Erschienen in: The European Physical Journal Special Topics ; 225 (2016), 8-9. - S. 1463-1481
https://dx.doi.org/10.1140/epjst/e2016-60147-8

Representing environment-induced helix-coil transitions in a coarse grained peptide model*

Cahit Dalgicdir[1], Christoph Globisch[1], Mehmet Sayar[2,a], and Christine Peter[1,b]

[1] Department of Chemistry, University of Konstanz, 78547 Konstanz, Germany
[2] Chemical and Biological Engineering Dept. & Mechanical Engineering Dept., College of Engineering, Koç University, 34450 Istanbul, Turkey

Abstract. Coarse grained (CG) models are widely used in studying peptide self-assembly and nanostructure formation. One of the recurrent challenges in CG modeling is the problem of limited transferability, for example to different thermodynamic state points and system compositions. Understanding transferability is generally a prerequisite to knowing for which problems a model can be reliably used and predictive. For peptides, one crucial transferability question is whether a model reproduces the molecule's conformational response to a change in its molecular environment. This is of particular importance since CG peptide models often have to resort to auxiliary interactions that aid secondary structure formation. Such interactions take care of properties of the real system that are per se lost in the coarse graining process such as dihedral-angle correlations along the backbone or backbone hydrogen bonding. These auxiliary interactions may then easily overstabilize certain conformational propensities and therefore destroy the ability of the model to respond to stimuli and environment changes, i.e. they impede transferability. In the present paper we have investigated a short peptide with amphiphilic EALA repeats which undergoes conformational transitions between a disordered and a helical state upon a change in pH value or due to the presence of a soft apolar/polar interface. We designed a base CG peptide model that does not carry a specific (backbone) bias towards a secondary structure. This base model was combined with two typical approaches of ensuring secondary structure formation, namely a C_α-C_α-C_α-C_α pseudodihedral angle potential or a virtual site interaction that mimics hydrogen bonding. We have investigated the ability of the two resulting CG models to represent the environment-induced conformational changes in the helix-coil equilibrium of EALA. We show that with both approaches a CG peptide model can be obtained that is environment-transferable and that correctly represents the peptide's conformational response to different stimuli compared to atomistic reference simulations. The two types of auxiliary interactions lead to different kinetic behavior as well as to different structural properties for fully formed helices and folding intermediates, and we discuss the advantages and disadvantages of these approaches.

[a] e-mail: `msayar@ku.edu.tr`
[b] e-mail: `christine.peter@uni-konstanz.de`

1 Introduction

Molecular simulations can elucidate the molecular processes and driving forces behind biomolecular structure formation processes such as peptide aggregation, formation of amyloid fibrils, or peptide-based nanostructures [1–5]. Unfortunately, simulations at an atomistic level of resolution are often not able to access the length- and time-scales required for these nanoscale systems. Therefore, quite a large number of different coarse graining approaches have emerged that share the same goal: reproducing the essential features of the biomolecular system while extending the time- and length-scale limits [6–11]. In top-down coarse-grained (CG) models, interaction functions are parameterized so that known (macroscopic) thermodynamic, conformational, or structural properties of the system of interest are reproduced. In bottom-up coarse graining approaches, CG interaction functions are systematically parameterized so that microscopic properties in atomistic reference simulations are reproduced. Especially in the context of the latter type of models, the question of transferability is being discussed as one of the most important challenges to coarse graining [12]. Because of the reduction of number of degrees of freedom, CG models are by construction more intimately connected to the state point of reference where they were parametrized. This means they cannot easily be used at different thermodynamic state points, different phases, different system compositions, etc.

One particular aspect of transferability is the conformational preference of a molecule where molecules adopt different conformational behaviors upon a change in environment. One possibility could be the chemical surroundings: a change in pH value or in salt or co-solvent concentration can drastically affect the conformational preference of a polymer or biopolymer [13–20]. Another factor that can influence the behavior of a biopolymer is the presence of a surface, an interface or a growing aggregate (the latter can be viewed as a microscopic interface) [15,21,22]. Reproducing equilibria between different conformations is already very challenging for a CG model. Reproducing the shift in such an equilibrium due to a change in the environment of the molecule is even more problematic. All interactions in the system (bonded and most importantly nonbonded interactions) need to interplay in a way that the conformational equilibrium between two states is tipped in either of the two directions when the environment changes [23]. This implies that the CG model correctly encompasses the different physical (thermodynamic) driving forces that keep the conformational equilibrium in place (depending on the environmental conditions).

For peptides and proteins, environment-induced conformational changes are among the most important phenomena that determine the richness of their structural landscapes. Molecules that are intrinsically disordered can transform into highly ordered β-sheet rich aggregates or into α-helices – depending on their molecular and chemical environment [24–28].

CG models often resort to auxiliary interactions to reproduce conformational properties, especially in the case of collective ones such as secondary structures. These auxiliary interactions take care, that helices form or that the molecule adheres to a perfect β-sheet structure. Often such interactions cover correlations between underlying atomistic torsional angles that are characteristic for a given secondary structure element. On an atomistic level, regular backbone torsions are not uncorrelated, and these correlations are intimately linked to secondary structure formation, as can best be seen in Ramachandran plots. Unfortunately, many of the physical reasons for these correlations are lost upon coarse graining and need to be brought back by auxiliary interactions in CG models such as longer-range pseudotorsion angles. Another type of frequently used auxiliary interactions that lie at the heart of secondary structure formation are virtual site interactions which mimic the formation of hydrogen bonds between amide groups. One problem of such supportive/auxiliary interactions is that they are usually parameterized to represent a desired conformational feature of a given target system, e.g. the formation of a helix, and that they are only a (poor) substitute for the true physical origin of the conformational behavior. Thus, they are not necessarily able to represent more delicate conformational changes where the system responds to a change in environment, e.g. the presence of an interface or a pH change that perturbs the interactions within the peptide. In the present paper we try to construct a CG model for a system that exhibits such environment-induced conformational changes and investigate, under which conditions models with auxiliary interactions such as pseudotorsion angles or virtual site interactions represent the conformational behavior.

We revisit EALA (Ace-EAALAEALAEALAE-Nme), a short peptide that had been derived from a designed model peptide sequence (the so-called GALA peptide [29]) which undergoes conformational transitions between a disordered and a helical state upon several environment changes: Fig. 1 shows the different states of EALA found in atomistic reference simulations [15]. In bulk aqueous solution at pH 5, the Glutamic acid side chains are protonated, the EALA peptide is overall neutral and adopts an equilibrium between α-helical and partially unfolded structures. At pH 7 and above, the Glutamic acid side chains are deprotonated and the peptide is negatively charged (with a total charge of -4). The resulting electrostatic repulsion between the side chains tips the conformational equilibrium further towards disordered structures and the overall helical propensity of the molecule is reduced.

Due to its amphiphilic nature, the EALA peptide is interface-active. At soft polar/apolar interfaces such as the air/water interface, EALA forms a stable α-helix where the hydrophobic Leucine side chains are neatly partitioned into the air and the Glutamic acid side chains are immersed in the water phase. At the interface, the α-helical form of EALA is even stable at pH 7, i.e. when the side chains are deprotonated. This means that the partitioning of the side chains that stabilizes the helix overrides the destabilization due to the electrostatic repulsion. This rather delicate balance between different driving forces that stabilize/destabilize the α-helix make EALA a suitable test system to investigate, whether a CG model can reproduce such environment-dependent conformational transitions.

The aim of this study is to compare different coarse-graining approaches in light of representing this environment dependence. In particular, we investigate different types of auxiliary interactions that are used to aid the formation of secondary structures. Our goal is to compare pseudodihedral (PD) [30,31] and virtual site (VS) [32,33] interactions in terms of their capability to capture the environment-driven conformational changes observed in the atomistic system (see Fig. 1) when they are balanced with the remaining bonded and nonbonded interactions in a CG model for the EALA peptide. We evaluate the representation of the conformational equilibrium as well as the dynamic and kinetic characteristics of the peptide system. We aim for a single CG model with a unique set of interaction potentials, that should be able to capture the peptide's conformational behavior in bulk aqueous solution at low and high pH as well as in contact with a soft polar/apolar interface as observed in the atomistic system. We address the questions: Can one achieve this at a CG level? Is a pseudotorsion or a virtual-site based approach equally well-suited in this respect? Are there other characteristic differences between the two auxiliary-interaction approaches?

2 Methods

2.1 Computational details

CG molecular dynamics (MD) simulations were carried out with GROMACS 4.5.7 [34]. All nonbonded interactions between the CG beads have been implemented to

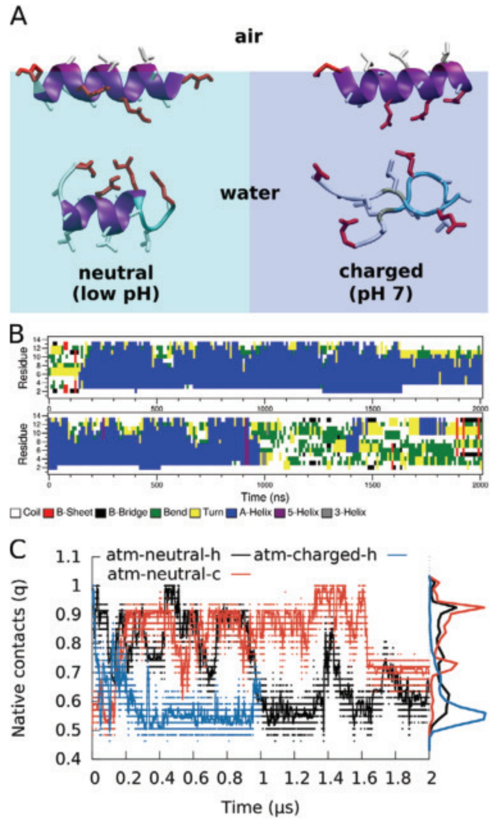


Fig. 1. A) The conformational preference of the EALA peptide depends on the environment: the disordered structure (bottom-right) found in aqueous solution at pH 7 (labeled as charged) can be driven towards a partially stable α-helix conformation (bottom-left) at low pH (labeled as neutral) due to the protonation of the Glutamic acid side chains. Interaction with a hydrophobic/hydrophilic interface stabilizes the α-helix conformation regardless of the pH (top left and right). B) The partially stable α-helix structure in bulk water at low pH displays slow dynamics: random (top) and α-helix (bottom) initial conformations display microsecond scale transitions as shown with the DSSP[47] analysis of secondary structure evolution. C) Comparison with an ideal α-helix conformation via native contact analysis displays the increased helical propensity for the neutral molecule at low pH (black curve: helical initial conformation; red curve: random initial conformation) compared to the charged state at pH 7 (blue curve).

GROMACS as tabulated potentials with a cutoff distance of 3.0 nm. Details regarding the different interaction potentials are given in the following sections. The leap-frog stochastic dynamics integrator [35] was used with a time step of 2 fs. To be able to correctly sample the vibrations of the stiff bond between the CA and A beads, the time step was not increased compared to the atomistic model. The system temperature was maintained at 298 K with a friction constant of 1 ps^{-1}. For Hamiltonian replica exchange simulations, PLUMED 2.1 with the Hamiltonian replica exchange fix and GROMACS 4.6.7 were used as simulation tools [36,37]. The VOTCA package [38] was used to map the CG beads onto the atomistic structure and extract the probability distributions from the CG simulations.

Analysis and the methodology for the MD simulations for all-atom reference states were already published in our earlier study [15]. Reference simulations of the

Table 1. Mapping of atoms from the atomistic reference to the CG beads.

atomistic	coarse-grained
C_α	*CA*
backbone C-O-N-H	*CN*
leucine sidechain	*L*
alanine sidechain	*A*
glutamic acid CH_2-CH_2	*G*
glutamic acid COOH	*GO*

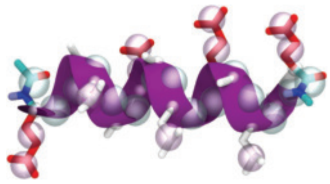


Fig. 2. CG mapping of EALA.

EALA peptide (charged and neutral) in bulk aqueous solution and in contact with an air/water interface were $2\,\mu$s long and the peptides were initially α-helical unless otherwise stated. The atomistic system was simulated on 40 cores (≈200 ns/day) amounting to 4650 CPUh per system. For the CG systems, we have accumulated at least $5\,\mu$s in single core simulations (≈$1\,\mu$s/day). For select systems we have improved sampling by Hamiltonian Replica Exchange amounting to $60\,\mu$s. In total the reported CG simulations sum up to $200\,\mu$s.

For visualization and production of molecular representations the VMD [39] package was used and the secondary structures were determined by the STRIDE algorithm [40].

2.2 Mapping between CG and atomistic resolution

The mapping scheme used for the CG model is listed in Table 1 and the superposition of the CG model onto the atomistic structure is shown in Fig. 2. The backbone of the CG model is composed of *CA* and *CN* beads. Leucine and alanine side chains are represented via a single bead (*L* and *A*, respectively), whereas glutamic acid side chains are mapped as *G* and *GO* beads, where the latter carries the charge at pH 7. All CG beads are mapped to the center of mass of their respective atom groups.

2.3 Bonded and nonbonded interactions

The basic, underlying CG interactions (from now on referred to as base model) were parameterized to reproduce structural descriptors such as bond, angle, torsion distributions for a generic peptide without a specific secondary structure. Bonded interaction potentials (bonds, angles, and dihedrals) were obtained by Boltzmann inversion (BI) of reference distributions [41,42] from atomistic simulations of EALA. Harmonic potentials were used for the bond-stretching potentials, where the spring constant and equilibrium position were obtained from a fit to the potential form obtained from BI. The angular interaction potentials were obtained from BI of the probability distribution of the angle between the relevant triplet of beads in the atomistic simulation of an α-helical EALA peptide at the air/water interface. All CG beads separated

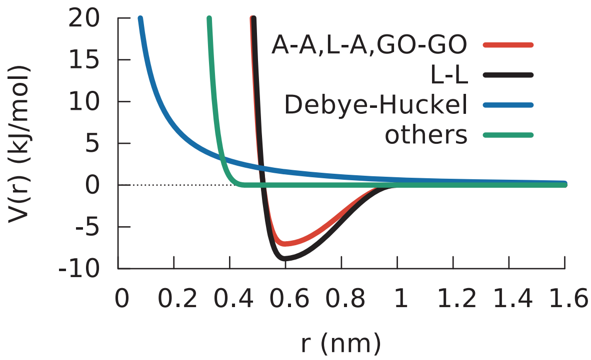


Fig. 3. Non-bonded potentials for the base CG model that were subsequently combined with the PD and VS interactions (see text).

by two covalent bonds interact via an angular potential, and in total nine different potential types were used.

Dihedral interaction potentials were used for the backbone beads only, whereas the torsional stiffness of the side chain beads were represented via improper dihedral potentials. For glutamic acid side chains the *GO* bead was not connected via a dihedral interaction, since the atomistic reference yields a very broad distribution. In order not to bias the base CG model towards any secondary structure, the dihedral potentials were obtained from an all-atom simulation of a single neutral EALA peptide in implicit water. In this simulation nonbonded interactions were used only for atoms that are part of CG beads which interact in the CG representation, and all other nonbonded interactions were excluded. In the presence of these exclusions the peptide samples random conformations, with no bias towards any secondary structure.

In Fig. 3 nonbonded interaction potentials for the CG beads are shown. Hydrophobic forces acting on leucine and alanine side chains were represented via Lennard-Jones cosine potentials (Eqn. 1), which are shown in Fig. 3.

$$U = \begin{cases} \frac{27}{4}\epsilon\left[\left(\frac{\sigma}{r}\right)^9 - \left(\frac{\sigma}{r}\right)^6\right] & \text{if } 0 \le r \le r_{min} \\ -\epsilon\, cos^2\left[\frac{\pi(r-r_{min})}{2(r_c-r_{min})}\right] & \text{if } r_{min} < r \le r_c \\ 0 & \text{if } r > r_c \end{cases} \tag{1}$$

where $r_c = 1.0$ is the cutoff value for the nonbonded potentials and r_{min} is the minimum of the Lennard-Jones potential [43].

For the *GO-GO* interactions at low pH an attractive Lennard-Jones interaction was used, shown in Fig. 3. At pH 7, when the *GO* side chains are deprotonated and charged, this attractive interaction is replaced with the Debye Hückel potential

$$u_{ij}^{el}(r) = q_i q_j \frac{e^{-r/\xi}}{4\pi D r} \tag{2}$$

where q_i is the charge of the bead, D is the dielectric constant of the solvent medium and ξ is the Debye screening length. Normally $D = 80$ for dielectric constant of water at near ambient conditions and $\xi = 1$ nm at physiological salt concentration of approximately 100 mM. In our simulations $q_i q_j = 1$ because the glutamic acid side chains are charged with -1. For all other nonbonded interactions generic repulsive Weeks-Chandler-Andersen potentials with $\sigma = 0.408$ were used (see Fig. 3, labeled as *others*) [44].

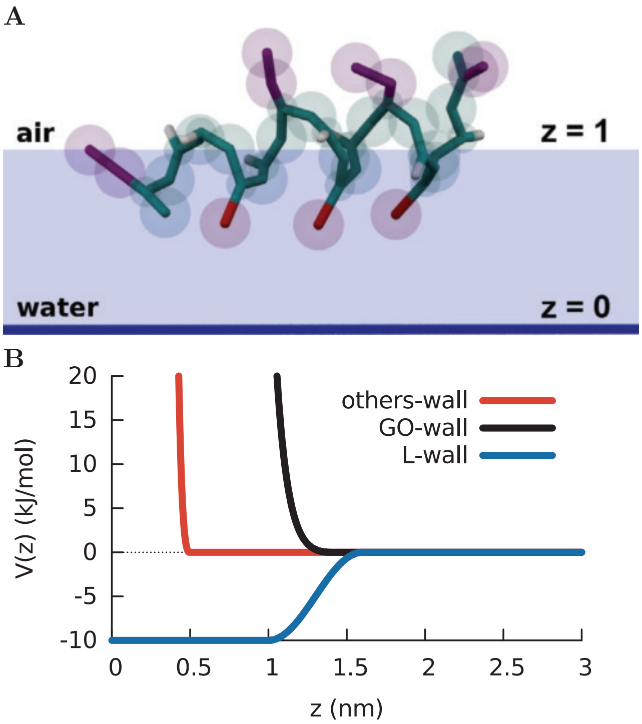


Fig. 4. A) CG representation of the partitioning effect of the hydrophobic/hydrophilic interface. The interface between air and water resides at $z = 1$ nm. B) Interaction potentials between the CG beads and the wall.

2.4 Representation of the polar/apolar interface

In order to represent an interface which enforces partitioning of the hydrophobic and hydrophilic residues at the CG level a soft wall was used [45]. Periodic boundary conditions were used in the x-y plane, and two walls were positioned at the bottom and top of the box in z direction. In Fig. 4A the wall at $z = 0$, the position of the interface at $z = 1$, and the EALA peptide's partitioning at the interface are depicted.

The interactions for the hydrophobic and charged particles were kept identical to the original parametrization of the LK peptide ($(\mathrm{LKKLLKL})_2$), that also displays environment-dependent conformational transition [21,45]. An attractive potential was used for the L-wall interactions (corresponding to the hydrophobic side chains) and a (soft) repulsive potential was used for the GO-wall interaction (corresponding to the hydrophilic side chains). The attractive interaction gradually increases from 0.5 nm above the interface and remains flat underneath it. The remaining CG beads do not interact with the interface, except a repulsive potential within 0.5 nm to the wall to keep the beads inside the simulation box.

2.5 Native contact analysis

The similarity of the CG conformations to an ideal α-helix was quantified by the number of native contacts between the CN beads within 0.55 nm. The native contact parameter q is normalized with the contact number for an ideal helix. This normalization resulted in a range of $q \approx 0.5$ to $q \approx 1.0$, which correspond to the random and

α-helix conformations, respectively. CN beads were chosen since the atoms which participate in backbone hydrogen bonding in the atomistic system are mapped on to these beads in the CG model. Using the timeline data the folding and unfolding events were monitored and distributions were computed. Figure 1 shows for two atomistic simulations of the neutral molecule in bulk aqueous solution both a standard secondary structure analysis (DSSP) and the native contact data. The figure nicely demonstrates that the native contact analysis is able to distinguish between fully formed helices, partially formed helices and random coil structures.

For one of the CG simulations of the neutral molecule in bulk water we analysed in detail which conformations were hidden behind certain q values: Representative conformations for different ranges of q values in Fig. 5B were obtained by a two stage clustering. First, by performing the native contact analysis for the $5\mu s$ simulation conformations were binned into q-values of width 0.1. Next, the conformations collected under each bin were structurally clustered based on the CA beads' RMSD. The structural clustering was performed with the g_cluster tool. Jarvis Patrick algorithm was used, where a new structure is added to a cluster if it shares at least 3 neighbors with another structure already present in the cluster. Neighbors are defined as the 10 closest structures. We report the largest representative clusters for all native contact bins, giving their relative size in percent of the fraction.

3 Results

In this study we compare two alternative methods for construction of a CG model in order to represent the stability of the helix conformation in EALA in four atomistic reference states. First, we investigate a CG model which relies on a pseudo dihedral (PD) potential to stabilize the helix conformation. In the second CG model, the helix conformation is stabilized by mimicking backbone hydrogen bonding via a virtual site (VS) interaction. In both PD and VS based CG models, the nonbonded interactions are tuned in a structural manner, to capture the atomistic system behavior. The equilibrium position of the potentials is chosen according to the atomistic references, and the strength is tuned to mimic the four atomistic reference systems.

3.1 The pseudo dihedral CG model in bulk water

The helix propensity of EALA peptide is strongly coupled to the environment as discussed in the introduction. The slow dynamics and computational limitations in all-atom simulations do not allow us to obtain an equilibrium probability distribution for the helix propensity in neither the neutral nor the charged state of the EALA peptide (see Dalgicdir et al. 2015 [15] and Supplementary Material). However, the native contact analysis shown in Fig. 1C indicates that the charge repulsion among Glutamic acid side chains significantly reduce the helix propensity.

In order to construct a CG model which displays a similar dependence on pH and presence of an air/water interface, we utilized a pseudo-dihedral potential which acts on four consecutive CA beads in the CG representation, which we will refer to as CA-CA-CA-CA pseudo-dihedral (PD). As we have demonstrated in our earlier study [15], the helix conformation is fully stable at the air/water interface. Therefore, by Boltzmann inversion of the probability distribution for the CA beads (obtained from the all-atom simulation at the air/water interface) the general form for the PD potential is obtained (see Supplementary Material). Next, by scaling this dihedral potential by 0.6 we optimized the bulk behavior of the CG model, such that together

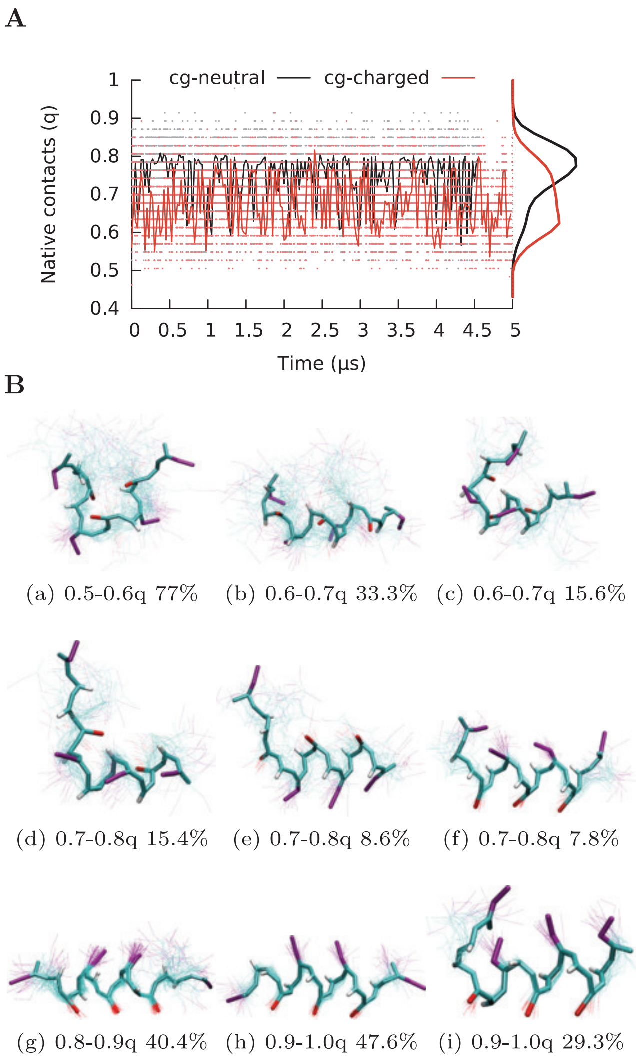


Fig. 5. The pseudo dihedral CG model captures the pH-dependent conformational change in bulk water: A) Comparison of the native contacts for the PD-CG model in the neutral and charged state. The probability distribution for the native contacts (on the right side) displays the reduced helix propensity for the charged peptide. B) Representative structures obtained from the clustering analysis of the simulation structures within certain ranges of native contact values q (see Methods section).

with the bonded and nonbonded interactions (as explained in the Methods section) the CG model reproduces the pH-dependent helix-propensity change.

The native contact analysis of the PD-CG model for the neutral and charged EALA molecule is displayed in Fig. 5. During the 5 μs simulation the CG molecule displays rapid changes in the native contacts. The probability distribution shown on the right hand side, clearly demonstrates the reduced helical propensity in the molecule when the Glutamic acid side chains are charged. Also shown in Fig. 5 is the clustering analysis of the native contact data. In both the neutral and charged states the peptide displays a variety of different conformations, ranging from full or

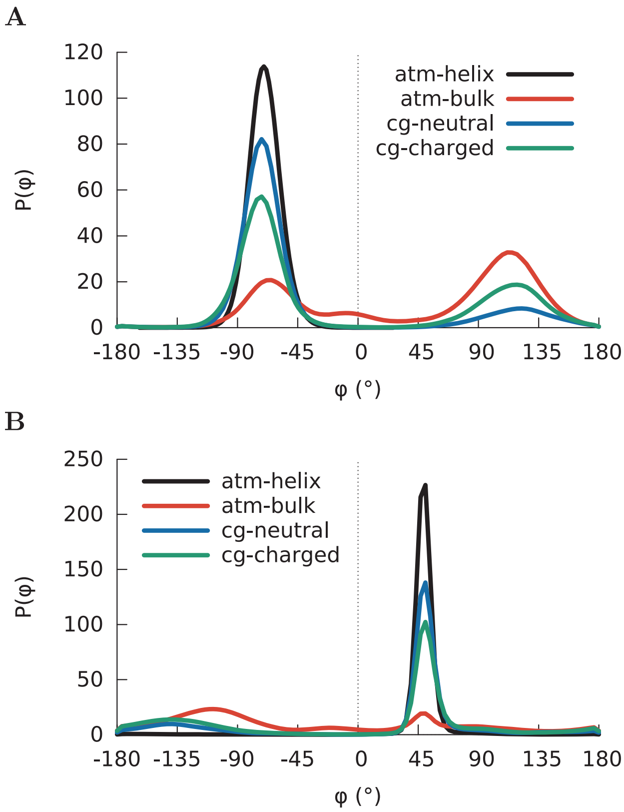


Fig. 6. *CN-CA-CN-CA* (panel A) and *CA-CA-CA-CA* (panel B) dihedral distributions for the PD-CG model in bulk water. *atm-helix* represents the atomistic reference at the air/water interface where the peptide remains α-helical and *atm-bulk* refers to the atomistic case where the charged peptide unfolds and yields a disordered structure. The repulsive interactions between the glutamic acid beads decrease the helicity.

partially helical structures, to completely unfolded conformations. However, the neutral molecule clearly has an increased helical propensity in agreement with the all-atom simulations described in our earlier study [15]. In the charged state a certain level of helicity remains, but fully formed helices are hardly observed. The destabilization of helical structures in the charged state arises from the Coulomb repulsion between the Glutamic acid side chains, which is represented by a Debye-Hückel potential in the CG model. Note that due to the slow kinetics the all-atom simulations do not yield the equilibrium probability distributions. Hence, in the absence of a quantitative all-atom reference data, our aim is to capture the change in conformational behavior in a qualitative manner with the PD-CG model. This model has a higher transition rate between different conformations, which enables a thorough sampling of the conformational phase space. However, these rapid transitions are potentially unphysical due to the phenomenological nature of the PD potential.

In addition to the native contact analysis, we also evaluate the backbone conformations via the backbone dihedral angles. In Fig. 6 the probability distributions for the backbone dihedrals between the *CN-CA-CN-CA* beads and the pseudo-dihedral between the *CA-CA-CA-CA* beads are presented. Unlike the native contact analysis, which provides a global major for the conformational preference, the dihedral distributions enable a comparison of the all-atom and CG simulations at the local

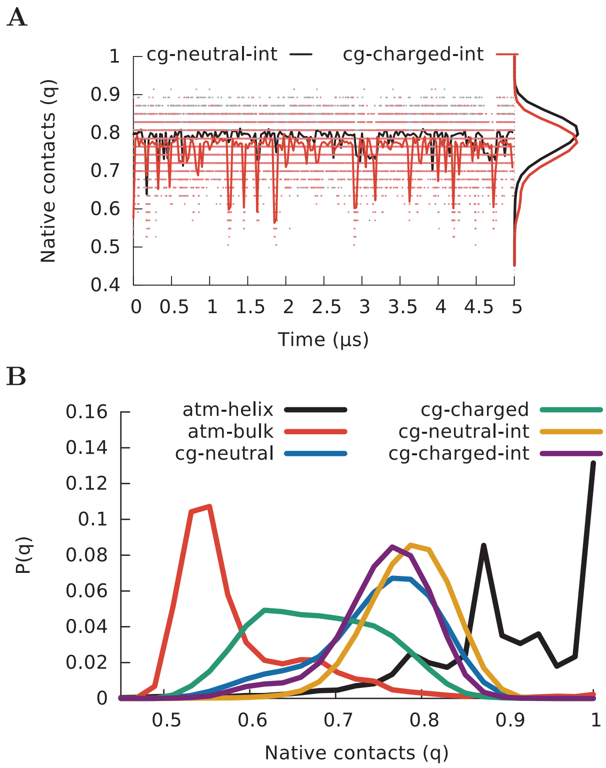


Fig. 7. Native contact analyses for the PD-CG model in simulations with an interface. A) Timeline data for simulations with the PD-CG model for the charged and the neutral state. B) Native contact distributions for simulations of the PD-CG model with and without an interface for the charged and the neutral state. In addition atomistic reference data for a fully-formed α-helix (black curve) and an unfolded/disordered molecule in bulk aqueous solution (red curve) are given.

structure level. Once again, the all-atom simulation results for the fully-formed helix and the charged state in bulk are shown to represent the two possible extremes for the conformational behavior, and not necessarily the equilibrium distribution for the charged and neutral states of the EALA peptide in bulk water. In this regard, our PD-CG model also captures the local structural features, and mimics the reduced helical propensity with these two dihedral angles.

3.2 The pseudo dihedral CG model at the interface

In this section, we turn our attention to the behavior of the PD-CG model at the interface. As explained in the model section, the interaction potentials between the peptide and the interface are chosen to mimic the partitioning of the hydrophobic and hydrophilic residues observed in all-atom simulations [15]. In Fig. 7A the native contact analysis of the simulations for a single CG peptide at the interface are shown for the charged and neutral peptide. A comparison of the probability distribution for the PD-CG model in bulk and at the interface, as well as the atomistic reference for the helix and the unfolded charged state are also shown in Fig. 7B. Unlike the bulk behavior, at the interface both the neutral and charged peptides adopt a helix

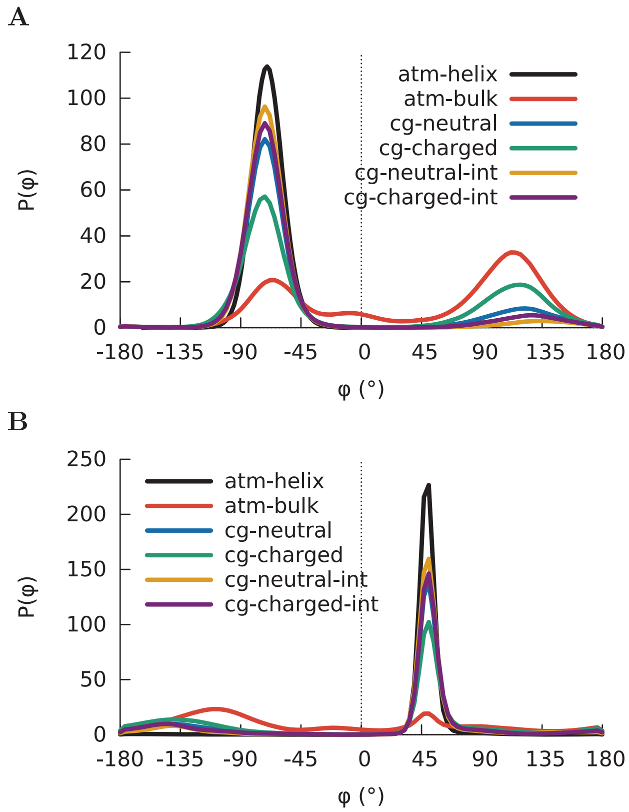


Fig. 8. Comparison of the *CN-CA-CN-CA* (panel A) and *CA-CA-CA-CA* (panel B) dihedral distributions in bulk water and in the presence of an interface for the neutral and the charged cases for the PD-CG model. The atomistic references *atm-helix* and *atm-bulk* refer to the cases at the air/water interface and in explicit bulk water respectively. In the former case the peptide is folded into an α-helix, whereas in the latter one it is disordered.

conformation, in agreement with the all-atom results [15]. Figure 8 shows the backbone dihedral distributions of the peptide with the PD-CG model in all four states (charged/neutral; bulk/interface) and further confirms the stabilization of the helical conformation at the interface. This stabilization is a direct result of the partitioning effect of the interface, where hydrophobic leucine and alanine residues are attracted, and hydrophilic *GO* beads are repelled by the wall. This effect even overcomes the electrostatic repulsion between the deprotonated Glutamic acid side chains, i.e. the repulsion between the *GO* beads in the charged state. The timeline data shown in Fig. 7 hints at the marginal stability of the helix in the case of the charged molecule, where frequent drops in the native contact value are observed. On the contrary, the neutral molecule displays a stable α-helix for the entire 5 μs simulation.

3.3 The virtual site CG model (VS-CG)

In the previous section we introduced the PD-CG model of EALA, where the peptide's conformational behavior and secondary structure propensities are enabled by the use of a pseudo torsion angle. The PD-CG model successfully represents environment induced conformational transitions such as pH-induced unfolding of a helix or

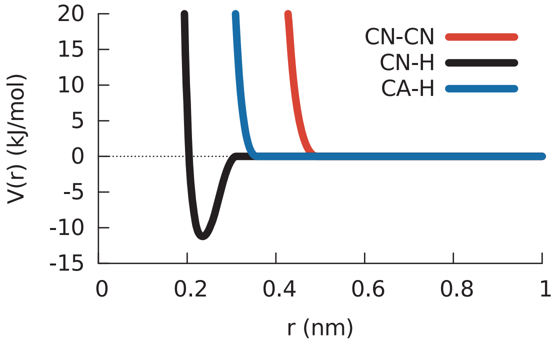


Fig. 9. Tabulated non-bonded potentials between the backbone beads for the VS-CG model. The H-bead represents the virtual site.

interface induced stabilization of a helix. In the present section we will use the existing model and replace this supportive/auxiliary pseudo torsion angle by a virtual site interaction which mimics the hydrogen bond interaction between amide groups. Here, a virtual site, H_i was reconstructed from a non-linear combination of three beads, CN_i, CA_i and CA_{i-1} and out of the plane of these atoms [45]. The position of the virtual site is calculated using,

$$r_s = r_i + ar_j + br_k + c\,(r_j \times r_k) \tag{3}$$

where r_s is the position of the virtual site, and i, j and k represents the beads CN, CA_{i-1} and CA_i respectively. $a = 1.8167$, $b = 1.6320$ and $c = 1.6503$ [46].

The virtual site, H, has a favorable interaction with the CN backbone beads at distances typical for a hydrogen bond. Due to the mapping used in our CG model, the hydrogen bonds between the i'th and $i+1$'th residues along the backbone in the atomistic system correspond to an interaction between virtual site H_i and CN_{i+3} in the CG model. The initial strength of the VS interaction is shown in Fig. 9. The remaining sets of CG interactions (i.e. the interactions within the neutral peptide, the electrostatic repulsion in the case that the side chains are deprotonated, and the interactions with the interface) have not been modified. With this virtual site interaction, the EALA peptide forms a very stable helix.

Next, we attempted to fine-tune this interaction strength. Unfortunately, folding/unfolding transitions become very slow when VS interactions between amide groups are employed to aid the formation of a helix. Therefore, equilibration problems impede a careful parametrization of the VS interaction strength in such a way that the helix is just stable, and that it unfolds when a physically reasonable (electrostatic) repulsion between the side chains is added. Enhanced sampling methods are required to exactly assess the degree of helicity for those interaction strengths where the system is (very slowly) folding and unfolding. To this end Hamiltonian Replica Exchange MD (H-REMD) simulations of the charged peptide have been used. By turning off the virtual site contribution to the Hamiltonian linearly across 30 replicas, the range of interaction strengths where the system tips from a fully folded helix to an unfolded coil structure could be determined. For details and results of the H-REMD simulations see the Supplementary Material.

Based on the H-REMD simulations we concluded that with a virtual site interaction strength around 12 kJ/mol one obtains a reasonable level of charge induced unfolding. Unfortunately, H-REMD could not fully resolve the equilibration problem: the timescale of H-REMD appears still insufficient, since even with relatively weak virtual site interactions folding transitions are slow on the timescale available

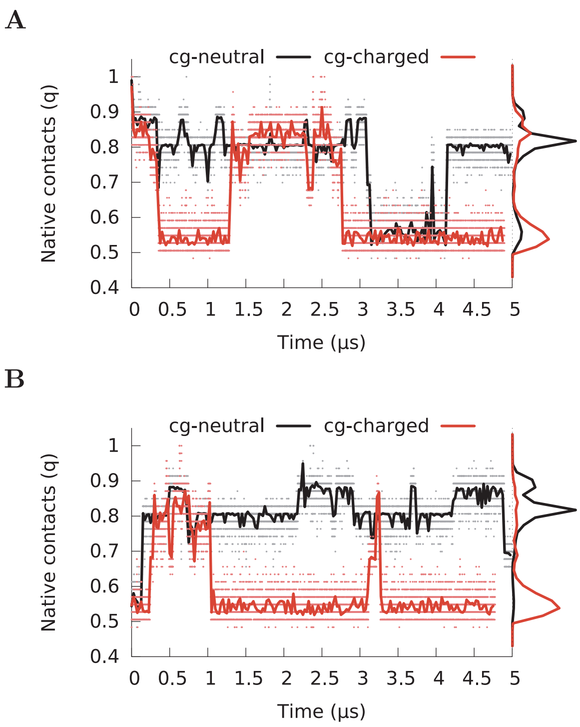
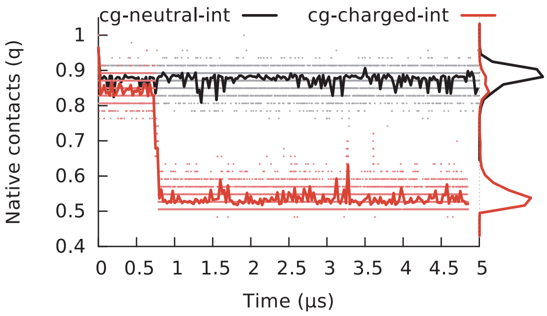


Fig. 10. Native contacts timeline data for the VS-CG model in bulk water with different initial conformations: helical (panel A) and extended (panel B).

to H-REMD simulations. Therefore we carried out long ($5\,\mu$s) MD simulations of the neutral and charged peptide both in bulk water and in contact with the interface at VS interaction strengths around 12 kJ/mol, i.e. 11.2, 11.9 and 12.6 kJ/mol. In order to get an impression of the folding/unfolding equilibrium in spite of the slow dynamics that allows only few folding/unfolding transitions per simulation, for each system two sets of simulations with different initial conditions were performed, namely a helical and an extended starting structure. A complete set of results can be found in the Supplementary Material, showing dihedral distributions and native contact analysis for all three VS interaction strengths. Here we only present native contact data for a VS interaction strength of 11.2 kJ/mol (black curve in Fig. 9) since for this value we find phenomenologically the correct level of charge induced unfolding.

Figure 10 shows that – independent of the starting structure – folding/unfolding transitions are observed in both peptide states (neutral and charged). This VS-CG model correctly distinguishes between the neutral and charged states, i.e. the helix is substantially more stable in the neutral form (possibly more stable than in the comparable CG model with pseudo torsion angle). The figure shows that conformational transitions in the VS-CG model more closely resemble the atomistic behavior than in the PD-CG model, i.e. transitions are rare (typically on the microsecond scale), sharp, and often occur via well-defined half-formed helices. This is not entirely unsuspected, since physical nature (and the resolution) of the short-range H-bond interaction in the VS-CG model is almost atomistic. One could suspect that cooperativity effects and collective motions in helix formation are better represented in the VS-CG compared

A



B

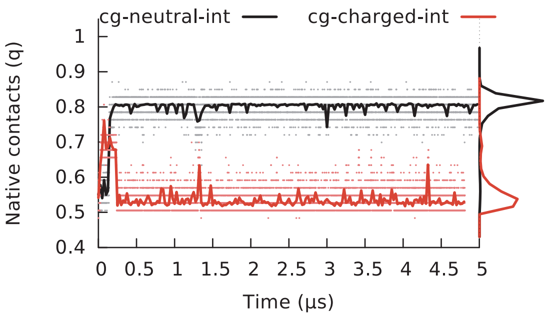


Fig. 11. Native contacts timeline data for the VS-CG model at the interface with different initial conformations: helical (panel A) and extended (panel B).

to the PD-CG model. Therefore, while the extremely long timescales are a disadvantage compared to the PD approach, the VS approach probably is the more physical one in terms of transitions and transition states.

Next, the behavior of the peptide with the VS-CG model at the interface reveals that the VS model in its present form does not fully reproduce the atomistic behavior. Figure 11 shows that for the neutral system, the wall has the expected stabilizing effect of the helix. The unfolded molecule rapidly folds into a helix and in general helices, once formed, do not unfold within the probed timescale. The charged system, however does not fold (or remain folded) at the interface. The data (including the data for the slightly stronger VS interaction of 11.9 kJ/mol that are presented in the Supplementary Material) suggest that the – already very slow – dynamics is further slowed down by the interface, possibly the partitioning at the interface hinders collective moves that are required for conformational transitions in this model. Thus, it is almost impossible to decide, whether the data at hand correctly reflects the helical propensity/folding equilibrium of the peptide (especially in the charged state shown in the Supplementary Material) at the interface. By further increasing the VS interaction strength to 12.58 kJ/mol the helix was observed to be too stable and the peptide was locked in a helical state (see Supplementary Material).

Summarizing, the VS model correctly represents the conformational helicity changes upon turning on/off the electrostatic repulsion. However, in its present parametrization the model fails to capture the helix stabilization at the interface for the charged state. Note that so far only the VS interaction strength has been varied, the remaining CG interactions have been taken from the PD model since the aim of this study was to compare the behavior of the system and its response to environmental changes between the two approaches, not to optimize the models. Further fine tuning of the VS model by better balancing the relative interaction strengths (most importantly the ones between the peptide beads and the wall, the charge repulsion, and the virtual site interaction) could be done. Given the long timescales, however, this is best done using different types of advanced sampling methods (e.g. probing the mechanical stability of the peptide by pulling). We are confident that this parametrization could be successful.

4 Discussion and conclusion

In this paper we present a CG model for the EALA peptide that correctly represents the conformational response of the peptide to different environmental stimuli. We combined an underlying base CG peptide model that does not yet carry a specific (backbone) bias towards a secondary structure with two typical approaches of ensuring secondary structure formation and investigate their ability to represent environment-induced conformational changes.

The base CG interactions have been parameterized to reproduce structural descriptors such as bond, angle, torsion distributions for a generic peptide without a specific secondary structure. The model was on purpose kept rather generic as far as non-bonded interactions are concerned, with analytical interaction potentials that were only moderately tuned to represent system-specific properties (excluded volume; hydrophobicity/hydrophilicity). Without further auxiliary interactions this is a generic peptide model without hard-encoded conformational preferences – except of course for nonbonded partitioning effects due to attractions between the side chains and electrostatic repulsion in case of the deprotonated side chains. Since neither backbone hydrogen bonding nor typical backbone correlations were explicitly accounted for, one would not expect that the peptide with this base CG model easily adopts secondary structures, and indeed, it didn't.

To this base model, two types of auxiliary/supportive interactions were added that are frequently employed in CG peptide models. We used a *CA-CA-CA-CA* pseudodihedral potential that adds secondary-structure-related backbone correlations or a virtual-site interaction that mimics backbone hydrogen bonding. We investigated, if the CG models with these two auxiliary interactions are capable of reproducing the shifts in the helix/coil equilibrium of EALA that are induced by deprotonating/protonating the side chains (i.e. turning on/off electrostatic repulsion) or by the presence of a soft polar/apolar interface that leads to a strong partitioning of the side chains.

We found that in both cases, the CG model is in principle capable of reproducing a shift in the conformational equilibrium. The PD model captures the behavior of atomistic EALA both in bulk and at the interface, i.e. with one set of CG potentials one is able to mimic the behavior of the molecule in four different states. The PD model not only reproduces the conformations of the two extreme cases of the folding equilibrium, i.e. the fully formed helix and the coil state, it also gives a realistic depiction of half-helical intermediate states (comparable to the states found in the atomistic reference ensemble).

We further found that the PD model displays rather rapid transitions between different conformations of the molecule, i.e. coarse graining leads to a significant speed-up of the dynamics compared to an atomistic simulation – possibly at the cost of the conformational transitions not being always physical. The VS model also qualitatively correctly represents the shift in the helicity upon turning on/off electrostatic repulsion. In addition, it captures a further stabilization of the helix conformation when the protonated (neutral) peptide is in contact with the interface. However, the

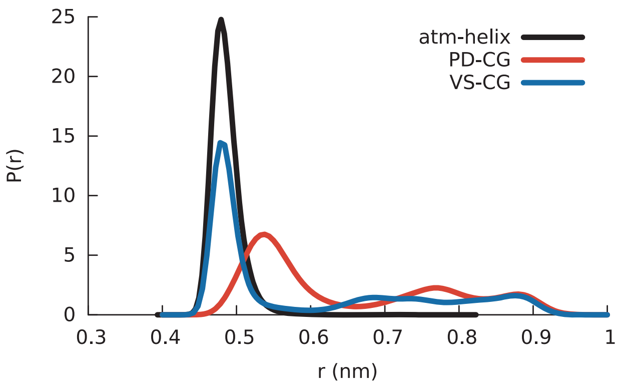


Fig. 12. Comparison of the histograms for the distances between the CN_i-CN_{i+3} beads for the different coarse graining approaches. CG simulations are shown for a neutral peptide in bulk water. For comparison the CN_i-CN_{i+3} distances for an atomistic α-helix are shown (black curve).

model fails to capture the helix stabilization at the interface for the charged state. Here, a further fine tuning by better balancing the relative interaction strengths – including those of the base peptide model, that had deliberately not been altered compared to the PD model – would have to be done and would likely be successful. This has not been done here, since this was not the primary aim of this study. In the present parametrization, the dynamics of the VS model is very slow, considerably slower than that of the PD model. Conformational transitions in the VS model were rare (typically on the microsecond scale) and more distinct compared to the PD model. Therefore the transitions in the VS model more closely resemble the atomistic behavior. We suggest that this is due to the fact that the short-ranged, directional H-bond (VS) interaction more closely resembles an atomistic interaction. The fact that the PD potential is a "phenomenological fix" to "repair" for missing backbone correlations on the CG level of resolution does also show in other descriptors that are sensitive to the precise helical structure. One example is the distribution of the distance between the CN_i-CN_{i+3} backbone beads (see Fig. 12). These distances are better reproduced by a model that has a hydrogen bonding term such as the VS model compared to the PD model.

In summary, the different approaches of the CG models to reproduce the helicity lead to similar conformational behavior with different dynamics. The dynamics of the PD-model is faster than the VS-model. The slow folding/unfolding kinetics in the VS model are a clear disadvantage compared to the PD approach, where obtaining a well-equilibrated ensemble is easily possible, both in bulk and at the interface. On the other hand, the VS approach is probably the more physical one in terms of the transitions and transition states. Still, with both approaches a CG peptide model can be obtained that is environment-transferable and that correctly represents the peptide's conformational response to different stimuli. In particular the results related to the interactions with interfaces are promising since the representation of interface-induced folding is a necessary prerequisite to use these models for the investigation of aggregation and nanostructure formation processes. We expect that the methodology presented here for the parametrization of these auxiliary potentials and the conclusions regarding the extent of transferability are valid beyond the EALA sequence investigated here.

M. Sayar thanks TÜBİTAK (grant no. 212T184) and TÜBA Distinguished Young Scientist Award (2012 awardee) for financial support. C. Peter gratefully acknowledges financial support by the German Research Foundation (grant PE 1625/3). Part of this work was performed on the computational resources funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC.

This paper is part of the Special Issue to honor Kurt Kremer's 60th birthday. Kurt has always been a source of inspiration and motivation in our research with his vision about multiscale simulation of soft matter systems. We are thankful for his continuous support since our early careers, providing guidance as well as encouraging independence. The Max Planck Partner Group Agreement between M. Sayar at Koc University and Kurt's Theory Group at MPIP, Mainz and his hosting C. Peter's Emmy Noether group has laid the foundation to our current collaboration.

References

1. R.O. Dror, R.M. Dirks, J.P. Grossman, H. Xu, D.E. Shaw, Annu. Rev. Biophys. **41**, 429 (2012)
2. A. Morriss-Andrews, J.-E. Shea, Annu. Rev. Phys. Chem. **66**, 643 (2015)
3. J.-E. Shea, C.L. Brooks III, Annu. Rev. Phys. Chem. **52**, 499 (2001)
4. C. Wu, J.-E. Shea, Curr. Opin. Struct. Biol. **21**, 209 (2011)
5. J.A. Lemkul, D.R. Bevan, ACS Chem. Neurosci. **3**, 845 (2012)
6. E. Brini, E.A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, N.F.A. van der Vegt, Soft Matter **9**, 2108 (2013)
7. S. Riniker, J.R. Allison, W.F. van Gunsteren, Phys. Chem. Chem. Phys. **14**, 12423 (2012)
8. M.G. Saunders, G.A. Voth, Curr. Opin. Struct. Biol. **22**, 144 (2012)
9. S. Takada, Curr. Opin. Struct. Biol. **22**, 130 (2012)
10. V. Tozzini, Quart. Rev. Biophys. **43**, 333 (2010)
11. W.G. Noid, J. Chem. Phys. **139**, 090901 (2013)
12. O. Engin, A. Villa, C. Peter, M. Sayar, Macromol. Theory Simul. **20**, 451 (2011)
13. F. Rodríguez-Ropero, N.F.A. van der Vegt, Phys. Chem. Chem. Phys. **17**, 8491 (2015)
14. F. Rodríguez-Ropero, T. Hajari, N.F.A. van der Vegt, J. Phys. Chem. B **119**, 15780 (2015)
15. C. Dalgicdir, C. Globisch, C. Peter, M. Sayar, PLoS Comput. Biol. **11**, e1004328 (2015)
16. S. Prajapati, V. Bhakuni, K.R. Babu, S.K. Jain, Europ. J. Biochem. **255**, 178 (1998)
17. M. Golczak, The FASEB J. (2001)
18. H.J. Dyson, P.E. Wright, Curr. Opin. Struct. Biol. **12**, 54 (2002)
19. D.E. Draper, D. Grilley, A.M. Soto, Annu. Rev. Biophys. Biomol. Struct. **34**, 221 (2005)
20. V. Receveur-Bréchot, J.-M. Bourhis, V.N. Uversky, B. Canard, S. Longhi, Proteins: Struct. Funct. Bioinf. **62**, 24 (2005)
21. C. Dalgicdir, M. Sayar, J. Phys. Chem. B **119**, 15164 (2015)
22. O. Engin. M. Sayar, J. Phys. Chem. B **116**, 2198 (2012)
23. C. Dalgicdir, O. Sensoy, C. Peter, M. Sayar, J. Chem. Phys. **139**, 234115 (2013)
24. I. Dikiy, D. Eliezer, Biochim. Biophys. Acta – Biomembr. **1818**, 1013 (2012)
25. H.A. Lashuel, C.R. Overk, A. Oueslati, E. Masliah, Nat. Rev. Neurosci. **14**, 38 (2012)
26. L. Jean, C.F. Lee, C. Lee, M. Shaw, D.J. Vaux, The FASEB J. **24**, 309 (2009)
27. L. Jean, C.F. Lee, D.J. Vaux, Biophys. J. **102**, 1154 (2012)
28. A. De Simone, C. Kitchen, A.H. Kwan, M. Sunde, C.M. Dobson, D. Frenkel, Proc. Natl. Acad. Sci. USA **109**, 6951 (2012)
29. W Li. Gala, Adv. Drug Delivery Rev. **56**, 967 (2004)
30. V. Tozzini, W. Rocchia, J.A. McCammon, J. Chem. Theory Comput. **2**, 667 (2006)
31. O. Bezkorovaynaya, A. Lukyanov, K. Kremer, C. Peter, J. Comput. Chem. **33**, 937 (2012)
32. A.V. Smith, C.K. Hall, Proteins **44**, 344 2001

33. T. Bereau, M. Deserno, J. Chem. Phys. **130**, 235106 (2009)
34. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, et al., Bioinformatics, btt055 (2013)
35. W.F. Van Gunsteren, H.J.C. Berendsen, Mol. Simul. **1**, 173 (1988)
36. M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, M. Parrinello, Comput. Phys. Commun. **180**, 1961 (2009)
37. G. Bussi, Mol. Phys. **112**, 379 (2014)
38. V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, D. Andrienko, J. Chem. Theory Comput. **5**, 3211 (2009)
39. W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. **14**, 33 (1996)
40. D. Frishman, P. Argos, Proteins: Struct. Funct. Bioinf. **23**, 566 (1995)
41. W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, O. Hahn, Acta Polym. **49**, 61 (1998)
42. V.A. Harmandaris, N.P. Adhikari, N.F.A. van der Vegt, K. Kremer, Macromolecules **39**, 6708 (2006)
43. Th. Soddemann, B. Dünweg, K. Kremer, Eur. Phys. J. E **6**, 409 (2001)
44. J.D. Weeks, D. Chandler, H.C. Andersen, J. Chem. Phys. **54**, 5237 (1971)
45. F. Ramezanghorbani, A transferable coarse-grained model for peptides that display an environment driven conformational transition. Master's thesis, Koc University, Koc Universitesi Sariyer Istanbul Turkey, 7 (2015)
46. B. Hess, D. van Der Spoel, E. Lindahl, Gromacs user manual version 4.5.7. (University of Groningen, Netherland, 2010)
47. W. Kabsch, C. Sander, Biopolymers **22**, 2577 (1983)