JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

PACSAB: Coarse-Grained Force Field for the Study of Protein−Protein Interactions and Conformational Sampling in Multiprotein Systems

Agustí Emperador,[*,†,‡] Pedro Sfriso,[†,‡] Marcos Ariel Villarreal,[§] Josep Lluis Gelpí,[†,‡,∥,⊥] and Modesto Orozco[*,†,‡,∥,⊥]

[†]Institute for Research in Biomedicine (IRB Barcelona), Baldiri i Reixac 10, Barcelona 08028, Spain

[‡]Joint BSC-IRB Research Program in Computational Biology, IRB Barcelona, Barcelona 08028, Spain

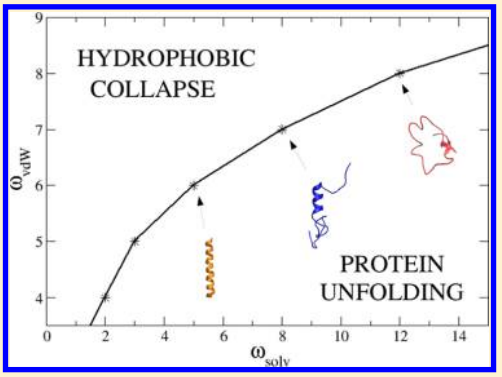
[§]Instituto de Investigaciones en Fisicoquímica de Córdoba - Departamento de Matemática y Física, CONICET-Universidad Nacional de Córdoba, University City, Córdoba 5000, Argentina

[∥]Barcelona Supercomputing Center, Jordi Girona 29, Barcelona 08034, Spain

[⊥]Departament de Bioquímica, Facultat de Biologia, Avgda Diagonal 645, Barcelona 08028, Spain

Supporting Information

ABSTRACT: Molecular dynamics simulations of proteins are usually performed on a single molecule, and coarse-grained protein models are calibrated using single-molecule simulations, therefore ignoring intermolecular interactions. We present here a new coarse-grained force field for the study of many protein systems. The force field, which is implemented in the context of the discrete molecular dynamics algorithm, is able to reproduce the properties of folded and unfolded proteins, in both isolation, complexed forming well-defined quaternary structures, or aggregated, thanks to its proper evaluation of protein−protein interactions. The accuracy and computational efficiency of the method makes it a universal tool for the study of the structure, dynamics, and association/dissociation of proteins.



I. INTRODUCTION

The theoretical representation of systems of interacting proteins presents major challenges due to the need to simulate very large systems (often above millions of atoms) for very long periods of time (in some cases on the time scale of days[1]). Despite the impressive advance of atomistic molecular dynamics,[2] the representation of protein structure, dynamics, and interactions still needs of the use of simplified models that allow a more efficient sampling of the protein conformational space. Coarse-grained (CG) models increase computational efficiency by using implicit solvent models[3] and by collapsing groups of atoms on beads.[4] This results in a reduction of the number of degrees of freedom of the system, which combined with more efficient motion propagation schemes accelerates the simulations with respect to atomistic molecular dynamics.

Most transferable CG force fields for proteins were fitted to reproduce the folded state of a protein,[5−7] or at most to reproduce the transition from unfolded to folded state,[8−12] but they cannot reproduce the behavior of intrinsically disordered proteins (IDPs). Attempts to develop IDP CG models yield to functionals which are unable to represent folded proteins,[13,14] highlighting the problems to represent with a single functional folded and unfolded states of proteins. Furthermore, existing CG force fields were created to study isolated proteins and are not prepared to reproduce well-ordered protein complexes. Some of the most successful coarse-grained models used in molecular dynamics simulations of proteins have been PaLaCe[5] and PRIMO,[7] that give an excellent description of the structure and dynamics of folded proteins, and also OPEP[11] which, apart from that, was able to fold several peptides and sample conformational changes in small aggregates.[12] In summary, despite decades of effort, there are not general CG methods able to represent correctly the dynamics of proteins both in its folded and unfolded conformations, and the association/dissociation dynamics in multiprotein systems. This lack of methodology hampers our ability to describe theoretically the dynamics, interactions, and association of proteins.

We present here a **p**airwise **a**dditive potential for **c**oarse-grained **s**ide chains and **a**tomistic **b**ackbone protein model (PACSAB) with a transferable force field for the simulation of many-protein systems. Contrary to many CG models which are based on knowledge rules on folded proteins,[15] our approach is based on a contraction of an implicit solvent classical atomistic model, which makes possible transferability to different sce-

Received: July 10, 2015
Published: November 10, 2015

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

narios and systems. The force field is adapted to the framework of discrete molecular dynamics (DMD),[16] which allows a very efficient sampling of large protein systems. The parameters defining the potential energy function in the model were fitted by exploring the phase diagram for a solution of Aβ40 peptides. The resulting force field was then tested in DMD simulations of IDPs, folded proteins, and protein−protein complexes. In summary, we present here a universal coarse-grained simulation model to explore the conformational space and interactions in multiprotein systems.

II. METHODS

Mapping of the Proteins. The aim of our model is to study different conformations and aggregation states of proteins, which means that the coarse-graining strategy should be designed to reproduce accurately excluded volume effects, side chain packing and backbone hydrogen bonding. Following Marrink's strategy,[17,18] we have placed beads at all C_αs to define the protein trace, plus a variable number of beads to describe the side chains using the mapping defined in ref 19 (typically each bead represents four heavy atoms; see Figure 1A).

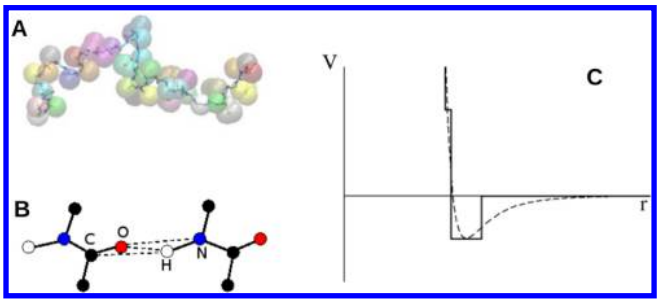


Figure 1. (A) Extended conformation of a Aβ40 peptide in our coarse-grained model. Each residue is represented with a different color. (B) Pseudobonds used to define the hydrogen bond (see main text). (C) Schematic picture of the construction of the discretized potential (solid line). The potential well is centered around $R^*_{AB} = R^*_A + R^*_B$, the sum of the bead radii (see main text). The dashed line is the continuous potential.

We have concentrated all the atoms of the bead on its center of mass, therefore all the atom−atom distances become equal to the bead−bead distances. Additionally, to represent explicitly hydrogen bonds we have added the backbone atoms N, H, C, and O.[5] We have also added a dummy atom bound to the C_α of each residue in order to keep the proper chirality of the amino acids. Solvent effects were reproduced using an implicit solvent model, which increases computational efficiency and sampling in the study of diluted systems.

DMD Simulations and Sampling. In DMD simulations the particles are considered as hard spheres interacting through discontinuous potentials, therefore moving at constant velocity until a collision (event) happens.[16] Events occur when pairwise distance equals the distance of a discontinuity in the interaction potential (see Figure 1C). No forces have to be calculated, and it is not necessary to integrate the equations of motion, speeding up the simulation as compared with conventional molecular dynamics (MD). Hardcore potentials preventing steric clashes are defined between unbound particles, and infinite square wells are defined between bound particles to keep the proper bond distances. Additional square wells are used to preserve the side chain geometry (pseudobonds).

According to DMD, the trajectory of the particles between collisions is

$$\vec{r}_i(t + t_c) = \vec{r}_i(t) + \vec{v}_i(t)t_c$$

where $t_c = \min(t_{ij})$ is the next collision time and

$$t_{ij} = \frac{-\vec{r}_{ij}\cdot\vec{v}_{ij} \pm \sqrt{(\vec{r}_{ij}\cdot\vec{v}_{ij})^2 - v_{ij}^{\,2}(r_{ij}^{\,2} - d^2)}}{v_{ij}^{\,2}}$$

being d the sum of the radii of particles i and j.

When two particles collide there is a transfer of linear momentum

$$m_i\vec{v}_i = m_i\vec{v}_i' + \Delta\vec{p}$$

$$m_j\vec{v}_j + \Delta\vec{p} = m_j\vec{v}_j'$$

Conservation of momentum and energy is imposed at each event, and from this, the velocity of each particle after the collision is found

$$m_i\vec{v}_i + m_j\vec{v}_j = m_i\vec{v}_i' + m_j\vec{v}_j'$$

$$\frac{1}{2}m_i v_i^{\,2} + \frac{1}{2}m_j v_j^{\,2} = \frac{1}{2}m_i v_i'^{\,2} + \frac{1}{2}m_j v_j'^{\,2} + \Delta V$$

Simulations were performed in the canonical ensemble, using the Andersen thermostat (for more details, see ref 16). The sampling obtained in implicit solvent DMD simulations is much higher than that expected from atomistic explicit solvent MD, due to the lack of collisions with solvent molecules. In practice, this implies that simulation time defined in a DMD trajectory corresponds to roughly 2−3 orders of magnitude longer real time.[20] The speed of the DMD simulations with the PACSAB model for different systems studied in this work is shown in Table S1 of the Supporting Information.

Construction of the Force Field. The force field consists of bonded and nonbonded terms. The first set of terms is used to maintain covalent structure, while the second accounts for intra- and intermolecular interactions. In all the cases, the different terms of the interaction potential are expressed by means of square well potentials to make possible their implementation with the DMD algorithm.

Bonded Terms. Square potentials are used to maintain all chemical bonds and to fix the bond angles. We also use a pseudobond to fix the dihedral angle of the peptide bonds in order to enforce its planar geometry, but we do not implement any other dihedral in the PACSAB model. Bonds and pseudobonds are defined as narrow square wells (with infinite depth to prevent bond breaking), whose center is at the equilibrium distance corresponding to each covalent bond, angle or dihedral.[21]

Nonbonded Terms. The non-bonded interactions comprise hydrogen bonding, defined only between the amide N, H, C, and O atoms and interactions between nonbonded beads (van der Waals and implicit solvation), affecting only C_α and side chain beads.

Hydrogen Bonds. They are represented by means of square wells of depth E_{hb} and are defined for the pairs O−H, O−N, and C−H, whenever these four atoms fulfill a geometry corresponding to the correct alignment and distance between the two dipoles N−H and O−C (see Figure 1B). Following the ideas in ref 22, we increase the stability of hydrogen bonds that are buried inside the protein, therefore not distorted by interactions with water. With this purpose the hydrogen bond

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

energy is defined as $E_{hb} = E_{hb}\gamma_i\gamma_j + E_{hb}^{core}(1 - \gamma_i\gamma_j)$, where $E_{hb}^{core} > E_{hb}$. γ is a structurally dependent shifting function that helps to smoothly move from fully exposed ($\gamma = 1$) to buried ($\gamma = 0$):

$$\gamma(n) = \frac{1}{1 + \exp((n - \alpha)/\beta)} \quad (1)$$

where n is an integer quantity related to the level of exposition to the solvent of the particle (see Appendix A), α is the limit value between exposed and buried, and β is the sharpness of the step. The values of α and β were adjusted from simulations on our training set of folded proteins (see below).

Interactions between Nonbonded Beads. The interaction between any pair of nonbonded beads A and B is defined as

$$V_{AB}(r) = \omega_{vdW}V_{AB}^{vdW}(r) + \omega_{solv}V_{AB}^{solv}(r) \quad (2)$$

where ω_{vdW} and ω_{solv} are the weights of the optimized van der Waals and implicit solvation terms (see below). In this work we have considered that electrostatic effects are properly included[22] through the hydrogen bonding and the implicit solvation terms Such an approach was used with success in the ab initio folding of several small proteins.[22]

To construct the interactions between non bonded coarse-grained beads, we assume that all the nonbonded interaction potential terms are pairwise additive in terms of the atomistic interactions:

$$V_{AB}(r) = \sum_{i\in A}\sum_{j\in B} V_{ij}(r) \quad (3)$$

where r is the distance between beads A and B and $V_{ij}(r)$ is the atomistic interaction. We use the van der Waals parameters ε_i^* of the atomistic CHARMM19 force field[23] to construct the coarse-grained van der Waals interactions, plus the atomistic EEF1 effective energy function of Lazaridis and Karplus[24] to derive the implicit solvent coarse-grained model. Constructing our potentials from atomistic interactions allows us to avoid biases[25] due to the use of statistical potentials derived from databases of crystal structures,[26,27] opening the possibility to study disordered proteins.

The *van der Waals interaction* between the coarse-grained beads is defined as

$$V_{AB}^{vdW}(r) = \varepsilon_{AB}^*\left[\left(\frac{R_{AB}^*}{r}\right)^{12} - 2\left(\frac{R_{AB}^*}{r}\right)^{6}\right] \quad (4)$$

r being the distance between beads A and B. $R_{AB}^* = R_A^* + R_B^*$, the sum of the radii of beads A and B. To compute the bead radii, we consider that the volume of each bead is proportional to the sum of the volumes of each atom included into the bead, leading to the relation $R^* = \rho(\sum R_i^{*3})^{1/3}$, R_i^* being the radius of each atom, ρ being fitted to 0.9 after inspection of atomistic residue−residue interaction profiles.

The interaction hardness ε_{AB}^* is computed extrapolating from atomistic van der Waals interactions (see Appendix B):

$$\varepsilon_{AB}^* = -\sqrt{\sum_{i\in A}\varepsilon_i^* \sum_{j\in B}\varepsilon_j^*}\left[\left(\frac{2/\rho}{N_A^{1/3} + N_B^{1/3}}\right)^{12} - 2\left(\frac{2/\rho}{N_A^{1/3} + N_B^{1/3}}\right)^{6}\right] \quad (5)$$

where N_A (N_B) is the number of atoms included by bead A (B) and ε_i^* are the atomistic van der Waals interaction hardnesses.

The *implicit solvation term* between the coarse-grained beads has been constructed from the atomistic EEF1 functional[24]

$$V_{ij}^{solv}(r) = -\int_{v_i} f_j \, d\vec{r} - \int_{v_j} f_i \, d\vec{r} \approx -f_j(r)v_i - f_i(r)v_j \quad (6)$$

where v_i is the volume and $f_i(r) = C\Delta G_i \exp(-(r/\lambda)^2)/r^2$ is the solvation free energy density of particle i, ΔG_i being the solvation free energy of the isolated atom i, λ a correlation length and $C = 1/(2\pi^{3/2}\lambda)$.[24] Both ΔG_i and v_i for each particle type are determined from experimental data.[24] The previous equation can be rewritten as

$$V_{ij}^{solv}(r) \approx -C(\Delta G_i v_j + \Delta G_j v_i)\exp(-(r/\lambda)^2)/r^2 \quad (7)$$

Assuming additivity, the solvation term affecting beads A and B is then defined as

$$V_{AB}^{solv}(r) \approx -C\sum_{i\in A}\sum_{j\in B}(\Delta G_i v_j + \Delta G_j v_i)\exp(-(r/\lambda)^2)/r^2 \quad (8)$$

We have included more information about the parameters of the atomistic force fields CHARMM19 and EEF1 in the Supporting Information.

The EEF1 implicit solvation functional assumes[24] that any "nonprotein space" is "solvent space", even if it is inside the protein. Thus, this model does not take into account that water has a finite size and cannot fit inside the core of the protein. This can be quite realistic when using an atomistic representation of the protein, but this is not a good approximation for coarse-grained representations of the system, where packing in the interior of the protein cannot be as dense. To correct this spureous effect we have modulated the implicit solvation term by including the factor γ (eq 1) in eq 8:

$$V_{AB}^{solv}(r) \approx -C(\gamma_A\sum_{i\in A}\Delta G_i\sum_{j\in B}v_j + \gamma_B\sum_{j\in B}\Delta G_j\sum_{i\in A}v_i)\exp(-(r/\lambda)^2)/r^2 \quad (9)$$

$$V_{AB}^{solv}(r) \approx -C(\Delta G_A v_B + \Delta G_B v_A)\exp(-(r/\lambda)^2)/r^2 \quad (10)$$

where $\Delta G_A = \gamma_A\sum_{i\in A}\Delta G_i$ and $v_A = \sum_{i\in A} v_i$.

Discretization of the Total Interaction Potential between Nonbonded Coarse-Grained Beads. To transform the potential described above to a discretized functional which can be inserted in the DMD algorithm, we define a well located at $R_{AB}^* = R_A^* + R_B^*$ (the minimum of the coarse-grained van der Waals potential term; see Figure 1). The well depth is computed as the sum of the two terms at distance $r = R_{AB}^*$:

$$V_{AB}(R_{AB}^*) = \omega_{vdW}V_{AB}^{vdW}(R_{AB}^*) + \omega_{solv}V_{AB}^{solv}(R_{AB}^*) \quad (11)$$

To reduce the computational cost of the simulations we approximate the nonbonded potential of mean force to a discretized potential with two energy steps, that form a potential well (or barrier) if the total potential is attractive (or repulsive). The inner and outer step distances are $0.9R_{AB}^*$ and $1.1R_{AB}^*$, respectively, while the hardcore repulsion distance was placed at $0.88R_{AB}^*$ (see Figure 1C).

Parametrization of the Force Field. We refined the parameters of the force field by analyzing the behavior in water of a single Aβ40 peptide, a 30 μM solution of Aβ40 peptides,[28] and a small folded protein (PDB id 1FAS). Our objective was

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

to find a parametrization able to represent correctly the three states (unfolded, aggregated, and folded).

We used the simulations of the protein 1FAS to adjust the hydrogen bonding strengths E_{hb} = 3 kcal/mol and E_{hb}^{core} = 4 kcal/mol as well as the parameters α = 10 and β = 0.5 used to define the factor γ (see eq 1).

The values of ω_{solv} and ω_{vdW} in eq 11 were selected to get a proper balance between aggregation and dissociation rates in simulations of a solution of Aβ40 peptides at a concentration of 30 μM.

III. RESULTS AND DISCUSSION

Force Field Calibration. The macroscopic solution was modeled by placing four Aβ40 peptides in a cubic box of the size corresponding to 30 μM concentration and with periodic boundary conditions. We observed that in this system the solvation term prompts dissociation and the van der Waals term prompts association. In order to have a good statistics of the association process we ran eight long DMD simulations for each point in the two-dimensional space (ω_{solv}, ω_{vdW}). We scanned the range of ω_{vdW} from 0 to 10 and ω_{solv} from 0 to 18 (mesh density of one unit per dimension) to build the phase diagram shown in Figure 2. Above the phase boundary line,

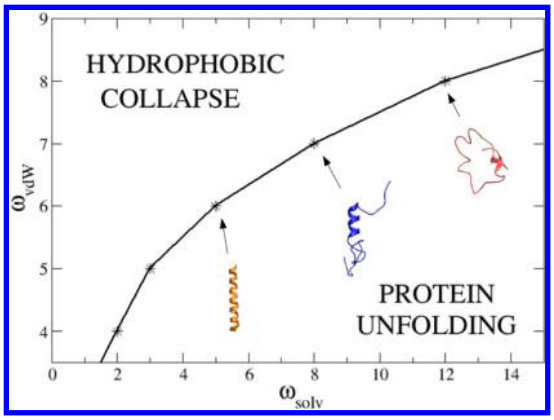


Figure 2. Phase diagram for the 30 μM concentration Aβ40 solution. Above the phase boundary line the solution precipitates. Small pictures show typical secondary structures obtained for monomeric Aβ40 in the simulations at certain points on the phase boundary line.

aggregation happens due to hydrophobic collapse, and below it, there is equilibrium between peptide associations and dissociations. This stationary regime was achieved when the trajectories reached 3 μs, but to make sure that the oligomer size distribution was stabilized we extended the simulations up to 5 μs. Equilibrium is rapidly reached in the DMD simulations due to the enhanced conformational sampling of the implicit solvent model we use, making 1 μs equivalent to 1 ms of real time (see Methods).

We chose the point ω_{solv} = 8, ω_{vdW} = 7 on the phase boundary line, that gives the secondary structure in better agreement with the conformational sampling obtained in the explicit solvent atomistic MD simulation of a single Aβ40 peptide in ref 29 (see structures in Figure 2), as well as a realistic aggregation profile in the Aβ40 solution. We refer the reader to ref 30 for a recent review about simulations of the Aβ40 peptide. We show in Figure S1 of the Supporting Information the secondary structure evolution as a function of time for an Aβ40 peptide. We started the simulation from a completely extended conformation, and rapidly an α-helix region is formed between residues 12 and 24. We show the α-helix and β-strand propensities in Figure S2. These results are in agreement with those obtained from united-atom implicit solvent simulations in refs 31 and 32 that give a higher stability of these secondary structure regions as compared with the explicit solvent simulations of ref 33.

In order to test the stability of α-helix and β-sheet motifs with the PACSAB force field, we have made simulations of an α-helix peptide (EK peptide) and of a β-sheet peptide (the Gly5-Trp29 segment of the protein with PDB code 1I6C), both starting from completely extended conformations. PACSAB folded these peptides to their native conformation, as shown in the Supporting Information (Figure S5).

We have shown in Figure 3 the evolution with time of the population of each oligomer order (computed as the average

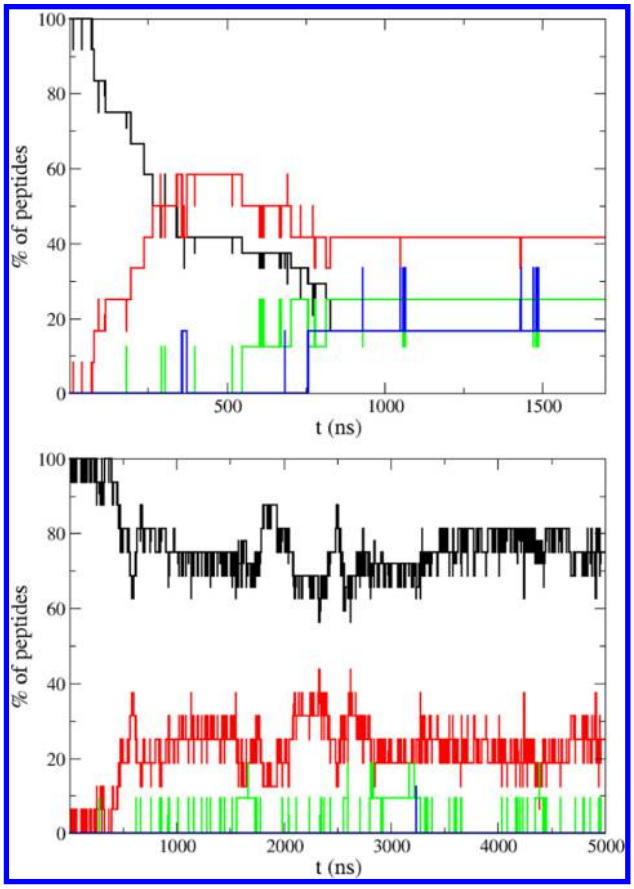


Figure 3. Evolution of the percentage of peptides in each oligomeric state during the trajectory (black line monomers; red line dimers; green line trimers; blue line tetramers): (upper panel) evolution for the point at coordinates (6,7) in the phase diagram; (lower panel) same for the point at (8,7).

over the eight simulations) for this point and for a point above the phase boundary line, where the dynamics of aggregation tends to populate higher order oligomers. We selected the protein 1FAS as a training system for fine-tuning of the (ω_{solv}, ω_{vdW}) values. However, as can be observed in Figure 4, no reoptimization was necessary, since the chosen parametrization reproduces correctly the structure of this folded protein. If a (ω_{solv}, ω_{vdW}) value below the phase boundary line in Figure 2 is chosen, the protein unfolds due to the underestimated hydrophobicity with such parametrization.

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

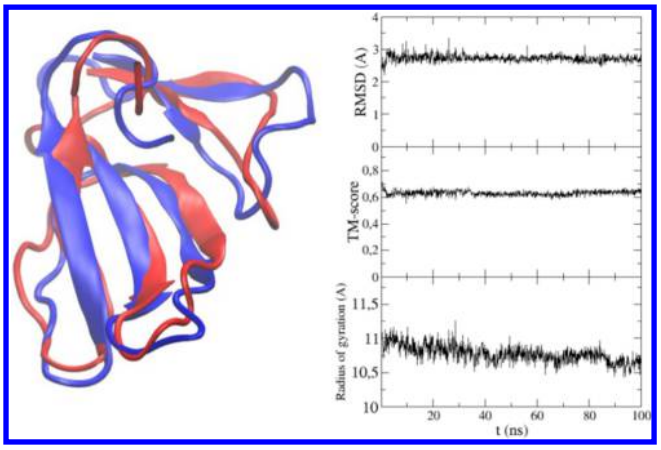


Figure 4. Structure of the protein with PDB id 1FAS after a DMD simulation of 100 ns (red cartoon), compared with the crystallographic structure (blue). Also shown are the RMSD, TM-score, and radius of gyration.

Amyloid Aggregation Dynamics. For each point in the phase diagram of Figure 2 we started the simulations from completely extended conformations of the peptides. We observed that at the beginning of the simulation, the peptides experience a fast collapse that drive their structure to a fold intermediate between a helical structure and a molten globule (see Figure 5), in good agreement with previous explicit solvent atomistic MD simulations of Aβ40.[29]

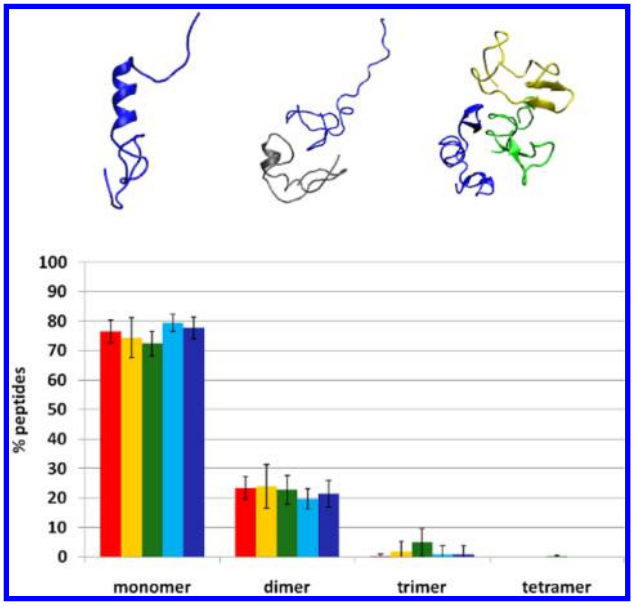


Figure 5. Oligomerization in the 30 μM Aβ40 peptide solution. (upper panel) Structures of a monomer, a dimer, and a trimer obtained during the simulations. (lower panel) Percentage of peptides in each oligomeric state observed at different DMD simulation times using the optimal force field parametrization (see main text): 1 μs (red), 2 μs (yellow), 3 μs (green), 4 μs (blue), and 5 μs (dark blue).

As simulation progresses, intermolecular collisions happen, some of them leading to peptide association. At 30 μM concentration peptides collide every 0.1 μs on average, but only ~10% of these collisions are productive (leading to the formation of a stable dimer). The low frequency of association/dissociation events requires an extensive sampling than cannot be achieved by standard explicit solvent atomistic MD simulations, but that was accessible with our implicit solvent DMD simulations (note in Figure 5 that a stationary regime had been reached within the simulation window). The association of monomers with dimers led to the formation of trimers, much less abundant due to the low population of dimers. For the same reason the existence of tetramers was residual. Our oligomer size distribution is coincident with the experimental distributions observed in a very recent work by Pujol-Pina et al.[34] (see Figures S3 and S4 of the Supporting Information)

We made simulations at higher concentrations, finding that higher order oligomers become more abundant as the concentration increases. Figure 6 shows the oligomer size distribution

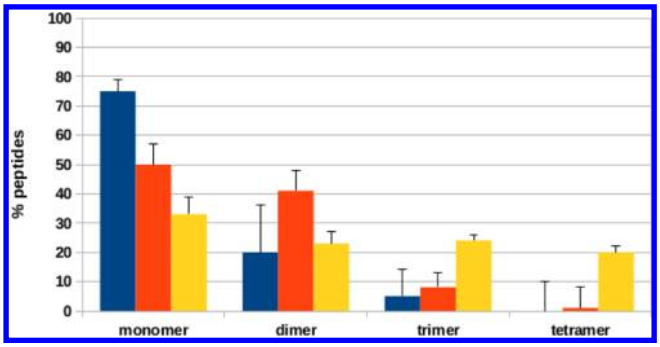


Figure 6. Percentage of peptides in each oligomeric state at different concentrations: 50 μM (blue), 100 μM (orange), and 240 μM (yellow).

obtained after 1 μs DMD simulations at different concentrations. Eight simulations were performed for each concentration. The oligomer size distribution at 50 μM fits well with the distribution at 30 μM (see Figure 5), but at 100 μM it has changed clearly with an evident increase of dimers. At 240 μM similar populations are found for monomers, dimers, trimers, and tetramers. This tendency is consistent with the results of very recent atomistic molecular dynamics simulations for solutions of β-amyloid peptides at very high concentration.[35]

Test Systems. In order to evaluate the quality and universality of the force field, we performed a comprehensive evaluation for folded proteins, intrinsically disordered proteins and protein−protein complexes.

Folded Proteins. We explored the ability of the coarse grained force field to reproduce the structure of folded proteins in long simulation time scales. For this purpose we selected a set of 25 proteins representative of the most prevalent protein folds[36] and performed DMD simulations of 500 ns (this gives, as explained above, a sampling corresponding to multimicrosecond trajectories in explicit solvent atomistic MD). Results in Figure 7 show that all the trajectories are stable, without any evident signal of unfolding as illustrated in the evolution of the radii of gyration. The RMS deviations from experimental structure are typically in the range 2−8 Å, higher than those found in atomistic MD simulations,[36−38] but matching the level of accuracy of state-of-the-art CG force fields designed specifically to reproduce the folded state of proteins[5,7,12] (see comparison with other coarse-grained models in Table S2 of the Supporting Information)

As demonstrated by the TM-score value,[36] the flexible loops are the main origin of the deviation of DMD samplings from experimental structures, while the protein core is fully

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

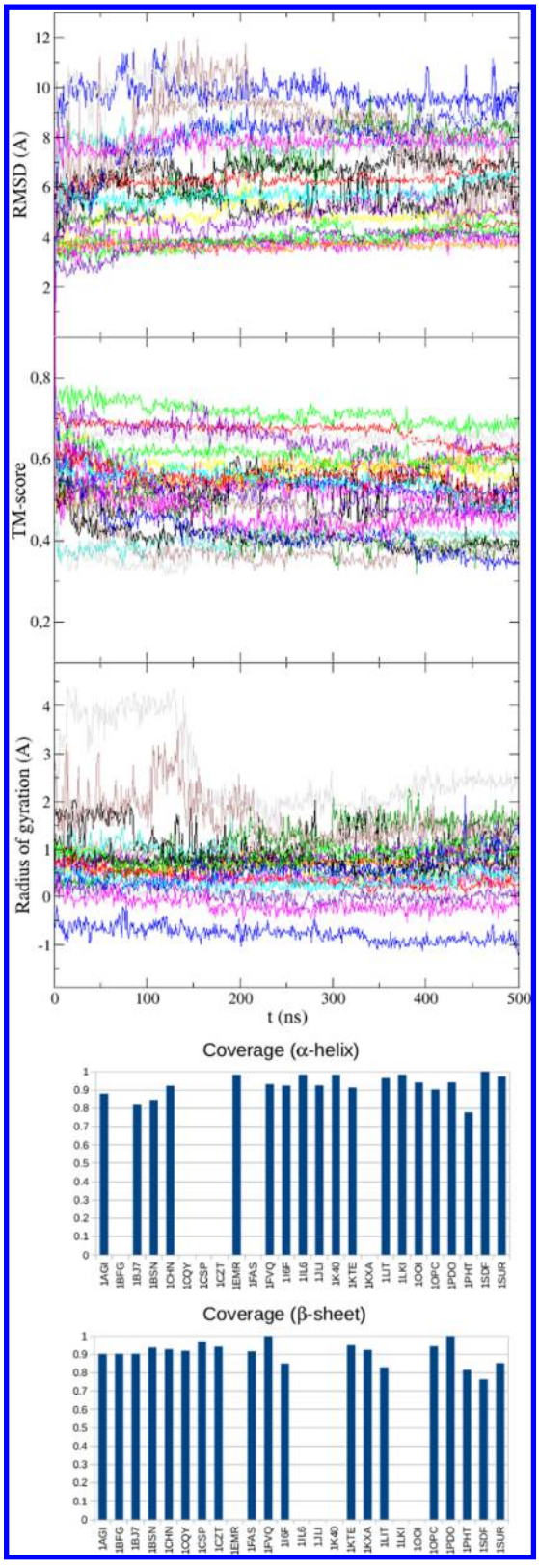


Figure 7. Structural properties of the folded protein benchmark after a DMD simulation of 500 ns. (top to bottom) RMSD respect to the native conformation, TM-score, change in the radius of gyration during the trajectory, conserved α-helix, and conserved β-sheet. The value of RMSD, TM-score, and radius of gyration is plotted in a different color for each protein. Proteins without α-helices are not shown in the α-helix coverage plot, and the same for β-sheets.

preserved during the simulations. We found conservation of native secondary structure (Figure 7), which is remarkable considering that our force field does not introduce, like others, specific restraints or backbone dihedral terms favoring the stability of usual secondary structure elements. Despite this lack of restrictions in the backbone dihedrals, the Ramachandran plots are well reproduced, as shown in Figure S6 of the Supporting Information. Finally, fold recognition algorithms detected in all the cases the real structure (or that of a very close homologue) from the sampled structure at the end of the DMD simulation (see Figures S7 and S8 of the Supporting Information)

In summary, despite the lack of specific training for folded proteins, the absence of restrictions on secondary structure, and the lack of structure potentials biasing the simulations toward the native state, our extremely simple CG model is able to sample properly the structure of folded proteins in very long simulations.

Intrinsically Disordered Proteins. To test the generality and universality of the force field we also explored the dynamics of two intrinsically disordered proteins (IDPs): ACTR and α-synuclein. ACTR is an IDP that folds in a well-defined structure only in the presence of its macromolecular partner,[39] while in its absence appears as a random coil with a residual percentage of α-helix. DMD simulations recognized the IDP nature of ACTR, sampling a wide variety of conformational states in a 1 μs DMD trajectory (see Figure 8). The only common feature between the conformations of the ensemble is the formation of residual secondary structure, in good agreement with circular dichroism measurements.[40]

Similar success is obtained with α-synuclein, which is stable when embedded in a lipid environment, while it is disordered in water,[41] except for some residual contacts between the residues around position 50 and the residues around position 120 in the sequence. As shown in Figure 8, the model is able to recognize the protein as an IDP, with a very low percentage of secondary structure, and no distinct contacts others than the robust interaction between residues ~50 and ~120, in good agreement again with experimental information.[41]

Protein−Protein Complexes. Finally, we tested the ability of our simulation procedure to recognize experimental structures of protein−protein complexes. We used here as test set the strong binding complexes of the Weng's protein−protein docking benchmark 4.0.[42] Following the standard criterion we considered as strong binding cases those complexes with a binding free energy $\Delta G < -10$ kcal/mol[43] (the complexes of the test set are listed in Table S3 of the Supporting Information). We evaluated the ability of the force field to distinguish experimental complexes from false positive structures generated by protein−protein docking calculations. We want to stress that, instead of refining protein docking poses,[44] we just want to use the PACSAB simulations as a filter to discard nonnative docking poses. Thus, for each complex in the test set we performed a 1 ns DMD simulation of the experimental structure and the best scored false positive docking pose generated in a previous study.[44] We found that dissociation happened in the first few picoseconds of the trajectory, so 1 ns simulations were long enough to filter the docking poses, that had been scored using pyDock,[45] a state-of-the-art scoring funtion for protein−protein docking.[46]

In average 85% of experimental structures remained stable during the simulations (see Figure 9A), while 80% of the false positives deviated significantly from its initial conformation,

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

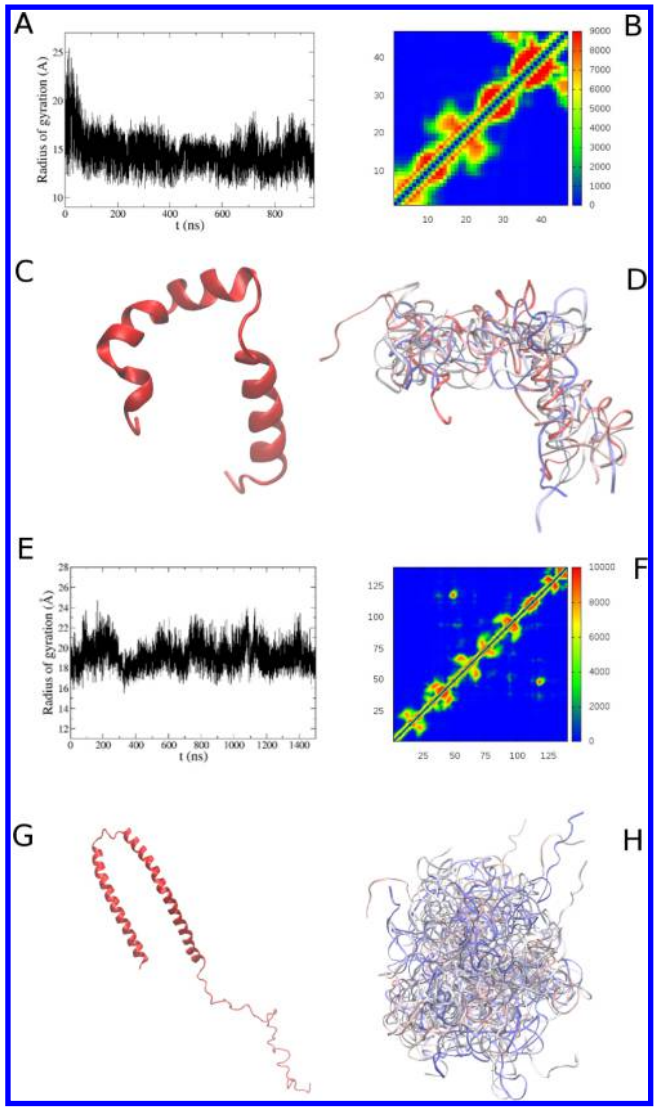


Figure 8. Structural data on ACTR (A–D) and α-synuclein (E–H). (A, E) Radius of gyration during the trajectory. (B, F) Contact map. The axes are the residue index. The color scale is from blue to red (arbitrary units). Contacts between consecutive residues have been neglected. (C) Structure of ACTR when bound to its macromolecular partner (starting structure for the DMD simulations). (G) Structure of α-synuclein when embedded in a lipid environment. (D, H) Ensemble of structures obtained during the simulation. A time-increasing colorscale (from red to blue) has been used here.

many of them leading to a complete disruption of the ligand–receptor complex (Figure 9C). In 90% of the complexes the RMS deviation from the starting structure is higher for the best scored false positive than for the experimental structure (see histogram in Figure 9B). Therefore, we have found that despite the lack of specific parametrization or the use of statistical potentials our simple DMD-based method is able not only to maintain the geometry of experimental protein–protein complexes, but to identify incorrect structures, even those that are given a strongly attractive interaction energy in docking calculations. The ability of the method to keep stable experimental complex structures while producing dissociation of nonbinding ligand–receptor orientations suggests us that the method could give good results in cross-docking[47] of proteins for which experimental information about possible binding is not available.

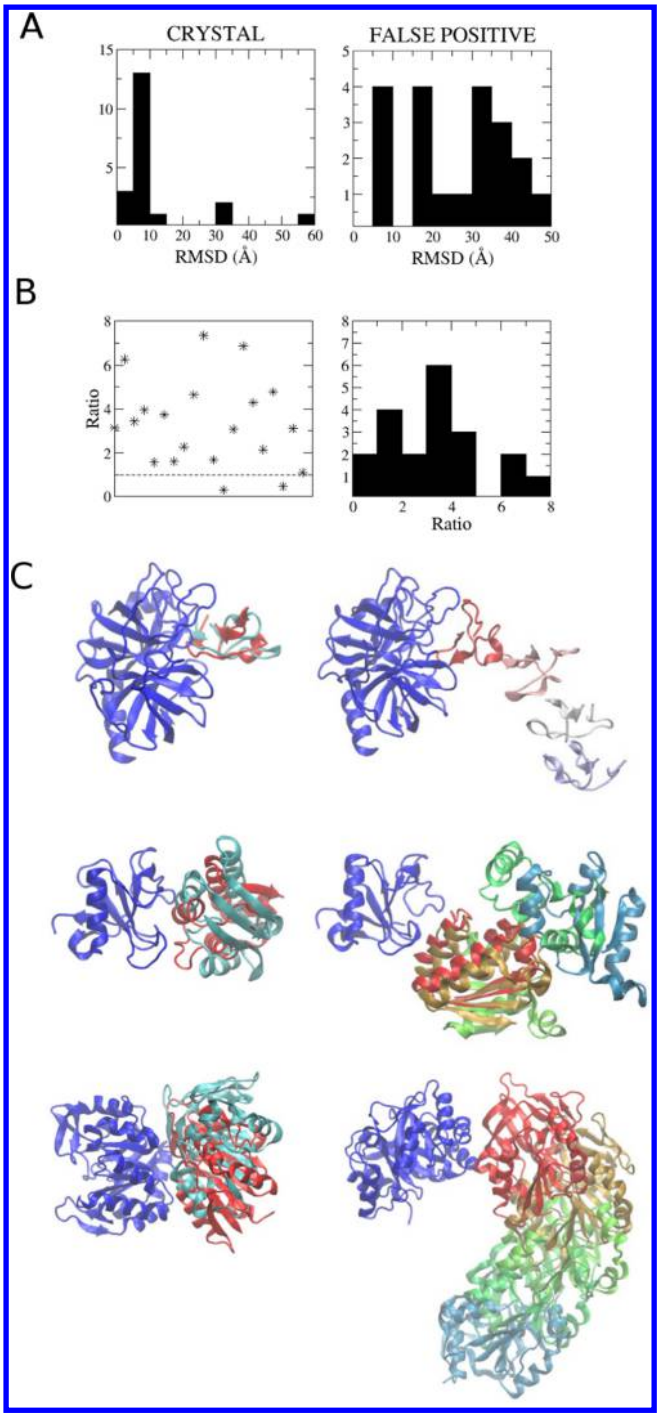


Figure 9. (A) Histogram of $RMSD_{exp}$, the RMSD with respect to the initial structure for the experimental complex (left), and histogram of $RMSD_{fp}$, the RMSD with respect to the initial structure for the best scored false positive (right). The RMSDs are calculated after a DMD simulation of 1 ns. (B) Values of the ratio $RMSD_{fp}/RMSD_{exp}$. At the left figure, symbols above the dashed line correspond to complexes for which $RMSD_{fp} > RMSD_{exp}$. At right is shown the corresponding histogram. (C) Structure of the crystal (left) and best scored false positive docking pose (right) for the complexes 1PPE, 1AY7, and 1GPW (from top to bottom). The receptor is colored in dark blue and the ligand in red. At left is shown position of the ligand (cyan) after a simulation of 1 ns; at right is shown the movement of the ligand at the beginning of the trajectory (in the frame of reference of the receptor). Several snapshots in a time-increasing colorscale (from red to blue) are shown for the ligand.

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929–5938

IV. CONCLUSIONS

We have constructed a physics-based discretized coarse-grained force field to represent the conformational space of proteins in solution, but also aggregated and complexed with other proteins. The force-field is implemented in a highly efficient discrete molecular dynamics algorithm which allowed us inexpensive simulations in huge systems, which would be inaccessible to standard atomistic molecular dynamics simulations. Exhaustive testing of the method shows that it is able to reproduce correctly the stability of both structured and intrinsically disordered proteins, to reproduce properly aggregation of β-amyloid peptides, and to recognize the correct structure of protein−protein complexes when compared with alternative ligand−receptor orientations which where highly scored by state-of-the-art protein−protein docking algorithms. To our knowledge, this is the first coarse-grained model able to represent both the conformational variability and interactions of proteins, including association, dissociation, and aggregation.

APPENDIX A

The index of packing n of particle i, used in the calculation of the factor γ

$$\gamma_i(n) = \frac{1}{1 + \exp((n - \alpha)/\beta)}$$

is computed as the number of faces in a truncated cube centered on particle i such that its center is near to any other particle j. The maximum value is $n = 14$, the total number of faces. $n = 14$ would correspond to a completely buried particle, $n = 0$ to a completely isolated particle. We have fitted $\alpha = 10$, the n value at which γ changes from the exposed particle ($\gamma \approx 1$) to the buried paticle ($\gamma \approx 0$) value (see Figure 10)

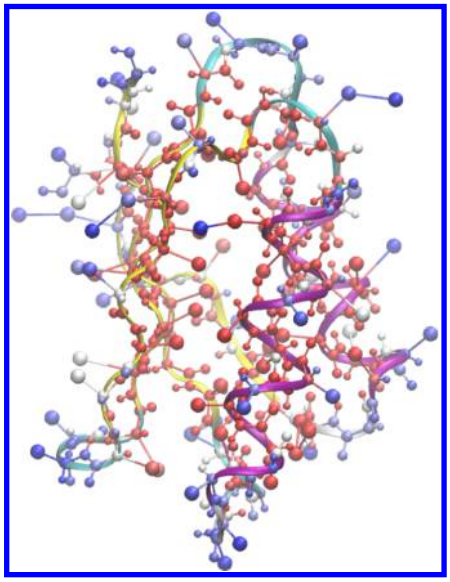


Figure 10. Structure of protein 1FVQ where particles are given a color scaled according to the value of γ (blue: exposed; red: buried).

APPENDIX B

We assume that van der Waals term of the interaction between beads A and B at the distance $r = R_{AB}$, such that V_{AB}^{vdW} has its energy minimum, is equal to the sum of atomistic van der Waals interactions at $r = R_{AB}$. The atomistic van der Waals interaction between atoms i and j is

$$V_{ij}^{at}(r) = \varepsilon_{ij}^*\left[\left(\frac{R_{ij}^*}{r}\right)^{12} - 2\left(\frac{R_{ij}^*}{r}\right)^{6}\right]$$

being r the distance between atom i and atom j. $\varepsilon_{ij}^* = \sqrt{\varepsilon_i^*\varepsilon_j^*}$ and $R_{ij}^* = R_i^* + R_j^*$. Supposing that all the atoms have the same van der Waals radii R_0^*, $R_{ij}^* \approx 2R_0^*$. Thus

$$R_A^* = \rho(\sum_i R_i^{*3})^{1/3} \approx \rho(N_A R_0^{*3})^{1/3}$$

and

$$\begin{aligned} R_{AB}^* &= R_A^* + R_B^* \\ &\approx \rho(N_A^{1/3} + N_B^{1/3})R_0^* \\ &= \rho(N_A^{1/3} + N_B^{1/3})R_{ij}^*/2 \end{aligned}$$

where N_A (N_B) is the number of atoms included by bead A (B).

Therefore, the value of the atomistic van der Waals interacion between atoms i and j at the distance R_{AB}^* is

$$\begin{aligned} V_{ij}^{at}(R_{AB}^*) &= \varepsilon_{ij}^*\left[\left(\frac{R_{ij}^*}{R_{AB}^*}\right)^{12} - 2\left(\frac{R_{ij}^*}{R_{AB}^*}\right)^{6}\right] \\ &= \varepsilon_{ij}^*\left[\left(\frac{2/\rho}{N_A^{1/3} + N_B^{1/3}}\right)^{12} - 2\left(\frac{2/\rho}{N_A^{1/3} + N_B^{1/3}}\right)^{6}\right] \end{aligned}$$

The value of the van der Waals term of the coarse-grained potential at the distance R_{AB}^* is the sum of the terms corresponding to the interactions between all the atoms included in bead A and bead B:

$$V(R_{AB}^*) = \sum_{i\in A}\sum_{j\in B} V_{ij}^{at}(R_{AB}^*) = \sum_{i\in A}\sum_{j\in B} \varepsilon_{ij}^*(x^{12} - 2x^6)$$

where we have defined $x = 2/[(N_A^{1/3} + N_B^{1/3})\rho]$.

Taking into account $\varepsilon_{ij}^* = \sqrt{\varepsilon_i^*\varepsilon_j^*}$ and assuming $\sum_{i\in A}\sum_{j\in B}\sqrt{\varepsilon_i^*\varepsilon_j^*} \approx \sqrt{(\sum_{i\in A}\varepsilon_i^*)(\sum_{j\in B}\varepsilon_j^*)}$ one finally obtains

$$V_{AB}^{vdW}(R_{AB}^*) = -\varepsilon_{AB}^* = \sqrt{(\sum_{i\in A}\varepsilon_i^*)(\sum_{j\in B}\varepsilon_j^*)}\,(x^{12} - 2x^6)$$

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00660.

Details on the atomistic force fields used to constrcut the PACSAB force field; tables describing the protein−protein complexes used, simulation speed for several systems and the backbone RMSD after PACSAB simulation for proteins tested with other coarse-grained models; figures for the secondary structure changes during a trajectory for Aβ40, secondary structure propensities for monomeric and oligomeric Aβ40, intramolecular and intermolecular contact maps for Aβ40, folding trajectories of EK peptide and a β-sheet

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

peptide, Ramachandran plots of the structures after the PACSAB simulations, and contact maps and structures of the folded proteins in the benchmark after the simulations, compared to the data obtained from experimental structures (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: agusti.emperador@irbbarcelona.org (A.E.).
*E-mail: modesto.orozco@irbbarcelona.org (M.O.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Natalia Carulla for enlightening comments about the aggregation process of β-amyloid peptides and Juan Fernandez-Recio for useful comments and information on binding energies of protein complexes. We thank the Spanish Ministry of Science (Grants BIO2012-32868), the Instituto de Salud Carlos III (INB), and the European Research Council (ERC-simDNA) for support. M.O. is an ICREA Academia Fellow.

REFERENCES

(1) Morriss-Andrews, A.; Shea, J. E. Computational studies of protein aggregation: methods and applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 643−666.

(2) Orozco, M. A theoretical view of protein dynamics. *Chem. Soc. Rev.* **2014**, *43*, 5051−5066.

(3) Kleinjung, J.; Fraternali, F. Design and application of implicit solvent models in biomolecular simulations. *Curr. Opin. Struct. Biol.* **2014**, *25*, 126−134.

(4) Saunders, M. G.; Voth, G. A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.

(5) Pasi, M.; Lavery, R.; Ceres, N. PaLaCe: a coarse-grain protein model for studying mechanical properties. *J. Chem. Theory Comput.* **2013**, *9*, 785−793.

(6) Hills, R. D., Jr.; Lu, L.; Voth, G. A. Multiscale coarse-graining of the protein energy landscape. *PLoS Comput. Biol.* **2010**, *6*, e1000827.

(7) Kar, P.; Gopal, S. M.; Cheng, Y. M.; Predeus, A.; Feig, M. RIMO: A Transferable Coarse-grained Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3769−3788.

(8) Kapoor, A.; Travesset, A. Folding and stability of helical bundle proteins from coarse-grained models. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1200−1211.

(9) Bereau, T.; Deserno, M. Generic doarse-grained model for protein folding and aggregation. *J. Chem. Phys.* **2009**, *130*, 235106.

(10) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Oldziej, S.; Wachucik, K.; Scheraga, H. A. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B* **2007**, *111*, 260−285.

(11) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J.-F.; Kalimeri, M.; Barducci, A.; Laurin, Y.; Tek, A.; Baaden, M.; Nguyen, P.-H.; Derreumaux, P. The OPEP protein model: from single molecule-s,amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **2014**, *43*, 4871−4893.

(12) Chebaro, Y.; Pasquali, S.; Derreumaux, P. The coase-grained OPEP force field for non-amyloid and amyloid proteins. *J. Phys. Chem. B* **2012**, *116*, 8741−8752.

(13) Auer, S. Phase diagram of polypeptide chains. *J. Chem. Phys.* **2011**, *135*, 175103.

(14) Urbanc, B.; Cruz, L.; Yun, S.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. In silico study of amyloid beta-protein folding and oligomerization. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17345−17350.

(15) Orozco, M.; Orellana, L.; Hospital, A.; Naganathan, A. N.; Emperador, A.; Carrillo, O.; Gelpi, J. L. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. *Adv. Protein Chem. Struct. Biol.* **2011**, *85*, 183−215.

(16) Emperador, A.; Meyer, T.; Orozco, M. Protein flexibility from discrete molecular dynamics simulations using quasi-physical potentials. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 83−94.

(17) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.

(18) 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* **2007**, *111*, 7812−7824.

(19) 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.* **2008**, *4*, 819−834.

(20) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. *Biophys. J.* **2015**, *108*, 1153−1164.

(21) Emperador, A.; Meyer, T.; Orozco, M. United-Atom Discrete Molecular Dynamics of Proteins Using Physics-Based Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 2001−2010.

(22) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. Ab initio folding of proteins with all-atom discrete molecular dynamics. *Structure* **2008**, *16*, 1010−1018.

(23) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.

(24) Lazaridis, T.; Karplus, M. Effective energy function for proteins in solution. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133−152.

(25) Moore, T. C.; Iacovella, C. R.; McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **2014**, *140*, 224104.

(26) Miyazawa, S.; Jernigan, R. L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **1996**, *256*, 623−644.

(27) Zhang, C.; Vasmatzis, G.; Cornette, J. L.; DeLisi, C. Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* **1997**, *267*, 707−26.

(28) Bitan, G.; Vollers, S. S.; Teplow, D. B. Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. *J. Biol. Chem.* **2003**, *278*, 34882−34889.

(29) Olubiyi, O. O.; Strodel, B. Structures of the amyloid β-peptides Aβ1−40 and Aβ1−42 as influenced by pH and a D-peptide. *J. Phys. Chem. B* **2012**, *116*, 3280−3291.

(30) Nasica-Labouze, J.; Nguyen, H.; Sterpone, F.; Berthomieu, O.; Buchete, N.-V.; Cote, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; Laio, A.; Li, M. S.; Melchionna, S.; Mousseau, N.; Mu, Y.; Paravastu, A.; Pasquali, S.; Rosenman, D. J.; Strodel, B.; Tarus, B.; Viles, J. H.; Zhang, T.; Wang, Ch.; Derreumaux, P. Amyloid β protein and Alzheimer's disease: when computer simulations complement experimental studies. *Chem. Rev.* **2015**, *115*, 3518−3563.

(31) Takeda, T.; Klimov, K. Probing the effect of amino-terminal truncation for Aβ40 peptides. *J. Phys. Chem. B* **2009**, *113*, 6692−6702.

(32) Kim, S.; Takeda, T.; Klimov, K. Mapping conformational ensembles of Aβ oligomers in molecyular dynamics simulations. *Biophys. J.* **2010**, *99*, 1949−1958.

(33) Lin, Y.-S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S. Investigating how peptide length and a pathogenic mutation modify the structural ensemble of amyloid beta monomer. *Biophys. J.* **2012**, *102*, 315−324.

(34) Pujol-Pina, R.; Vilaprinyo-Pascual, S.; Mazzucato, R.; Arcellla, A.; Vilaseca, M.; Orozco, M.; Carulla, N. *Sci. Rep.* **2015**, *5*, 14809.

(35) Barz, B.; Olubiyi, O. O.; Strodel, B. Early amyloid β-protein aggregation precedes conformational change. *Chem. Commun.* **2014**, *50*, 5373−5375.

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938

(36) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. A consensus view of protein dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 796−801.
(37) Meyer, T.; D'Abramo, M.; Hospital, A.; Rueda, M.; Ferrer-Costa, C.; Perez, A.; Carrillo, O.; Camps, J.; Fenollosa, C.; Repchevsky, D.; Gelpi, J. L.; Orozco, M. MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories. *Structure* **2010**, *18*, 1399−1409.
(38) Candotti, M.; Perez, A.; Ferrer-Costa, C.; Rueda, M.; Meyer, T.; Gelpi, J. L.; Orozco, M. Exploring early stages of the chemical unfolding of proteins at the proteome scale. *PLoS Comput. Biol.* **2013**, *9*, e1003393.
(39) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **2002**, *415*, 549−553.
(40) Kjaergaard, M.; Norholm, A. B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein Sci.* **2010**, *19*, 1555−1564.
(41) Esteban-Martin, S.; Silvestre-Ryan, J.; Bertoncini, C. W.; Salvatella, X. Identification of fibril-like tertiary contacts in soluble monomeric α-synuclein. *Biophys. J.* **2013**, *105*, 1192−1198.
(42) Hwang, H.; Vreven, T.; Janin, J.; Weng, Z. Protein-protein docking benchmark version 4.0. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3111−3114.
(43) Kastritis, P. L.; Moal, I. H.; Hwang, H.; Weng, Z.; Bates, P. A.; Bonvin, A. M.; Janin, J. A structure-based benchmark for protein-protein binding affinity. *Protein Sci.* **2011**, *20*, 482−491.
(44) Emperador, A.; Solernou, A.; Sfriso, P.; Pons, C.; Gelpi, J. L.; Fernandez-Recio, J.; Orozco, M. Efficient relaxation of protein-protein interfaces by discrete molecular dynamics. *J. Chem. Theory Comput.* **2013**, *9*, 1222−1229.
(45) Cheng, T. M.; Blundell, T. L.; Fernandez-Recio, J. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins: Struct., Funct., Genet.* **2007**, *68*, 503−515.
(46) Fernandez-Recio, J.; Totrov, M.; Abagyan, R. Soft protein-protein docking in internal coordinates. *Protein Sci.* **2002**, *11*, 280−291.
(47) Sacquin-Mora, S.; Carbone, A.; Lavery, R. Identification of protein interaction partners and protein-protein interaction sites. *J. Mol. Biol.* **2008**, *382*, 1276−1289.

DOI: 10.1021/acs.jctc.5b00660
J. Chem. Theory Comput. 2015, 11, 5929−5938