PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

A polarizable coarse-grained protein model for Dissipative Particle Dynamics†

Emanuel K. Peter,[a] Kirill Lykov,[a] and Igor V. Pivkin*[a,b]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

We present a new coarse-grained polarizable protein model for Dissipative Particle Dynamics (DPD) method. This method allows large timesteps in particle-based systems and speeds up sampling by many orders of magnitude. Our new model is based on the electrostatic polarization of the protein backbone and a detailed representation of the sidechains in combination with a polarizable water model. We define our model parameters using the experimental structures of two proteins, TrpZip2 and TrpCage. Backmapping and subsequent short replica-exchange molecular dynamics runs verify our approach and show convergence to the experimental structures on the atomistic level. We validate our model on five different proteins : GB1, the WW-domain, the B-domain of Protein A, the Peripheral binding subunit and Villin headpiece.

1 Introduction

Computer simulations on an atomistic scale are suitable for studying biomolecular systems and can be used complementary to experiments for the exploration of free energy landscapes [1,2] and binding free energies [3]. However, at present all-atom approaches are limited in their timescale and can be computationally expensive. Coarse-graining (CG) is a method capable to reach long time-scales by solving the time-scale problem through a reduction of the degrees of freedom in the system [4–7]. The CG-method has been used extensively in the popular MARTINI approach [8]. Systematic approaches for coarse-graining have been developed, including the iterative Boltzmann inversion approach [9], inverse Monte Carlo [10] and the force-matching method [11,12]. These methods rely on a construction of coarse-grained potentials from all-atomistic trajectories.

Coarse-graining of proteins ranges over different length and time-scales and several methods have been used for the description of these biomolecular systems, including Gō-models in protein folding [13], where the native structure is biased using restraint potentials, and other models [14–16]. Very recently, polarizability effects have been incorporated into coarse-grained forcefields, especially for the coarse-grained representation of water [17–21]. In addition, charge-charge interactions have been added for coarse-grained protein simulations to optimize the stability of secondary and tertiary structure formation [22,23]. Atomistic and coarse-grained force fields have also been connected in multiscale simulations [4,24–27].

Dissipative Particle Dynamics (DPD) is a method which employs soft repulsive potentials and which enables sampling of particle based systems with large timesteps, which cannot be used in conventional molecular dynamics [28–30]. In combination with suitable parameters, this method enhances sampling many orders of magnitude faster than other simulation techniques. In this paper, we present a novel polarizable coarse grained forcefield for the description of proteins on the basis of DPD method. We mention that protein conformations have been previously described using Dissipative Particle Dynamics in references [31,32], but not at the level of detail as presented here. Our model uses a polarizable description for water [33], as well as a polarizable backbone of the protein. The polarizability in the system is described by Drude oscillators, represented by 2 opposite charges connected by an harmonic potential, while the electrostatic interactions are calculated using a modified particle-particle particle mesh ewald (P3ME) formalism for DPD [34,35]. We define the parameters of the new forcefield using the experimental structures of TrpZip2 and TrpCage [36,37], and validate it on 5 different proteins, where the timescales of folding exceed the present capabilities of the all-atom simulation method [38]: GB1 [39], the WW-domain [40], the B-domain of Protein A [41], the Peripheral binding subunit [42] and Villin headpiece [43]. The new model further extends the scope of applications of Dissipative Particle Dynamics to simulations of biomolecular systems. We emphasize that our model is free from elastic network restraints needed for instance in the MARTINI coarse-grained model to keep the secondary structure stable, which we consider as a strong advantage of the presented here forcefield.

[a] Insitute of Computational Science, Faculty of Informatics, University of Lugano, Lugano, Switzerland. E-mail: igor.pivkin@usi.ch
[b] Swiss Institute of Bioinformatics, Lausanne, Switzerland.

Physical Chemistry Chemical Physics Accepted Manuscript

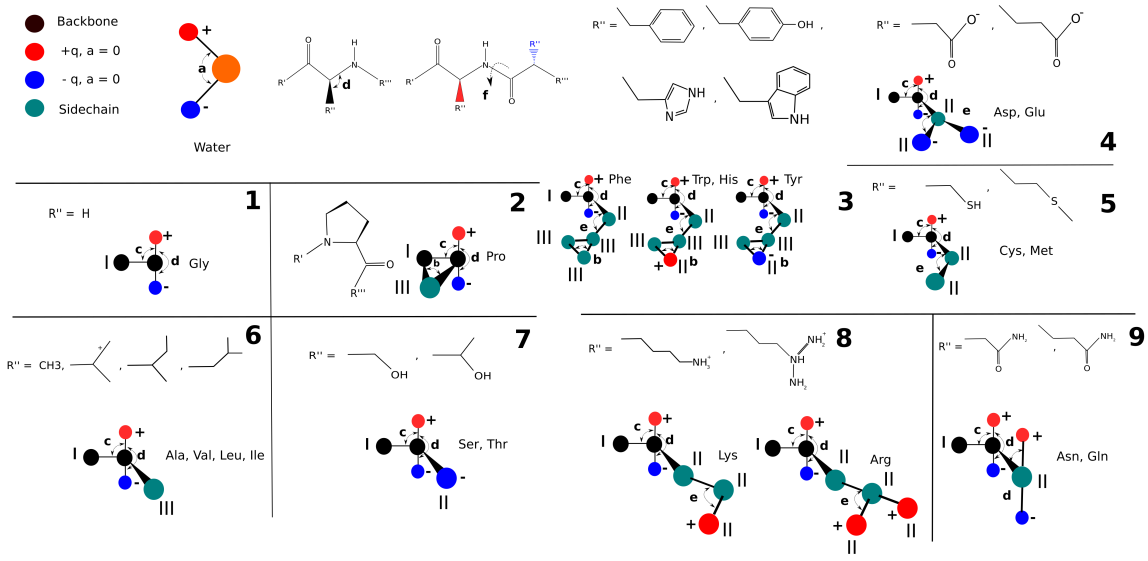

Fig. 1 Coarse-grained model of polarizable amino-acids and water. The peptide backbone is represented by 2 hydrophilic I type beads (black) and a dipole (negative charges - blue, positive charges - red), while the sidechain is attached perpendicular to the 2 backbone atoms. The polarity of sidechains is modeled by point charges, while the aliphatic groups are represented by hydrophobic beads (cyan). We use a polarizable water model consisting of one central bead and 2 Drude particles carrying opposite charges. Interaction parameters between different types of particles (Drude, type I, type II, type III, water) and equilibrium angles (a,b,c,d,e,f) employed in the model are listed in Table 2 and Table 3, respectively. Each of the different aminoacids has been distributed into different classes of 9 aminoacid group models with different representations of the sidechain - one bead, aromatic sidechains and hydrophilic sidechains (See section Methods). The numbers indicate the aminoacid group indexes.

2 Methods

2.1 Dissipative Particle Dynamics

For the simulations presented here, we use the Dissipative Particle Dynamics method [29,44]. In this particle-based method, the time-evolution of the system is described by Newton's law of motion,

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i \ , \tag{1}$$

and

$$\frac{d\mathbf{v}_i}{dt} = \mathbf{f}_i \ , \tag{2}$$

where $\mathbf{r}_i$ and $\mathbf{v}_i$ are position and velocity of particle i, respectively. The Velocity Verlet algorithm is used to propagate the system in time [29]. The force $\mathbf{f}_i$ which applies on each particle i consists of three additive parts,

$$\mathbf{f}_i = \sum_{j \neq i} (\mathbf{F}^C_{ij} + \mathbf{F}^D_{ij} + \mathbf{F}^R_{ij}) \ , \tag{3}$$

which are non-zero within a cutoff radius $R_c = 1$, which defines the length scale in the DPD system. The conservative force, $\mathbf{F}^C_{ij}$, is a soft repulsive force, which acts along the vector between particles i and j. An additional parameter a_{ij} defines the maximal repulsion between two particles

$$\mathbf{F}^C_{ij} = \begin{cases} a_{ij}(1 - r_{ij}/R_c)\hat{\mathbf{r}}_{ij}, & r_{ij} < R_c \\ 0, & r_{ij} \geq R_c \end{cases} \ , \tag{4}$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, $r_{ij} = |\mathbf{r}_{ij}|$ and $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/|\mathbf{r}_{ij}|$. The dissipative force, $\mathbf{F}^D_{ij}$, and the random force, $\mathbf{F}^R_{ij}$, are defined as

$$\mathbf{F}^D_{ij} = -\gamma w^D(r_{ij})(\hat{\mathbf{r}}_{ij}\mathbf{v}_{ij})\hat{\mathbf{r}}_{ij} \ , \tag{5}$$

and

$$\mathbf{F}^R_{ij} = \sigma w^R(r_{ij})\theta_{ij}\hat{\mathbf{r}}_{ij} \ , \tag{6}$$

where $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$, $w^D(r_{ij}) = \left[w^R(r_{ij})\right]^2$ and $w^R(r_{ij}) = 1 - r_{ij}/R_c$ are weight functions that depend on the inter-particle distance r_{ij} and vanish at $r_{ij} > R_c$, $\sigma^2 = 2\gamma k_B T$, and $\theta_{ij}(t)$ is a randomly fluctuating variable with a Gaussian distribution [29]. The dissipative and the random forces, which form the DPD thermostat, are linked through the fluctuation-dissipation theorem [45]. Additionally, we consider two harmonic bonded forces: the force between bonded particles, $\mathbf{F}^B_{ij}$, and the harmonic angular force, $\mathbf{F}^A_{ij}$, given by

$$\mathbf{F}^B_{ij} = dU^B/dr_{ij} \ , \text{with } U^B = \frac{1}{2}k_D(r_{ij} - d_0)^2 \ , \tag{7}$$

and

$$\mathbf{F}^A_{ij} = dU^A/d\theta_{ijk} \ , \text{with } U^A = \frac{1}{2}k_\theta(\theta_{ijk} - \theta_0)^2 \ , \tag{8}$$

where d_0 is the equilibrium bond length, θ_{ijk} stands for the instantaneous angle between 3 particles, while θ_0 is the equilibrium angle. The same potential is applied for dihedrals in the system, while each dihedral angle θ_{ijkl} is determined between 4 beads in the system. The description of electrostatics is implemented in the form of a non-bonded Coulombic force

$$\mathbf{F}^{Coul}_{ij} = dU^{Coul}/dr_{ij} \ . \tag{9}$$

Physical Chemistry Chemical Physics Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

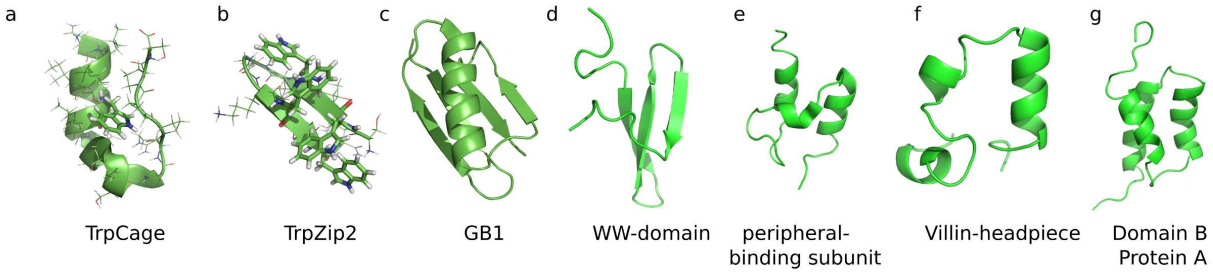

Fig. 2 Experimental structures of TrpCage (a), TrpZip2 (b), GB1 (c) , the WW-domain (d), the Peripheral binding subunit (e), Villin headpiece (f) and the B-domain of Protein A (g) simulated in the present work. TrpCage and TrpZip2 were used for the calibration, so that the coarse-grained model could reliably describe formation of α-helical and β-stranded peptides. GB1, the WW-domain, the peripheral binding subunit, Villin headpiece and the B-domain of Protein A were used for the validation of our new model. TrpCage (Sequence: NLYIQWLKDGGPSSGRPPPS[37]) PDB : 1L2Y. TrpZip2 (Sequence: SWTWENGKWTWKX[36]) PDB : 1LE1. GB1 (Sequence: MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDAATKTFTVTE) PDB : 2J52,[39]. (d) WW-domain (sequence : GATAVSEWTEYKTADGKTYYYNNRTLESTWEKPQELK), PDB : 1E0L[40]. (e) Peripheral binding subunit (sequence : VIAMPSVRKYAREKGVDIRLVQGTGKNGRVLKEDIDAFLAGGA), PDB : 2PDD[42]. (f) Villin headpiece (sequence : MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF), PDB : 1VII[43]. (g) B domain, Protein A (sequence : TADNKFNKEQQNAFYEILHLPNLNEEQRNGFIQSLKDDPSQSANLLAEAKKLNDAQAPKA), PDB : 1BDC[41]. For the NMR structures, we chose Model # 1 as reference structure.

2.2 Particle-Mesh Ewald electrostatics

The electrostatic interactions are described by the Particle Particle-Particle Mesh Ewald (P3ME) method in combination with a Slater type smearing out of charges[34,35,46]. The electrostatic potential between N point charges is described by

$$U(\mathbf{r}^N)^{Coul} = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\sum_i\sum_{j>i}\frac{\rho_i\rho_j}{r_{ij}}\,, \tag{10}$$

where ρ_i and ρ_j are the charge density values of a pair of DPD particles, r_{ij} is the distance between the charges, ε_0 and ε_r are the dielectric constants of vacuum and water ($\varepsilon_r = 78$) at room temperature. In this P3ME treatment, the long-range electrostatic energy given by equation 10 is decomposed into a real space and a reciprocal space[47,48]. This expression is then written as

$$U(\mathbf{r}^N)^{Coul} = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\left(\sum_i\sum_{j>i}\rho_i\rho_j\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} + \frac{2\pi}{V}\sum_{\mathbf{k}\neq 0}^{\infty}Q(k)S(\mathbf{k})S(-\mathbf{k}) - \frac{\alpha}{\sqrt{\pi}}\sum_i^N\rho_i^2\right)\,, \tag{11}$$

with

$$Q(k) = \frac{e^{-k^2/4\alpha^2}}{k^2}\,, \tag{12}$$

$$S(\mathbf{k}) = \sum_{i=1}^{N} q_i e^{i\mathbf{k}\mathbf{r}_i}\,, \tag{13}$$

$$\mathbf{k} = \frac{2\pi}{L}(m_x, m_y, m_z)\,, \tag{14}$$

$$k = |\mathbf{k}| \tag{15}$$

where α is the parameter that controls the contribution in real space, k is the magnitude of the reciprocal vector $\mathbf{k}$, while m_x, m_y, m_z are integer numbers. Since the soft conservative force would allow a full overlap of particles at $r_{ij} = 0$ and infinite ion-pair formation, we use the approach of Melchor *et al.*, and apply the charge distribution $\rho(|\mathbf{r}|)$ over a 3 dimensional mesh[35,46,49]

$$\rho(|\mathbf{r}|) = \frac{q}{\pi\lambda^3}e^{-2|\mathbf{r}|/\lambda}\,, \tag{16}$$

with $\lambda = 0.7R_c$ and $|\mathbf{r}|$ standing for the distance between the particle center and each grid point. Thus, the reduced interaction potential u_{ij} between two charge distributions in DPD, separated by a distance r_{ij} from center to center is given by[46]

$$\frac{4\pi u_{ij}(r_{ij})}{\Gamma} = \frac{\rho_i\rho_j}{r_{ij}}[1-(1+\frac{R_c}{\lambda}r_{ij}e^{-2R_c/\lambda r_{ij}}]\,, \tag{17}$$

with

$$\Gamma = \frac{e^2}{k_BT\varepsilon_0\varepsilon_r R_c}\,. \tag{18}$$

Table 1 Parameters of the Drude oscillator DPD water model, i.e. angular constant k_θ, equilibrium angle θ_0, bond force constant k_D, charge of Drude particles $|q|$ and equilibrium bond length d_0. Detailed description of the water model can be found in reference[33]

k_θ	θ_0	k_D	$\|q\|$	d_0
7.5 k_BT/rad^2	0^o	$1*10^5$ k_BT	0.75 e	0.2 R_c

2.3 Polarizable water for DPD simulations

In our simulations we employ the recently developed DPD model of polarizable water. For detailed description we refer to reference[33] (see also Figure 1). In this model, a central uncharged bead is connected to 2 oppositely charged Drude particles. A harmonic angle potential between the 2 Drude particles is applied, so that there is a match with the dielectric constant of water. Only

Physical Chemistry Chemical Physics Accepted Manuscript

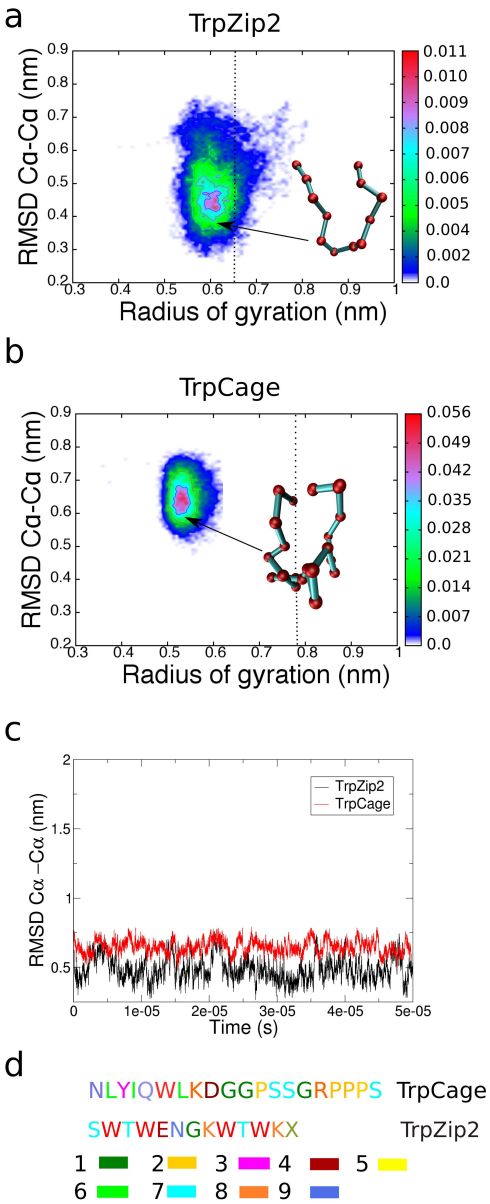

Fig. 3 (a) Probability plot of coarse-grained simulation of TrpZip2 as function of $RMSD_{C\alpha-C\alpha}$ to the native structure (given by forward mapped PDB: 1LE1) and the radius of gyration, Rg. (b) Probability plot of coarse grained simulation of TrpCage as function of $RMSD_{C\alpha-C\alpha}$ to the native structure (given by forward mapped PDB: 1L2Y) and the radius of gyration, Rg. The sidechains have been omitted in this representation and only the backbone atoms are shown for clarity. We do note that the RMSD to the backbone of the native structure is affected by an error of $\pm 0.4\ nm$ due to coarse-graining approach as described in the text.The experimental radii of gyration are indicated by a dashed line (c) $RMSD_{C\alpha-C\alpha}$ to the native structure as function of simulation time. (d) Color-assigned sequences of TrpCage and TrpZip2. The colors indicate the assignment of each aminoacid to the different groups in the coarse-graining approach (See Figure 1).

the central particle interacts with the rest of the system through a conservative force, while the Drude particles have a conservative force parameter of $a_{ij} = 0$. Through this modeling approach, the effective dielectric behavior of water is considered and polarization effects are included, which play a role in the dynamics of proteins. We use the numerical density of central water particles $N_p = 3$, a dissipative force coefficient $\gamma = 4$, $\sigma^2 = 2\gamma k_B T$ as random force coefficient, and a DPD temperature $k_B T = 1$. For the water model[33], we modified the point charges and increased the harmonic angular constant using a $\varepsilon_r = 78$ as electrostatic screening parameter (see Table 1). Each DPD particle representing water in our simulations corresponds to $N_m = 3$ water molecules, defining the unit of length as $R_c = 0.646\ nm$[29,50]. We chose a conservative force coefficient of $a_{ij} = 35$ for interactions between central water particles to match the dimensionless compressibility of water. The time scale is defined by matching the self diffusion constant of water[29] and equal to $\tau = 25 \times 10^{-9}\ s$[29,34]. Thus, one timestep of $dt = 0.001$ in the DPD framework used in our simulations, corresponds to $25 \times 10^{-12}\ s$.

We use these parameters for water model in all simulations and analyses presented in this work.

Table 2 Conservative force parameter a_{ij} used in the simulations. See Figure 1 for the model chosen in the coarse-grained approach.

type	Drude	type I	type II	type III	water
Drude	0.0	0.0	0.0	0.0	0.0
type I	0.0	5.0	5.0	5.0	15.0
type II	0.0	5.0	10.0	10.0	20.0
type III	0.0	5.0	10.0	5.0	30.0
water	0.0	15.0	20.0	30.0	35.0

Table 3 Equilibrium angles θ_0 for the internal coordinates of the sidechains used in the coarse-grained model. Dihedral angle potentials (type f) with an angle of 180^o are applied between the backbone dipole and the sidechain of the nearest neighbor along the polypeptide chain. See Figure 1 for the model chosen in the coarse-grained approach.

type	a	b	c	d	e	f
angle θ_0	0^o	60^o	71^o	90^o	120^o	180^o
$k_\theta(k_B T/(rad^2))$	1.0	1.0	1.0	1.0	1.0	1.0

2.4 Polarizable protein model

2.4.1 Protein backbone.

In the coarse-graining approach of the polypeptide, we take into account that peptide dynamics, secondary and tertiary structure formation are governed by the motion of the protein backbone, i.e. the formation and breaking of contacts along the chain[51]. The different accessibility of the peptide backbone to water and its mobility mainly determine the tendency of the protein to fold into defined tertiary structures[52,53]. In the parametrization procedure of the protein backbone it is our goal to display with coarse-grained model the formation and the unfolding of secondary structure elements, i.e. β-strands and α-helical elements. In contrast to different systematic approaches of coarse-graining[9,12,54], we aimed on a correct representation of probability distributions, i.e. free energy surfaces as close as possible

Physical Chemistry Chemical Physics Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

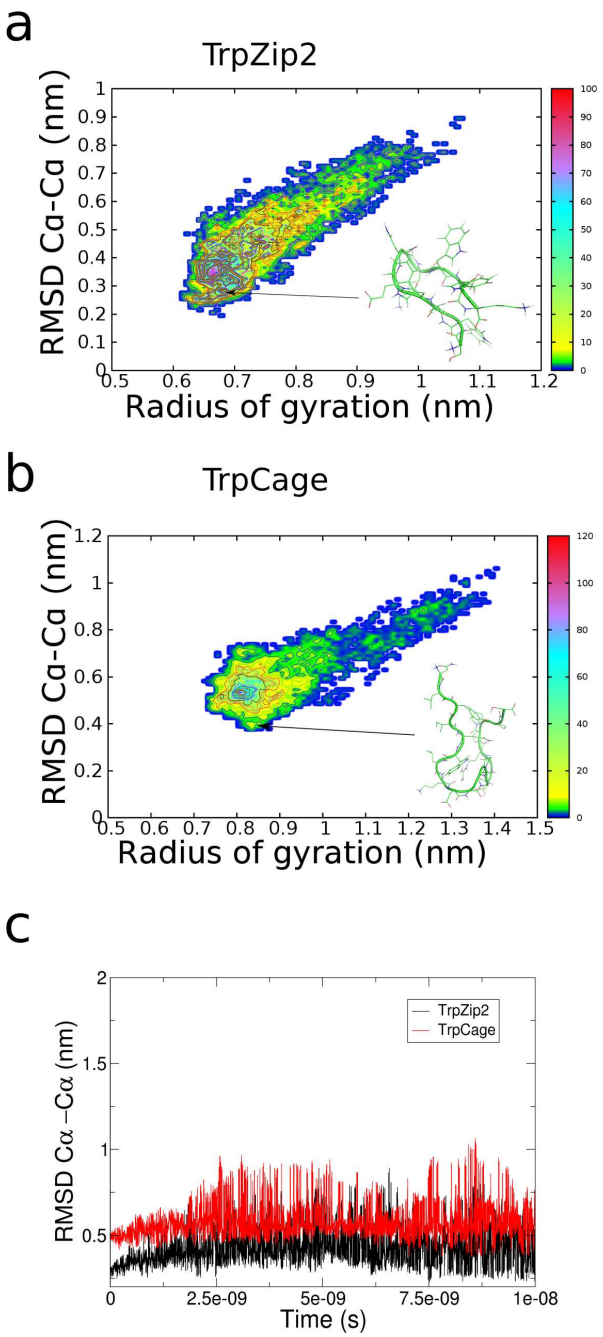

Fig. 4 Results from 10 ns all-atom replica exchange MD simulation with backmapped structures obtained from DPD coarse-grained simulations as starting structures. (a) Probability plot of all-atom simulation of TrpZip2 as function of $RMSD_{C\alpha-C\alpha}$ to the native structure (given by PDB: 1LE1) and the radius of gyration, Rg. (b) Probability plot of all-simulation of TrpCage as function of $RMSD_{C\alpha-C\alpha}$ to the native structure (given by PDB: 1L2Y) and the radius of gyration, Rg. (c) $RMSD_{C\alpha-C\alpha}$ to the native structure as function of simulation time.

to the native state (while the unfolded ensemble is partially neglected). In order to display the equilibrium between folding and unfolding of secondary structure elements, we model the attractive part of the peptide backbone as an electrostatic dipole, which is oriented perpendicular to the peptide sidechain, mimicking the dynamics of hydrogen bonding along the backbone and the solvation dynamics by polarizable water[22,23]. As in the case of polarizable water model, we use a Drude oscillator for the dipolar contribution of the protein backbone. Each of the Drude particles carries a charge of $\pm 0.25e$. It is bonded to the central particle by a harmonic bond with a force constant of $1*10^5\ k_BT$ and equilibrium length of $0.1\ R_c$. We applied an harmonic angle constant of $1.0\ k_BT/(rad^2)$ and equilibrium angle of 90 degrees for the angle between the 2 Drude particles and the central backbone particle. In addition, we applied harmonic angles with constant $1.0\ k_BT/(rad^2)$ for the internal coordinates of each sidechain as shown in Table 3. All beads representing the protein sidechains and the backbone are connected by a harmonic bond potential with an equilibrium length of $0.3\ R_c$ and force constant of $1*10^5\ k_BT$. Finally, the N-terminus is charged with $+1.0e$, while the C-terminus carries the opposite charge.

We used a knowledge-based approach for the parametrization of the protein model[14,55–58]. Specifically, we started by considering a general topology for each aminoacid similar to other approaches[12,55,56], while we additionally included a polarizable backbone and charges for specific aminoacids as shown in Figure 1. We then added internal coordinates in the form of angular and dihedral potentials as implemented in common all-atom forcefields [59]. Finally, we adjusted the non-bonded conservative force parameters for the interactions within the protein and the interactions with water, as well as the polarizability of the protein backbone through scaling the 2 opposite charges, until we reached a converged behavior in folding of an α-helical and a β-stranded peptide. Specifically, we used TrpZip2[36] and TrpCage[37] as model systems, to test the ability of our approach to fold these proteins into their native structure.

For the sidechains of each aminoacid, we mainly distinguished between hydrophilic and hydrophobic aminoacids. We then further made a differentiation between aromatic aminoacids and aliphatic aminoacids. In the following, we describe the approaches for modeling the different types of aminoacids.

2.4.2 One bead sidechain.

As aminoacids with one single bead representing the sidechain, we selected hydrophobic aminoacids alanine, valine, leucine and isoleucine, and modeled each of them with one single hydrophobic bead of type III (see Figure 1 and Table 2). For hydrophilic aminoacids with one single bead in the sidechain, we selected serin and threonin, and represented each of them with a bead of type II and one negative charge in the center of this bead. Glycine is the only aminoacid with no sidechain, while proline has an additional bond between the sidechain bead with a particle type III and the backbone bead, so that the cyclic ring of this aminoacid is resembled. The central Cα bead to which the sidechain bead is connected carries both Drude particles and is connected to a second bead in the backbone which represents the amide nitrogen.

Physical Chemistry Chemical Physics Accepted Manuscript

2.4.3 Two bead sidechain.

For the aminoacids cysteine and methionine, we modelled the sidechain with 2 beads of hydrophobic type II, which should resemble the sulfur with a larger vdW radius than in the case of serine and threonine with oxygen in its sidechain.

2.4.4 Aromatic aminoacids.

For the aminoacids with an aromatic sidechain, we described the aromatic ring as a ring consisting of 3 beads and one additional bead which represents $C\beta$ connected to $C\alpha$. In the case of phenylalanine, all particles in the ring are of bead type III, while we included an additional bead of type II and a positive charge of $q = +0.4e$ for tryptophan and histidine with one additional nitrogen in the ring structure (see Figure 1 and Table 2). In contrast to the latter 2 aminoacids, we modeled the hydroxy-group of tyrosine as an additional bead of type II carrying a negative charge of $q = -0.4e$.

2.4.5 Hydrophilic aminoacids.

For the hydrophilic aminoacids, we chose a more detailed differentiation between the different types of aminoacids, since polarization effects play an essential role in the system due to the electrostatics and the polarizable water model. For lysine, we modeled a sidechain with 3 beads, while the last bead carries a full positive charge of $1e$. For arginine, we modeled the guanidine sidechain by a branched sidechain consisting of 3 beads, with 2 beads representing the amines with a positive charge of $q = +0.4e$. For aspartic and glutamic acid, the sidechain is shorter in its length by 1 bead and the beads in both endings carry a negative charge of $q = -0.4e$. Finally, for the aminoacids with an amide group in their sidechain, i.e. asparagine and glutamine, we modelled the sidechain by 3 beads, with 2 beads of opposite charge of $\pm 0.25e$. These 2 charged beads have DPD conservative force parameter equal to 0.

2.4.6 Levels of hydrophobicity and hydrophilicity.

In our knowledge-based coarse-grained model, we made an estimate of the relative level of hydrophobicity and hydrophilicity based on the average effective vdW-radius of each amino-acid sidechain. Through this procedure, we made empirical estimates on each conservative force parameter in our model. A conservative force parameter value of 5 corresponds to a high level of attraction between beads of the same type, while 10 represents moderate repulsion. The overall reference value is defined by the interaction with water, where parameter of $a_{ij} = 15$ corresponds to hydrophilic interactions with water, while $a_{ij} = 30$ represents hydrophobic amino-acid properties. In this way, aminoacids with aromatic sidechains (hydrophobic) have a larger effective radius than hydrophilic residues (hydrophilic). For the aliphatic $C\beta$ atoms, we selected a non-bonded neutral type, to resemble partial hydrophobic properties of this part of the sidechain. A similar type of classification of particle types has been made in the prominent MARTINI approach[8].

2.5 Implementation, Program and System Preparation

All simulations have been done using the modified version of the LAMMPS simulation package[60]. For the analysis of the trajectories and the replica exchange MD simulations, we used modified parts of the GROMACS-4.5 simulation suite and in-house programs[61]. The interactions in the REMD simulations on an atomistic scale were described with the GROMOS 53a5 forcefield. The starting configurations and the protein topology were produced using in-house code. We employed the forward mapping procedure, implemented and distributed by Tielemann and coworkers[62], to map the experimental structures of proteins to our coarse-grained representation, which we used for the comparison of our simulations with the experiment (root mean square deviation from the native structure, $RMSD_{C\alpha-C\alpha}$). In this mapping procedure, the center of mass of the backbone and the sidechain of each individual amino-acid is mapped on the center of mass of the coarse-grained structure by geometrical assignments. The same process was applied for the reverse mapping. We mapped each experimental structure into the coarse-grained representation and then calculated the RMSD (in the coarse-grained trajectory) to the central atoms of the backbone of the forward mapped structure, which represent the $C\alpha$ carbon atoms. Then we mapped this quantity forward to real parameters. In the same way, we transformed the radius of gyration. For the calculation of the free energies in our protein folding simulations, we used the relation $F = -k_B T \ln P$.

2.6 Estimation of absolute error

The error in the comparison of our simulation data from the coarse-grained trajectory with the forward mapped structure (on the coarse-grained scale) from the experiment is associated with following factors:

- The centers of mass of each amino-acid in our coarse-grained model deviates slightly from the center of mass of the corresponding amino-acid on the all-atom level. This error is minimal for residues with a small sidechain, such as Glycine, but comparably large for Lysine or for amino-acids with larger sidechains, such as Trptophan. We tried to minimize this error through our parameterization and model development. For each aminoacid the error is on the order of ≈ 0.2 *nm* for the backbone atoms ($C\alpha$), which were used for the analysis of our trajectories. For the sidechains in the system the deviations are larger on the order of ≈ 0.4 *nm*.

- In our modeling approach, the bond distances between the backbone beads and along the sidechains are equal to $0.3\ R_c$, while this is not the case for the inter-particle distance of $C\alpha$ for the experimental structure which we used as reference structure, where the distances are of varying length. We estimate the error connected to the variance of the reference native structure in the case of the RMSD calculation with an associated value of ≈ 0.2 *nm*.

Finally we estimate a total error in the calculation of the $RMSD_{C\alpha-C\alpha}$ and Rg to be about 0.4 *nm*, where we only take the backbone of each protein into account. Thus, each quantity X_c is

Physical Chemistry Chemical Physics Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

mapped backwards to real units depending on the length scale of the system according to

$$X(r) = X_c * R_c \ \pm 0.4 \ [nm] \ . \tag{19}$$

After the model parameters were defined, we additionally simulated the experimental structures of TrpCage and TrpZip2 mapped forward from folded state to our coarse grained description. We found that these 2 proteins remained at $RMSD_{C\alpha-C\alpha}$ $0.5-0.7$ nm from the experimental structure, in agreement with the error analysis presented above.

2.7 Simulation details

For the simulation of TrpZip2, we centered the extended polypeptide chain (81 beads) into a box with dimensions $10\times10\times10$ DPD units filled with 3064 polarizable coarse-grained waters. We simulated the protein for a total of 2000000 steps, with a timestep of $dt = 0.001$, which corresponds to a total simulation time of 50×10^{-6} s. For the simulation of TrpCage, we centered the extended coarse grained polypeptide chain consisting of 119 beads into a box with dimensions $12.5\times12.5\times12.5$ DPD units and filled the box with 5863 polarizable CG waters. We simulated the protein for a total of 2000000 steps, with a timestep of $dt = 0.001$ which corresponds to a simulation time of 50×10^{-6} s. For GB1, we modeled the peptide as an extended chain with 309 beads in a box with dimensions $20\times20\times20$ DPD units filled with 23279 coarse-grained waters. After an equilibration for 20000 steps, we centered the peptide into a smaller box with dimensions $14\times14\times14$ DPD units and filled the box with 8433 CG waters. We then simulated the system for 5340000 steps corresponding to 0.1335×10^{-3} s. For the simulation of the WW-domain, Villin-headpiece and the Peripheral binding subunit, we centered each extended coarse grained peptide (WW-domain - 236 beads, Villin-headpiece - 185 beads, Peripheral binding subunit - 215 beads) into a box with dimensions $15\times15\times15$ DPD units and added 10612 coarse grained waters. We simulated each system for 4800000 steps. For the simulations of the B-domain of Protein A, we centered the extended polypeptide chain (368 beads) in a box with dimensions $15\times15\times15$ and filled it with 10606 coarse-grained waters. In all of our simulations, we used an electrostatic cutoff of 1.5 R_c and a cutoff for the other DPD forces of 1 R_c. We integrated the equations of motion with a timestep of 0.001 τ. For the all atom simulations of the backmapped structures of TrpCage and TrpZip2, we used the GROMOS 53a5 forcefield. AA-TrpZip2 was centered in a box with dimensions $3.34\times3.34\times3.34$ nm^3 and the box was filled with 1158 SPC/E waters. AA-TrpCage was centered in a box with dimensions $3.39\times3.39\times3.39$ nm^3 and the box was filled with 1184 SPC/E waters. We applied PME electrostatics to the system, while we used a shift function for the description of the Lennard-Jones interactions. Both potentials were truncated at a cutoff of 1.0 nm. We used no constraints and applied a timestep $dt = 1.0\times10^{-15}$ s to integrate the equations of motion. We simulated the system for a total of 10×10^{-9} s using the replica exchange MD (REMD) methodology[63]. We used 12 replicas in the temperature range from 300 to 322 K. The exchange statistics of this comparably short run was sufficient.

3 Results

3.1 Calibration simulations used to define model parameters

In the following, we describe the results of simulations where we applied our coarse-grained model on folding of TrpZip2 and TrpCage. Both proteins have been extensively studied in detail, both experimentally and computationally, and serve as model systems for α-helical and β-strand folding. We chose these systems for the calibration, since both proteins cover the different aminoacid groups in our coarse-grained model, except group 5 containing cysteine and methionine (See Figures 1 and 3d). We start with the simulations of TrpZip2 using the model described above. In Figure 3a, we show the probability distribution of TrpZip2 as function of $RMSD_{C\alpha-C\alpha}$ and the radius of gyration (Rg). It is important to note that interpretation of the results should be different from typical molecular dynamics analysis. Due to the backmapping procedure (see section 2.6), the reported quantities have an associated error of ±0.4 nm, which we also see in the comparison with the experimental radii of gyration. This error is connected to the overall coarse-graining approach which we use. We observe that TrpZip2 resides with a higher probability between $RMSD_{C\alpha-C\alpha}$ of 0.32 nm and 0.7 nm and Rg ranging from 0.55 to 0.75 nm (experimental value 0.66 nm). The maxima in the probability is located at $RMSD_{C\alpha-C\alpha}$ ranging from 0.34 to 0.55 nm and Rg between 0.44 and 0.68 nm, where we find near-native configurations in which the loop is already formed, as well as collapsed configurations in which contacts exist between Trp-residues 2 and 11. In the native state, we find that the hairpin is aligned in a β-stranded form, while the Trp-residues 4 and 9 form a hydrophobic contact, which is in agreement with experimental data[36]. As a folding mechanism, we observe a zipper-like closure of the hairpin, which agrees well with a study by Swope and coworkers[64]. We observe that TrpZip2 folds along a 2-state like mechanism, with no favored intermediate near-native state. The lowest free energy is at $RMSD_{C\alpha-C\alpha}$ of 0.4 nm with a value of approximately -3.9 k_BT. We note that the free energy value obtained with our coarse-grained model is in good agreement with results by Yang *et al.* ($-4k_BT$ for the native minimum)[65]. Finally, we emphasize that our coarse-grained model is not meant to be used as an alternative for all atomistic methods for *detailed* investigation the protein folding pathway, nevertheless it decays into the native state of this protein.

We continue with simulations of an α-helical peptide TrpCage[37]. In Figure 3b, we show the probability distribution as function of $RMSD_{C\alpha-C\alpha}$ and the radius of gyration. We observe that the main conformer is located within $RMSD_{C\alpha-C\alpha}$ of 0.55 nm and 0.75 nm and Rg between 0.5 and 0.65 nm (experimental value 0.79 nm), which corresponds to the folded state of TrpCage. In this state, which is very close to the native state of TrpCage, the α-helical element between residues Asn1 and Asp9 is formed and Trp6 points into the hydrophobic cleft of the protein. We find that the potential of mean force is approximately -4.5 k_BT for the folded state of the protein. This value is in good agreement with results from all-atom simulations[66,67]. Finally, we mention that we observe a diffusion collision folding mechanism of TrpCage

Physical Chemistry Chemical Physics Accepted Manuscript

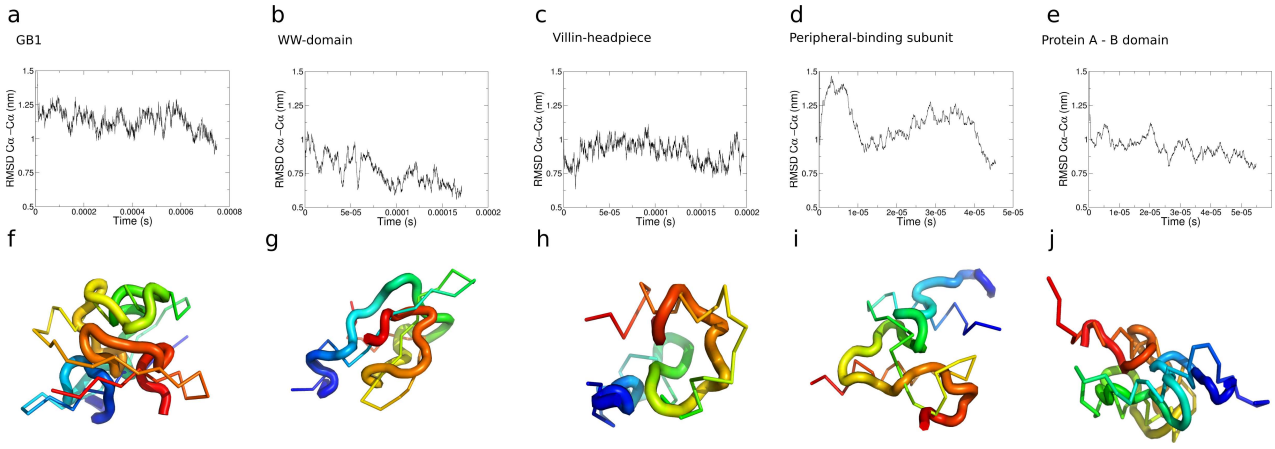

Fig. 5 Results from the folding simulations on GB1, the WW-domain, Villin-headpiece, the peripheral binding subunit and the B-domain of ProteinA. (a-e) $RMSD_{C\alpha-C\alpha}$ as function of simulation time. (f-j) Comparison of least square fitted backward-mapped structures from the simulations (cartoon) to the experimental structures (ribbon) of all 5 proteins.

with formation of the α-helical element prior to the final collapse into the native structure. Similar pathways have been observed in experiments and simulations as major folding pathways of TrpCage [68–70]. We note though that the timescale and dynamics are faster in our coarse-grained approach than in all-atom simulations. Nevertheless, the model describes self-assembly of proteins in solution in good agrement with the experiment, without any restraints on the system. In the $RMSD_{C\alpha-C\alpha}$ as function of simulation time of both proteins, we see that the configurations are stable and the initial collapse occurs within the first steps of the simulation (approximately 50000 steps) (see Figure 3c). This indicates that the unfolded region of both proteins is not sampled accurately, but rather the native state of both peptides.

3.2 Simulation of backmapped structures in all-atom MD simulations

We mapped the lowest RMSD structures of TrpCage and TrpZip2 back to the atomistic scale in order to verify our coarse-graining approach and to compare our measures (absolute errors) of the $RMSD_{C\alpha-C\alpha}$ and Rg with the explicit scale. For the subsequent simulations, we only chose a small simulation time of 10×10^{-9} s and a low number of 12 replicas, so that we could ensure that we sample mainly the structure of the probability maximum in the coarse-grained simulation. For the simulations of TrpZip2 and TrpCage, we extracted the lowest RMSD structure from the CG-trajectory and mapped this structure to the all-atom (AA) description as implemented by Wassenaar *et al.* [62].

We start with the description of TrpZip2, where we find that the starting structure (backmapped from the coarse grained trajectory) already has a RMSD to the backbone of the native structure of 0.29 nm. In the REMD simulation, we find that the structure mainly resides at $RMSD_{C\alpha-C\alpha}$ of 0.3 nm, while also slightly unfolded conformations in $RMSD_{C\alpha-C\alpha}$ exist in the range between 0.5 nm and 0.8 nm (see Figure 4a). The lowest RMSD structures correspond to $RMSD_{C\alpha-C\alpha}$ of 0.22 nm and Rg of 0.65 nm. In these lowest RMSD conformations, we find that the hydrophobic core itself is not as compact as in the native structure of TrpZip2 [36], and Trp2 is exposed to the solvent.

We continue with the description of TrpCage, where we see that the starting structure is only near native at $RMSD_{C\alpha-C\alpha}$ of 0.5 nm. We also observe that the TrpCage remains mainly at this RMSD value, while also lower RMSD values occur throughout this short REMD run (see Figure 4b). We notice that in these lowest RMSD structures the 3-10 helix and the polyproline helix are well described, while the α-helical part is loosely packed. We mention that Trp6 is correctly aligned and prone for the formation of the hydrophobic core of this protein. The same counts for residues Asp9 and Arg16, which are on the same perimeter and are able to form the salt-bridge.

When we compare the measures in the all-atom and the coarse-grained simulations, we directly see that our estimate of the errors (see section 2.6) are suitable for our model. We observe that the deviation in Rg is on the order of $\approx$ 0.4 nm while the deviation in the RMSD is $\approx$ $0.2-0.3$ nm. Additionally, we find that the RMSD curves as function of simulation time do not have a drift towards unfolded conformers, indicating that the coarse-grained structures represent a stable free energy minimum (see Figure 4c).

We conclude that our calibration simulations using the new DPD based parameter set is suitable for the description of both proteins, TrpZip2 and TrpCage.

3.3 Validation on 5 different proteins

In the following, we present the results of the validation of our new methodology on folding of 5 different proteins : GB1, the WW-domain, domain B from Protein A, the peripherial binding subunit and Villin Headpiece. We start with describing our simulations on folding of GB1, a mixed α-β peptide. In this simulation, we observe that the protein collapses to a $RMSD_{C\alpha-C\alpha}$ of 0.9 nm, into a state which has the same tertiary structure as the native state of GB1, while the protein is in a more compact state than the native structure (see Figure 5a,f). Here we observe that 3 of the 4 β-strands of the native structure are present with the α-helix between residues Asp22 and Gly37. The outer β-strand

Physical Chemistry Chemical Physics Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

between residues Glu42 and Asp48 is not preserved as well as the other motifs of GB1.

In the simulation of the WW-domain, we observe 3 collapses into a near-native structure followed by a slow decrease in the $RMSD_{C\alpha-C\alpha}$ at $1.7*10^{-4}$ s to $RMSD_{C\alpha-C\alpha}$ of 0.65 nm. In the tertiary structure of this protein connected to the lowest $RMSD_{C\alpha-C\alpha}$, we observe 3 turns in the structure and 3 longitudinal extensions which represent the 3 β-stranded parts of this domain well. As in the case of GB1, we observe that the near native structure has the same tertiary structure as the native state, while the protein is not as compact as the native state (see Figure 5b,g). In the case of Villin-headpiece, we observe that the protein constantly resides between $RMSD_{C\alpha-C\alpha}$ of 0.65 nm and 1 nm. In the analysis of the near native state of this protein obtained from our simulation, we observe that the tertiary structure is identical with the structure of the native state within the associated error (See Figure 5c,h). As in the case of the WW-domain, the 3 turns between the alpha-helical regions of this protein are well described in the structure. Slight deviations from the (forwarded) native structure occur in the α-helical regions themselves, leading to a more compact state. Again, we emphasize that during the forward mapping procedure and due to our coarse-grained model a certain error is associated with the structural comparison between both structures (see section 2.6). For the peripheral binding subunit, we observe a final collapse to a near native structure with $RMSD_{C\alpha-C\alpha}$ of 0.77 nm at a simulation time of $4.5*10^{-5}$ s. For this lowest RMSD structure, we observe a N-terminal part with 2 turns and 2 α-helical regions are well described, while the C-terminal region (residues 62 to residue 76) is extended in our structure and only partially forms an α-helix in the range between residues leucine 69 and Phe 76. In that case the structure is less compact than the native state, but describes the tertiary structure and secondary structure well (see Figures 5d,i).

In the simulation of the B-domain of Protein A, we observe that the structure collapses in the native state with $RMSD_{C\alpha-C\alpha} = 0.77$ nm at a simulation time of $5.5*10^{-5}$ s . In this state, the protein has all 3 turns formed with the 3 helices in the regions (Gln11-His19, Asn24-Asp37 and Gln41-Gln56). In that case the 3 helices are packed more strongly, while the 3 helices in the native structure are more loosely packed (See Figures 5e,j).

In summary, we have validated our model on 5 different proteins. In the cases of the WW-domain and the peripheral binding subunit, we obtained good agreement within the error associated with our forward mapping procedure. For the B-domain of protein A, we observed a higher packing compared to the native structure, while all 3 helices and the turns are preserved. In the case of GB1, we observe correct arrangement of secondary structure elements and close alignment in the tertiary structure, while one outer β-sheet is not conserved.

4 Conclusions

In this work, we presented a novel coarse-grained polarizable protein model for Dissipative Particle Dynamics simulation technique. Our model is based on polarizable backbone dipoles which enable polypeptides to form their native structure. We define the model parameters based on folding of the 2 peptides TrpCage and TrpZip2, with the condition that the main conformers of the coarse-grained model agree with the experimental structures[36,37]. We validate the model by simulating folding of 5 additional single domain proteins, i.e. GB1, the WW-domain, the peripheral binding subunit, Villin headpiece and the B-domain of Protein A. We believe that the presented model is suitable for the coarse-grained description of proteins and has the potential to improve sampling of native states in coarse-grained protein simulations. Currently, we are extending the model to simulations of peptide aggregation, where self-assembly processes play an important role.

5 Acknowledgements

This work was supported by Swiss National Science Foundation grant 200021_138231 and grant from Swiss Platform for Advanced Scientific Computing. Simulations were carried out at Swiss National Supercomputer Center (CSCS) under the project u4.

References

1 R. Zhou, B. J. Berne and R. Germain, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 14931–14936.
2 J.-E. Shea and C. L. Brooks III, *Annu. Rev. Phys. Chem.*, 2001, **52**, 499–535.
3 Y. Deng and B. Roux, *J. Phys. Chem. B*, 2009, **113**, 2234–2246.
4 M. Neri, C. Anselmi, M. Cascella, A. Maritan and P. Carloni, *Phys. Rev. Lett.*, 2005, **95**, 218102.
5 H. J. Risselada, S. J. Marrink and M. Mueller, *Phys. Rev. Lett.*, 2011, **106**, 148102.
6 R. Potestio, S. Fritsch, P. Espanol, R. Delgado-Buscalioni, K. Kremer, R. Everaers and D. Donadio, *Phys. Rev. Lett.*, 2013, **110**, 103801.
7 R. Potestio, P. Espanol, R. Delago-Buscalioni, R. Everaers, K. Kremer and D. Donadio, *Phys. Rev. Lett.*, 2013, **111**, 060601.
8 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.
9 H. Wang, C. Junghans and K. Kremer, *Eur. Phys. J. E. Soft Matter*, 2009, **28**, 221–229.
10 T. Murtola, E. Falck, M. Karttunen and I. Vattulainen, *J. Chem. Phys.*, 2007, **126**, 075101.
11 T. Yan, C. J. Burnham, M. G. D. Popolo and G. A. Voth, *J. Phys. Chem. B*, 2004, **108**, 11877–11881.
12 R. D. Hills, L. Lu and G. A. Voth, *PLOS Comput. Biol.*, 2010, **6**, e1000827.
13 S. Takada, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 11698–11700.
14 V. Tozzini, *Curr. Opin. Struct. Biol.*, 2005, **15**, 144–150.
15 C. Wu and J.-E. Shea, *Curr. Opin. Struct. Biol.*, 2011, **21**, 209–220.
16 A. Morriss-Andrews and J.-E. Shea, *J. Phys. Chem. Lett.*, 2014, **5**, 1899–1908.
17 S. O. Yesylevskyy, L. V. Schaefer, D. Sengupta and S. J. Marrink, *PLoS Comput. Biol.*, 2010, **6**, e1000810.

Physical Chemistry Chemical Physics Accepted Manuscript

18 S. J. Marrink and D. P. Tieleman, *Chem. Soc. Rev.*, 2013, **42**, 6801–6822.

19 Z. Wu, Q. Cui and A. Yethiraj, *J. Phys. Chem. B*, 2010, **114**, 10524–10529.

20 S. Riniker and W. F. van Gunsteren, *J. Chem. Phys.*, 2011, **134**, 084110.

21 M. Masella, D. Borgis and P. Cuniasse, *J. Comput. Chem.*, 2008, **29**, 1707–1724.

22 T. Terakawa and S. Takada, *J. Chem. Theory Comput.*, 2014, **10**, 711–721.

23 E. Spiga, D. Alemani, M. T. Degiacomi, M. Cascella and M. D. Peraro, *J. Chem. Theory Comput.*, **9**, 3515–3526.

24 A. J. Rzepiela, M. Louhivouri, C. Peter and S. J. Marrink, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10437–10448.

25 Q. Shi, S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2006, **110**, 15045–15048.

26 J. Michel, M. Orsi and J. W. Essex, *J. Phys. Chem. B*, 2008, **112**, 657–660.

27 S. Riniker, A. P. Eichenberger and W. F. van Gunsteren, *Eur. Biophys. J.*, 2012, **41**, 647–661.

28 P. J. Hoogerbrugge and J. M. V. A. Koelman, *Europhys. Lett.*, 1992, **19**, 155–160.

29 R. D. Groot and P. B. Warren, *J. Chem. Phys.*, 1997, **107**, 51335.

30 E. G. Flekkoy and P. V. Coveney, *Phys. Rev. Lett.*, 1999, **83**, 1775–1778.

31 A. Vishnyakov, D. S. Talaga and A. V. Neimark, *J. Phys. Chem. Lett.*, 2012, **3**, 3081–3087.

32 G. Guigas, D. Morozova and M. Weiss, *Adv. Protein Chem. Struct. Biol.*, 2011, **85**, 143–182.

33 E. K. Peter and I. V. Pivkin, *J. Chem. Phys.*, 2014, **141**, 164506.

34 R. D. Groot, *J. Chem. Phys.*, 2003, **118**, 11265.

35 C. Sagui and T. Darden, *J. Chem. Phys.*, 2001, **114**, 6578.

36 A. G. Cochran, N. J. Skelton and M. A. Starovasnik, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **98**, 5578–5583.

37 J. W. Neidigh and R. M. Fesinmeyer, *Nat.Struct.Biol.*, 2002, **9**, 425–430.

38 J. Kubelka, J. Hofrichter and W. A. Eaton, *Curr. Opin. Struct. Biol.*, 2004, **14**, 76–88.

39 D. J. Wilton, R. B. Tunnicliffe, Y. O. Kamatari, K. Akasaka and M. P. Williamson, *Proteins*, 2008, **71**, 1432.

40 M. J. Macias, V. Gervais, C. Civera and H. Oschkinat, *Nat. Struct. Biol.*, 2000, **7**, 375.

41 H. Gouda, H. Torigoe, A. Saito, Y. Arata and I. Shimada, *Biochemistry*, 1992, **31**, 9665–9672.

42 Y. N. Kalia, S. M. Brocklehurst, D. S. Hipps, E. Appella, K. Sakaguchi and R. N. Perham, *J. Mol. Biol.*, 1993, **230**, 323–341.

43 C. J. McKnight, P. T. Masudaira and P. S. Kim, *Nat. Struct. Biol.*, 1997, **4**, 180–184.

44 I. V. Pivkin, B. Caswell and G. E. Karniadakis, *Reviews in Computational Chemistry*, John Wiley & Sons, Inc., 2010, pp. 85–110.

45 P. Espanol and P. Warren, *Europhys Lett*, 1995, **30**, 191–196.

46 M. Gonzalez-Melchor, E. Mayoral, M. E. Velazquez and J. Alejandre, *J. Chem. Phys.*, 2006, **125**, 224107.

47 P. P. Ewald, *Ann. Phys.*, 1921, **64**, 253.

48 D. Frenkel and B. Smit, *Academic, New York*, 1996.

49 H. Saint-Martin, J. Hernandez-Cobos, M. I. Bernal-Uruchurtu, I. Ortega-Blake and H. J. C. Berendsen, *J. Chem. Phys.*, 2000, **113**, 10899.

50 E. E. Keaveny, I. V. Pivkin, M. Maxey and G. E. Karniadakis, *J Chem Phys*, 2005, **123**, year.

51 K. W. Plaxco, K. T. Simons and D. Baker, *J. Mol. Biol.*, 1998, **277**, 985–994.

52 G. D. Rose, P. J. Fleming, J. R. Banavar and A. Maritan, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 16623–16633.

53 D. W. Bolen and G. D. Rose, *Annu. Rev. Biochem.*, 2008, **77**, 339–362.

54 S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469–2473.

55 M. Jamroz, M. Orozco, A. Kolinski and S. Kmiecik, *J. Chem. Theory Comput.*, 2013, **9**, 119–125.

56 M. Friedel and J.-E. Shea, *J. Chem. Phys.*, 2004, **120**, 5809.

57 M. Baaden and S. J. Marrink, *Curr. Opin. Struct. Biol.*, 2013, **23**, 878–886.

58 J. W. Mullinax and W. G. Noid, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 19867–19872.

59 P. A. Kollman, *Acc. Chem. Res.*, 1996, **29**, 461–469.

60 S. Plimton, *J. Comp. Phys.*, 1995, **117**, 1–19.

61 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

62 T. A. Wassenaar, K. Pluhackova, R. A. Boeckmann, S. J. Marrink and D. P. Tielemann, *J. Chem. Theory Comput.*, 2014, **10**, 676–690.

63 K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.*, 1996, **65**, 1604–1608.

64 J. W. Pitera, I. Haque and W. C. Swope, *J. Chem. Phys.*, 2006, **124**, 141102.

65 W. Y. Yang, J. W. Pitera, W. C. Swope and M. Gruebele, *J. Mol. Biol.*, 2004, **336**, 241–251.

66 R. Zhou, *Proc. Natl. Acad. Sci. U.S.A*, 2003, **100**, 13280–13285.

67 R. Day, D. Paschek and A. E. Garcia, *Proteins*, 2010, **78**, 1889–1899.

68 J. Juraszek and P. G. Bolhuis, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 15859–15864.

69 J. Juraszek and P. G. Bolhuis, *Biophys. J.*, 2008, **95**, 4246–4257.

70 H. Meuzelaar, K. A. Marino, A. Huerta-Viga, M. R. Panman, L. E. J. Smeenk, A. J. Kettelarij, P. T. J. H. van Maarseveen, P. G. Bolhuis and S. Woutersen, *J. Phys. Chem. B*, 2013, **117**, 11490–11501.

Physical Chemistry Chemical Physics Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]