THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Soluble State of Villin Headpiece Protein as a Tool in the Assessment of MD Force Fields

Published as part of The Journal of Physical Chemistry virtual special issue "Carol K. Hall Festschrift".

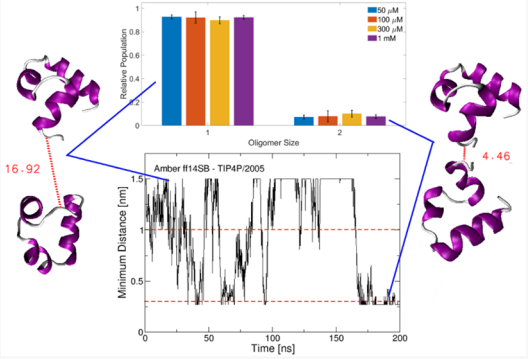
Brian Andrews, Kaho Long, and Brigita Urbanc*

Cite This: *J. Phys. Chem. B* 2021, 125, 6897–6911

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Protein self-assembly plays an important role in cellular processes. Whereas molecular dynamics (MD) represents a powerful tool in studying assembly mechanisms, its predictions depend on the accuracy of underlying force fields, which are known to overly promote protein assembly. We here examine villin headpiece domain, HP36, which remains soluble at concentrations amenable to MD studies. The experimental characterization of soluble HP36 at concentrations of 0.05 to 1 mM reveals concentration-independent 90% monomeric and 10% dimeric populations. Extensive all-atom MD simulations at two protein concentrations, 0.9 and 8.5 mM, probe the HP36 dimer population, stability, and kinetics of dimer formation within two MD force fields, Amber ff14SB and CHARMM36m. MD results demonstrate that whereas CHARMM36m captures experimental HP36 monomer populations at the lower concentration, both force fields overly promote HP36 association at the higher concentration. Moreover, contacts stabilizing HP36 dimers are force-field-dependent. CHARMM36m produces consistently higher HP36 monomer populations, lower association rates, and weaker dependence of these quantities on the protein concentration than Amber ff14SB. Nonetheless, the highest monomer populations and dissociation constants are observed when the TIP3P water model in Amber ff14SB is replaced by TIP4P/2005, showcasing the critical role of the water model in addressing the protein solubility problem in MD.



INTRODUCTION

Protein misfolding and aberrant aggregation are known to underlie many systemic and neurodegenerative diseases as prevalent as Alzheimer's, Parkinson's, and type-II diabetes,[1] for which there is currently no cure. Understanding protein aggregation mechanisms and the characterization of self-assembly pathways would allow for therapeutic approaches that could prevent aggregation or redirect the aggregation process from the pathogenic pathway.[2] Molecular dynamics (MD) simulations represent a powerful tool in discerning protein self-assembly mechanisms.[3] Despite many advances in MD simulations of protein aggregation,[4] the accuracy of the commonly used MD force fields with respect to reproducing the experimentally observed characteristics of protein aggregation has been placed under scrutiny.[5,6] Petrov and Zagrovic reported that current MD force fields underestimate protein solubility, leading to protein self-assembly even for proteins that are soluble *in vitro* at high concentrations.[5] Similarly, recent studies have showed that MD force fields do not sufficiently solvate proteins and overly promote their self-assembly.[7,8] Best et al. addressed this problem by enhancing protein−water interactions without altering water−water or protein−protein interactions in variants of the Amber ff03 force field.[7] Following the same protocol, Nawrocki et al. adjusted CHARMM36[9] by increasing the strength of protein−water interactions.[8] Similarly, in one of the latest parametrizations of the CHARMM force field, CHARMM36m, Huang et al. modified the TIP3P water model, which makes the dispersion part of protein−water interactions more favorable.[10] Zhang et al. examined the self-assembly of GAG in water/ethanol mixtures and concluded that of the three MD force fields under study, Amber ff14SB,[11] OPLS-AA/M,[12] and CHARMM36m,[10] only CHARMM36m reproduced the available experimental data, whereas OPLS-AA/M showed large discrepancies, and the Amber ff14SB results were inconclusive.[13]

One of the major obstacles in the development of an accurate MD force field (and the water model) is a lack of experimental constraints on protein self-assembly under conditions that are accessible to MD studies.[14] First, only

Received: May 25, 2021
Published: June 18, 2021

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

relatively high protein concentrations above 1 mM are accessible to explicit-solvent MD studies of protein self-assembly pathways, with the initial conditions corresponding to multiple spatially separated unstructured protein molecules. Experimental data at such high protein concentrations are not available because most proteins are not soluble at such high concentrations. Second, there is a large gap between time scales that can be sampled by MD simulations (even if the multiscale MD or enhanced sampling techniques[15] are applied) and time scales of *in vitro* protein aggregation into amyloid fibrils, which takes a significantly longer time and can span hours or even days. For this reason, it is advantageous to focus on the soluble state of the protein, specifically on oligomer formation, which is known to occur on a subsecond time scale for proteins at micromolar concentrations[16−18] and proteins that remain soluble at high protein concentrations amenable to MD studies.

We here examine the 36-residue-long chicken villin headpiece domain protein, HP36, which is known to remain soluble up to protein concentrations of 1 to 2 mM and above.[19−21] The native fold of HP36 is well characterized by two short and one longer α-helix, and this fold is preserved even when this domain (without the first amino acid residue) is embedded into the 76-residue-long villin headpiece protein.[22,23] HP36 folds on time scales of ∼1 μs[22,24] and is expected to be predominantly monomeric, although it may form assemblies at higher concentrations (>10 mM).[19,22,25] We characterize the soluble state of HP36 and its dependence on the protein concentration within the range of 0.05 to 1 mM by the application of photoinduced cross-linking of unmodified proteins (PICUP),[26] followed by sodium dodecyl sulfate gel electrophoresis (SDS-PAGE). PICUP/SDS-PAGE was previously applied to characterize soluble oligomers formed by amyloid β-protein (Aβ)[16,27] and, more recently, insulin.[18] We then use these experimental data to calibrate implicit solvent parameters of the four-bead protein model[28,29] with residue-specific interactions and perform discrete (discontinuous) molecular dynamics (DMD)[30−32] simulations within the DMD4B-HYDRA force field,[33] which was previously used to examine the oligomer formation of several variants of Aβ[34−37] and stefin B.[38] We perform extensive explicit-solvent MD simulations probing the HP36 dimer stability and kinetics of dimer formation at two different proteins concentrations, starting from weakly bound HP36 dimers and from two randomly placed HP36 proteins with or without the native fold. Our study evaluates two MD force fields, Amber ff14SB and CHARMM36m, with respect to their ability to capture the soluble state of HP36.

METHODS

Experimental Techniques. *Preparation of HP36 Solution.* The HP36 peptide in a dry powder form (Genscript, Piscataway, NJ) was dissolved in a 10 mM sodium phosphate buffer at neutral pH (7.4). The sample was incubated at 37 °C for 10 s and vortexed for 10 s to ensure that the protein was fully dissolved in the buffer. The protein concentration was determined by measuring absorbance at 280 nm, accounting for the molar extinction coefficient of 5500 M^{-1} cm^{-1}.

Photoinduced Cross-Linking of Unmodified Proteins. PICUP was applied to HP36 solution, following the published protocol used for Aβ and insulin.[17,18,39] In brief, stock solutions of each 1 mM tris(2,2-bipyridyl)dichlororuthenium-(II) hexahydrate (Sigma-Aldrich, St. Louis, MO) and 20 mM ammonium persulfate (APS) (Sigma-Aldrich) were prepared in a 10 mM sodium phosphate buffer. One μL of each stock solution was added to 18 μL of solubilized HP36 solution. The mixture was thoroughly vortexed and inserted into a macro camera bellows, which was placed a distance of 10 cm away from an SLR K1000 Pentamax camera. The PICUP reaction was initiated by irradiating the mixture with light from a XGY-II (B) cold halogen light source for 1 s, whereby the camera shutter was used to control the time of irradiation. After irradiation, the mixture was transferred to an Eppendorf tube with 10 μL of tricine SDS sample buffer solution containing 0.5 mM ethylenediaminetetraacetic acid (EDTA) to quench the cross-linking reaction. The process was repeated at several incubation times to monitor the time evolution of the oligomer size distribution.

SDS-PAGE and Silver Staining. The samples were heated to 95 °C for 4 min, followed by centrifugation at 16 000 rpm at 4 °C for 30 s. Five μL of the resulting samples was inserted into the wells of a Novex 10−20% tricine 1.0 mm × 12 well gel, whereby one of the lanes was reserved for a Mark 12 protein ladder. The gel was run in the XCell SureLock Mini-Cell using a PowerEase 500 power supply at 90 V for an average time of 160 min. Silver staining of the gel was performed with a Silverquest staining kit. The gel was initially rinsed with Milli-Q water and then fixed with a 10% acetic acid, 40% ethanol, and 50% water solution for 60 min. The gel was then rinsed with a 30% ethanol solution for 10 min, incubated in a 10% sensitizer solution for 10 min, washed again with a 30% ethanol solution for 10 min, stained with a 1% staining solution for 15 min, washed with Milli-Q water for 10 s, and developed with a 10% developer solution until a clear image emerged on the gel. At that point, 10 mL of a stopper was directly added to the gel and allowed to act for 10 min. Finally, the gel was washed with Milli-Q water for an additional 10 min, followed by scanning at 48-bit color with a resolution of 800 dpi, whereupon gel densitometry was performed using ImageJ.[40]

DMD4B-HYDRA Simulations. *Discrete Molecular Dynamics.* DMD can be utilized to avoid the time-consuming numerical integration of Newton's equations required in MD with continuous interparticle potentials. The DMD4B-HYDRA force field[33,34] is based on a four-bead protein model,[29] in which each amino acid is represented by four beads (except for three-bead glycine). The beads represent the amide (N), the α carbon (C_α), the carbonyl (C′), and the side-chain β carbon (C_β) groups. The energy of a hydrogen bond (HB), E_{HB}, is used as the unit of energy. The temperature is then expressed in units of E_{HB}/k_B. Backbone HBs can be formed between the amide N of an amino acid and the carbonyl C of another amino acid. Amino acid residues are given residue-specific hydropathic properties that facilitate effective hydrophobic/hydrophilic interactions between C_β beads. These hydropathic properties are based on the phenomenological hydropathy scale by Kyte and Doolittle.[41] The implicit solvent parameter that controls the strength of hydropathic interactions, E_{HP}, was set to E_{HP} = 0.3, as in our previous studies.[34,35,38,42,43] The strength of effective electrostatic interactions, E_{CH}, was used as the force-field calibration parameter.

DMD4B-HYDRA Simulation Protocol. A cubic simulation box of edge length L was populated by 32 HP36 molecules arranged into a cubic box followed by DMD simulations at a high temperature (T = 4.0), during which all interactions were turned off and conformations at subsequent simulation times were recorded to obtain 32 distinct initial conformations, each

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

containing 32 unstructured monomeric HP36 molecules. Protein concentrations of 0.1, 0.3, and 3 mM were obtained by setting the edge length L to 75, 50, and 25 nm, respectively. Each set of production runs consisted of 32 DMD trajectories that were 40×10^6 DMD simulation time units long. Nine sets of DMD simulations at $T = 0.13$ ($E_{HB} = 2.4$ kcal/mol) were performed for each of the three values of E_{CH} (0, 0.01, and 0.1) combined with each of the three protein concentrations (0.1, 0.3, and 3 mM). Five additional sets of DMD simulations at the implicit solvent parameter $E_{CH} = 0$ and protein concentration 0.1 mM were acquired at five simulation temperatures, 0.13, 0.16, 0.19, 0.22, and 0.25, corresponding to E_{HB} values of 4.6, 3.8, 3.2, 2.7, and 2.4 kcal/mol, respectively. Finally, two more sets of DMD simulations were conducted at $E_{CH} = 0$ and $T = 0.25$ at the two higher protein concentrations, 0.3 and 3 mM, respectively. In total, we acquired 512 DMD trajectories of HP36 dimer formation, each 40×10^6 DMD simulation time units long.

All-Atom MD Simulations in Explicit Solvent. *MD Simulation Protocol.* All MD simulations were performed with GROMACS 5.1.2.[44−50] MD simulations were conducted with the following force-field water model combinations: Amber ff14SB with the TIP3P model[11,51] CHARMM36m with the modified TIP3P model,[9,10,52,53] and both force fields with the TIP4P/2005 water model.[54] The N- and C-termini of each HP36 molecule correspond to NH_3^+ and COO^-, respectively, in both force fields. Because each simulation consists of a pair of HP36 proteins, four Cl^- ions are added to the simulation box to neutralize the system. The Verlet cutoff scheme and a time step of 2 fs are used during the equilibration and production steps. Energy minimization is performed with the steepest descent method for 100 000 steps and is followed by a 200 ps equilibration step at 300 K and 1.0 bar. All production runs use the velocity rescale thermostat[55] and a Berendsen barostat.[56]

MD Simulations of HP36 Dimers From DMD4B-HYDRA Simulations. Five HP36 dimer conformations derived from DMD4B-HYDRA simulations at 0.1 mM ($T = 0.25$, $E_{CH} = 0$) were converted from the four-bead model representation to a united-atom representation via replacing the C_β bead with the side-chain heavy atoms. This conversion was performed with an in-house software package, protsView. The dimers were then placed in a $(15\ \text{nm})^3$ box (corresponding to a concentration of 0.9 mM). The addition of hydrogen atoms and atom collisions were then resolved within GROMACS during the preparation and energy minimization steps, respectively. The root-mean-square deviation (RMSD) values of dimer conformations before and after the equilibration step were ≤0.1 nm. Another set of five initial conformations was prepared using the same initial HP36 dimer conformation placed in a $(7.3\ \text{nm})^3$ box (corresponding to a protein concentration of 8.5 mM). For each of the two force fields, Amber ff14SB and CHARMM36m, two sets of five 200 ns long MD trajectories of HP36 dimer dynamics (for two proteins concentrations) were acquired, resulting in 4 μs of MD simulations.

MD Simulations of HP36 Dimer Formation From Randomly Placed Proteins. Additional MD trajectories were acquired at protein concentrations of 0.9 and 8.5 mM to examine the effect of the water model and the initial tertiary structure on HP36 dimer formation. Five distinct initial conformations were produced within GROMACS by randomly placing two folded HP36 molecules (PDB[57] ID 1VII[58]) into $(15\ \text{nm})^3$ and $(7.3\ \text{nm})^3$ simulation boxes, respectively. An additional five distinct initial conformations of two unfolded HP36 molecules, which were randomly selected from monomers derived by DMD4B-HYDRA simulations and converted into fully atomistic structures, were randomly placed into a $(7.3\ \text{nm})^3$ box. All preparation steps were identical to those previously described. For each force field (Amber ff14SB and CHARMM36m), each water model (force-field-specific TIP3P and TIP4P/2005), and each initial condition (folded versus unfolded), we acquired five 200 ns long MD trajectories, resulting in 10 μs of MD simulations.

Analysis. *DMD Oligomer Size Distribution.* All visualization and structural analysis of DMD simulations were performed with the Visual Molecular Dynamics (VMD) software.[59] After 40×10^6 simulation steps, each system was checked to fix any HP36 units that had partially crossed the periodic boundary condition. Then, two peptides were considered to be a part of an oligomer if any of the beads of the peptides were ≤0.5 nm apart. Additional analysis was performed, in which the cutoff distance that defines an oligomer was increased from 0.5 to 0.6 nm. For each trajectory, the number of oligomers for each oligomer order was entered into a histogram. After each of the 32 trajectories are considered, the entire histogram was normalized by the total number of oligomers in all 32 trajectories. Histograms for each individual trajectory were similarly constructed and were then used to calculate the errors for each oligomer order as the standard error of the mean (SEM) values.

Solvent-Accessible Surface Area. The solvent-accessible surface area (SASA) of dimers derived from DMD4B-HYDRA simulations at 0.9 mM concentration was calculated using GROMACS 5.1.2. Additionally, the SASA of the hydrophobic residues (A, L, I, V, F, M) was also calculated over the course of the simulations.

Root-Mean-Square Deviation. Before the analysis, each trajectory was analyzed with GROMACS 5.1.2. The RMSD was then calculated for simulations of DMD-derived dimers of 0.9 and 8.5 mM concentration with respect to the initial dimer conformation, explicitly using GROMACS 5.1.2 to monitor the simulation convergence.

Distance Maps. Distance maps, similar to those used by Harada et al.,[20] were constructed using minimum residue pair distances with GROMACS 5.1.2 using times of 50−200 ns. Configurations were recorded every 2 ps (75 000 configurations), and the minimum distance between each residue was calculated considering only heavy atoms, recorded and averaged over all recorded configurations. These average minimum distances were then plotted where the x and y axes were the residue number (1−72 for dimers) and the z axis was the minimum distance, which was truncated at 1.5 nm for clarity. (Only close contacts were highlighted visually.) These graphs could be separated into quadrants where those along the main diagonal showed the secondary and tertiary structures (intrapeptide contacts) of each individual HP36, whereas the off-diagonal quandrants showed the proximate interpeptide contacts. SASA and RMSD values converged within the first 50 ns, thus conformations within 50−250 ns of each MD trajectory were used in distance map calculations.

Contact Maps. Intrapeptide and interpeptide contact maps are calculated within VMD to elucidate the tertiary and quaternary structures of HP36 dimers. Two amino acid residues are in contact if the respective distance between C_α atoms is below a cutoff distance of 0.6 nm. The contact maps

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

show the contact probabilities in the top left triangle and the SEM values of these probabilities in the lower right triangle. In the four-bead model of DMD4B-HYDRA simulations, two proteins form a dimer if the distance between any two beads belonging to different protein molecules is ≤0.5 nm. Contact maps of dimers derived from DMD4B-HYDRA simulations were based on all dimer conformations formed between (20 and 40) × 10^6 simulation time units of 32 DMD4B-HYDRA trajectories obtained at protein concentration 3 mM, resulting in 20 485 dimer conformations in total. In all-atom simulations, two proteins form a dimer if the minimum distance between any two heavy atoms belonging to different proteins is ≤0.3 nm. Note that the definition of a fully atomistic dimer differs from the definition of DMD4B-HYDRA dimer conformations, where the cutoff distance that defines an oligomer is applied to distances between pairs of beads in the four-bead model, and a larger cutoff distance of 0.5 nm is required to account for the coarse-grained representation of the amino acid residues. Contact maps of dimer conformations derived from all-atom MD simulations are extracted from each MD trajectory using time frames between 175 to 200 ns (with time frames 2 ps apart).

Kinetic and Equilibrium Parameters. The minimum distance between the heavy atoms of the two HP36 proteins was taken from distances maps constructed every 100 ps from 0 to 200 ns. Following the protocol of Best et al.,[7] an association event occurs when the minimum distance is ≤0.3 nm, and a dissociation event occurs when the minimum distance becomes ≥1 nm after a binding event. If the proteins are initially ≥0.3 nm apart, then they are considered unbound until the condition for a binding event occurs. Using the definition of a fully atomistic dimer described above, the association (binding) and dissociation (unbinding) rates can be calculated using the mean residence times. The mean residence t_{on} (ns)[1] is calculated by recording the time between an unbinding event (monomers are ≥1 nm apart) and a binding event (monomers are within 0.3 nm). With multiple binding and unbinding events, these times are all recorded and averaged to produce t_{on}. The mean residence time t_{off} (ns)[2] is similarly calculated by recording the time after a binding event to an unbinding event. If the proteins are initially in a bound state (≤0.3 nm) or an unbound state (≥0.3 nm), then the time until a binding or unbinding event contributes to the averages t_{on} and t_{off}, respectively. The unbinding and binding rates are defined as $k_{off} = 1/t_{off}$ (ns^{-1}) and $k_{on} = 1/(t_{on}c_o)$ (ns mM)$^{-1}$, where c_o is the protein concentration in millimoles. The dissociation constant, $K_D = k_{off}/k_{on}$ (mM), is then calculated to examine the solubility of HP36 in these force fields. Errors correspond to the SEM values.

Native Contacts. Native contacts were identified using the nuclear magnetic resonance (NMR) structure of the wild-type chicken villin subdomain HP36[58] from the Protein Data Bank[57] (PDB ID 1VII). The distance map of this native structure was calculated within GROMACS 5.1.2. Then, the number of intrapeptide distances ≤0.3 nm was counted as native contacts. The residue pairs that contribute to these contacts were recorded. Additional distance maps were constructed for all HP36 pair systems in the simulation, including hydrogen atoms, using times of 175−200 ns (12 500 configurations in each simulation). Each HP36 protein in the simulation was then compared with the list of native state contacts. The total number of intrapeptide contacts and the number of these contacts that corresponded to native contacts were both separately counted for both proteins. In both the native state and the simulation-derived states, the diagonal entries were not considered, and thus the minimum number of contacts possible was 35 for each protein. Errors correspond to SEM values.

RESULTS AND DISCUSSION

HP36 is one of the most studied proteins due to its small size, unique folded structure, and rapid folding kinetics, making it attractive for experimental and computational investigations.[5,7,20,22,60−64] The sequence of HP36 is

MLSDE^{5}DFKAV^{10}FGMTR^{15}SAFAN^{20}LPLWK^{25}QQNLK^{30}KEKGL^{35}F

The sequence starting with L2 and ending with F36 corresponds to the 35 amino acid long HP35 subdomain of the 76 amino acid long C-terminal villin headpiece domain. HP35 was reported to retain the conformation it adopts within the intact villin headpiece domain, making it the shortest autonomously folded protein composed entirely of naturally occurring amino acids. Importantly, HP35 and HP36 possess similar physical properties.[22] HP35 corresponds to the region L42−F76 of the 76 amino acid headpiece domain, which occupies the 791−825 region of intact chicken villin. Hereafter, we use the numbering 1−36 when referring to specific amino acids in the HP36 sequence. The native fold of HP36 consists of two short helices located at D4−K8 and R15−F18 alongside a longer helix spanning residues L23−K32.[58] Three phenylalanines (F7, F11, and F18) were reported to stabilize the hydrophobic core of the folded structure.

Previously reported infrared (IR) and NMR spectroscopy data revealed that HP36 remains soluble at least up to concentrations of 1 to 2 mM.[19−21] A comparison of NMR spectra of HP36 samples at concentrations of 1 and 32 mM reported by Harada et al. showed that most of the signals were present at both concentrations; however, significant deviations of chemical shifts for several amino acid residues (S3, V10, L23, W24, G34, and L35) were noted, indicating the emergence of protein−protein interactions at the higher protein concentration.[20] Similar observations were made in studies of HP35 samples using NMR,[65] CD,[66] and ultrafast 2D IR vibrational echo spectroscopy.[25] In the latter study, Chung et al. showed that HP35 is soluble at concentrations up to 6 mM, whereas at 18 mM, the Fourier transform infrared (FTIR) spectrum of the CN stretching mode changes indicated a potential emergence of HP35 assemblies, likely dimers.[25]

High solubility makes HP36 an ideal system to use as a means of an assessment of MD force fields with respect to their ability to accurately reproduce protein self-assembly, provided that the soluble state of the protein is experimentally characterized. Soluble protein exists in a solution as a mixture of monomers and oligomers of various sizes. An example of a soluble globular protein, which does not aggregate at neutral pH yet forms significant amounts of oligomers, is insulin. Mawhinney et al. used combined PICUP/SDS-PAGE techniques to characterize soluble insulin at neutral pH (nonamyloidogenic conditions) and acidic pH (amyloidogenic conditions).[18] Whereas insulin oligomers were present at neutral and acidic pH, oligomers dominated the soluble state of insulin at neutral pH under nonamyloidogenic conditions.[18] Because oligomers form within a subsecond time scale, whereas the process of aggregation requires significantly longer time scales (hours, days, or even months), the oligomer size

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

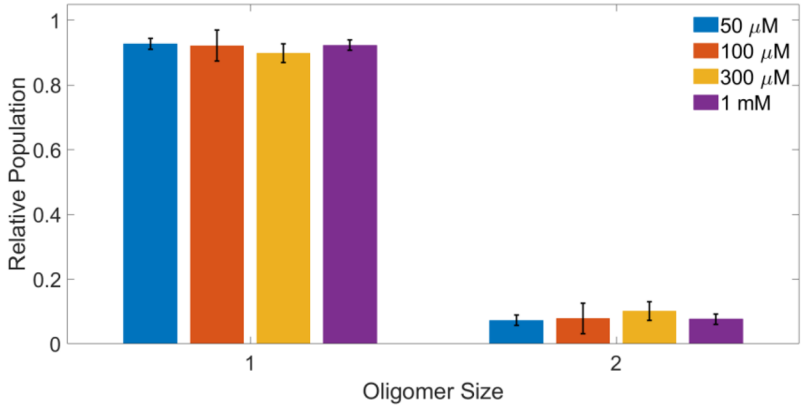


Figure 1. Results of the gel densitometry analysis for three PICUP/SDS-PAGE trials of 50, 100, and 300 μM and 1 mM of HP36 quiescently incubated at pH 7.4 for 0 h.

distribution derived from PICUP/SDS-PAGE experiments can be used to directly evaluate MD force fields with respect to their capacity to capture the soluble state of a protein.

In this work, we derive the oligomer size distribution of HP36 using PICUP followed by SDS-PAGE on HP36 samples at different protein concentrations up to and including 1 mM. Using the PICUP/SDS-PAGE oligomer size distribution of HP36, we then calibrate the implicit solvent DMD4B-HYDRA simulations to reproduce the experimentally derived HP36 oligomer size distribution. Five HP36 dimers derived from DMD4B-HYDRA simulations are converted to fully atomistic conformations, and their stability is examined by all-atom MD in explicit water using two force fields, Amber ff14SB and CHARMM36m. Additional MD simulations of HP36 dimer formation, starting from two randomly placed HP36 molecules with or without the native fold, are acquired to assess the capacity of these two force fields to capture the soluble state of HP36 at different protein concentrations.

In Vitro Measurements Reveal Predominantly Monomeric HP36 with ∼10% dimers. The efficiency of intermolecular cross-linking via PICUP relies on the presence of tyrosines or tryptophans in the primary sequence of the protein.[26] There is a single tryptophan, W24, in the HP36 sequence, which can facilitate the formation of covalent bonds among proximate tryptophans during the PICUP reaction. Whereas it is well documented that cross-linking occurs under conditions that allow for radical formation, the atomistic details of this process are not fully understood. In addition to tyrosine and tryptophan, several other amino acids are expected to undergo modifications under radical forming conditions[67] and may also contribute to cross-linking, as recently examined by Zhang et al. for Aβ oligomer formation.[68,69] During PICUP, covalent bonds (cross-links) form among proximate protein molecules. For two amino acid residues to form a covalent bond, their respective C_β atoms need to be at a distance of a few angstroms. Because radical formation is a fast process relative to conformational changes in proteins, some "instantaneous" collisions between two protein molecules are expected to result in covalent bond formation. We induced a cross-linking reaction by PICUP to examine HP36 samples at concentrations of 0.05, 0.10, 0.30, and 1 mM. (See the Methods for details.) After the PICUP reaction was quenched, SDS-PAGE resulted in gel bands corresponding to HP36 monomer and dimers, which were converted into oligomer size distributions after the application of the densitometry analysis. The PICUP/SDS-PAGE results in Figure 1 show that HP36 remains predominantly monomeric (∼90%) with an admixture of ∼10% dimers. Whereas the experimental data in Figure 1 correspond to 0 h of incubation, this distribution did not change with incubation time for up to 8 h of incubation.

One might expect an increase in the dimer fraction and the emergence of larger oligomers at higher protein concentrations. However, the results in Figure 1 clearly demonstrate that the fraction of HP36 dimers in these samples does not significantly depend on the protein concentration. To verify that HP36 does not aggregate, we applied a thioflavin (ThT) fluorescence assay. Figure S1 shows the ThT fluorescence intensity of quiescently incubated 0.05 mM HP36 versus incubation time, which remains low for incubation times of >140 h (∼6 days), indicating that HP36 does not form amyloid fibril-like aggregates. Thus HP36 remains 90% monomeric with 10% dimers at concentrations between 0.05 and 1 mM.

DMD4B-HYDRA Simulations of HP36 Self-Assembly. Examining protein self-assembly by fully atomistic MD simulations in explicit solvent is challenging. For this reason, the DMD4B-HYDRA approach, which combines an intermediate-resolution peptide model (with four beads per amino acid residue) with an implicit-solvent force field, accounting for residue-specific hydropathy and charge, has been developed and applied to elucidate the folding and early stages of self-assembly of an intrinsically disordered Aβ and variants[34−37,42] as well as globular protein stefin B[38] and the effect of pH on the folding of two mucin domains.[43] Because of the coarse-grained nature of the protein model, the research strategy we use is to calibrate the implicit solvent parameters in the four-bead model under the experimental conditions of interest. Here HP36 oligomer size distributions derived by PICUP/SDS-PAGE are used to calibrate DMD4B-HYDRA simulations using the protocol described in the Methods section. The results are described in detail in the Supporting Information. Figure S2 shows that effective electrostatic interactions promote HP36 self-assembly in a concentration-dependent way, which is not consistent with the previously described experimental results. When the HB strength is decreased in the absence of effective electrostatic interactions, our results in Figure S3 reveal that the oligomer size distribution of HP36 is approximately independent from the protein concentration and is thereby more in line with experimental findings. Five

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

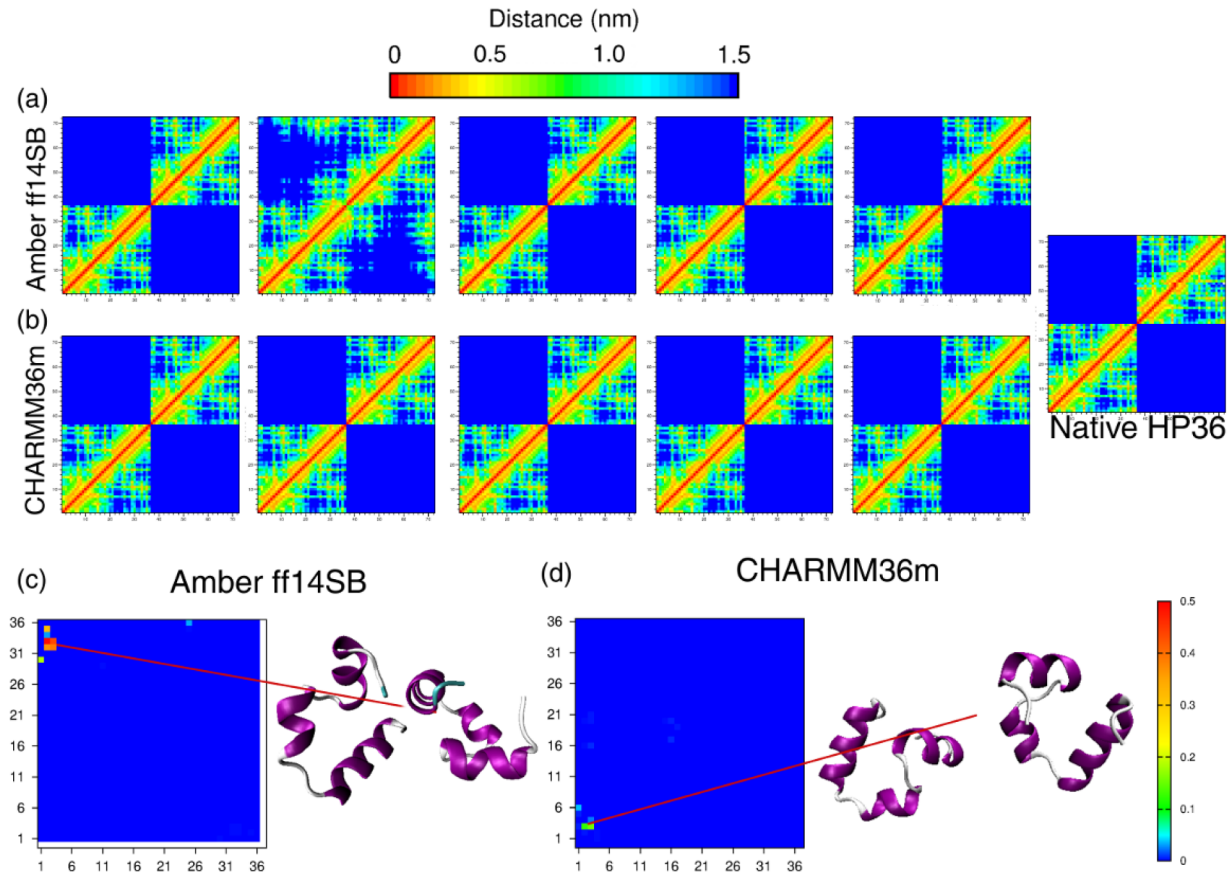


Figure 2. Distance maps of individual MD trajectories of HP36 dimer formation at protein concentration of 0.9 mM with randomly placed native-state HP36 monomers as initial conformations in (a) Amber ff14SB and (b) CHARMM36m. The distance map representing two natively folded HP36 monomers is displayed on the right. Interpeptide contact maps and proximate representative structures are produced for (c) Amber ff14SB and (d) CHARMM36m using 12 500 (0, 12 500, 0, 0, 0 from trajectories 1−5, respectively) and 10 426 (0, 0, 7676, 2751, 0 from trajectories 1−5, respectively) dimer structures, respectively, chosen based on the criterion for all-atom dimers described in the Methods section. The contact maps show the probability (color scale) for C_α atoms from each molecule to be within a 6 Å cutoff.

HP36 dimers derived from DMD4B-HYDRA simulations and converted into fully atomistic structures are shown in Figure S4 as mostly unstructured and weakly bound. Because HP36 is known for its stable native fold, which is expected to persist even at high protein concentrations, weakly bound HP36 dimers without tertiary contacts, predicted by DMD4B-HYDRA simulations, are not representative of native HP36 in water under neutral pH conditions. Whereas the DMD4B-HYDRA force field is well suited to model nonspecific hydrophobic collapse in water, which drives the folding and self-assembly of intrinsically disordered proteins that lack a stable fold, its applicability is limited for proteins with a native structure, in particular, from protein folds, which are stabilized by specific atomistic structure-dependent side-chain−side-chain contacts. Nonetheless, the weakly bound HP36 unstructured dimers are well suited as initial conformations for fully atomistic MD studies aimed at a comparison of different MD force fields.

Stability of All-Atom HP36 Dimers Derived from DMD4B-HYDRA Simulations. We here examined five weakly bound dimers derived from DMD4B-HYDRA simulations (Figure S5) by explicit solvent MD to study whether either of the two MD force fields can capture the observed preference of HP36 for the monomeric state at a concentration of 0.9 mM, which is comparable to the 1 mM concentration used in the PICUP/SDS-PAGE experiment previously described. The results of these MD simulations are described in detail in the Supporting Information. In brief, in either MD force field, the HP36 dimer dissociates in only one of the five trajectories, which indicates that the HP36 dimer is overly stabilized. Figure S6 shows that SASA and hydrophobic SASA decrease with the simulation time, indicating that the weakly bound and unstructured HP36 dimers undergo hydrophobic collapse; however, this collapse is significantly more pronounced in Amber ff14SB than in CHARMM36m. Figure S7 shows per-trajectory distance maps, whereby each distance map is split into four quadrants: The two quadrants along the main diagonal elucidate the proximate intrapeptide residues within each HP36 molecule (tertiary structure), and the two off-diagonal quadrants show proximate interpeptide residues that contribute to the dimer stability (quaternary structure). A comparison of distance maps corresponding to distinct MD trajectories in Figure S7 reveals that tertiary and quaternary structures acquired during fully atomistic MD simulations are strongly force-field-dependent. This can be directly observed in contact maps (Figure S8), which show tertiary and quaternary contacts averaged over the five MD trajectories for each set of MD simulations. The native contacts that form within the three helical regions of HP36 are present in the intrapeptide contact map of Amber ff14SB dimers but not CHARMM36m

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

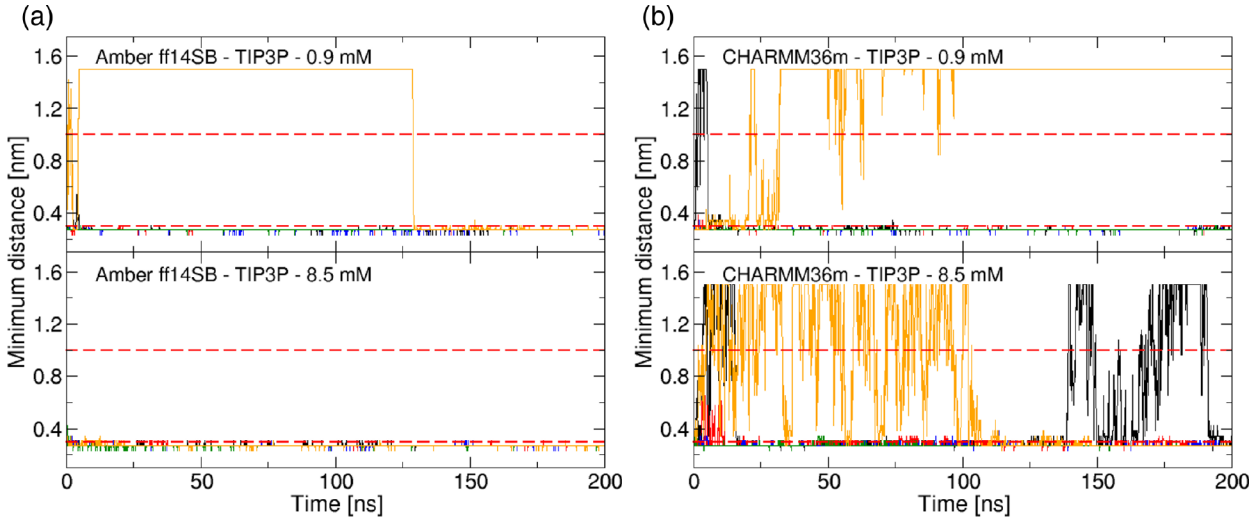


Figure 3. Temporal evolution of the minimum interpeptide distance between heavy atoms of two HP36 molecules in MD simulations using DMD4B-HYDRA dimer conformations as initial conformations. The minimum interpeptide distance versus time is displayed for two sets of 10 MD trajectories, 5 within each force field, (a) Amber ff14SB/TIP3P and (b) CHARMM36m/TIP3P, acquired at protein concentrations of 0.9 (top graphs) and 8.5 mM (bottom graphs), respectively. Distinct trajectories are distinguished by color. The red dashed lines represent the thresholds for dimer formation (≤0.3 nm) and dissociation into monomers (≥1 nm).

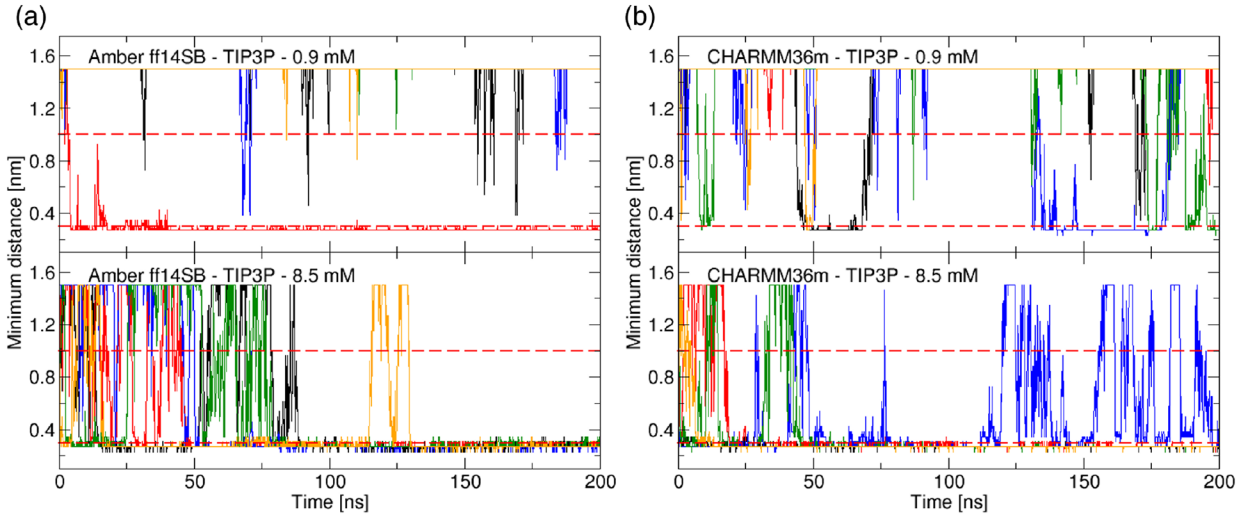


Figure 4. Temporal evolution of the minimum interpeptide distance between heavy atoms of two HP36 molecules in MD simulations using two randomly placed natively folded HP36 conformations as initial conformations. The minimum interpeptide distance versus time is shown for two sets of 10 MD trajectories, 5 within each force field, (a) Amber ff14SB/TIP3P and (b) CHARMM36m/TIP3P, acquired at protein concentrations of 0.9 (top graphs) and 8.5 mM (bottom graphs), respectively. Distinct trajectories are distinguished by color. The red dashed lines represent the thresholds for dimer formation (≤0.3 nm) and dissociation into monomers (≥1 nm).

dimers (Figure S8). Amber ff14SB dimers are held together by contacts between two distal N-terminal and C-terminal regions, whereas CHARMM36m dimers are stabilized by interpeptide contacts among regions coinciding with two shorter helices of the native fold (Figure S8). These findings suggest that the HP36 folding and self-assembly pathways are distinct in the two MD force fields.

MD Simulations of Dimer Formation from Natively Folded HP36 Proteins. The selection of the initial HP36 conformations for MD simulations of HP36 dimer formation is expected to affect the resulting HP36 dimer formation dynamics because simulation times significantly longer than the folding time of HP36, which would ensure adequate sampling, are not easily accessible. For this reason, five 200 ns long MD trajectories in each force field (10 MD trajectories in total) were acquired at a protein concentration 0.9 mM using two randomly placed natively folded HP36 monomers as initial conformations, as HP36 is expected to be mostly monomeric with a native fold. Figure 2a,b shows distance maps for each of the five MD trajectories derived from Amber ff14SB and CHARMM36m alongside the distance map corresponding to two natively folded HP36 proteins separated by a distance greater than 1.5 nm as a reference. The two quadrants along the main diagonal of distance maps in Figure 2 indicate that both MD force fields well stabilize the native structure of HP36. The two off-diagonal quadrants are mostly free of quaternary contacts, indicating that HP36 remains predominantly monomeric, in line with experimental findings. Only one Amber ff14SB trajectory (Figure 2a, the second map from the left) corresponds to a HP36 dimer, which is stabilized by

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

quaternary contacts between the N- and C-terminal regions, $_{1}$MLSD$_{4}$ and $_{31}$KEKGLF$_{36}$. These quaternary contacts also contribute to the HP36 dimer stability in Amber ff14SB trajectories with initial DMD4B-HYDRA dimers (Figure S8). CHARMM36m trajectories did not produce any stable HP36 dimers (Figure 2b). Short-lived HP36 dimer conformations extracted from CHARMM36m trajectories exhibited quaternary contacts between pairs of N-terminal $_{1}$MLSD$_{4}$ regions, which were not present in HP36 dimers extracted from CHARMM36m trajectories with initial DMD4B-HYDRA dimer conformations. The analysis shows that whereas both MD force fields favor the HP36 monomeric state at 0.9 mM, Amber ff14SB promotes HP36 dimer formation more than CHARMM36m.

Does the Protein Concentration Affect HP36 Dimer Formation in MD Simulations? We here ask to what extent the protein concentration affects the HP36 dimer formation in MD simulations using Amber ff14SB (with TIP3P water model) and CHARMM36m (with TIP3P water model). We acquired 10 MD trajectories (5 within each force field) at a protein concentration of 8.5 mM, starting from the same initial HP36 dimer conformations derived from DMD4B-HYDRA simulations (Figure S6) that were used in the corresponding simulations at 0.9 mM. The concentration of 8.5 mM was selected to facilitate a comparison to the previous study of Best and collaborators.[7] The RMSD values of these MD trajectories converged within the first 50 ns (Figure S9c,d), with the exception of a single CHARMM36m trajectory (the fifth initial dimer conformation in Figure S9b from the left), in which the dimer dissociated. Figure 3 shows the minimum interpeptide distance between heavy atoms of the two HP36 molecules in the simulations, which we use to distinguish between the monomeric and dimeric states of the two HP36 proteins in the simulation box. The system is considered to be in a monomeric state if this minimum distance is >1 nm (Figure 4, the top red dashed line). Conformations with a minimum distance below 0.3 nm are defined as dimer conformations (Figure 4, the bottom red dashed line).

Only one of the five Amber ff14SB trajectories at 0.9 mM produced monomeric HP36 for most of the simulation time (Figure 3a, top panel). In contrast, all five Amber ff14SB trajectories at 8.5 mM produced stable dimer conformations. The association and dissociation dynamics appear quite different in CHARMM36m trajectories (Figure 3b), which indicate more frequent association and dissociation events at both concentrations. To quantify HP36 monomer populations in these simulations, we calculated the total simulation time that the system is found in a monomeric state across all five MD trajectories, that is, the sum of time intervals between a dissociation event (minimum distance $\geq$1 nm) and the subsequent association event (minimum distance $\leq$0.3 nm), divided by the total simulation time, 1000 ns. The resulting monomer propensities of HP36 at 0.9 mM are 12.76 and 17.84% in Amber ff14SB and CHARMM36m, respectively. The monomer propensity decreases in MD simulations at 8.5 mM to 0.08 and 14.01% in Amber ff14SB and CHARMM36m, respectively. Whereas this decrease is large in Amber ff14SB (~150-fold), the monomer propensity decrease in CHARMM36m is only modest. Although MD-derived monomer propensities are still below the experimental value of ~90%, CHARMM36m consistently produces higher monomer propensities at both protein concentrations compared with Amber ff14SB.

It could be argued that unfolded, weakly bound HP36 dimers derived from DMD4B-HYDRA simulations as initial conformations for MD simulations are structurally too distinct from the expected folded HP36 monomers, and the simulation times are too short to allow these dimers to dissociate and fold. We thus acquired additional 10 200 ns long MD trajectories (five within each force field and the respective TIP3P water model) at 8.5 mM by using randomly placed natively folded HP36 monomers as the initial conformations. The resulting per-trajectory distance maps are displayed in Figure S10a,b for Amber ff14SB and CHARMM36m, respectively. The effect of the protein concentration on the distance maps can be elucidated by comparing Figure S10a,b (8.5 mM) with the corresponding distance maps Figure 2a,b (0.9 mM) for Amber ff14SB and CHARMM36m, respectively. This comparison reveals significantly more proximal quaternary contacts at the higher concentration in both force fields, indicating a strong concentration dependence of HP36 dimer stability. Figure S10c,d shows the average quaternary contacts formed at 8.5 mM in each MD force field, which are not only force-field-dependent but also distinct from quaternary contacts formed at 0.9 mM (compare with Figure 2c,d). In Amber ff14SB, the most prevalent quaternary contacts at 8.5 mM form between pairs of N-terminal regions $_{1}$MLSDED$_{6}$. In addition to this dominant cluster of quaternary contacts, interpeptide contacts between $_{8}$KAVFGM$_{13}$ (the region between the two short helices of the native fold) and the C-terminal region $_{33}$KGLF$_{36}$ are observed. In contrast, in CHARMM36m, HP36 dimers are stabilized at 8.5 mM by quaternary contacts between pairs of C-terminal regions, $_{33}$KGLF$_{36}$. Thus interactions stabilizing HP36 dimers are both concentration- and force-field-dependent.

The effect of the protein concentration on HP36 dimer formation is apparent in Figure 4. The calculated monomer propensities at 0.9 mM are 80.4 and 90.1% in Amber ff14SB and CHARMM36m, respectively, which is consistent with PICUP/SDS-PAGE results. The top graphs in Figure 4a,b demonstrate that the agreement with experimental results is not due to a lack of intermolecular interactions, as the two HP36 molecules are frequently within the range of effective electrostatic interactions; however, most of these interactions do not result in dimer stabilization. As the protein concentration increases from 0.9 to 8.5 mM, the monomer propensities drop from 80.4 to 22.11% in Amber ff14SB and from 90.1 to 22.48% in CHARMM36m, indicating a strong concentration dependence of the HP36 dimer population, which cannot be easily reconciled with experimental data reported here and in previous studies.[20,25]

Effect of Water Model and Initial Conditions on HP36 Dimer Formation in MD Simulations. We here explored dimer formation by MD at a fixed protein concentration of 8.5 mM by using initial conditions in which the two HP36 molecules are randomly placed in the simulation box. Two sets of such initial conditions with HP36 in the native folded state or unfolded state are explored. In addition to comparing Amber ff14SB and CHARMM36m for these two sets of initial conditions, we examine whether the water model affects the solubility of HP36 in MD simulations in either of the two force fields. These two force fields were both developed using the TIP3P water model; however, the more recent water model, TIP4P/2005, was reported to lead to significantly improved properties of water,[54,70,71] which may directly affect the protein solubility. To examine the effect of the water model on HP36

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

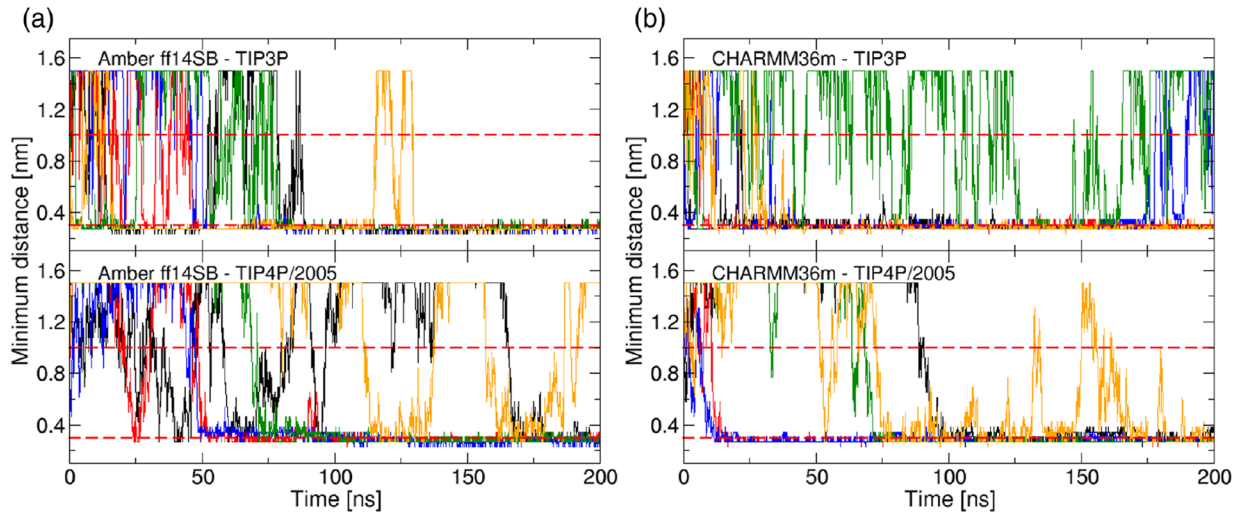


Figure 5. Temporal evolution of the minimum interpeptide distance between heavy atoms of two HP36 molecules in MD simulations using two randomly placed natively folded HP36 conformations as initial conformations. 10 MD trajectories (5 within each force field, Amber ff14SB and CHARMM36m) at a protein concentration of 8.5 mM were acquired using two different water models, TIP3P and TIP4P/2005. Distinct trajectories are distinguished by color. The red dashed lines represent the thresholds for dimer formation (≤0.3 nm) and dissociation into monomers (≥1 nm). The top panels of (a) and (b) show the same data as the bottom panels in Figure 4.

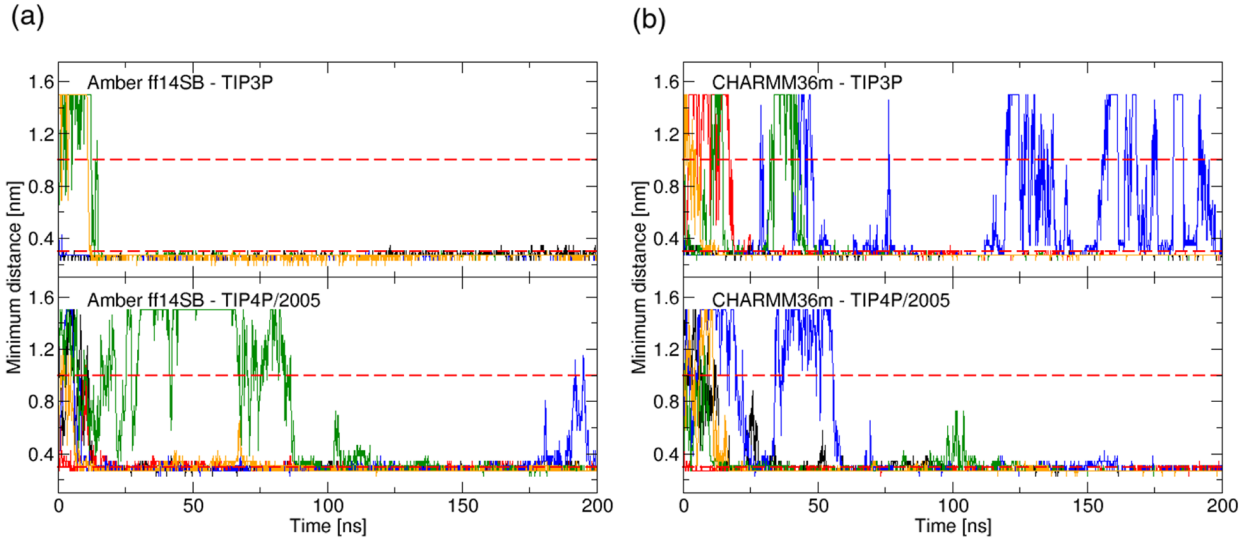


Figure 6. Temporal evolution of the minimum interpeptide distance between heavy atoms of two HP36 molecules in MD simulations using two randomly placed unfolded HP36 conformations as the initial conformations. Two sets of 10 MD trajectories (5 within each force field, Amber ff14SB and CHARMM36m) at a protein concentration of 8.5 mM were acquired using two different water models, TIP3P and TIP4P/2005. Distinct trajectories are distinguished by color. The red dashed lines represent the thresholds for dimer formation (≤0.3 nm) and dissociation into monomers (≥1 nm).

dimer formation, two sets of 10 MD trajectories (5 for each force field) were acquired at a protein concentration of 8.5 mM, starting first from the initial state with two randomly placed folded HP36 monomers. The minimum interpeptide distance was then monitored over the simulation time, as shown in Figure 5.

The results in Figure 5a demonstrate that replacing TIP3P by TIP4P/2005 in Amber ff14SB increases the amount of time the system is found in the monomer state. In contrast, Figure 5b shows that replacing TIP3P by TIP4P/2005 in CHARMM36m does not affect the dimer formation in any obvious way. The monomer propensity in Amber ff14SB increases from 22.11 to 46.59% when TIP3P is replaced by TIP4P/2005, which is more than a two-fold increase. Replacing TIP3P by TIP4P/2005 in CHARMM36m only modestly increases the monomer propensity from 22.48 to 28.67%.

Two additional sets of 10 MD trajectories (5 for each force field) were then acquired at 8.5 mM, with the initial conformations corresponding to two randomly placed unfolded HP36 monomers. The resulting minimum interpeptide distance versus simulation time is displayed in Figure 6. In the absence of the native fold in initial HP36 conformations, Amber ff14SB trajectories (Figure 6a) are characterized by decreased monomer populations of 2.67 (TIP3P) and 12.44% (TIP4P/2005) in comparison with the respective conformations from trajectories that started from natively folded HP36 molecules (Figure 5a) with monomer populations of 21.11

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

Table 1. Average Association Rates, Dissociation Rates, and Dissociation Constants for the 12 Sets of MD Trajectories of HP36 Dimer Formation Dynamics[a]

force field	water model	k_{on} (M^{-1} ns^{-1})	k_{off} (ns^{-1})	K_D (mM)
		native state HP36 (0.9 mM)		
Amber ff14SB	TIP3P	- ± -	- ± -	- ± -[b]
CHARMM36m	TIP3P	- ± -	- ± -	- ± -[b]
		native state HP36 (8.5 mM)		
Amber ff14SB	TIP3P	13.57 ± 1.85	0.024 ± 0.002	2.19 ± 0.295
Amber ff14SB	TIP4P/2005	3.07 ± 0.63	0.032 ± 0.016	10.21 ± 4.21
CHARMM36m	TIP3P	23.76 ± 8.90	0.051 ± 0.037	5.88 ± 5.11
CHARMM36m	TIP4P/2005	5.68 ± 1.85	0.015 ± 0.009	4.28 ± 1.62
		unfolded HP36 (8.5 mM)		
Amber ff14SB	TIP3P	500.0 ± 284.1	0.007 ± 0.001	0.34 ± 0.14
Amber ff14SB	TIP4P/2005	32.52 ± 21.48	0.008 ± 0.001	1.68 ± 1.28
CHARMM36m	TIP3P	247.0 ± 232.3	0.02 ± 0.01	1.18 ± 0.67
CHARMM36m	TIP4P/2005	10.20 ± 2.71[c]	0.008 ± 0.001	0.91 ± 0.48
		DMD-derived dimers (0.9 mM)		
Amber ff14SB	TIP3P	4296 ± 2594[c]	0.009 ± 0.003	0.32 ± 0.31
CHARMM36m	TIP3P	5607 ± 2455	0.020 ± 0.014	0.79 ± 0.79
		DMD-derived dimers (8.5 mM)		
Amber ff14SB	TIP3P	672.5 ± 233.5[c]	0.005 ± 0	0.007 ± 0.004
CHARMM36m	TIP3P	138.1 ± 98.4[c]	0.011 ± 0.006	2.12 ± 1.50

[a]Errors correspond to SEM values. [b]Trajectories in which a dimer did not associate within 200 ns are excluded from the calculated average and SEM values. See Table S1 for more details. [c]Trajectories in which a dimer did not dissociate within 200 ns are excluded from the calculated average and SEM values of k_{on}. See Table S1 for details.

(TIP3P) and 46.59% (TIP4P/2005). Again, replacing TIP3P with TIP4P/2005 resulted in significantly increased monomer populations in Amber ff14SB. In the absence of the native fold in initial HP36 conformations, CHARMM36m trajectories (Figure 6b) exhibit decreased monomer populations 10.72 (TIP3P) and 8.83% (TIP4P/2005) when compared with the respective conformations from trajectories that started from natively folded HP36 molecules (Figure 5b) with monomer populations of 22.48 (TIP3P) and 28.67% (TIP4P/2005). Interestingly, replacing TIP3P with TIP4P/2005 in trajectories derived from unfolded HP36 peptides yields a slightly decreased monomer population in CHARMM36m. Thus, unlike in CHARMM36m, replacing TIP3P by TIP4B/2005 in Amber ff14SB significantly increases the solubility of HP36. In both force fields, starting from two initially folded HP36 molecules leads to the highest monomer populations, regardless of the water model, indicating that the stability of the native fold directly contributes to HP36 solubility.

Kinetics of HP36 Dimer Formation Exhibits Force-Field- and Water-Model-Dependent Characteristics. The results described thus far compared the thermodynamic aspects of HP36 dimer formation in Amber ff14SB and CHARMM36m. The minimum distance versus the simulation time of individual MD trajectories in Figures 3, 5, and 6, however, reflects not only the thermodynamic but also the kinetic differences in dimer formation between the two force fields, such as the frequency of dimer association and dissociation events. We here calculated the association (k_{on}) and dissociation (k_{off}) rates and the dissociation constant, K_D (as defined in the Methods section), for every MD trajectory (Table S1). The average values of k_{on}, k_{off}, and K_D alongside the respective SEM values for each set of MD trajectories are listed in Table 1.

MD trajectories at 0.9 mM starting from randomly placed initially folded HP36 molecules frequently exhibited no association events. In Table S1, these events are assigned an infinitely large dissociation constant value. Consistent with the calculated monomer propensities at this concentration, both force fields capture the kinetics of HP36 dimer formation fairly well. Eight of the 10 trajectories result in K_D > 10 which is consistent with previously reported experimental results.[20,25] The details are force-field-dependent. In line with the analysis of contacts (Figure 2) and the calculation of monomer propensities, the kinetic rates also indicate a more transient nature of the HP36 dimers in CHARMM36m than in Amber ff14SB. The dissociation constants derived from MD trajectories at 8.5 mM with initially folded HP36 molecules: 2.19 mM for Amber ff14SB/TIP3P and 5.88 mM for CHARMM36m/TIP3P (Table 1, rows 3 and 5) are significantly smaller than those at the lower protein concentration. Similarly, the kinetic rates k_{on} derived from MD trajectories at 8.5 mM are significantly higher than the corresponding rates at the lower concentration.

The last four rows in Table 1 list the rates calculated from MD trajectories, which used DMD4B-HYDRA dimers as the initial conformations at two different protein concentrations. At the low concentration of 0.9 mM, the two force fields (each combined with TIP3P) result in comparable k_{on} values. In contrast, k_{off} in CHARMM36m is larger than that in Amber ff14SB by a factor of ∼2, and so is the resulting K_D; however, although this is in line with data in Figure 3, this effect is not statistically significant. Increasing the protein concentration to 8.5 mM results in a significant decrease in k_{on} values by a factor of ∼6 and ∼40 in Amber ff14SB and CHARMM36m, respectively, whereas k_{off} decreases by a factor of 2 in both force fields. The concentration-induced changes in K_D are not statistically significant in either force field. Nonetheless, at a protein concentration of 8.5 mM, CHARMM36m results in a significantly higher value of K_D (2.12 mM) compared with Amber ff14SB, although the available experimental data[20,25] suggest that this value is still about an order of magnitude too low.

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

We then explored the effect of the initial conformations on the kinetic rates by examining MD trajectories acquired at a protein concentration of 8.5 mM, which started from randomly placed initially folded HP36 molecules (Table 1, rows 3 and 5), and compared then with the respective data for MD trajectories at the same concentration, which started from dimers derived from DMD4B-HYDRA simulations (Table 1, rows 13 and 14). Starting MD simulations from folded HP36 molecules more closely approximates the expected experimental conditions under this concentration, where the vast majority of H36 is expected to be monomeric. In line with these expectations, the resulting k_{on} rates are ∼50 times (Amber ff14SB/TIP3P) or ∼5 times (CHARMM36m/TIP3P) lower, and the k_{off} values are approximately 5 times higher (both force fields), resulting in more realistic albeit still underestimated K_D values: 2.19 (Amber ff14SB/TIP3P) and 5.88 mM (CHARMM36m/TIP3P).

Next, we asked if replacing the water model affects the association versus dissociation rate and the K_D at 8.5 mM in MD trajectories with initially folded HP36 molecules (Table 1, rows 3−6). Notably, when TIP3P is replaced by TIP4P/2005 in Amber ff14SB simulations, an approximately four-fold decrease in k_{on} is observed without any significant change in k_{off}, resulting in the highest value of K_D = 10.21 mM among all sets of MD trajectories. No statistically significant changes in these rates or K_D are noted when TIP3P is replaced by TIP4P/2005 in CHARMM36m simulations. Does the effect of the water model on the rates and the K_D in Amber ff14SB depend on the initial conditions? To answer this question, we also calculated the above quantities for MD trajectories at 8.5 mM with initially unfolded HP36 molecules (Table 1, rows 7−10). Upon replacing TIP3P by TIP4P/2005 in Amber ff14SB simulations, a statistically significant ∼15-fold decrease in k_{on} was observed without any significant change in k_{off}. CHARMM36m simulations resulted in an even larger ∼15-fold decrease in k_{on} but also produced a 2.5-fold decrease in k_{off}; however, these changes were only marginally significant. The changes in K_D values induced by replacing TIP3P by TIP4P/2005 for these sets of simulations were not statistically significant (Table 1, rows 9−12).

The above analysis showcases the force field specificity of the kinetic rates and K_D in HP36 dimer formation. When combined with its respective TIP3P water models, CHARMM36m produces consistently lower k_{on} rates and higher K_D values than Amber ff14SB, although these differences between the two force fields are not statistically significant, except for MD sets of trajectories with initial dimer conformations derived from DMD4B-HYDRA simulations at a protein concentration of 8.5 mM (Table 1, rows 13 and 14). This trend is consistent with the larger average monomer propensities observed in CHARMM36m simulations relative to the Amber ff14SB simulations, as previously discussed, and may be a result of the reparameterization of the TIP3P water model to make the dispersion part of the protein−water interactions more favorable and thereby address the general problem of overly compact ensembles of intrinsically disordered proteins.[10] The above results demonstrate that the water model can have an immense impact on the protein solubility, kinetic rates, and K_D.

The force-field differences described above are larger than expected given that these MD force fields were developed relatively recently. We here asked if these differences are related to how well these force fields stabilize the native HP36 fold, as the above results show that MD trajectories, which started from fully folded HP36 molecules, produced more realistic kinetic rates and K_D values than the other two types of initial conditions. To this end, we identified the native contacts from the experimentally resolved HP36 folded structure (PDB ID 1VII;[58] see the Methods section for a definition of a contact) and examined how many of the total of 114 native contacts are present as intrapeptide contacts in all-atom conformations across all sets of MD trajectories. The results of this calculation, displayed in Table S2 for each individual trajectory and summarized in Table 2, indicate that fully

Table 2. Average Number of Native Intrapeptide Contacts in Each of the Two HP36 Molecules for Each Set of MD Trajectories[a]

force field	water model	protein 1 av native contacts	protein 2 av native contacts
native state HP36 (0.9 mM)			
Amber ff14SB	TIP3P	106.6 ± 1.22	101.8 ± 1.07
CHARMM36m	TIP3P	102.8 ± 1.07	101.6 ± 0.93
native state HP36 (8.5 mM)			
Amber ff14SB	TIP3P	102.4 ± 0.68	101.4 ± 0.75
Amber ff14SB	TIP4P/2005	102.4 ± 0.81	101.8 ± 0.49
CHARMM36m	TIP3P	100.6 ± 1.72	100.4 ± 0.60
CHARMM36m	TIP4P/2005	102 ± 0.71	102.4 ± 0.69
unfolded HP36 (8.5 mM)			
Amber ff14SB	TIP3P	50.0 ± 2.26	61.2 ± 2.63
Amber ff14SB	TIP4P/2005	58.0 ± 2.24	52.6 ± 2.24
CHARMM36m	TIP3P	43.2 ± 2.15	38.8 ± 0.37
CHARMM36m	TIP4P/2005	43.0 ± 1.97	40.4 ± 0.98
DMD-derived dimers (0.9 mM)			
Amber ff14SB	TIP3P	50.2 ± 3.06	49.4 ± 3.30
CHARMM36m	TIP3P	37.6 ± 1.44	38.4 ± 1.89
DMD-derived dimers (8.5 mM)			
Amber ff14SB	TIP3P	50.2 ± 2.96	46.8 ± 2.08
CHARMM36m	TIP3P	42.4 ± 1.40	37.4 ± 0.75

[a]Averages were calculated using conformations from all time frames within 175−200 ns of each trajectory. The experimentally resolved HP36 folded structure (PDB ID 1VII[58]) is associated with 114 native contacts, whereby 35 of these correspond to covalently bonded nearest-neighbor residues. The trajectory-specific average number of native intrapeptide contacts is listed in Table S2. Errors correspond to SEM values.

atomistic HP36 conformations from MD trajectories with initial DMD4B-HYDRA dimer conformations are characterized by a relatively low fraction of native contacts at both protein concentrations: 15/79 = 0.19 and 3/79 = 0.04 within Amber ff14SB and CHARMM36m, respectively (excluding the 35 covalently bonded neighboring contacts) (Table 2, rows 11−14). There are significantly fewer native contacts present in CHARMM36m than in Amber ff14SB conformations, which is consistent with the distance maps in Figures S7 and S8.

Interestingly, the Amber ff14SB and CHARMM36m conformations from MD trajectories with initially folded HP36 molecules retained a comparable fraction of native contacts of 0.91 and 0.84 at 0.9 and 8.5 mM, respectively, regardless of the water model (Table 2, rows 1−6). MD conformations from MS trajectories, which were initiated from unfolded conformations, show, on average, more native contacts than those initiated from DMD4B-HYDRA conformations with 0.27 and 0.08 for Amber ff14SB and CHARMM36m, respectively (Table 2, rows 7−10), with

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

Amber ff14SB resulting in a significantly higher fraction of native contacts than CHARMM36m. These results show that the folding and assembly pathways of HP36 strongly depend on the force field; however, the native HP36 fold appears to be equally well reproduced in both force fields, and the water model does not affect the fraction of native contacts formed, regardless of the initial conditions.

CONCLUSIONS

Villin headpiece domain HP36 is one of the most studied protein domains, both experimentally and computationally. Its high solubility allows computer simulations to examine the HP36 assembly dynamics at concentrations that are also experimentally accessible. In this study, we examined the soluble state of villin headpiece domain HP36 *in vitro* at four concentrations between 0.05 and 1 mM using PICUP followed by SDS-PAGE to quantify the HP36 monomer population and its dependence on the protein concentration. Our results indicate that the soluble state of HP36 is independent of the protein concentration and is characterized by 90% monomeric with 10% dimeric populations. These experimental findings are qualitatively consistent with previously published experimental findings on HP36[19,20,22,25,65] and provide quantitative data that can be used in the assessment of MD force fields with respect to their capacity to accurately reproduce the soluble state of HP36. To the best of our knowledge, our study is the first to experimentally characterize the soluble state of HP36 by the application of PICUP/SDS-PAGE and show that the HP36 oligomer size distribution does not depend on the protein concentration.

Using the above experimental characterization of the soluble state, we then calibrated implicit solvent parameters in the intermediate resolution protein model combined with the DMD4B-HYDRA force field to obtain the experimentally observed oligomer size distribution, which only weakly depends on the protein concentration. Dimer conformations from DMD4B-HYDRA simulations were the converted into fully atomistic structures and used as initial conformations for explicit-solvent all-atom MD simulations. The main objective of MD simulations was to assess and compare two MD force fields, Amber ff14SB and CHARMM36m, with respect to their ability to capture the soluble state of HP36. This is a challenging task because the time scales of HP36 folding and self-assembly, while short from an experimental point of view, are still not easily accessible to explicit-solvent fully atomistic MD simulations. Another problem stems from the trajectory-to-trajectory variability, which can be observed even in a minimalistic model of self-assembly with competing interactions of comparable strengths.[72] Consequently, inferring the soluble state from MD trajectories is challenging, and inevitably, the results will depend to some extent on the initial state of the system because the necessary conditions for the ergodocity theorem to apply are not met. Given these considerations, we acquired 14 sets of explicit-solvent MD simulations, each set comprising of five 200 ns long trajectories, to examine the trajectory-to-trajectory variability and to feed the statistical analysis, resulting in a total simulation time of 14 μs.

Our analysis of the dimer dissociation and association dynamics revealed a consistent trend of CHARMM36m (combined with its native TIP3P) producing larger HP36 monomer populations than Amber ff14SB (combined with TIP3P). CHARMM36m well reproduces the HP36 monomer propensity at 0.9 mM when randomly placed folded HP36 monomers are used as initial conformations, and Amber ff14SB does not lag far behind. In MD simulations at 8.5 mM, HP36 monomer populations are vastly underestimated within both force fields. When comparing the monomer populations and kinetic rates at two different protein concentrations, Amber ff14SB (with TIP3P) HP36 dimer simulations exhibited more concentration dependence that the respective CHARMM36m simulations. We also noted striking differences in the tertiary and quaternary structure of Amber ff14SB (TIP3P) versus CHARMM36m (TIP3P) HP36 dimer conformations. Notably, the contacts stabilizing HP36 dimers are both concentration- and force-field-dependent.

We found a significant improvement in the HP36 solubility in terms of the monomer populations and K_D values upon replacing TIP3P by TIP4P/2005 in Amber ff14SB. This result is surprising if we take into consideration the results of our previous study on the intrinsic conformational dynamics of the alanine residue in GAG peptide, which evaluated several MD force-field/water model combinations with respect to their capacity to capture available spectroscopic data on this short peptide.[73] Although the effect of the water model in Amber ff14SB was not examined, no significant changes in the conformational dynamics of the alanine reside in GAG upon replacing TIP3P by TIP4P/2005 in OPLS-AA/L[74,75] and OPLS-AA/M[12] were reported.[73] We also verified that replacing the water model had no effect on the number of native contacts that formed in the MD simulations described in this work. The set of Amber ff14SB (with TIP4P/2005) trajectories, which used two randomly placed folded HP36 molecules as initial configurations, produced the highest value of $K_D \approx 10$ mM. Taking into account the meaning of K_D, the concentration at which half of the system is dissociated, this value may be in the correct order of magnitude. FTIR spectroscopy data for HP35 (HP36 without the N-terminal methionine) reported by Chung et al. revealed changes in the spectrum of the CN stretching model consistent with HP35 self-assembly at protein concentration of ~18 mM but not ~6 mM.[25] Similarly, Harada et al. reported that most chemical shifts in NMR measurements were conserved when the HP36 concentration increased from 1 to 32 mM,[20] which places the lower bound on K_D to the latter.

In summary, this work provides a unique evaluation of Amber ff14SB and CHARMM36m with respect to their capacity to capture the soluble state of HP36. Our findings show that an improved water model, such as TIP4P/2005, combined with Amber ff14SB has the potential to solve the problem of underestimated protein solubility in MD simulations. Importantly, we propose a new strategy for MD force-field development by experimentally characterizing the soluble state of a protein via PICUP/SDS-PAGE and using the resulting experimental data for a direct comparison to MD results with an overarching goal of improving MD force fields with respect to their capacity to capture protein aggregation, which will, in the long run, help to improve our understanding of many human diseases that result from aberrant protein aggregation.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c04589.

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

Table S1: Trajectory-specific association rates, dissociation rates, and dissociation constants. Table S2: Number of trajectory-specific native contacts. Figure S1: ThT fluorescence intensity versus incubation time. Figure S2: DMD4B-HYDRA oligomer size distributions at different protein concentrations and effective electrostatic interaction strengths. Figure S3: DMD4B-HYDRA oligomer size distributions at different temperatures. Figure S4: DMD4B-HYDRA oligomer size distributions at different temperatures and effective electrostatic interaction strengths. Figure S5: DMD4B-HYDRA dimers in the four-bead and all-atom representations. Figure S6: Time evolution of hydrophobic SASA and SASA values for MD simulations at 0.9 mM with initial DMD4B-HYDRA dimer conformations. Figure S7: Distance maps with representative structures from MD simulations with initial DMD4B-HYDRA dimer conformations. Figure S8: Contact maps of the native HP36 fold, DMD4B-HYDRA dimers, and fully atomistic dimers from MD trajectories with initial DMD4B-HYDRA dimer conformations. Figure S9: Time evolution of RMSD values for four sets of MD simulations with initial dimer conformations derived from DMD4B-HYDRA simulations. Figure S10: Distance maps and contact maps with representative dimer conformations from MD trajectories with initially folded HP36 molecules (PDF)

AUTHOR INFORMATION

Corresponding Author

Brigita Urbanc − *Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104, United States;* orcid.org/0000-0001-9159-7698; Phone: +123 (0)123 4445556; Email: bu25@drexel.edu

Authors

Brian Andrews − *Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104, United States*

Kaho Long − *Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c04589

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation through the grant numbers MCB-1817650 and OAC-1919691. MD simulations were performed on hardware supported by Drexel's University Research Computing Facility. We thank Luis Cruz for help with the ProtsView compilation, Mats Dahlgren for version 1 of `achemso`, and Donald Arseneau for the code taken from `cite` to move the citations after the punctuation. Many users have provided feedback on the class, which is reflected in all of the different demonstrations shown in this document.

ADDITIONAL NOTES

[1]Read as time until bound or "on" state.
[2]Read as time until unbound or "off" state.

REFERENCES

(1) Chiti, F.; Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.

(2) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch. Biochem. Biophys.* **2008**, *469*, 132−141.

(3) Shea, J.-E.; Urbanc, B. Insights into Aβ aggregation: A molecular dynamics perspective. *Curr. Top. Med. Chem.* **2013**, *12*, 2596−2610.

(4) Carballo-Pacheco, M.; Strödel, B. Advances in the simulation of protein aggregation at the atomistic scale. *J. Phys. Chem. B* **2016**, *120*, 2991−2999.

(5) Petrov, D.; Zagrovic, B. Are current atomistic force fields accurate enough to study proteins in crowded environments? *PLoS Comput. Biol.* **2014**, *10*, e1003638.

(6) Carballo-Pacheco, M.; Ismail, A. E.; Strödel, B. On the applicability of force fields to study the aggregation of amyloidogenic peptides using molecular dynamics simulations. *J. Chem. Theory Comput.* **2018**, *14*, 6063−6075.

(7) Best, R. B.; Zheng, W.; Mittal, J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.

(8) Nawrocki, G.; Wang, P.-h.; Yu, I.; Sugita, Y.; Feig, M. Slow-down in diffusion in crowded protein solutions correlates with transient cluster formation. *J. Phys. Chem. B* **2017**, *121*, 11072−11084.

(9) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.

(10) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.

(11) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(12) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.

(13) Zhang, S.; Andrews, B.; Schweitzer-Stenner, R.; Urbanc, B. Intrinsic conformational dynamics of alanine in water/ethanol mixtures: An experiment-driven molecular dynamics study. *J. Phys. Chem. B* **2020**, *124*, 11600−11616.

(14) den Engelsman, J.; Garidel, P.; Smulders, R.; Koll, H.; Smith, B.; Bassarab, S.; Seidl, A.; Hainzl, O.; Jiskoot, W. Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. *Pharm. Res.* **2011**, *28*, 920−933.

(15) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16*, 163−199.

(16) Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B. Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 330−335.

(17) Williams, T. L.; Serpell, L. C.; Urbanc, B. Stabilization of native amyloid β-protein oligomers by copper and hydrogen peroxide induced cross-linking of unmodified proteins (CHICUP). *Biochim. Biophys. Acta, Proteins Proteomics* **2016**, *1864*, 249−259.

(18) Mawhinney, M. T.; Williams, T. L.; Hart, J. L.; Taheri, M. L.; Urbanc, B. Elucidation of insulin assembly in acidic and neutral pH: Characterization of low molecular weight oligomers. *Proteins: Struct., Funct., Genet.* **2017**, *85*, 2096−2110.

(19) Brewer, S. H.; Vu, D. M.; Tang, Y. F.; Li, Y.; Franzen, S.; Raleigh, D. P.; Dyer, R. B. Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 16662−16667.

(20) Harada, R.; Tochio, N.; Kigawa, T.; Sugita, Y.; Feig, M. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J. Am. Chem. Soc.* **2013**, *135*, 3696−3701.
(21) Nagarajan, S.; Xiao, S.; Raleigh, D. P.; Dyer, R. B. Heterogeneity in the folding of villin headpiece subdomain HP36. *J. Phys. Chem. B* **2018**, *122*, 11640−11648.
(22) McKnight, C. J.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S. A thermostable 35-residue subdomain within villin headpiece. *J. Mol. Biol.* **1996**, *260*, 126−134.
(23) Vugmeyster, L.; Do, T.; Ostrovsky, D.; Fu, R. Effect of subdomain interactions on methyl group dynamics in the hydrophobic core of villin headpiece protein. *Protein Sci.* **2014**, *23*, 145−156.
(24) Buscaglia, M.; Kubelka, J.; Eaton, W. A.; Hofrichter, J. Determination of ultrafast protein folding rates from loop formation dynamics. *J. Mol. Biol.* **2005**, *347*, 657−664.
(25) Chung, J. K.; Thielges, M. C.; Fayer, M. D. Dynamics of the folded and unfolded villin headpiece (HP35) measured with ultrafast 2D IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3578−3583.
(26) Fancy, D. A.; Kodadek, T. Chemistry for the analysis of protein-protein interactions: rapid and efficient cross-linking triggered by long wavelength light. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6020−6024.
(27) Bitan, G.; Vollers, S. S.; Teplow, D. B. Elucidation of primary structure elements controlling early amyloid β-protein oligomerization. *J. Biol. Chem.* **2003**, *278*, 34882−34889.
(28) Smith, A. V.; Hall, C. K. Protein refolding versus aggregation: Computer simulations on an intermediate-resolution protein model. *J. Mol. Biol.* **2001**, *312*, 187−202.
(29) Ding, F.; Borreguero, J. M.; Buldyrey, S. V.; Stanley, H. E.; Dokholyan, N. V. Mechanism for the α-helix to β-hairpin transition. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 220−228.
(30) Rapaport, D. C. *The Art of Molecular Dynamics Simulation*; Cambridge University Press: Cambridge, U.K., 1997.
(31) Smith, S. W.; Hall, C. K.; Freeman, B. D. Molecular dynamics for polymeric fluids using discontinous potentials. *J. Comput. Phys.* **1997**, *134*, 16−30.
(32) Zhou, Y.; Karplus, M.; Wichert, J. M.; Hall, C. K. Equilibrium thermodynamics of homopolymers and clusters: Molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.* **1997**, *107*, 10691−10708.
(33) Urbanc, B.; Borreguero, J. M.; Cruz, L.; Stanley, H. E. *Ab initio* discrete molecular dynamics approach to protein folding and aggregation. *Methods Enzymol.* **2006**, *412*, 314−338.
(34) Urbanc, B.; Cruz, L.; Yun, S.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *In silico* study of amyloid β-protein folding and oligomerization. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17345−17350.
(35) Urbanc, B.; Betnel, M.; Cruz, L.; Bitan, G.; Teplow, D. B. Elucidation of amyloid β-protein oligomerization mechanisms: Discrete molecular dynamics study. *J. Am. Chem. Soc.* **2010**, *132*, 4266−4280.
(36) Meral, D.; Urbanc, B. Discrete molecular dynamics study of oligomer formation by N-terminally truncated amyloid β-protein. *J. Mol. Biol.* **2013**, *425*, 2260−2275.
(37) Žganec, M.; Kruczek, N.; Urbanc, B. Amino acid substitutions [K16A] and [K28A] distinctly affect amyloid β-protein oligomerization. *J. Biol. Phys.* **2016**, *42*, 453−476.
(38) Žganec, M.; Žerovnik, E.; Urbanc, B. Self-assembly of globular protein stefin B into polymorphic oligomers probed by discrete molecular dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2355−2366.
(39) Long, K.; Williams, T. L.; Urbanc, B. Insulin inhibits Aβ42 aggregation and prevents Aβ42-induced membrane disruption. *Biochemistry* **2019**, *58*, 4519−4529.
(40) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671−675.
(41) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105−132.
(42) Lam, A.; Teplow, D. B.; Stanley, H. E.; Urbanc, B. Effects of the Arctic (E22 → G) Mutation on Amyloid β-Protein Folding: Discrete Molecular Dynamics Study. *J. Am. Chem. Soc.* **2008**, *130*, 17413−17422.
(43) Barz, B.; Turner, B. S.; Bansil, R.; Urbanc, B. Folding of pig gastric mucin non-glycosylated domains: A discrete molecular dynamics study. *J. Biol. Phys.* **2012**, *38*, 681−703.
(44) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(45) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306−317.
(46) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(47) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(48) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(49) Pall, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations With GROMACS. *Lect. Notes Comput. Sci.* **2015**, *8759*, 3−27.
(50) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(51) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(52) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(53) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(54) Abascal, J.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
(55) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(56) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(57) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(58) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol.* **1997**, *4*, 180−184.
(59) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(60) Ding, F.; Jha, R. K.; Dokholyan, N. V. Scaling behavior and structure of denatured proteins. *Structure* **2005**, *13*, 1047−1054.
(61) Lei, H.; Wu, C.; Liu, H.; Duan, Y. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 4925−4930.
(62) Lei, H.; Duan, Y. Two-stage folding of HP-35 from ab initio simulations. *J. Mol. Biol.* **2007**, *370*, 196−206.

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911

(63) Harada, R.; Takano, Y.; Baba, T.; Shigeta, Y. Simple, yet powerful methodologies for conformational sampling of proteins. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6155−6173.
(64) Wu, K.; Xu, S.; Wan, B.; Xiu, P.; Zhou, X. A novel multiscale scheme to accelerate atomistic simulations of bio-macromolecules by adaptively driving coarse-grained coordinates. *J. Chem. Phys.* **2020**, *152* (11), 114115.
(65) Havlin, R. H.; Tycko, R. Probing site-specific conformational distributions in protein folding with solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3284−3289.
(66) Bunagan, M. R.; Gao, J.; Kelly, J. W.; Gai, F. Probing the folding transition state structure of the villin headpiece subdomain via side chain and backbone mutagenesis. *J. Am. Chem. Soc.* **2009**, *131*, 7470−7476.
(67) Cheignon, C.; Tomas, M.; Bonnefont-Rousselot, D.; Faller, P.; Hureau, C.; Collin, F. Oxidative stress and the amyloid β peptide in Alzheimer's disease. *Redox Biol.* **2018**, *14*, 450−464.
(68) Zhang, S.; Fox, D. M.; Urbanc, B. Insights into formation and structure of Aβ oligomers cross-linked via tyrosines. *J. Phys. Chem. B* **2017**, *121*, 5523−5535.
(69) Zhang, S.; Fox, D. M.; Urbanc, B. Elucidating the role of hydroxylated phenylalanine in formation and structure of cross-linked Aβ oligomers. *J. Phys. Chem. B* **2019**, *123*, 1068−1084.
(70) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discuss.* **2009**, *141*, 251−276.
(71) Vega, C.; Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663−19688.
(72) Barz, B.; Urbanc, B. Minimal model of self−assembly: Emergence of diversity and complexity. *J. Phys. Chem. B* **2014**, *118*, 3761−3770.
(73) Zhang, S.; Schweitzer-Stenner, R.; Urbanc, B. Do molecular dynamics force fields capture conformational dynamics of alanine in water? *J. Chem. Theory Comput.* **2020**, *16*, 510−527.
(74) 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.* **1996**, *118*, 11225−11236.
(75) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

https://doi.org/10.1021/acs.jpcb.1c04589
J. Phys. Chem. B 2021, 125, 6897−6911