemble (from 200 conformations of CG with additional 1,3,5-angle potential; analyzed in CG coordinates).

each  $CAB_n$ -PEP<sub>(n+1)</sub>-CAB<sub>(n+1)</sub> angle and the torsion around the preceding bond  $PEP_n$ -CAB<sub>n</sub>-PEP<sub>(n+1)</sub>-CAB<sub>(n+1)</sub> (see Fig. 5 upper left panel). Although the CG sampling reproduces the distributions of each individual DOF very well, it fails to capture the correlations (upper right panel CG model with pressure and PEP-CAB-PEP angle refinement). The figure also shows that application of the additional 1,3,5 PEP-PEP-PEP angle potential described above (lower left panel) does not "repair" these missing correlations. However, in the "backmapped" ensemble, that is, 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-again evaluated in CG DOFs-back (lower right panel). The same observation could be made for all other correlated sets of CG DOFs along the peptide chain. This means that after backmapping, a representation of the CG conformation exists, that is at the same time compatible with the geometric restraints imposed by the CG coordinates and still exhibits the correlations found in the original atomistic sampling. A possible interpretation is that the correlations are caused by the backbone geometry on the atomistic level or local steric effects due to the side chain beads and are therefore by construction lost in the linear CG representation. However, as these correlations are affecting the chain geometry only very locally—within short segments of the chain—the overall conformational properties of the chain segments are well enough represented by angle and dihedral potentials so that the correlations can be restored by backmapping when the local chain geometry is inserted again.

# Atomistic properties from CG sampling: analyzing the backmapped ensemble

We also analyzed atomistic properties of the ensemble of backmapped structures (obtained from up to 200, for a few CG models up to 400, randomly chosen structures from the CG trajectories) and compared them to those of the original ensemble from (free) atomistic simulations. We focus on the atomistic backbone torsional angles at the C<sub>a</sub> atoms (Ramachandran angles  $\phi$ and  $\psi$ ) which characterize the peptide chain conformations and secondary structure formation. Figure 6 (upper panels) shows the dihedral angle distributions for two of the  $\phi$  and  $\psi$  angles for the backmapped ensemble of the original CG model with only pressure and PEP-CAB-PEP angle refinement, and for the new CG model with additional 1,3,5 PEP-PEP-PEP angle potential. We see that the agreement with the original atomistic sampling



**Figure 6.** Atomistic backbone dihedral angles (Ramachandran angles) in original atomistic (black solid line) and backmapped ensembles (based on 200 structures) from different CG simulations. Upper panels: CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent (green dotted line) and CG model with 1,3,5 PEP-PEP-PEP angle potential (blue squares); Lower panels: CG with 1,3,5 PEP-PEP-PEP angle potential (blue squares); CG with 1,3,5 PEP-PEP-PEP angle potential (blue squares), CG with 1-5 PEP-PEP distance potential (red crosses), CG with both 1,3,5 PEP-PEP-PEP angle and 1-5 PEP-PEP distance potentials (pink circles). Panels a and c :  $\Phi$  (C—N—C<sub> $\alpha$ </sub>—C) angle; panels b and d:  $\Psi$  (N—C<sub> $\alpha$ </sub>—C—N) angle. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

is already reasonable for the original CG model and improves even further for the CG model with additional 1,3,5 PEP-PEP-PEP angle potential. In the lower panels, we compare the sampling obtained with the CG models with 1,5 PEP-PEP distance potential, 1,3,5 PEP-PEP-PEP angle potential, and a case with both additional potentials. We see that the sampling is very similar in all cases, with the 1,3,5 PEP-PEP-PEP angle giving overall the best agreement with the atomistic reference. The comparison for the other torsions along the backbone is not shown, the results are qualitatively identical. Also not shown are data from backmapped atomistic ensembles based on a larger set (400) of CG structures which gave qualitatively the same results. The good agreement shows that the local chain properties of the atomistic peptide chain are well conserved (or can be very well recovered) in the coarse-graining and subsequent backmapping process.

Figure 7 shows the Ramachandran plot, that is, the correlation of the  $\phi$  and  $\psi$  angles averaged over all C<sub> $\alpha$ </sub> atoms for the original atomistic and the backmapped ensembles for the different CG models (with and without the additional 1,3,5 PEP-PEP-PEP angle potential). Here, we observe very good agreement, with a significant improvement for the newer CG model (with additional 1,3,5 PEP-PEP-PEP angle potential) in the region of  $\phi \approx 60$  which should not be sampled very often according to the atomistic reference. We do observe that the CG peptides sample the  $\alpha$ -helical regions slightly too infrequently, but here we would like to point out that at the moment the nonbonded interactions in the CG model are still preliminary. Most importantly they do not include an explicit-hydrogen bonding term, which will be relevant to observe a larger amount of  $\alpha$ -helical conformations (also for our Ala<sub>3</sub> peptide which contains altogether four peptide groups due to the capped ends), a problem which will be addressed in the near future.

#### Extension of the model to Ala<sub>4</sub>

To test if the bonded potentials obtained for alanine in Ala<sub>3</sub> can be transferred to other oligomers, we applied them without reparameterization to Ala<sub>4</sub>, that is, a peptide with nine CG beads. For Ala<sub>3</sub>, we had treated all bonds, angles, and torsions individually (with the exception of all PEP-CAB-PEP angles, for which we used the iteratively refined potential described above), but the potentials for equivalent sites along the chain were very similar. For Ala<sub>4</sub>, we used one potential for each type of interaction which was determined by averaging over the respective equivalent potentials in Ala<sub>3</sub>. Originally, three CG simulations of Ala<sub>4</sub> were carried out, one with the original CG model with only pressure and PEP-CAB-PEP angle refinement, one with the additional 1,5 PEP-PEP distance potential, and one with the additional 1,3,5 PEP-PEP-PEP angle potential. For the regular bond and angle distributions, no differences between the CG models were found and all CG models correspond well to the atomistic sampling.



**Figure 7.** Ramachandran plots for original atomistic simulation (upper left panel), and backmapped ensembles (based on 200 structures) from different CG simulations: CG with iterated PEP-CAB-PEP potential and pressure-corrected solvent (upper right panel); CG model with 1,3,5 PEP-PEP angle potential (lower left panel); CG model with both 1-5 PEP-PEP distance potential and 1,3,5 PEP-PEP-PEP angle potential (lower right panel). The arrows indicate the  $\Psi$ ,  $\Phi$  angles corresponding to the  $\alpha$ -helix and  $\beta$ -strand model structures used in the analysis of the 1,3,5 PEP-PEP angle (Figure 3).



**Figure 8.** Distributions sampled by Ala<sub>4</sub> (nine beads; CG potentials which were parameterized for the seven beads model). Atomistic reference: black solid line; CG with iterated PEP-CAB-PEP potential, and pressure-corrected solvent: green dotted line; CG with additional 1-5 PEP-PEP distance potential: red crosses; CG with additional 1,3,5 PEP-PEP-angle potential: blue squares. Panel a: CAB1-PEP2-CAB2-PEP3 angle; panel b: PEP3-CAB3-PEP4-CAB4 angle; panel c: PEP2-CAB2-PEP3-CAB3 angle; panel d: end-to-end distance (pink circles: CG model with additional 1-5 PEP-PEP distance potential and 1,3,5 PEP-PEP-PEP angle potential simultaneously; blue dot-dashed line: implicit-solvent CG model with additional 1,3,5 PEP-PEP-PEP angle potential).

Results for selected CG dihedral angles and the end-to-end distance are displayed in Figure 8. For the two CAB-PEP-CAB-PEP dihedral angles, the CG simulations reproduce the atomistic distribution well (Fig. 8a), with the best, almost perfect agreement for the CG model with 1,3,5 PEP-PEP-PEP angle potential. For the PEP1-CAB1-PEP2-CAB2 and the PEP3-CAB3-PEP4-CAB4 dihedral angles, the CG simulations reproduce the atomistic distribution well (panel b). However, for the analogous torsion in the center of the peptide (panel c), the situation is slightly different. Here, only the simulation with the 1,3,5 PEP-PEP-PEP angle potential manages to capture the change in the peak heights in the dihedral distribution which is observed in the atomistic simulation. This is quite remarkable as this change in the population of the two states had not been accounted for in the parameterization (the atomistic distribution on which the parameterization of the corresponding torsion potential had been performed was very similar to the one shown in panel b). Figure 8d shows the end-to-end distribution. The range of end-to-end distances is reasonably well reproduced by all CG models with the highest peak somewhat underestimated, whereas the shoulders at shorter distances are overestimated. Here, the CG model with 1,3,5 PEP-PEP-PEP angle potential performs remarkably better than the other models, both in the explicit-solvent and the implicit-solvent case. This again confirms that a potential of this type is essential to reproduce the conformational behavior even of short peptides.<sup>[10,54]</sup> To better understand what is needed to even better reproduce the end-to-end distance distribution of Ala<sub>4</sub>, we performed one quick test with an explicit-solvent CG model where both additional potentials (1,5 PEP-PEP distance and 1,3,5 PEP-PEP-PEP angle) were turned on simultaneously despite the fact that this superposition of potentials over the same DOF is highly questionable [eq. (3)]. Surprisingly, the result for the end-to-end distribution of Ala4 is very good, almost exactly reproducing the atomistic reference (Fig. 8d). It should be noted though that for Ala<sub>3</sub>, the CG model with the combination of 1,5 PEP-PEP distance and 1,3,5 PEP-PEP-PEP angle did not in general produce a better agreement with the atomistic reference (see e.g. Figs. 3d and 7), therefore, we conclude that in the explicit-solvent case a combination of specifically (simultaneously) parameterized 1,3,5 PEP-PEP-PEP angle and 1,5 PEP PEP distance interactions might be an advisable route. We will follow up on this when we address the nonbonded interactions and hydrogen bonding in the future, also taking into account that the exact nature of 1,3,5 PEP-PEP-PEP angle and 1,5 PEP PEP distance interactions is coupled to the intrachain nonbonded interactions (which act between beads that are separated by four or more CG bonds).

## Conclusions

We showed how a CG peptide model can be parameterized based on conformational properties determined from an atomistic simulation in such a way that the CG model samples conformations consistently with the atomistic reference. Consequently, one can reintroduce atomistic coordinates into the CG structures and thereby recover the original atomistic conformational equilibrium. The original structure-based approach of Boltzmann inversion of distribution functions of bond lengths, angles and torsions gives an implicit-solvent CG peptide model that very well reproduces the atomistic results. If one wants to use these bonded interaction functions in a CG model with a CG explicitsolvent representation, an iterative correction for a few selected DOFs is required. Here, it is important to note that we on purpose do not iteratively refine several interaction functions at once since we find that they are coupled and mutually influence each other. In our case, correcting one angular interaction function was sufficient to "repair" both deviations in angle as well as in dihedral distributions. We show that the resulting CG model not only reproduces those properties that had been used for parameterization but also gives a reasonable representation of overall conformational properties such as end-to-end distance or internal angles which characterize the conformations of the peptide chain. However, we found that to really reproduce the end-to-end distance, an additional potential in the 1,3,5 PEP-PEP-PEP chain segments is necessary which is intimately linked to the secondary structures adopted by the chain. This shows that in the case of biological chain molecules with complex conformational equilibria and correlated DOFs, an approach with decoupled bond, angle, and torsion potentials is possibly insufficient depending on the degree of coarse-graining. In some cases, microscopic structural/conformational details are, by construction, lost in the coarse-graining procedure. We show that these "lost" properties can be recovered in the backmapping procedure if the overall conformational sampling of the chain is correctly represented in the CG level of resolution, and that the backmapped ensemble agrees well with the original atomistic one, that is, in a multiscale approach switching between CG and atomistic level is possible.

Even though we use structure-based coarse-graining to obtain potentials for bonds, angles, and torsions and use structurebased potentials for nonbonded interactions, the two types of interactions are treated in a well separated manner. This implies that the results on the capability of such CG models to represent conformational properties of biomolecules in aqueous medium are independent from the question which is the best method to parameterize intermolecular/nonbonded interactions. It is very well possible to use the present strategy to parameterize bonds, angles, and torsions in connection with other methods to determine nonbonded interactions, for example, based on forces or thermodynamic arguments.

One should note that the CG model had been specifically parameterized to reproduce the atomistic ensemble obtained from a specific atomistic force field. If one wanted to set up a multiscale simulation project with a different force field one would have to go through the parameterization again, but it would of course be possible to start from the known potentials from this project and simply add a few refinement steps to adjust to the different reference sampling. This means that the major result of this article is not so much a fixed set of CG interaction functions that is transferable to other atomistic references or to other CG models with nonbonded interactions coming from a different source. We rather provide an analysis of which intramolecular interaction types are required (and how to obtain the potentials from Boltzmann inversion and in some cases iterative Boltzmann inversion) for reproducing the conformational sampling of the peptide and we provide an analysis of the extent one can expect the atomistic and CG ensembles to agree, which is important if one wants to use the two levels of resolution simultaneously or sequentially in a multiscale framework.

The most important next step is to revisit the nonbonded interactions in the CG model. In the present model, we have not yet taken care of special nonbonded interactions along the peptide chain which represent hydrogen bonding interactions and are essential to reproduce secondary structures. This will be crucial to be able to move toward modeling peptide-based (biological or bioinspired) materials with a multiscale modeling approach.

Another most pressing group of questions in CG modeling are those regarding the transferability of CG potentials. In this manuscript, we focused on bonded interaction functions and on correctly representing the local characteristics of the peptide chain. We show that these interaction functions which are determined for short oligopeptides can be transferred to slightly longer peptide sequences. For intramolecular (bond, angle, torsions) interactions one question is to which extent it is possible to obtain CG potentials based on exhaustive atomistic sampling of (reasonably short) oligopeptides and use the resulting interaction functions also for longer sequences where an atomistic reference could not as easily be obtained.<sup>[11]</sup> This is a question of particular importance for problems in the field of biological materials or unstructured proteins, where no well-defined structural reference data exist. A different, very promising route toward transferable potentials explicitly includes a multitude of references in the coarse-graining process resulting in a set of potentials that can be applied to all molecules in the original reference data set<sup>[27]</sup>. We plan to study these and other aspects of transferability more closely in the future.

## Acknowledgments

The authors thank Victor Rühle, Christoph Junghans, Christoph Globisch, Mara Jochum, and Alessandra Villa for many helpful discussions.

**Keywords:** molecular dynamics simulation • coarse graining • multiscale models • backmapping • peptides • oligoalanine

- G. A. Voth, Ed., Coarse-Graining of Condensed Phase and Biomolecular Systems; Chapman and Hall/CRC Press, Taylor and Francis Group, Boca Raton, FL, 2008.
- [2] W. Tschöp, K. Kremer, J. Batoulis, T. Burger, O. Hahn, Acta Polym. 1998, 49, 61–74.

- [3] J. Baschnagel, K. Binder, P. Doruker, A. Gusev, O. Hahn, K. Kremer, W. Mattice, F. Müller-Plathe, M. Murat, W. Paul, S. Santos, U. Suter, V. Tries, *Adv. Polym. Sci.* 2000, 152, 41–156.
- [4] F. Müller-Plathe, ChemPhysChem. 2002, 3, 754–769.
- [5] M. Klein, W. Shinoda, *Science* **2008**, *321*, 798–800.
- [6] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, M. Karttunen, Phys. Chem. Chem. Phys. 2009, 11, 1869–1892.
- [7] C. Peter, K. Kremer, *Soft Matter* **2009**, *5*, 4357–4366.
- [8] T. Head-Gordon, S. Brown, Curr. Opin. Struct. Biol. 2003, 13, 160–167.
- [9] N. Buchete, J. Straub, D. Thirumalai, Curr. Opin. Struct. Biol. 2004, 14, 225–232.
- [10] V.Q.Tozzini, Rev. Biophys. 2010, 43, 333–371.
- [11] O. Engin, A. Villa, C. Peter, M. Sayar, *Macromol. Theory Simul.* 2011, 20, 451–465.
- [12] C. Clementi, Curr. Opin. Struct. Biol. 2008, 18, 10–15.
- [13] N.Go, H. Abe, Biopolymers 1981, 20, 991–1011.
- [14] A. Arkhipov, P. L. Freddolino, K. Schulten, Structure 2006, 14, 1767–1777.
- [15] Y. C. Kim, C. Tang, G. M. Clore, G. Hummer, Proc. Natl. Acad. Sci. USA 2008, 105, 12855–12860.
- [16] V. Krishna, G. S. Ayton, G. A. Voth, *Biophys. J.* **2010**, *98*, 18–26.
- [17] M. Schor, B. Ensing, P. Bolhuis, Faraday Discuss. 2009, 144, 127–141.
- [18] S. Brown, N. Fawzi, T. Head-Gordon, Proc. Natl. Acad. Sci. USA 2003, 100, 10712–10717.
- [19] G. Bellesia, J. Shea, J. Chem. Phys. 2007, 126, 245104.
- [20] H. Nguyen, C. Hall, *Biophys. J.* 2004, 87, 4122–4134.
- [21] T. Bereau, M. Deserno, J. Chem. Phys. 2009, 130, 235106.
- [22] S. Gopal, S. Mukherjee, Y. Cheng, M. Feig, Proteins 2010, 78, 1266–1281.
- [23] M. R. Betancourt, J. Skolnick, J. Mol. Biol. 2004, 342, 635–649.
- [24] J. Maupetit, P. Tuffery, P. Derreumaux, Proteins 2007, 69, 394–408.
- [25] A. Liwo, M. Khalili, H. Scheraga, Proc. Natl. Acad. Sci. USA 2005, 102, 2362–2367.
- [26] P. Màjek, R. Elber, Proteins: Struct Funct Bioinform 2009, 76, 822-836.
- [27] J.W. Mullinax, W.G. Noid, Proc. Natl. Acad. Sci. USA 2010, 107, 19867–19872.
- [28] A. Lyubartsev, A. Laaksonen, Phys. Rev. E 1995, 52, 3730–3737.
- [29] D. Reith, M. Puetz, F. Müller-Plathe, J. Comput. Chem. 2003, 24, 1624–1636.
- [30] S. Izvekov, M. Parrinello, C. Burnham, G. Voth, J. Chem. Phys. 2004, 120, 10896–10913.
- [31] J. Zhou, I. Thorpe, S. Izvekov, G. A. Voth, Biophys. J. 2007, 92, 4289–4303.
- [32] C. Peter, L. Delle Site, K. Kremer, Soft Matter 2008, 4, 859–869.
- [33] W. Noid, J. Chu, G. Ayton, V. Krishna, S. Izvekov, G. Voth, A. Das, H. Andersen, J. Chem. Phys. 2008, 128, 244114.
- [34] L. Monticelli, S. Kandasamy, X. Periole, R. Larson, D. Tieleman, S. Marrink, J. Chem. Theory Comput. 2008, 4, 819–834.
- [35] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, D. Andrienko, J. Chem. Theory Comput. 2009, 5, 3211–3223.
- [36] R. DeVane, W. Shinoda, P. B. Moore, M. L. Klein, J. Chem. Theory Comput. 2009, 5, 2115–2124.
- [37] C.F. Abrams, K. Kremer, Macromolecules 2003, 36, 260-267.
- [38] Q. Sun, R. Faller, Comput. Chem. Eng. 2005, 29, 2380–2385.
- [39] D. Fritz, V. A. Harmandaris, K. Kremer, N. F. A. van der Vegt, *Macromolecules* 2009, 42, 7579–7588.
- [40] A. Lyubartsev, A. Mirzoev, L. Chen, A. Laaksonen, Faraday Discuss. 2010, 144, 43–56.
- [41] A. Villa, C. Peter, N. van der Vegt, Phys. Chem. Chem. Phys. 2009, 11, 2077–2086.
- [42] A. Villa, N. van der Vegt, C. Peter, Phys. Chem. Chem. Phys. 2009, 11, 2068–2076.
- [43] M.R. Betancourt, J. Phys. Chem. B 2008, 112, 5058-5069.
- [44] M. R. Betancourt, S. J. Omovie, J. Chem. Phys. 2009, 130, 195103.
- [45] I.F.Thorpe, J. Zhou, G. A. Voth, J. Phys. Chem. B 2008, 112, 13079–13090.
- [46] J. Zhou, I. F. Thorpe, S. Izvekov, G. A. Voth, Biophys. J. 2007, 92, 4289–4303.
- [47] S. Xiao, W. Stacklies, M. Cetinkaya, B. Markert, F. Gräter, *Biophys. J.* 2009, 96, 3997–4005.

- [48] A. Nova, S. Keten, N. M. Pugno, A. Redaelli, M. J. Buehler, Nano Lett. 2010, 10, 2626–2634.
- [49] Y.Wang, S. Izvekov, T. Yan, G. A. Voth, J. Phys. Chem. B 2006, 110, 3564–3575.
- [50] V. Rühle, C. Junghans, Macromol. Theory Simul. 2011, 20, 472–477.
- [51] V. A. Harmandaris, D. Reith, N. F. A. van der Vegt, K. Kremer, *Macromol. Chem. Phys.* **2007**, *208*, 2109–2120.
- [52] J. M. Sorenson, T. Head-Gordon, J. Comput. Biol. 2002, 9, 35–54.
- [53] A. Mukherjee, B. Bagchi, J. Chem. Phys. 2003, 118, 4733–4747.
- [54] V.Tozzini, W. Rocchia, J. McCammon, J. Chem. Theory Comput. **2006**, *2*, 667–673.
- [55] A. Savelyev, G. A. Papoian, *Biophys. J.* 2009, 96, 4044–4052.
- [56] D. Alemani, F. Collu, M. Cascella, M. D. Peraro, J. Chem. Theory Comput. 2010, 6, 315–324.
- [57] H. Wang, C. Junghans, K. Kremer, *Eur. Phys. J. E* 2009, *28*, 221–229.
- [58] K. Feenstra, B. Hess, H. Berendsen, J. Comput. Chem. **1999**, 20, 786–798.

- [59] H. Hess, H. Bekker, H. Berendsen, J. Fraaije, J. Comput. Chem. 1997, 18, 1463–1472.
- [60] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theory Comput. 2008, 4, 435–447.
- [61] H. Berendsen, J. Postma, W. van Gunsteren, A. DiNola, J. Haak, J. Chem. Phys. 1984, 81, 3684–3690.
- [62] T. Darden, D. York, L. Pedersen, J. Chem. Phys. 1993, 98, 10089–10092.
- [63] C. Oostenbrink, A. Villa, A. Mark, W. van Gunsteren, J. Comput. Chem. 2004, 25, 1656–1676.
- [64] H. Berendsen, J. Grigera, T. Straatsma, J. Phys. Chem. **1987**, *91*, 6269–6271.