A Discontinuous Potential Model for Protein–Protein Interactions

Qing Shao and Carol K. Hall

Abstract Protein–protein interactions play an important role in many biologic and industrial processes. In this work, we develop a two-bead-per-residue model that enables us to account for protein–protein interactions in a multi-protein system using discontinuous molecular dynamics simulations. This model deploys discontinuous potentials to describe the non-bonded interactions and virtual bonds to keep proteins in their native state. The geometric and energetic parameters are derived from the potentials of mean force between sidechain–sidechain, sidechain–backbone, and backbone–backbone pairs. The energetic parameters are scaled with the aim of matching the second virial coefficient of lysozyme reported in experiment. We also investigate the performance of several bond-building strategies.

Keywords Coarse-grained model · Protein–protein interactions · Discontinuous molecular dynamics · Square-well potential · Osmotic second virial coefficient

1 Introduction

Here, we report the development of a two-bead-per-residue protein model that can be used with discontinuous molecular dynamics (DMD) simulations to investigate protein–protein interactions in a multi-protein system. We expect that this new model will allow us to simulate multi-protein systems on longer timescales than what has heretofore been achievable, helping us to deepen our understanding of processes such as protein crystallization [1], protein recognition [2], and protein purification [3].

Protein models can be classified broadly into two types: all-atom if they describe every atom in the protein explicitly and coarse-grained if they group several atoms into one interactive site. All-atom force fields such as CHARMM [4], AMBER [5],

Q. Shao · C.K. Hall (✉)
Department of Chemical and Biomolecular Engineering,
North Carolina State University, Raleigh 27695, USA
e-mail: hall@ncsu.edu

R.Q. Snurr et al. (eds.), *Foundations of Molecular Modeling and Simulation*,
Molecular Modeling and Simulation, DOI 10.1007/978-981-10-1128-3_1

GROMOS [6, 7], and OPLS/AA [8] are very good at describing the behavior of a single protein and how it interacts with other molecules in explicit solvent. However, atomistic simulations are usually limited to one or several small proteins and timescales up to several hundred nanoseconds, effectively precluding the investigation of many interesting multi-protein problems. Coarse-grained models enable us to simulate larger systems for longer timescales using less computational resources. There are two major choices to be made in the development of coarse-grained models: (1) how to coarse-grain the protein geometry and (2) how to obtain the geometric and energetic parameters (see recent review papers [9–15] that summarize the various coarse-graining methods, coarse-grained protein models, and their applications). Coarse-grained protein models can be categorized based on how the atoms are grouped together to form the coarse-grained bead (four-bead-per-residue [16], two-bead-per-residue [17], one-bead-per-residue [18, 19], and ultra-coarse-grained [20]) and how the force field parameters are determined (e.g., Go-type [21], knowledge-based [22, 23], and physics-based [24]).

Coarse-grained models are usually more problem-specific than all-atom models because of transferability issues. Coarse-grained protein models are often developed with the goal of examining particular properties. Most of the current coarse-grained protein models focus on the folding/unfolding problem. It thus remains to be seen how well protein models developed based on such properties do in describing behavior that is a consequence of protein–protein interactions. For example, Stark et al. [25] found that the popular MARTINI force field predicts a second virial coefficient of lysozyme that differs considerably from the experimental value. This inconsistency between simulation and experiment points out the importance of developing protein models that are designed to apply to problems where protein–protein interactions play a major role.

It is also important that the method used to simulate multi-protein systems be fast. Most of the models used in simulating multi-protein systems are based on continuous intermolecular potentials like the Lennard–Jones potential. Simulations based on continuous potentials proceed by solving Newton equations at a uniformly spaced time intervals. They have an algorithm complexity of $O(N\log N)$, where N is the number of particles in the system. The big-O notation describes how the performance or complexity (referring to the number of operations) required to run an algorithm depends on the number of particles in the system. Therefore, the required computational time for continuous MD simulations increases dramatically with the number of beads in the system, limiting their application to relatively small systems.

Discontinuous molecular dynamics (DMD) simulations can be used to investigate large systems efficiently with moderate computational resources. DMD simulations were designed to be applicable to systems that interact via discontinuous potentials (square-well/square-shoulder and hard-sphere). They proceed by analytically calculating the next collision time. Several papers [26–28] describe the details of DMD simulations. The algorithm complexity of DMD simulations is $O(N\log N)$. (One paper by Paul [29] even claims a realization of the DMD method

with an algorithm complexity of $O(1)$.) The enhanced algorithm complexity of DMD simulations compared to continuous MD simulations make them suitable for the investigation of long-time processes like spontaneous formation of amyloid structure, which are still challenging for continuous MD simulations.

This work reports our effort to develop a coarse-grained protein model that can be used to study protein–protein interactions in multi-protein systems via DMD simulations. We deploy a two-bead-per-residue protein model: one bead for the backbone and the other for the sidechain. The parameters of our protein model are obtained by coarse-graining atomistic simulation results for backbone–backbone, backbone–sidechain, and sidechain–sidechain interactions in explicit water. The rest of the paper is organized as follows. Section 2 describes the protein model in detail; Sect. 3 describes the atomistic and DMD simulations; Sect. 4 discusses the analysis leading to the final choice of model parameters; and Sect. 5 summarizes the current status of the model.

2 Model Description

We deploy a two-bead-per-residue protein model to represent the 20 natural amino acid residues. Since computational efficiency was a major consideration here, we tried to minimize the number of beads in the system and at the same time represent the chemical heterogeneity of the individual amino acid residues. Although a one-bead-per-residue model minimizes the number of beads in the system, we found that it made it harder to represent the difference among the various types of amino acid residue in DMD simulations. Protein models with more than two beads per residue do a good job of representing the chemical heterogeneity of the 20 residues (see, e.g., our protein model, PRIME20 [16]), but this increases the required computational resources. The two-bead-per-residue model is a good compromise for large proteins.

The 18 amino acid residues except glycine and proline are represented by two beads: one bead at the position of the C-α atom to represent the backbone entity and the other at the sidechain center of mass to represent the sidechain entity. Glycine and proline residues are represented solely by a single bead at the positions of their C-α atoms because either they do not have a sidechain or the sidechain is closely linked with the backbone. Figure 1 shows a schematic of the two-bead model for alanine.

The potential energy of the system is the sum of the intermolecular potential energy, intramolecular potential energy, and virtual bond energy for all the beads in the system (Eq. 1).

$$U_{\text{total}} = \sum U_{\text{inter}}(r) + \sum U_{\text{intra}}(r) + \sum U_{\text{bond}}(r) \tag{1}$$

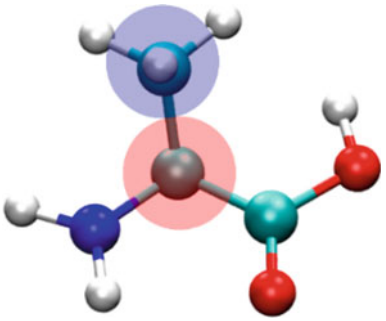

Fig. 1 Schematic of the two-bead-per-residue model. One bead is at Cα, and the other is at the center of mass of the sidechain

The intermolecular bead–bead interactions are represented by a single square well or single square shoulder potential as given in Eq. (2):

$$\begin{cases} U_{\text{inter}}(r) = \infty, & r \leq \sigma_1 \\ U_{\text{inter}}(r) = \epsilon, & \sigma_1 < r < \sigma_2 \\ U_{\text{inter}}(r) = 0, & r \geq \sigma_2 \end{cases} \tag{2}$$

where r is the bead–bead distance, σ_1 and σ_2 are geometric parameters, and ε is the energetic parameter. The geometric and energetic parameters (σ_1, σ_2, and ε) are derived from the potentials of mean force (PMFs) of sidechain–sidechain, sidechain–backbone, and backbone–backbone pairs from atomistic simulations in explicit water solvent as discussed in Sect. 4. A single square-well potential ($\varepsilon < 0$) indicates that the two entities attract each other in explicit water; a single square-shoulder potential ($\varepsilon > 0$) indicates that these two entities repel each other in explicit water; and a hard-sphere potential ($\varepsilon = 0$) indicates that the two entities just have an excluded volume interaction in water. The effect of water is taken into account in the parameters because the PMFs were obtained from the pair's interactions in explicit water solvent.

The intramolecular bead–bead non-bonded interactions consider excluded volume effects only. The hard-sphere potential is used to describe the intramolecular bead–bead non-bonded interactions (Eq. 3).

$$\begin{cases} U_{\text{intra}}(r) = \infty, & r \leq 0.8\sigma_1 \\ U_{\text{intra}}(r) = 0, & r > 0.8\sigma_1 \end{cases} \tag{3}$$

where r is the bead–bead distance and σ_1 is the geometric parameter in Eq. (2). The geometric parameters could, in principle, be obtained from the volumes of the individual beads, but to simplify the process, we choose to use $0.8\sigma_1$ as the geometric parameter. We found that this selection avoids overlap between beads in a protein and works well with the virtual bond setting, which is described in the next paragraph.

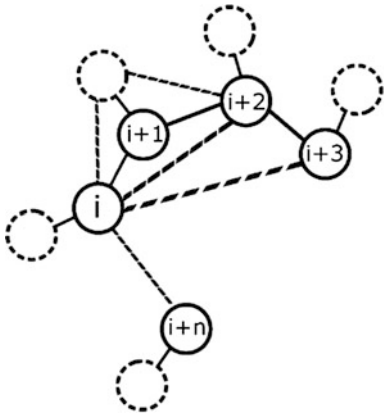

Fig. 2 Schematic describing virtual bonds. The *circles with solid borders* are backbone beads, and the *circles with dash-line borders* are sidechain beads. The virtual bonds connect these beads to keep the protein in its native state

We deploy virtual bead–bead bonds to maintain the protein in its native state. The virtual bond potential is a double hard wall (Eq. 4).

$$\begin{cases} U_{\text{bond}}(r) = \infty, & r \leq (1-\delta)\sigma \\ U_{\text{bond}}(r) = 0, & (1-\delta)\sigma < r < (1+\delta)\sigma \\ U_{\text{bond}}(r) = \infty, & r \geq (1+\delta)\sigma \end{cases} \tag{4}$$

where r is the bead–bead distance, σ is an equilibrium bead–bead distance obtained from the native state of the protein, and δ is the flexibility factor. Figure 2 shows a schematic describing the virtual bonds. The native state of a protein is its naturally folded structure. Here, we use the structure of a protein in the Protein Data Bank (PDB) as its native state. The virtual bonds can be divided into two categories depending on the indices of the connected beads along the amino acid sequence. The "local" category includes virtual bonds between beads whose index difference is less than four. They are used to maintain the protein local secondary structure. The other category (non-local) includes virtual bonds between beads far away from each other along the amino acid sequence. These bonds are used to maintain the tertiary and quaternary structures of a protein. Section 4.2 discusses the choice of the virtual bonds in detail.

3 Simulation Details

3.1 Atomistic Simulation

We conducted atomistic simulations to obtain the geometric and energetic parameters for the coarse-grained beads in the two-bead-per-residue model; these parameters are then used in the DMD simulations. The sidechain and backbone entities were generated from amino acid residues. Glycine and proline entities were generated by capping their N and C terminals with an acetyl group and an N-methyl

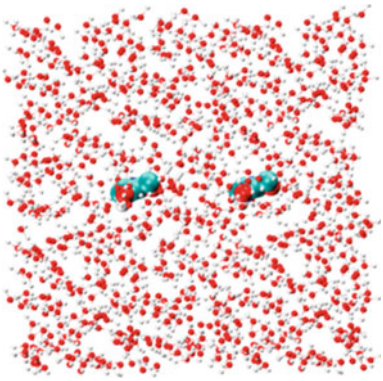

Fig. 3 Glycine–glycine pair in a 3.0 × 3.0 × 3.0 nm^3 box. Glycine molecules are represented in a VDW view, and water molecules are represented in CPK model view l. C atom *green*, N atom *blue*, O atom *red*, and H atom *white*

amide group. These caps prevent the two entities from associating with others through their N or C termini. The glycine entity was also used as the backbone entity because it is an amino acid without a sidechain. Sidechain entities were generated by detaching the sidechain of an amino acid residue from its backbone and replacing the CB atom with an H atom. Two sidechain or backbone entities were placed in a 3.0 × 3.0 × 3.0 nm^3 box filled with water molecules at a density of 1.0 g/nm^3. The initial entity–entity distance was at least 1.0 nm to avoid any artificial association. The GROMOS54a7 force field [7] was used to describe the sidechain and backbone entity, and the SPC model [30] was deployed to describe the water molecules since it is recommended for use with the GROMOS force field. Figure 3 shows the initial configuration of a glycine–glycine pair in a water box.

For each system, a 1-ns isothermal–isobaric ensemble (NPT, T = 300 K, P = 1 bar) MD simulation with a 1-fs time step was conducted after energy minimization to let the system reach the equilibrated density and potential energy. Then, a 100-ns canonical ensemble (NVT, T = 300 K) MD simulation with a 2-fs time step was conducted to collect data every 500 fs. The 12-6 Lennard–Jones interactions were treated with a 1.0-nm cutoff, and the electrostatic interactions were treated with particle mesh Ewald sum [31]. No bonds were constrained to their equilibrium length during the 1-ns NPT MD simulation. The bonds to the hydrogen atoms were constrained to their equilibrium length using LINCS algorithm [32] during the 100-ns NVT MD simulation. The desired temperature was maintained using the v-rescale algorithm [33], and the desired pressure was maintained using the Parrinello–Rahman algorithm [34]. The MD simulations and energy minimization were conducted using GROMACS-4.6.5 [35].

3.2 DMD Simulation

We conducted DMD simulations to test and scale the parameters obtained from atomistic simulations. The DMD simulations were conducted in the NVT ensemble

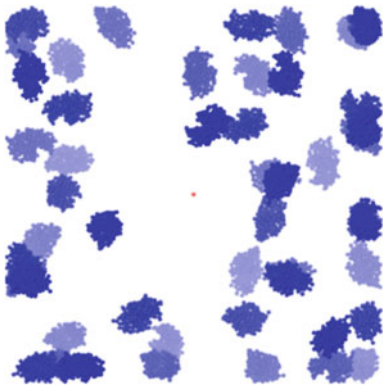

Fig. 4 The initial configurations of 50 lysozymes in a $40 \times 40 \times 40$ nm^3 box (1.38 μM)

using code developed in our group. For single-protein systems, the protein was placed in the center of a $10 \times 10 \times 10$ nm^3 box. The temperature of the system was maintained at 1.0 using the Andersen thermostat [36]. For multi-protein systems, 50 lysozyme proteins were placed at random positions in a $40 \times 40 \times 40$ nm^3 box (1.38 μM) using Packmol [37]. The initial protein–protein distance was at least 7 nm to avoid any artificial association. Figure 4 shows the initial configuration for 50 lysozyme proteins.

4 Parameter Development

4.1 Intermolecular Interaction

We use a pair of glycine (G) entities to illustrate how we get geometric and energetic parameters (σ_1, σ_2, and ε) from atomistic simulation results (Fig. 5). The radial distribution functions between the centers of mass of two glycine entities (Fig. 5a) were obtained from the MD simulation. Boltzmann inversion [38] was used to calculate the PMF (Fig. 5b). There are several ways to select the geometric and energetic parameter from a continuous potential [39, 40]. Here, we choose the geometric parameter σ_1 to be one root where the PMF = 0 (Fig. 5b) and the energetic parameter ε to be the lowest value of the PMF. Here, we choose σ_2 to be where $g(r)$ reaches the range of 1.0 ± 0.1. This method may result in a small energy perturbation (−0.1 kBT when $g(r) = 1.1$ and 0.1 kBT when $g(r) = 0.9$) but avoids the possibility of having an artificially large well/shoulder width when the PMF approaches zero slowly. The geometric parameter σ_1 for the square-shoulder is where the PMF starts to increase rapidly, and the energetic parameter ε is the

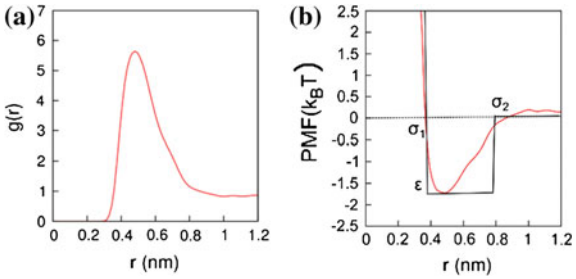

Fig. 5 **a** Radial distribution functions and **b** potential of mean force (PMF) and square-well potential for a G–G pair. The geometric (σ_1 and σ_2) and energetic (ε) parameters for the square-well potential were obtained from discretizing the PMF of entity pairs in atomistic simulations

value of PMF at σ_1. The geometric parameter σ_2 for the square-shoulder is determined in the same manner as for the square-well. If the value of the PMF is always between −0.1 and 0.1 k_BT, we deploy a hard-sphere potential for the bead–bead interaction. The geometric parameter σ_1 is selected in the same way as that for the square-shoulder.

It is of interest to ask whether or not the parameters (σ_1, σ_2, and ε) of the 210 pairs are physically meaningful and if they bias toward certain conformations. The value of the parameter σ_1 reflects how close the two entities can approach each other in water solvent. The majority of bead–bead pairs have σ_1 in the range of 0.33–0.45 nm, which is quite close to the size of the heavy atoms in the entities. These entities should be able to contact with each other directly in water solvent. Only five pairs have σ_1 larger than 0.45 nm: arginine (R)–arginine (R), arginine (R)–lysine (K), glutamic acid (E)–glutamic acid (E), tryptophan (W)–tryptophan (W), and tryptophan (W)–tyrosine (Y).

The value of parameter σ_2 reflects how far apart the two entity beads can be and still influence each other. The values of σ_2 for the 210 pairs range from 0.55 to 1.0 nm. This wide range illustrates the chemical dissimilarities among the 18 sidechain entities. The values of σ_2 for the hydrophilic and charged pairs (such as 0.85 nm for the asparagine–asparagine pair and 1.0 nm for the lysine–lysine pair) are generally larger than those for the hydrophobic pairs (0.55 nm for the valine–valine pair). This is expected because the former two are controlled by electrostatic interactions, which decrease much more slowly as a function of distance than the van der Waals interactions which control the hydrophobic associations.

The value of the energetic parameter ε reflects whether the two entities attract or repel each other. We first consider charged sidechain entities. The pairs of sidechain entities with the same sign charge have positive ε (a repulsive force), and the pairs with opposite sign charge have negative ε (an attractive force). Our atomistic MD

Table 1 The entity pairs that have a hard-sphere potential

G-S	T-R	S-H	R-I
V-D	T-H	D-Q	E-I
V-R	S-R	D-L	Y-Y
V-E	S-K	D-I	Y-W
T-D	S-E	D-M	W-W

simulations successfully capture how these charged entities interact with each other. Histidine (H) has a pKa similar to 7.0, so its net charge is quite weak. Therefore, we do not find a strong repulsive force between H and the negatively charged sidechains. Instead, we find a weak attraction, probably due the effect of water molecules.

We further examine the values of parameter ε for the other entity pairs. Two pairs, glycine–aspartic acid and glycine–proline, have a positive ε, which may be due to their different influences on the structure of water molecules. The other pairs have a negative ε, whose value depends on the chemistries of the entities and their individual effects on the structure of surrounding water molecules. For instance, the value of ε for the valine–valine pair is $-1.44\ k_BT$, and that for the serine–serine pair is only $-0.61\ k_BT$, consistent with the fact that hydrophobic substances associate more stably than hydrophilic substances in water. The glutamic acid–cysteine and lysine–tryptophan pairs have much lower ε than the others. The former may be due to an interaction between the S atoms and the charged group, and the latter may be due to a charged-group-π conjugation.

Twenty entity pairs (Table 1) have a hard-sphere potential because their interactions are judged to be very weak based on the criterion stated above. Some of these may be due to the different hydrophilicities of the entities (such as the sidechains of valine and aspartic acid). Some of these may be due to the effect of water molecules. Consider for instance, the glycine–serine sidechain pair. The serine sidechain has a hydroxyl group, which should be able to associate with the oxygen atom on glycine; however, these two entities can also form hydrogen bonds with water molecules. The water molecules around the two entities may make the glycine–serine sidechain association energetically comparable to the non-associated state. This weak interaction reminds us of the importance of taking the effect of water molecules into account when considering protein–protein interactions.

4.2 Virtual Bond

An ideal set of virtual bonds should be able to maintain the protein in its native state, while maximizing the timescale per simulation step. We investigated how this goal could be achieved by tuning the types of virtual bonds and the flexibility factor δ in Eq. (4). Table 2 lists the choice of virtual bond types and the values of δ for

Table 2 Three virtual bond sets. CA[i] means the ith backbone bead, and CB[i] means the ith sidechain bead

Local		Non-local	
Bond types	δ	Bond types	δ
Rigid			
CA[i]–CA[i + 1] CA[i]–CA[i + 2] CA[i]–CA[i + 3] CB[i]–CA[i] CB[i]–CA[i − 1] CB[i]–CA[i + 1]	0.05	CA[i]–CA[i + 10] CA[i]–CA[i + 20] ($i = i + 2$)[a] CB[i]–CB[i + 10] disulfide bonds	0.05
Moderate			
CA[i]–CA[i + 1] CA[i]–CA[i + 2] CA[i]–CA[i + 3] CB[i]–CA[i] CB[i]–CA[i − 1] CB[i]–CA[i + 1]	0.12	CA[i]–CA[i + 10] ($i = i + 2$) CA[i]–CA[i + 20] ($i = 2$, $i = i + 4$)[b] CA[i]–CA[i + 40] ($i = i + 8$) CA[i]–CA[i + 80] ($i = i + 16$) CB[i]–CB[i + 10] ($i = i + 4$) CB[i]–CB[i + 20] ($i = 2$, $i = i + 4$) disulfide bonds	0.05
Loose			
CA[i]–CA[i + 1], CA[i]–CA[i + 2] CA[i]–CA[i + 3] CB[i]–CA[i] CB[i]–CA[i − 1] CB[i]–CA[i + 1]	0.25	CA[i]–CA[i + 10] ($i = i + 2$), CA[i]–CA[i + 20] ($i = 2$, $i = i + 4$), CA[i]–CA[i + 40] ($i = i + 8$) CA[i]–CA[i + 80] ($i = i + 16$) CB[i]–CB[i + 10] ($i = i + 4$) CB[i]–CB[i + 20] ($i = 2$, $i = i + 4$) disulfide bonds CA[i]–CA[i + 120]($i = i + 16$)(myoglobin)	0.12 (lysozyme) 0.1 (myoglobin)

The value of $\delta\sigma$ is limited to be less than 0.1 nm

[a]$i = i + n$ means that this type of virtual bonds is set for beads i, i + n, $i + 2n$…

[b]$i = n$ means this type of virtual bonds starts from nth bead

three sets used for lysozyme and for myoglobin. These are labeled as "rigid," "moderate," and "loose" based on δ. The number of virtual bonds in the "loose" category is greater than that in the "moderate" category which is itself greater than that in the "rigid" category.

We select lysozyme (PDB ID: 193L) as our first test protein to evaluate the ability of these three virtual bond sets to maintain the protein in its native state. Lysozyme was chosen as our first test case because it is relatively small and rigid. The virtual bond set's ability to maintain the protein in its native state was measured using the root mean square deviation (RMSD) of all beads from the lysozyme native state conformation during a DMD simulation of 200 million collisions. As shown in Fig. 6, the small RMSD fluctations (0.15–0.30 nm for the rigid set, 0.18–0.30 nm for the moderate set, and 0.15–0.25 nm for the loose set) indicate that all three sets work well at maintaining lysozyme in its native structure. Interestingly,

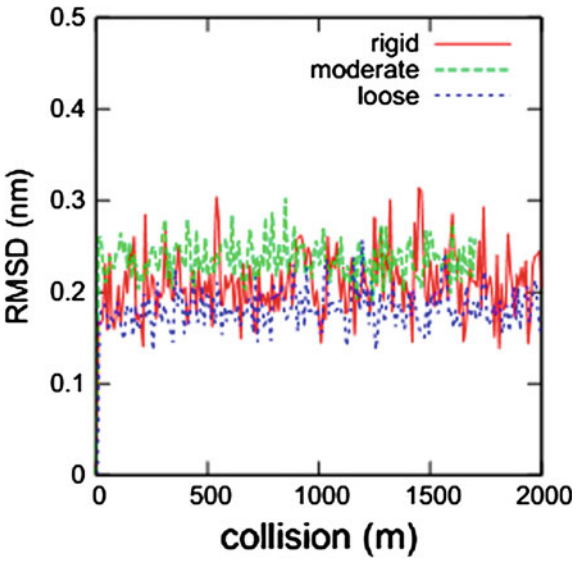

Fig. 6 Root mean square deviation (RMSD) of all beads in a lysozyme during a 2-billion-collision DMD simulation

Table 3 Simulation time advanced by a 100-million-collision DMD simulation of 50 lysozymes

	Reduced time
Rigid	619
Moderate	1457
Loose	2072

the "loose" set works as well as the other two even though its δ is much higher than the other two.

The "loose" virtual bond set works best at maximizing the simulation timescale per million collisions. As listed in Table 3, a simulation of 50 lysozymes shows that for a 100-million-collision simulation, the simulation-reduced time achieved with the "loose" virtual bond set is 1.4 times that with the "moderate" virtual bond set and 3.3 times that with the "rigid" virtual bond set. DMD simulations of complex molecules such as proteins spend more than 90 % of the simulation time in collisions between the bonded beads, indicating that the timescale of a DMD simulation heavily depends on the bond flexibility. A "loose" virtual bond set allows the simulation to advance much faster than a "rigid" one. However, the "loose" set may also increase the risk that the protein will deform. This risk should be taken into account when selecting proper virtual bonds.

We then tested the performance of these three virtual bond settings for a more flexible protein: myoglobin (PDB ID: 1YMB). The RMSD of myoglobin well illustrates the importance of choosing the virtual bond types and flexibility factor more carefully. The "rigid" and "moderate" virtual bond sets for myoglobin are the

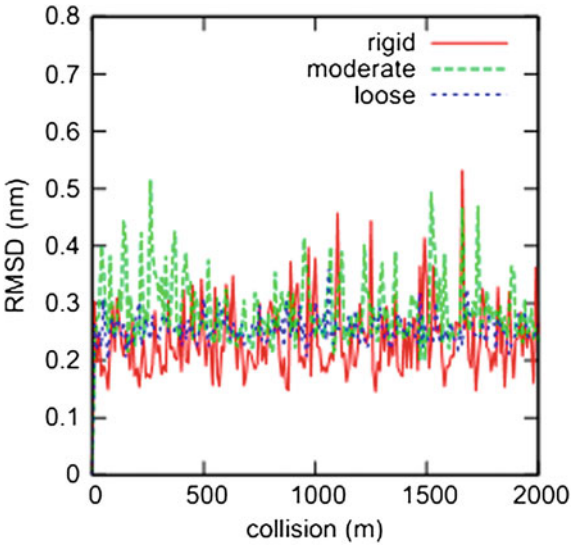

Fig. 7 RMSD of all beads in myoglobin during a 2-billion-collision DMD simulation

same as for lysozyme. As shown in Fig. 7, the RMSDs with these two virtual bond sets fluctuate from 0.15 to 0.5 nm (rigid) and 0.2 to 0.5 nm (moderate). The large RMSD fluctuations indicate that we need to select virtual bonds carefully. We thus set a "loose" set for myoglobin, which has a new type of virtual bond: CA[i]−CA[i + 120] ($i = i + 16$) because myoglobin is larger than lysozyme. In this set, the value of δ for the local virtual bonds is chosen to be 0.25; however, unlike our choice for lysozyme, we set the non-local δ to be 0.10 instead of 0.12. As shown in Fig. 7, the increase in the number of "non-local" virtual bonds improves the ability of the model to hold myoglobin in its native state (0.25 < RMSD < 0.3 nm).

The comparison among the three virtual bond sets shows the importance of the non-local virtual bonds in maintaining the protein in its native state. A protein usually keeps its tertiary and quaternary structures with the help of non-bonded intramolecular interactions such as hydrogen bonds and hydrophobic associations, which are not considered in the current version of our model. All the virtual bonds that connect the beads whose index difference is less than four are there to maintain the bonds, angles, and dihedral angles between the beads. They work well at maintaining the local secondary structure of a protein but have little influence on the tertiary and quaternary structures of the protein. Thus, the model depends on the non-local virtual bonds between beads that are far away in the sequence to hold different regions of a protein together. The selection of "non-local" virtual bonds is still an art. Using the virtual bonds that connect the beads whose index difference is 10, 20, 40, 80 …, up to the total number of the beads in the protein, works well. The value of δ for the "non-local" virtual bonds should be less than that for the "local"

ones due to their large equilibrated bond length. Although the total number of "non-local" virtual bonds is much less than that for "local" virtual bonds, the non-local bonds are very important as they greatly enhance the ability of a protein to stay in its native state. For instance, having six virtual bonds between CA[i] and CA[i + 120] results in a decrease of the RMSD of myoglobin from around 0.2–0.5 nm (rigid and moderate) to 0.25–0.3 nm (loose) (Fig. 7). Such a strategy could be useful for the simulation of other complex systems.

4.3 Energetic Parameter Adjustment

The values of the force field parameters need to be adjusted to ensure that the coarse-grained model gives reasonable results in comparison with experiment. Tables 4, 5, and 6 show the values of σ_1, σ_2 and ε obtained from the PMFs for the 210 bead-bead pairs. Here, we select the osmotic second virial coefficient (B_{22}) of lysozyme as the reference property because it well represents the strength of lysozyme–lysozyme interactions in a solution. Lenhoff and his colleagues [41–44] measured B_{22} for lysozyme in a variety of solutions. Lysozyme is expected to have a positive B_{22} in water because it is positively charged. Their data indicate that B_{22} of lysozyme in water at pH 7 is around 5×10^{-4} mol ml/g^2.

We use Eq. (5) to calculate B_{22} from radial distribution functions g(r) of the center of mass of lysozymes obtained from DMD simulations of our system of 50 lysozymes.

$$B_{22} = -\frac{2\pi}{M_w^2 N_A} \int_0^\infty (g(r) - 1) r^2 \mathrm{d}r \tag{5}$$

where M_w is the molecular weight of the protein, N_A is the Avogadro constant, and $g(r)$ is the radial distribution function. We then compare the simulation value of B_{22} to the experimental one; a simulation value of B_{22} larger than the experimental one would imply that the force field overestimates the attraction among proteins, while a simulation value of B_{22} smaller than the experimental one would imply that the force field overestimates the repulsion among proteins.

The energetic parameters are adjusted so that the value of B_{22} obtained from simulations approaches the experimental value. Ideally, all the geometric and energetic parameters should be adjusted individually, but this would require massive data which are not achievable now. Alternatively, if we fix the interaction ranges of the 210 pairs, i.e., the geometric parameters σ_1 and σ_2, and keep the ratio of the energetic parameters for any two pairs unchanged, we can adjust all the energetic parameters by multiplying them by a single factor f. This helps us to narrow the difference between the simulation and experiment results for B_{22}.

Table 4 Geometric parameter σ_1 of 210 interactive site pairs (nm)

	G	A	V	P	T	S	N	D	R	K	E	Q	L	I	F	Y	W	M	C	H
G	0.33																			
A	0.38	0.35																		
V	0.38	0.38	0.40																	
P	0.38	0.40	0.40	0.45																
T	0.38	0.38	0.40	0.40	0.40															
S	0.38	0.38	0.40	0.40	0.40	0.40														
N	0.38	0.38	0.38	0.40	0.38	0.38	0.38													
D	0.38	0.38	0.40	0.40	0.40	0.35	0.38	0.45												
R	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.40	0.50											
K	0.38	0.38	0.40	0.40	0.38	0.38	0.38	0.38	0.50	0.40										
E	0.38	0.38	0.45	0.40	0.40	0.38	0.38	0.45	0.38	0.40	0.50									
Q	0.38	0.38	0.38	0.40	0.38	0.38	0.38	0.40	0.38	0.38	0.38	0.38								
L	0.38	0.38	0.40	0.40	0.40	0.40	0.38	0.40	0.38	0.40	0.40	0.38	0.40							
I	0.38	0.38	0.40	0.38	0.38	0.38	0.38	0.40	0.38	0.40	0.40	0.38	0.40	0.38						
F	0.38	0.38	0.40	0.40	0.40	0.40	0.38	0.42	0.38	0.38	0.40	0.38	0.40	0.38	0.40					
Y	0.38	0.38	0.40	0.40	0.40	0.40	0.38	0.45	0.38	0.38	0.45	0.38	0.40	0.40	0.40					
W	0.38	0.38	0.40	0.40	0.45	0.40	0.38	0.45	0.38	0.38	0.45	0.38	0.40	0.40	0.42	0.60	0.60			
M	0.38	0.38	0.40	0.40	0.40	0.40	0.38	0.40	0.38	0.38	0.40	0.38	0.40	0.38	0.40	0.40	0.40	0.40		
C	0.38	0.38	0.38	0.40	0.38	0.38	0.38	0.30	0.38	0.38	0.35	0.38	0.40	0.40	0.35	0.40	0.38	0.38	0.38	
H	0.38	0.38	0.38	0.40	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38

Table 5 Geometric parameter σ_2 of 210 interactive site pairs (nm), "hs" means hard-sphere

	G	A	V	P	T	S	N	D	R	K	E	Q	L	I	F	Y	W	M	C	H
G	0.75																			
A	0.60	0.55																		
V	0.70	0.65	0.65																	
P	0.75	0.65	0.70	0.70																
T	0.65	0.60	0.65	0.65	0.65															
S	hs	0.60	0.65	0.65	0.60	0.60														
N	0.65	0.62	0.70	0.65	0.70	0.70	0.65													
D	0.60	0.65	hs	0.60	hs	0.40	0.55	0.85												
R	0.65	0.65	hs	0.80	hs	hs	0.50	0.80	0.65											
K	0.75	0.70	0.65	0.70	0.60	hs	0.70	0.80	0.80	1.00										
E	0.65	0.65	hs	0.65	0.80	hs	0.60	0.80	0.70	0.80	0.80									
Q	0.75	0.65	0.70	0.70	0.70	0.70	0.70	hs	0.65	0.70	0.60	0.75								
L	0.75	0.65	0.70	0.70	0.65	0.65	0.70	hs	0.70	0.70	0.80	0.75	0.70							
I	0.70	0.65	0.70	0.70	0.70	0.65	0.70	hs	hs	0.75	hs	0.70	0.70	0.70						
F	0.75	0.65	0.70	0.70	0.65	0.60	0.70	0.68	0.55	0.65	0.75	0.75	0.70	0.70	0.70					
Y	0.70	0.60	0.60	0.70	0.65	0.60	0.70	0.80	0.70	0.65	0.85	0.70	0.70	0.70	0.70	hs				
W	0.80	0.70	0.70	0.70	0.70	0.60	0.70	0.80	0.65	0.70	0.80	0.80	0.75	0.75	0.75	hs	hs			
M	0.75	0.65	0.65	0.70	0.65	0.65	0.70	hs	0.75	0.65	0.75	0.75	0.70	0.70	0.65	0.65	0.70	0.65		
C	0.60	0.60	0.65	0.65	0.70	0.65	0.65	0.45	0.65	0.65	0.50	0.65	0.70	0.70	0.70	0.75	0.75	0.70	0.65	
H	0.75	0.65	0.70	0.70	hs	hs	0.70	0.55	0.60	0.70	0.58	0.75	0.70	0.70	0.70	0.70	0.70	0.70	0.75	0.75

Table 6 Energetic parameter ε of 210 interactive site pairs (k_BT), "hs" means hard-sphere

	G	A	V	P	T	S	N	D	R	K	E	Q	L	I	F	Y	W	M	C	H
G	−1.61																			
A	−0.81	−1.44																		
V	−1.21	−1.01	−1.21																	
P	−1.42	−0.81	−0.81	−1.42																
T	−0.40	−0.81	−0.81	−0.81	−0.40															
S	hs	−1.13	−0.81	−0.61	−0.40	−0.61														
N	−0.81	−1.01	−1.13	−1.01	−0.61	−0.81	−0.81													
D	0.40	−0.20	hs	0.40	hs	−1.01	−1.01	0.81												
R	−0.40	−0.20	hs	−0.81	hs	hs	−0.20	−0.81	0.81											
K	−0.40	−0.40	−0.61	−0.81	−0.61	hs	−0.61	−0.81	0.81	0.40										
E	−0.40	−0.40	hs	−0.40	0.40	hs	−0.81	0.61	−0.61	−0.81	0.61									
Q	−1.21	−1.10	−0.81	−1.01	−0.81	−0.40	−0.81	hs	−0.61	−0.61	−0.40	−0.61								
L	−1.41	−1.21	−1.01	−1.01	−0.81	−1.01	−1.01	hs	−0.40	−0.81	−0.40	−1.21	−1.21							
I	−1.21	−0.81	−1.01	−1.21	−0.81	−0.81	−0.81	hs	hs	−0.40	hs	−0.81	−1.01	−0.81						
F	−1.62	−1.01	−0.81	−1.42	−0.81	−0.40	−1.21	−0.41	−1.62	−0.40	−0.40	−1.21	−1.21	−1.21	−1.21					
Y	−1.21	−0.81	−0.40	−0.61	−0.20	−0.40	−0.40	0.61	−2.22	−1.42	0.40	−0.81	−0.81	−0.61	−0.81	hs				
W	−1.21	−0.81	−0.81	−1.21	−0.40	−0.40	−0.61	0.61	−2.42	−2.83	0.40	−0.81	−1.01	−0.81	−0.81	hs	hs			
M	−1.41	−1.01	−1.01	−1.21	−0.40	−0.61	−1.21	hs	−1.01	−1.01	−0.40	−1.21	−1.13	−1.13	−1.01	−0.81	−1.01	−1.01		
C	−0.81	−0.81	−0.81	−0.81	−0.81	−0.61	−1.01	−3.03	−1.13	−0.81	−2.43	−0.73	−1.01	−0.81	−1.01	−0.48	−0.81	−0.81	−0.73	
H	−0.81	−0.81	−1.01	−1.01	hs	hs	−0.81	−1.54	−0.40	−0.40	−1.21	−1.01	−1.21	−1.21	−1.13	−1.01	−1.01	−0.81	−0.60	−0.61

The reduced temperature T^* is set to 1.0 when tuning the factor. For each f, the average value of B_{22} was obtained from three independent DMD simulations starting from different initial configurations. These simulations lasted for 120–170 billion steps with a total reduced time τ of around 1×10^6.

We find that f needs to be small in order to get our value of B_{22} to be close to experimental value. The value of B_{22} is $1.9 \pm 0.83 \times 10^{-4}$ mol ml/g^2 when $f = 0.15$ and increases to $7.1 \pm 3.18 \times 10^{-4}$ mol ml/g^2 when $f = 0.10$. These values are close to the value obtained experimentally (5×10^{-4} mol ml/g^2). We chose to set $f = 0.10$ as the scale factor because the simulation results for B_{22} with $f = 0.1$ straddle the experimental value. It is not surprising to find such a small f value. There are two possible reasons for the need for such drastic rescaling. First, coarse-graining smooths the free energy surface and makes it easier for proteins to aggregate. Second, the current coarse-graining method may not be able to address the effect of water molecules near the proteins well. Stark et al. [25] also found that they needed to drastically rescale the energetic parameters of the MARTINI force field [45] to match the experimental value for B_{22} of lysozyme. The necessity of scaling parameters to match experiment results was also observed for ionic liquids [46]. A possible reason for such necessity is that the current models use an additive two-body interaction system, which is an approximation to the many-body interactions. Parameter scaling may be an effective way to attenuate the error brought by the different interaction systems.

5 Conclusion

We have developed a discontinuous potential two-bead-per-residue protein model so that we can conduct DMD simulations to investigate protein–protein interactions in a multi-protein system. The current model focuses on proteins that are in their native states. We derive the intermolecular bead–bead interactions from the potential of mean force obtained from atomistic simulations. Examination of the geometric and energetic parameters shows that these parameters are physically meaningful. We also developed strategies to set the types and flexibility of the virtual bonds to constrain the proteins in their native state while maximizing the simulation timescale. Comparison of a variety of virtual bond sets illustrates that high bond flexibility (the loose set) improves the DMD simulation performance. We also scale the energetic parameters of our model to match experimental results on the osmotic second virial coefficient of lysozyme. We are using this model to investigate the formation of the corona of proteins that forms around a nanoparticle.

Acknowledgments This work was supported by National Science Foundation (CBET-1236053) and the National Institutes of Health (EB006006). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

References

1. Kastelic, M., Kalyuzhnyi, Y.V., Hribar-Lee, B., Dill, K.A., Vlachy, V.: Protein aggregation in salt solutions. Proc. Natl. Acad. Sci. U.S.A. **112**, 6766–6770 (2015)
2. Azzarito, V., Long, K., Murphy, N.S., Wilson, A.J.: Inhibition of α-helix-mediated protein-protein interactions using designed molecules. Nat. Chem. **5**, 161–173 (2013)
3. Hober, S., Nord, K., Linhult, M.: Protein A chromatography for antibody purification. J. Chrom. B **848**, 40–47 (2007)
4. Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., MacKerell, A.D.: Optimization of the additive CHARMM All-atom protein force field Targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. J. Chem. Theory Comput. **8**, 3257–3273 (2012)
5. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C.: Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Struct. Funct. Bioinf. **65**, 712–725 (2006)
6. Huang, W., Lin, Z., van Gunsteren, W.F.: Validation of the GROMOS 54A7 force field with respect to β-peptide folding. J. Chem. Theory Comput. **7**, 1237–1243 (2011)
7. Schmid, N., Eichenberger, A., Choutko, A., Riniker, S., Winger, M., Mark, A., van Gunsteren, W.: Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur. Biophys. J. **40**, 843–856 (2011)
8. Jorgensen, W.L., Maxwell, D.S., Tirado-Rives, J.: Development and testing of the OPLS All-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. **118**, 11225–11236 (1996)
9. Tozzini, V.: Coarse-grained models for proteins. Curr. Opin. Struct. Biol. **15**, 144–150 (2005)
10. Wu, C., Shea, J.-E.: Coarse-grained models for protein aggregation. Curr. Opin. Struct. Biol. **21**, 209–220 (2011)
11. Saunders, M.G., Voth, G.A.: Coarse-graining of multiprotein assemblies. Curr. Opin. Struct. Biol. **22**, 144–150 (2012)
12. Baaden, M., Marrink, S.J.: Coarse-grain modelling of protein–protein interactions. Curr. Opin. Struct. Biol. **23**, 878–886 (2013)
13. Noid, W.G.: Perspective: coarse-grained models for biomolecular systems. J. Chem. Phys. **139**, 090901(1–25) (2013)
14. Saunders, M.G., Voth, G.A.: Coarse-graining methods for computational biology. Annu. Rev. Biophys. **42**, 73–93 (2013)
15. Kar, P., Feig, M.: In Biomolecular modelling and simulations. In: Karabencheva Christova, T. (ed.) vol. 96, p. 143. Elsevier Academic Press Inc., San Diego (2014)
16. Cheon, M., Chang, I., Hall, C.K.: Extending the PRIME model for protein aggregation to all 20 amino acids. Proteins **78**, 2950–2960 (2010)
17. Arkhipov, A., Yin, Y., Schulten, K.: Four-scale description of membrane sculpting by BAR domains. Biophys. J. **95**, 2806–2821 (2008)
18. Head-Gordon, T., Brown, S.: Minimalist models for protein folding and design. Curr. Opin. Struct. Biol. **13**, 160–167 (2003)
19. Matysiak, S., Clementi, C.: Minimalist protein model as a diagnostic tool for misfolding and aggregation. J. Mol. Biol. **363**, 297–308 (2006)

20. Dama, J.F., Sinitskiy, A.V., McCullagh, M., Weare, J., Roux, B., Dinner, A.R., Voth, G.A.: The theory of ultra-coarse-graining. 1. general principles. J. Chem. Theory Comput. **9**, 2466–2480 (2013)
21. Best, R.B., Hummer, G., Eaton, W.A.: Native contacts determine protein folding mechanisms in atomistic simulations. Proc. Natl. Acad. Sci. U.S.A. **110**, 17874–17879 (2013)
22. Sippl, M.J.: Knowledge-based potentials for proteins. Curr. Opin. Struct. Biol. **5**, 229–235 (1995)
23. Thompson, J.J., Tabatabaei Ghomi, H., Lill, M.A.: Application of information theory to a three-body coarse-grained representation of proteins in the PDB: insights into the structural and evolutionary roles of residues in protein structure. Proteins, **82**, 3450–3465 (2014)
24. Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., de Vries, A.H.: The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B **111**, 7812–7824 (2007)
25. Stark, A.C., Andrews, C.T., Elcock, A.H.: Toward optimized potential functions for protein-protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the MARTINI coarse-grained force field. J. Chem. Theory Comput. **9**, 4176–4185 (2013)
26. Smith, S.W., Hall, C.K., Freeman, B.D.: Molecular dynamics for polymeric fluids using discontinuous potentials. J. Comput. Phys. **134**, 16–30 (1997)
27. Proctor, E.A., Ding, F., Dokholyan, N.V.: Discrete molecular dynamics. Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 80–92 (2011)
28. Shirvanyants, D., Ding, F., Tsao, D., Ramachandran, S., Dokholyan, N.V.: Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization. J. Phys. Chem. B **116**, 8375–8382 (2012)
29. Paul, G.: A complexity O(1) priority queue for event driven molecular dynamics simulations. J. Comput. Phys. **221**, 615–625 (2007)
30. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J.: Intermolecular Forces Reidel. Dordrecht (1981)
31. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G.: A smooth particle mesh Ewald method. J. Chem. Phys. **103**, 8577–8593 (1995)
32. Hess, B.: P-LINCS: a parallel linear constraint solver for molecular simulation. J. Chem. Theory Comput. **4**, 116–122 (2008)
33. Bussi, G., Donadio, D., Parrinello, M.: Canonical sampling through velocity rescaling. J. Chem. Phys. **126**, 014101(1–7) (2007)
34. Parrinello, M., Rahman, A.: Polymorphic transitions in single crystals: a new molecular dynamics method. J App. Phys. **52**, 7182–7190 (1981)
35. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J. C., Kasson, P.M., van der Spoel, D., Hess, B., Lindahl, E.: GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics **29**, 845–854 (2013)
36. Andersen, H.C.: Molecular dynamics simulations at constant pressure and/or temperature. J. Chem. Phys. **72**, 2384–2393 (1980)
37. Martínez, L., Andrade, R., Birgin, E.G., Martínez, J.M.: PACKMOL: A package for building initial configurations for molecular dynamics simulations. J. Comput. Chem. **30**, 2157–2164 (2009)
38. Reith, D., Pütz, M., Müller-Plathe, F.: Deriving effective mesoscale potentials from atomistic simulations. J. Comput. Chem. **24**, 1624–1636 (2003)
39. Thomson, C., Lue, L., Bannerman, M.N.: Mapping continuous potentials to discrete forms. J Chem. Phys. **140**, 034105(1–9) (2014)
40. Curtis, E.M., Hall, C.K.: Molecular dynamics simulations of dppc bilayers using "LIME", a new coarse-grained model. J Phys. Chem. B **117**, 5019–5030 (2013)
41. Neal, B.L., Lenhoff, A.M.: Excluded volume contribution to the osmotic second virial coefficient for proteins. AIChE J. **41**, 1010–1014 (1995)

42. Ruppert, S., Sandler, S.I., Lenhoff, A.M.: Correlation between the osmotic second Virial coefficient and the solubility of proteins. Biotechnol. Progr. **17**, 182–187 (2001)
43. Tessier, P.M., Lenhoff, A.M., Sandler, S.I.: Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography. Biophys. J. **82**, 1620–1631 (2002)
44. Tessier, P.M., Sandler, S.I., Lenhoff, A.M.: Direct measurement of protein osmotic second virial cross coefficients by cross-interaction chromatography. Protein Sci. **13**, 1379–1390 (2004)
45. Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., Marrink, S.-J.: The MARTINI coarse-grained force field: extension to proteins. J. Chem. Theory Comput. **4**, 819–834 (2008)
46. Marin-Rimoldi, E., Shah, J.K., Maginn, E.J.: Monte Carlo simulations of water solubility in ionic liquids: a force field assessment. Fluid Phase Equilib. (2015). doi:10.1016/j.fluid.2015.07.007