

Transferable and Polarizable Coarse Grained Model for Proteins—ProMPT

Abhilash Sahoo,^{||} Pei-Yin Lee,^{||} and Silvina Matysiak*Cite This: *J. Chem. Theory Comput.* 2022, 18, 5046–5055

Read Online

ACCESS |



Metrics & More



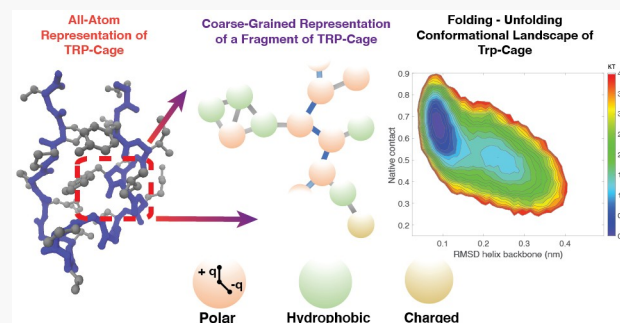
Article Recommendations



Supporting Information

ABSTRACT: The application of classical molecular dynamics (MD) simulations at atomic resolution (fine-grained level, FG), to most biomolecular processes, remains limited because of the associated computational complexity of representing all the atoms. This problem is magnified in the presence of protein-based biomolecular systems that have a very large conformational space, and MD simulations with fine-grained resolution have slow dynamics to explore this space. Current transferable coarse grained (CG) force fields in literature are either limited to only peptides with the environment encoded in an implicit form or cannot capture transitions into secondary/tertiary peptide structures from a primary sequence of amino acids. In this work, we present a transferable CG

force field with an explicit representation of the environment for accurate simulations with proteins. The force field consists of a set of pseudoatoms representing different chemical groups that can be joined/associated together to create different biomolecular systems. This preserves the transferability of the force field to multiple environments and simulation conditions. We have added electronic polarization that can respond to environmental heterogeneity/fluctuations and couple it to protein's structural transitions. The nonbonded interactions are parametrized with physics-based features such as solvation and partitioning free energies determined by thermodynamic calculations and matched with experiments and/or atomistic simulations. The bonded potentials are inferred from corresponding distributions in nonredundant protein structure databases. We present validations of the CG model with simulations of well-studied aqueous protein systems with specific protein fold types—Trp-cage, Trpzip4, villin, WW-domain, and β - α - β . We also explore the applications of the force field to study aqueous aggregation of A β 16–22 peptides.



INTRODUCTION

Physiological functions of protein molecules are closely intertwined with their associated structure and dynamics.^{1,2} This complex macro-organization is shaped through microscopic multibody interactions that define a protein molecule's conformational landscape. In this direction, computer simulations, primarily molecular dynamics, are being increasingly leveraged to understand protein biochemistry and biophysics.^{3–5}

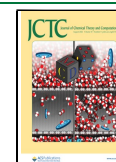
With recent advancements in dedicated high performance computing architectures and graphical processing units, it is now possible to run long simulations with small proteins for multiple microseconds, at conventional all-atom resolution.^{6–9} Such long molecular simulations could capture a small number of protein folding–unfolding transitions in an unbiased manner. But such processes require a significant amount of computing resources and are still not scalable to larger systems. Most biological systems of interest involve large biomolecules that interact over long spatiotemporal scales. To alleviate some of these shortcomings, several enhanced sampling approaches such as metadynamics and replica exchange molecular dynamics have been proposed to allow faster exploration of conformational space with limited resources.^{10–13} But such

approaches require extensive knowledge about the particular biomolecular system and/or can be infeasible for larger systems.

Coarse grained molecular dynamics (CG-MD) involves creating a simplified representation or minimal model of biomolecules that can capture the essential biophysics.¹⁴ This approach allows efficient access to long spatiotemporal scales by directly reducing computational complexity and allowing fast conformational transitions by smoothing the local free energy landscape. Early coarse grained (CG) models with simplified phenomenological potentials were instrumental in establishing the foundations of energy landscape theory of protein folding.^{15–17} The coarse grained models require creating interaction sites that are representative of a particular molecule and defining interaction schemes (potentials) that

Received: March 17, 2022

Published: July 6, 2022



allow these interaction sites to communicate. The interaction potentials, also known as *force fields*, can be variations of knowledge-based (developed from analysis of statistical databases) or physics-derived potentials (created, for example, to fit free energies, partition coefficients, and biomolecular phases).^{18–25} Coarse grained molecular simulations of proteins have played a significant role in shaping our understanding of physiological processes such as protein folding, protein aggregation, and membrane–protein interactions.^{17,26–32}

CG models with varying molecular resolutions and diverse coarse graining strategies have been proposed. The levels of coarse graining can vary from multiple residues represented by a single interaction site to models with representation for all the heavy atoms in a biomolecular system. Several of the CG models employ an implicit description of the solvent environment through interaction potentials.^{21,22,25,33–38} These models have been used to study several protein-based dynamical processes such as folding, adsorption, misfolding, and aggregation. While these models provide a significant reduction in system complexity, an implicit representation of the environment cannot be used to study heterogeneous environmental effects that can be essential in crowded physiological systems. Moreover, the role of solvent in governing thermodynamics and kinetics of protein folding is well documented.^{39–41} In some of the implicit-water CG models, the dynamics is biased toward the native state and cannot capture non-native states.^{16,42–44}

MARTINI is a popular coarse grained force field with an explicit description for solvents and a modular architecture.^{19,45} The interactions are directed through a set of interaction levels created by leveraging environment-dependent free energies (solvation, vaporization, and partitioning free energies). Due to its reasonable accuracy and ease of use, particularly in a heterogeneous and crowded environment, this force field has been widely adopted by the natural science and engineering communities. In the case of proteins, MARTINI has been used to study ligand binding, aggregation, surface absorption, and membrane translocation.^{46–51} However, the model relies on restraining a protein's secondary structures through artificial potentials, and the global/tertiary structure also needs to be restrained to prevent spontaneous unfolding. This prevents any study of dynamical changes to a protein's conformation over the simulation time using MARTINI. While the Gō model has been employed along with the MARTINI force field to study large-amplitude conformational changes, it can suffer from several shortcomings including loss of amino acid identity and insensitivity to point mutations and environmental changes.^{26,52–54}

In our previous publications, we introduced an explicit-solvent polarizable CG model for a selection of amino acids to study secondary structure transitions starting from the primary sequence in the presence of different environmental stimuli such as hydrophobic media, interfaces, and lipid bilayers.^{55–59} In this work, we have formalized our approach to model parametrization, introduced new bonded and nonbonded potentials, updated residue geometry, and extended the representation to all natural amino acids to capture accurate protein tertiary structures.

Our Protein Model with Polarizability and Transferability (ProMPT) consists of two types of interaction sites (beads)—primary coarse grained beads and dummy beads. The basic biomolecule structure is created with primary interaction sites which feature modular architecture and geometry similar to

those of the MARTINI model that allows for easy transferability and are parametrized along the MARTINI scales.^{19,60} Through additional off-center dummy charges to the primary sites that represent polarizable entities, we have introduced an explicit local dipole moment that can result in anisotropic Coulombic interactions similar to hydrogen bonds in a protein's secondary structure. The angle and dihedral potentials are derived from statistical distributions of these features from the protein data bank (PDB). ProMPT can capture secondary and tertiary conformational transitions, along with appropriate intermediate conformations and accurate folding free energy profiles. As such, this CG model can be applied to study spatiotemporally complex biological phenomena and processes involving proteins. In this paper, we present model validations using simulations of small protein structures in an aqueous solvent and aqueous protein aggregation. We have also discussed future applications and current shortcomings of this CG potential.

METHODS

The coarse grained interaction sites follow a basic *MARTINI-influenced* nomenclature, with atom types grouped into polar, neutral, charged, and hydrophobic beads.^{19,60} The polar beads have explicit electrostatic dipoles added through charged dummy particles, constrained to the main interaction site (Figure 1). For parametrization, we initially run 50 ns atomistic

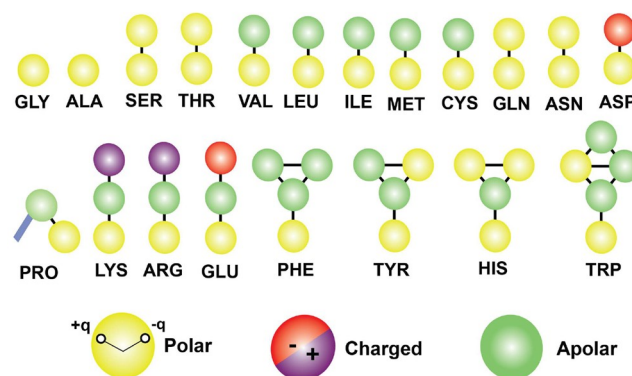


Figure 1. Schematic geometries of coarse grained amino acids.

simulations (using CHARMM36m/TIP3P force field^{61,62} and CHARMM-GUI⁶³ equilibration protocol) of relevant tripeptides that have the corresponding mapped regions in chemical space. The parameters (charge on the beads, q , and distance between the beads, r) are generated to match the dipole moment corresponding to the maximum in the dipole moment distribution of the mapped region from the atomistic simulation. The charge and relevant geometry of these charged dummies are provided in Table S1. The dummy charges interact with the local environment through electrostatic interactions (such as change in dielectric constant at the membrane–water interface), which then directs the protein's structural changes. As these charges are placed off-center to the main interaction site, they can introduce an asymmetry and directionality to interactions and introduce spatial heterogeneity in the local charge distribution. Here, a parallel can be drawn between the direct dipole–dipole and charge–dipole interactions in this CG model and electrostatic alignments in hydrogen bonds. Previous publications with a much simplified variant of this CG model could generate appropriate sequence-

specific secondary structures through alignment of these dipolar charges.^{55–59} Several experiment and simulation-based studies have also underlined the importance of these molecular dipoles in protein folding.

ProMPT has been parametrized to work with Yesylevskyy et al.'s polarizable water model, which maps four atomistic water molecules to three coarse grained beads.⁶⁴ This water model also features dummy charges that communicate exclusively through electrostatics and can interact with our protein model (Figure S1). We have added a brief note on this water model in the Supporting Information.

The amino acid geometry defining the structures of the CG amino acids is shown in Figure 1 and Figure S2. The interaction sites representing the protein backbone and polar side chains are assigned polarizable atom-types, with off-center dummy charges. The primary CG interaction sites that map to aromatic regions in the chemical space have a radius of 0.43 nm, compared to 0.47 nm for all other interaction sites.

Nonbonded Interactions. All the nonbonded Lennard-Jones (LJ) type interactions (potential between site i and site j , $V(r_{ij}) = \frac{c_{12,ij}}{r_{ij}^{12}} - \frac{c_{6,ij}}{r_{ij}^6}$, where r_{ij} is the separation between interaction site i and interaction site j and c_6 and c_{12} are constants) between the atom types are parametrized along the MARTINI interaction levels, which allows for easy transferability and can, therefore, be used with all biomolecular environments that can be represented with the MARTINI force field. A complete description of the nonbonded interaction parameters is presented in the Supporting Information through heatmaps (Figures S3–S8). While most of the cross-interactions between nonpolarizable interaction sites are directly borrowed from the MARTINI force field, the interactions between our polarizable groups had to be reparametrized to balance out the added electrostatic interactions from the charged dummies. Similar to the MARTINI force field, the parametrization here aimed to fit nonbonded interaction parameters to reproduce accurate free energies of solvation and partitioning. The interaction level between hydrophobic groups and the interaction sites representing water is reduced by up to 50% from the MARTINI level to reproduce appropriate environment-induced structural transitions. These reductions balance out issues of overpolarization by backbone dipoles and allow conformational switching from helices to β -strands. We refer readers to the previous publication by Ganesan et al. and Sahoo et al. regarding further details about nonbonded parametrization.^{57,58} The electrostatic charges on specific dummy beads are enumerated in the Supporting Information.

A set of special interactions were added in an *ad hoc* manner to capture specific protein–protein interactions. We applied specific attraction ($\epsilon = 3.0$ kJ/mol) between positively charged groups (such as ions and cationic amino acid side chains) and the aromatic rings to mimic cation– π effects.⁶⁵ Similarly, interactions between aromatic rings and proline side chains with aromatic rings were made attractive ($\epsilon = 3.0$ kJ/mol) to capture π – π stacking and CH– π interactions that have been highlighted in quantum mechanical calculations.^{66–68} These electronic effects are ubiquitous in protein folding. Additional dummy–dummy and dummy–charged group hard-core repulsions ($c_{12} \sim 10^{-7}$ nm) were added to prevent over-interaction between CG groups similar to methods to prevent polarization catastrophe in polarizable force fields.^{69,70} This

allows for fast switching between conformations through binding/unbinding among charged dummies.

Bonded Interactions. Bonded interactions can be broadly categorized as bonds, angles, and dihedrals. The features describing bonds in ProMPT—bond lengths (Table S2) and bond rigidity were borrowed directly either from the MARTINI force field or the previous iteration of this force field.^{55,57,58} The corresponding parameters are enumerated in the SI (Tables S2–S6). Some angular (between backbone beads and first side chain) and dihedral potentials (backbone only) were informed from statistical distributions of protein structures.

We used a nonredundancy p -value of 10^{-7} to create a database of about 14 000 protein structures from the protein data bank. The angular and dihedral potentials were based on their corresponding normalized distributions from this database. The angle between the protein backbone sites was universally set to 109° , restrained through a harmonic potential. The angular potential between the backbone and the side chain was created specific to each amino acid by applying Boltzmann inversion at 300 K to each amino acid specific distribution. These potential energy functions corresponding to each amino acid have been reported in the Supporting Information (Figures S9–S11). Finally, the dihedral potentials between the backbone interaction centers are secondary-structure specific, with different tabulated potentials for α -helix, 3-10 helix, and β -sheets. The functional forms of α -helix and 3-10 helix potentials were taken from previous publications of C α based coarse grained models.⁷¹

$$V(\phi; A, B, C, D) = A[1 + \cos(\phi + \phi_0)] + B[1 + \cos(\phi - \phi_0)] + C[1 + \cos(3(\phi + \phi_0))] + D\left[1 + \cos\left(\phi + \phi_0 + \frac{\pi}{4}\right)\right] \quad (1)$$

The fitted parameters are provided in the Supporting Information (Table S6). These periodic functions allow for introducing multiple local/global minimums, in contrast to single and deep global minima by Boltzmann inversion. We parametrized these functions to set the global functional minimum to match the maximum value in the dihedral probability distributions of relevant secondary structures, in addition to introducing several local minima. These local extremes introduce additional frustrations to the protein's conformations and have been adopted in previous coarse grained models.⁷¹ The tabulated β sheet potential is generated through a direct Boltzmann inversion (at 300 K) of dihedral distributions from regions specific to beta sheets in the protein structure database. Figure 2 shows the dihedral potentials employed in ProMPT.

Simulation Setup. All simulations are performed with GROMACS 2019.4.⁷² The starting conformation and the protein topology are generated with in-house codes, taking the amino acid sequence from the reference PDB structures listed in Table 1. The initial conformation for all the proteins is a random unfolded extended conformation. The setup for each protein system, including the number of protein beads, the number of polarizable water molecules, and the temperature details, is also listed in Table 1. An adequate number of ions are added to neutralize the system. We have used reduced

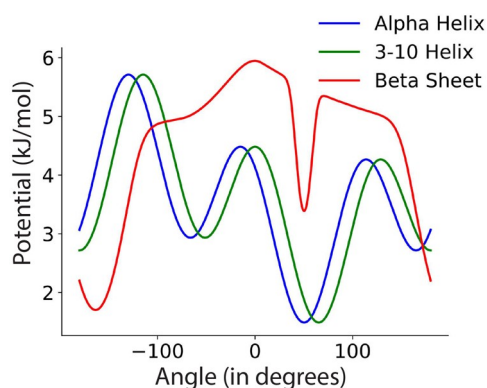


Figure 2. Secondary-structure specific dihedral potential used in the CG force field. The tabulated potentials are fitted (for α -helix and 3-10 helix) or derived (β -sheet) to capture the maximum value in the dihedral probability distributions.

temperature, $T^* = K_b T / \epsilon$, as our temperature scale, where ϵ is the highest interaction strength (5.6 kJ/mol) in our CG potential. The energy minimization is first conducted with the steepest descent. An NPT equilibration run at $T^* = 0.52$ for 50 ps is followed at 1 bar and with a time step 0.001 ps. The canonical production run is then performed with a time step 0.01 ps for the simulation time listed in Table 1 at different reduced temperatures. Since the box size is constant, the solvent density remains the same across various temperatures. The electrostatic cutoff is 1.6 nm, and the particle-mesh Ewald (PME) method is applied for long-range electrostatic interactions.⁷³ A Nosé-Hoover thermostat is used to maintain the system at the desired temperature.⁷⁴

Analysis. We use the potential of mean force (PMF) plots of Trp-cage, Trpzip4, and villin to validate the ability of our CG force field to capture a protein's conformational landscape. Backbone native contact and root-mean-squared deviation (RMSD) of peptide fragments are used as reaction coordinates for PMF calculations. The backbone native contact measures the fraction of backbone contacts with a cutoff of 7 Å that are replicated in our CG trajectories that are native to the PDB-converted-CG structure. The RMSD backbone (BB) is a measure of relative deviation of our CG backbone from the backbone beads in the PDB-converted CG structure. For all proteins, the first 80 ns of trajectory data is removed, and the equilibrated data are collected at all temperatures simulated and reweighted through Multistate Bennett Acceptance Ratio (MBAR) method.⁷⁵ Here, we present the final PMF at $T^* = 0.52$. For villin, we use a cluster analysis tool from GROMACS with a RMSD cutoff of 0.3 nm to generate representative conformation corresponding to each basin in the PMF. For the aggregation simulations, two peptides are considered part of

the β -sheet if more than three dummy–dummy interaction pairs are formed. A fraction of one means that all the peptides (8 of them) are forming a β -sheet.

RESULTS AND DISCUSSION

Miniproteins Trp-cage and Trpzip4 are selected to be our first test proteins because they are already well-studied both computationally and experimentally.^{76–80} Trp-cage has an α -helix starting from the N-terminus, a 3-10 helix in the middle, and an unstructured C-terminal tail, which collapse together forming a hydrophobic core (Figure 3c). Trpzip4 has a β -hairpin structure with four Trp residues packed together. Figure 3a shows the PMF for Trp-cage using native contact and RMSD helix BB as two reaction coordinates. A two-state folding/unfolding path can be observed from the PMF, which agrees with experiments and REMD simulation results.^{76,81} The folded basin at around the 0.1 nm RMSD helix BB indicates a perfect α -helix. The distribution of native contact is relatively broader, from 0.6 to 0.7. The unfolded basin has a native contact of 0.4 to 0.5 and RMSD helix BB from 0.2 to 0.3 nm, indicating some partial helicity in agreement with previous simulation reports.⁸² The PDB structure and a representative folded structure with our CG model for Trp-cage are shown in Figure 3c. Both the α -helix and 3-10 helix can be well reproduced at the correct positions. The hydrophobic core formed by Tyr3, Trp6, and three proline residues at the C-terminus is also well captured as seen in the experimental structure (Figure S12). Our estimation of ΔG for protein folding is 3.2 kJ/mol, close to the experimental estimate of 2.6 kJ/mol.⁸³ We also perform a set of replica exchange with solute tempering (REST) simulations to estimate the folding–unfolding free energy difference for Trp-cage; from the atomistic simulations, the ΔG is estimated to be around 2.81 kJ/mol at 290 K (Figure S13). The time series of the two reaction coordinates we use for PMF at $T^* = 0.82$ are shown in Figures S14 and S15. At this higher temperature ($T^* = 0.82$), our CG model could sample frequent transitions between the folded states and the unfolded states. Figure 3b shows the PMF for Trpzip4 with native contact and RMSD BB as the reaction coordinates. The PMF of Trpzip4 also shows a two-state folding/unfolding path, which agrees with experiments and atomistic simulations. The folded basin for Trpzip4 is located at 0.35 nm RMSD BB with a native contact of around 0.8. The unfolded basin has a broad RMSD BB distribution from 0.5 to 0.7 nm with a native contact of around 0.4. Figure 3d shows both the PDB structure (left) and a representative folded state for Trpzip4 with our model (right). The β -sheet content can be well captured. Four tryptophan residues are interacting with each other and forming a hydrophobic core as observed in the experimental structure (Figure S16). Here, our CG model estimated a ΔG value of about 7.3 kJ/mol for

Table 1. Simulation Setup for Each Protein

	Trp-cage	Trpzip4	villin	WW-domain	β - α - β	A β 16-22
PDB code	1L2Y	1LE3	1YRF	1E0L	2K10	^a
number of proteins	1	1	1	1	1	8
number of water molecules	1957	1610	26391	30136	24998	2584
simulation time per temperature (ns)	200	200	600	100	200	300
temperature range performed	0.52–1.04	0.52–1.04	0.52–1.04	0.52,0.82	0.52,0.82	0.52
total number of temperatures simulated	15	15	19	2	2	1

^aSequence: KLVFFAE.

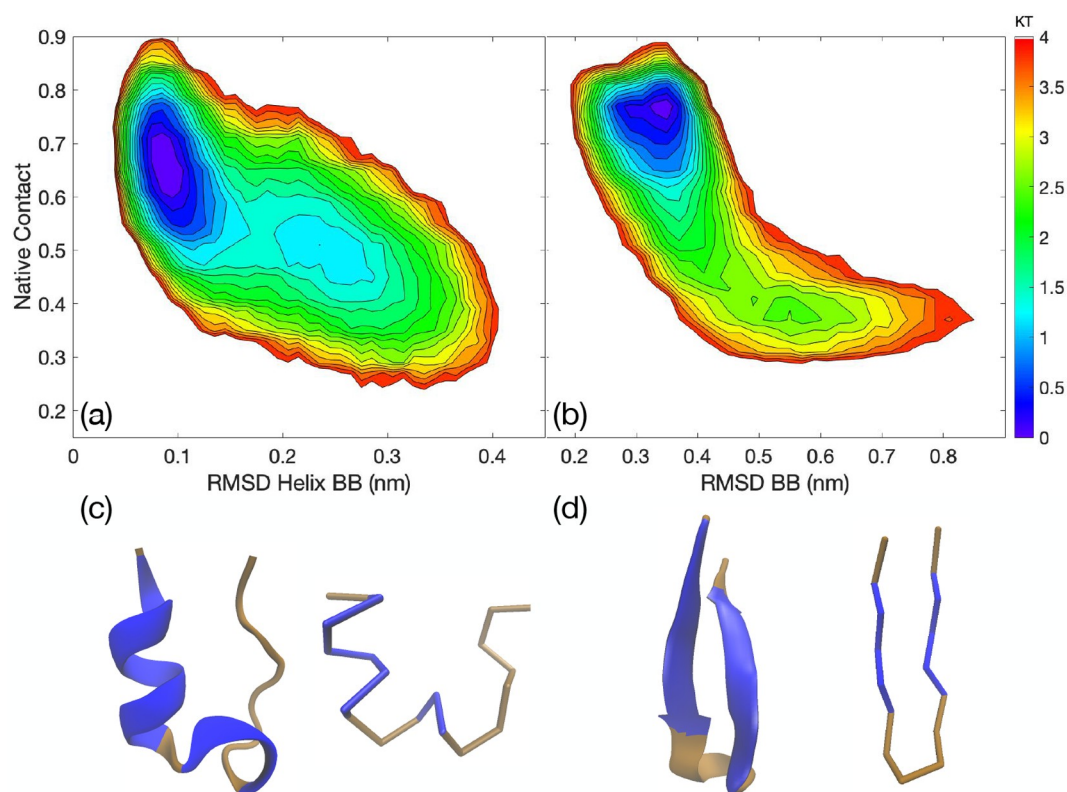


Figure 3. PMF for Trp-cage (a) with RMSD helix BB and native contact as the reaction coordinates. PMF for Trpzip4 (b) with RMSD BB and native contact as the reaction coordinates. Both PDB structure (left) and the representative structure (right) from our model for Trp-cage (c) and Trpzip4 (d) are shown. Blue indicates the specific secondary structure each protein exhibits.

protein folding, compared to 12.3 kJ/mol reported in experiments.⁸⁴ From the REST simulations of Trpzip4, a ΔG value of 4.53 kJ/mol is estimated (Figure S17), which is lower than both that of the experimental data and the estimation from our CG model. Again, the multiple folding–unfolding transitions at a higher temperature ($T^* = 0.82$) for the two reaction coordinates, as time series are shown in Figures S18 and S19. We use the simulations of Trp-cage to evaluate the computational efficiency improved by using our CG model compared to atomistic simulations using an enhanced sampling method. In order to get a converged PMF, we used different protocols for atomistic (REST) and CG (replicas at different temperatures). Therefore, the computational resources needed will also depend on the choices of difference parameters such as the number of replicas and the range of temperatures. Here, we present a simplified comparison of the computational resources used to get the estimate of folding–unfolding free energy. Our CG model improves the computational efficiency by at least 10-fold. This has not taken into account the extra simulations needed to fine-tune the parameters for the REST simulation protocol and the atomistic force field, that will lead to an even higher computational cost in practice. The details of the computational efficiency comparison are provided in the SI. A similar comparison can be made between the resources needed for a single folding event between our ProMPT CG simulations and the previous long time scale unbiased molecular dynamics simulations with all-atom resolution.⁸⁵ Here, the improvement in computational efficiency is almost 21 000 times. Please refer to the note (in the SI) on computational efficiency for further details.

Our next target protein—villin—has three helices with a hydrophobic core. Successful folding of villin validates our CG protein model for general applications to proteins with tertiary packing of homogeneous secondary structures. Figure 4 shows

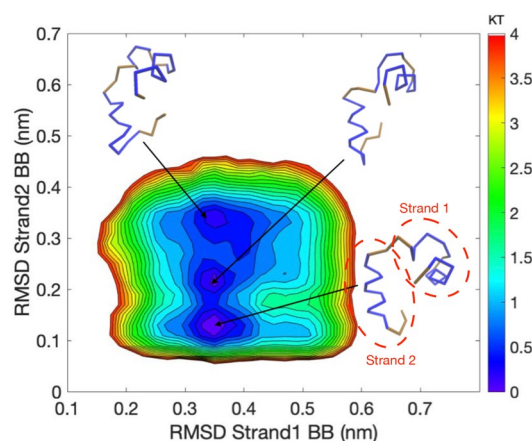


Figure 4. PMF for villin with RMSD S1 and RMSD S2 as the reaction coordinates at $T^* = 0.52$. The representative structures for each basin are shown as insets.

the PMF for villin with RMSD strand 1 (S1) and RMSD strand 2 (S2) as two reaction coordinates, where the locations of the two strands are highlighted in red. Strand 1 contains two helices from the N-terminus (residues 2–18), while strand 2 only contains one long helix (residues 21–32) from the C-terminus. There are one native state and two intermediate

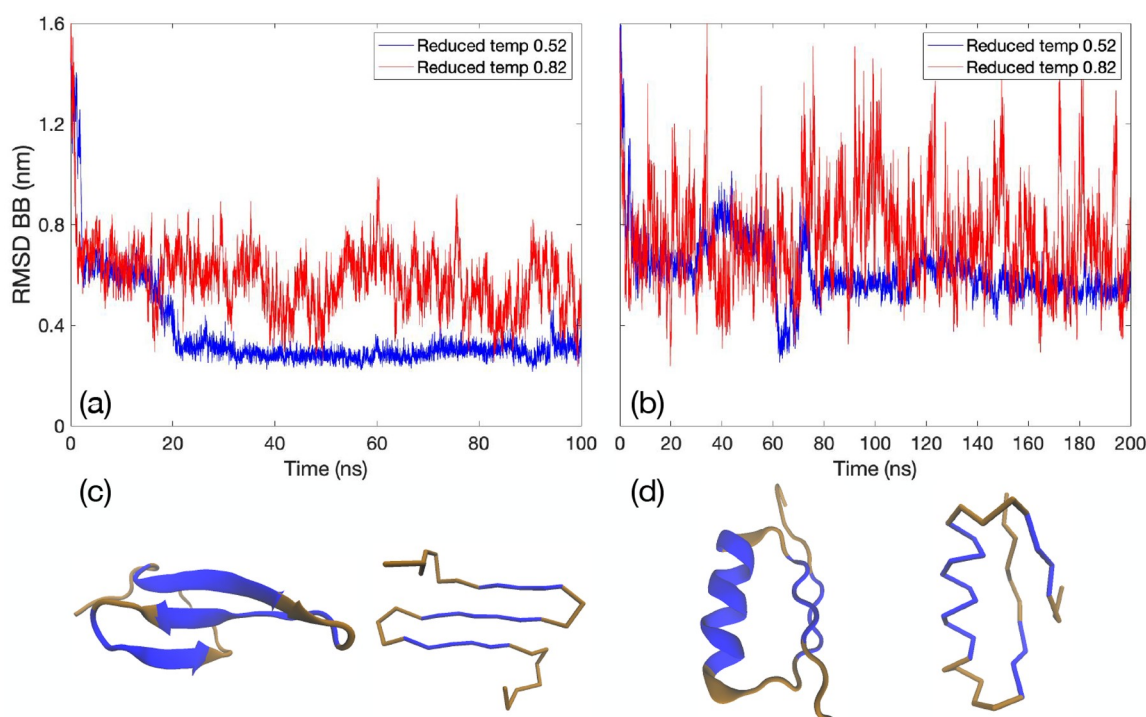


Figure 5. RMSD BB time series for (a) WW-domain and (b) β - α - β at $T^* = 0.52$ (blue) and $T^* = 0.82$ (red). The PDB structure (left) and the representative structure (right) are shown in (c) and (d) for WW-domain and β - α - β , respectively. Blue indicates the specific secondary structure each protein exhibits.

states populated at $T^* = 0.52$. At higher temperatures, the location of the unfolded basin shifts to more extended conformations (Figure S20). All of the states populated at $T^* = 0.52$ have similar RMSD S1 values but different RMSD S2 values, which correspond to different tertiary structures. Indeed, it has been reported in experiments that although there are fluctuations in the N-terminal part of villin, large-scale unfolding will not occur unless undergoing global unfolding.⁸⁶ In our model, global unfolding happens at higher temperatures, where disruption of strand 1 can be found as seen in Figure S20. The native state is centered at RMSD S1 0.35 nm and RMSD S2 0.1 nm. The representative structure of the folded basin with our model is very similar to the experimental structure from PDB. For the intermediate states, one is centered at RMSD S1 0.35 nm and RMSD S2 0.2 nm while the other one is centered at RMSD S1 0.35 nm and RMSD S2 0.35 nm. Both of the intermediate basins show a decrease in helicity for strand 2 and a loss of the hydrophobic core, while the helicity for the basin with larger RMSD S2 is reduced further. The intermediate states are similar to the states described in the experimental work from Reiner and co-workers, where the intermediate states have an unfolded strand 2 that is undocked from the hydrophobic core.⁸⁶ Both the native state and the intermediate states are on the folded side of the folding–unfolding landscape. The frequent transitions between folded states and unfolded states for RMSD S1 and RMSD S2 are shown at $T^* = 0.82$ in Figures S21 and S22.

For the proteins having a tertiary structure with homogeneous β -sheets, we used the WW-domain as our benchmark protein. The PDB structure is shown in Figure 5c with three antiparallel β -strands and two long tails at both ends. A representative folded structure with our model is shown in Figure 5c. In Figure 5a, the RMSD BB time series without considering the loop regions are shown at two temperatures.

From the time series at $T^* = 0.52$, we observe that the simulation converges at a value of RMSD BB around 0.3 nm. The simulation can reach the folded state in about 20 ns. We could note frequent transitions between folded and unfolded states at $T^* = 0.82$. The last protein we focus on is β - α - β (PDB code: 2K10), which is a designed protein that has a mixture of secondary structures (partially α -helix and β -sheet) with a defined tertiary structure. Using DSSP,⁸⁷ we assigned the secondary structures using one of the deposited NMR structures in the PDB database. The antiparallel β -strands only have four residues on each strand. Figure 5d shows a representative folded structure with our model; the α -helix can be well formed, while the β -sheets are formed but shifted. Figure 5b shows the RMSD BB time series for β - α - β at two temperatures, without considering the tails. The RMSD BB converges at around 0.5 nm at $T^* = 0.52$. The time series at a higher temperature ($T^* = 0.82$) is shown in red, which confirms the ability of ProMPT to capture frequent transitions between different states.

We have demonstrated that with our model we can fold a variety of proteins, which include pure β -sheets, pure helices, and proteins that have a defined tertiary structure composed of homogeneous secondary structures. For proteins that have mixed secondary structures such as β - α - β , the packing could not be perfectly reproduced. We also attempted to simulate the folding of another well-studied protein with heterogeneous secondary structures—GB1 (PDB code: 2J52). GB1 is a slightly larger protein with four β -strands and one α -helix. Although our CG model could generate an accurate secondary structure at appropriate regions, the overall tertiary packing was not reproduced (results not shown). We compared the molecular packing of amino acid side chains in ProMPT for the GB1 to GB1 PDB native structure mapped into our CG description. ProMPT simulations, with the backbone re-

strained to the native state, resulted in a protein core with packing defects and voids. On the other hand, the CG structure mapped directly from the PDB native state showed no such anomalies. Therefore, we postulate that appropriate modifications to side-chain size can achieve adequate protein-core packing required for correct tertiary structures. Proteins with mixed folds will be our next target proteins and could potentially be solved by adjusting the size and shape of each amino acid in the CG model to better mimic the side-chain geometry. A recent parametrization of the MARTINI 3 force field has adopted this strategy and included several smaller bead types to capture the accurate geometrical shapes of biomolecules.⁴⁵ This resulted in significant improvement in molecular packing and could capture correct open and close conformations in transmembrane proteins.

Beyond studying protein folding, ProMPT can also be applied to study peptide aggregation. Here we demonstrate the aggregation of A β 16–22 peptides in water at $T^* = 0.52$. The protein concentration in the system is 0.0426 M. Figure 6

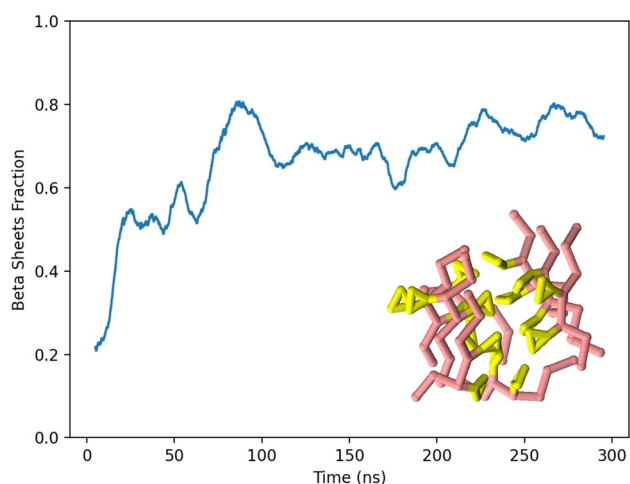


Figure 6. Time series for the number of A β 16–22 peptide forming β -sheets. An illustration of β -sheet aggregation is shown in the inset with yellow representing PHE residues.

shows the time series for the β -sheet fraction. The β -sheet fraction fluctuates between 60% and 80%, indicating significant fibrillation. Occasionally the β -sheet fraction can reach 100%, where all the A β peptides are involved in the β -sheet formation. Similar to previous experimental and simulation studies, peptides organize into stable, layered β -sheet rich structures.^{88–92} PHE residues are shown in yellow, and a hydrophobic core formed by PHE packing can be observed from the aggregate snapshot (inset of Figure 6). We observed antiparallel β -sheets in our simulations, similar to the structures reported by experiments.⁹³ Nguyen et al. applied replica exchange MD on the A β 16–22 monomer, dimer, and trimer to study the effects of different atomistic force fields on the formation of β -sheets.⁹⁴ While AMBER99 and OPLS generated a diversity of conformations, only GROMOS96 could successfully generate antiparallel β -sheets. With ProMPT, we observe aggregation into β -sheet-rich, fibril-like structures by 30 ns and can also qualitatively study the kinetics of this aggregation process.

Here we must note that although similar energy scales are used, comparisons should not be drawn with the MARTINI family of force fields as MARTINI external restraints are

instituted to prevent structural transitions and spontaneous unfolding.

The applications and validations of our CG model in this work focused on capturing an accurate protein conformational landscape and structural diversity in an *aqueous* environment. This explicit treatment of the environment can capture local electrostatic changes such as dielectric fluctuations more effectively, which is necessary for several biomolecular processes.⁶⁴ Moreover, the importance of these local electrostatic changes was noted in protein folding by previous reports.⁹⁵ Also, the use of the explicit environment allows for easy transferability of interactions while simulating proteins in different biomolecular environments. We do not have to reparametrize the protein force fields as the cross-interactions between the protein and the environment would automatically change with changing environments. Therefore, ProMPT can have special applications in studies involving chemically heterogeneous biomolecular systems due to its *explicit* description of the environment (and solvent). Because of similar parametrization of nonbonded interactions, ProMPT can be patched with our *in-house* coarse grained membrane model—water explicit polarizable membrane model (WEP-MEM)—to study membrane and detergent assisted conformational transitions.^{57–59,96} ProMPT can also be patched with our CG model for divalent ions to study metal-assisted protein folding.⁹⁷ Similarly, due to the use of the same nonbonded interaction scales, ProMPT can be used alongside the biomolecules parametrized in the MARTINI universe such as oligonucleotides, lipids, and biological polymers without explicit reparametrization. Finally, the modular architecture of the force field allows for easy parametrization of other biomolecules. Future work in our lab will be focused on studies to elucidate these environment-driven conformational changes in the protein structure and dynamics with this force field.

CONCLUSION

In this article, we developed a polarizable coarse grained model with explicit representation for the environment—ProMPT. The key objective here was to create a coarse grained molecular dynamics force field to capture tertiary folding of protein structures with minimal constraints. The CG mapping scheme follows closely the MARTINI coarse grained force field geometry. In ProMPT, the polar CG beads have explicit drude-like charges that can couple a protein's environment to its structure and dynamics. We parametrized the nonbonded interactions between CG beads through free energies of their interaction with the environment. The bonded interaction potentials were generated through analysis of corresponding normalized bonded-feature distributions generated from a database of nonredundant protein structures from the protein data bank. This CG force field was validated by reproducing the conformational free energy landscape of several well-studied small protein systems. While ProMPT can accurately fold protein structures with homogeneous secondary structures, we observed some side-chain packing defects in the folding of proteins composed with heterogeneous secondary structures. Our future research efforts will be focused on fixing side chain packing through accurate mapping of side chain geometry. Due to its transferability, ProMPT can be patched with our previous *in-house* lipid and divalent ion models to study protein folding assisted by the physiological environment.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00269>.

Details about the ProMPT (coarse grained architecture, bonded and nonbonded potentials) and model validations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Silvina Matysiak – Fischell Department of Bioengineering, University of Maryland, College Park, Maryland 20742, United States; orcid.org/0000-0003-3824-9787; Phone: +1 301-405-0313; Email: matysiak@umd.edu

Authors

Abhilash Sahoo – Biophysics Program, University of Maryland, College Park, Maryland 20742, United States; orcid.org/0000-0002-0115-3194

Pei-Yin Lee – Chemical Physics Program, University of Maryland, College Park, Maryland 20742, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.2c00269>

Author Contributions

[†]A.S. and P.-Y.L. contributed equally to this work.

Funding

We acknowledge financial support from the National Science Foundation under Grant CBET-1760879 and computational resources at the University of Maryland. A.S.'s contribution to this research was supported in part by National Science Foundation Award DGE-1632976.

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) McLachlan, A. D. Protein Structure and Function. *Annu. Rev. Phys. Chem.* **1972**, *23*, 165–192.
- (2) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84*, 551–575.
- (3) Salsbury, F. R., Jr Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr. Opin. Pharmacol.* **2010**, *10*, 738–744.
- (4) Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679–6685.
- (5) Dror, R. O.; Dirks, R. M.; Grossman, J.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429–452.
- (6) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M.; Xu, H.; Trbovic, N.; Friesner, R. A.; Palmer, A. G.; Shaw, D. E. Microsecond Molecular Dynamics Simulation Shows Effect of Slow Loop Dynamics on Backbone Amide Order Parameters of Proteins. *J. Phys. Chem. B* **2008**, *112*, 6155–6158.
- (7) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys. J.* **2008**, *94*, L75–L77.
- (8) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.
- (9) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Shaw, D. E. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J. Phys. Chem. B* **2016**, *120*, 8313–8320.
- (10) Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2020**, *2*, 200–212.
- (11) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (12) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826–843.
- (13) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (14) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898–7936.
- (15) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- (16) Hills, R.; Brooks, C. Insights from Coarse-Grained Gō Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889–905.
- (17) Clementi, C. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10–15.
- (18) Samudrala, R.; Moulton, J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction 1 Edited by F. Cohen. *J. Mol. Biol.* **1998**, *275*, 895–916.
- (19) 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.
- (20) Kar, P.; Gopal, S. M.; Cheng, Y.-M.; Predeus, A.; Feig, M. PRIMO: A Transferable Coarse-Grained Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3769–3788.
- (21) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503.
- (22) Chebaro, Y.; Pasquali, S.; Derreumaux, P. The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins. *J. Phys. Chem. B* **2012**, *116*, 8741–8752.
- (23) Zieba, K.; Ślusarz, M.; Ślusarz, R.; Liwo, A.; Czaplewski, C.; Sieradzan, A. K. Extension of the UNRES Coarse-Grained Force Field to Membrane Proteins in the Lipid Bilayer. *J. Phys. Chem. B* **2019**, *123*, 7829–7839.
- (24) Chan, H.; Cherukara, M. J.; Narayanan, B.; Loeffler, T. D.; Benmore, C.; Gray, S. K.; Sankaranarayanan, S. K. R. S. Machine learning coarse grained models for water. *Nat. Commun.* **2019**, *10*, 379.
- (25) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755–767.
- (26) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937–953.
- (27) Brown, S.; Fawzi, N. J.; Head-Gordon, T. Coarse-grained sequences for protein folding and design. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 10712–10717.
- (28) Morriss-Andrews, A.; Brown, F. L. H.; Shea, J.-E. A Coarse-Grained Model for Peptide Aggregation on a Membrane Surface. *J. Phys. Chem. B* **2014**, *118*, 8420–8432.
- (29) Das, P.; Matysiak, S.; Clementi, C. Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10141–10146.
- (30) Wu, H.; Dalal, Y.; Papoian, G. A. Binding Dynamics of Disordered Linker Histone H1 with a Nucleosomal Particle. *J. Mol. Biol.* **2021**, *433*, 166881.
- (31) Yang, D. S.; Saeedi, A.; Davtyan, A.; Fathi, M.; Sherman, M. B.; Safari, M. S.; Klindziuk, A.; Barton, M. C.; Varadarajan, N.; Kolomeisky, A. B.; Vekilov, P. G. Mesoscopic protein-rich clusters

host the nucleation of mutant p53 amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2015618118.

- (32) Danielsen, S. P. O.; McCarty, J.; Shea, J.-E.; Delaney, K. T.; Fredrickson, G. H. Molecular design of self-coacervation phenomena in block polyampholytes. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 8224–8232.
- (33) Barz, B.; Urbanc, B. Minimal Model of Self-Assembly: Emergence of Diversity and Complexity. *J. Phys. Chem. B* **2014**, *118*, 3761–3770.
- (34) Auer, S.; Meersman, F.; Dobson, C. M.; Vendruscolo, M. A Generic Mechanism of Emergence of Amyloid Protofilaments from Disordered Oligomeric Aggregates. *PLoS Comput. Biol.* **2008**, *4*, No. e1000222.
- (35) Yue, K.; Fiebig, K. M.; Thomas, P. D.; Chan, H. S.; Shakhnovich, E. I.; Dill, K. A. A test of lattice protein folding algorithms. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 325–329.
- (36) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. Coarse-Grained Peptide Modeling Using a Systematic Multiscale Approach. *Biophys. J.* **2007**, *92*, 4289–4303.
- (37) Mashayak, S. Y.; Jochum, M. N.; Koschke, K.; Aluru, N. R.; Rühle, V.; Junghans, C. Relative Entropy and Optimization-Driven Coarse-Graining Methods in VOTCA. *PLoS One* **2015**, *10*, No. e0131754.
- (38) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
- (39) Vaiana, S.; Manno, M.; Emanuele, A.; Palma-Vittorelli, M.; Palma, M. The Role of Solvent in Protein Folding and in Aggregation. *J. Biol. Phys.* **2001**, *27*, 133–145.
- (40) Oshima, H.; Kinoshita, M. Essential roles of protein-solvent many-body correlation in solvent-entropy effect on protein folding and denaturation: Comparison between hard-sphere solvent and water. *J. Chem. Phys.* **2015**, *142*, 145103.
- (41) Nick Pace, C.; Treviño, S.; Prabhakaran, E.; Martin Scholtz, J. Protein structure, stability and solubility in water and other solvents. *Philos. Trans. R. Soc. London, B, Biol. Sci.* **2004**, *359*, 1225–1235.
- (42) Klimov, D. K.; Thirumalai, D. Mechanisms and kinetics of beta-hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2544–2549.
- (43) Klimov, D. K.; Thirumalai, D. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J. Chem. Phys.* **1998**, *109*, 4119–4125.
- (44) Ueda, Y.; Taketomi, H.; Gō, N. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A Three-dimensional lattice model of lysozyme. *Biopolymers* **1978**, *17*, 1531–1548.
- (45) Souza, P. C. T.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388.
- (46) Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein–ligand binding with the coarse-grained Martini model. *Nat. Commun.* **2020**, *11*, 3714.
- (47) Souza, P. C. T.; Limongelli, V.; Wu, S.; Marrink, S. J.; Monticelli, L. Perspectives on High-Throughput Ligand/Protein Docking With Martini MD Simulations. *Front. Mol. Biosci.* **2021**, *8*, 657222.
- (48) Gobbo, C.; Beurroies, I.; de Ridder, D.; Eelkema, R.; Marrink, S. J.; De Feyter, S.; van Esch, J. H.; de Vries, A. H. MARTINI Model for Physisorption of Organic Molecules on Graphite. *J. Phys. Chem. C* **2013**, *117*, 15623–15631.
- (49) Arnarez, C.; Marrink, S. J.; Periole, X. Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes. *Chem. Sci.* **2016**, *7*, 4435–4443.
- (50) Sun, F.; Schroer, C. F.; Xu, L.; Yin, H.; Marrink, S. J.; Luo, S.-Z. Molecular Dynamics of the Association of L-Selectin and FERM Regulated by PIP2. *Biophys. J.* **2018**, *114*, 1858–1868.
- (51) Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Böckmann, R. A. High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2278–2291.
- (52) Karanikolas, J.; Brooks, C. L. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **2002**, *11*, 2351–2361.
- (53) Poma, A. B.; Cieplak, M.; Theodorakis, P. E. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. *J. Chem. Theory Comput.* **2017**, *13*, 1366–1374.
- (54) Mahmood, M. I.; Poma, A. B.; Okazaki, K.-i. Optimizing Gō-MARTINI Coarse-Grained Model for F-BAR Protein on Lipid Membrane. *Front. Mol. Biosci.* **2021**, *8*, 619381.
- (55) Ganesan, S. J.; Matysiak, S. Role of Backbone Dipole Interactions in the Formation of Secondary and Supersecondary Structures of Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2569–2576.
- (56) Ganesan, S. J.; Xu, H.; Matysiak, S. Effect of lipid head group interactions on membrane properties and membrane-induced cationic β -hairpin folding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 17836–17850.
- (57) Ganesan, S. J.; Matysiak, S. Interplay between the hydrophobic effect and dipole interactions in peptide aggregation at interfaces. *Phys. Chem. Chem. Phys.* **2016**, *18*, 2449–2458.
- (58) Sahoo, A.; Xu, H.; Matysiak, S. Pathways of amyloid-beta absorption and aggregation in a membranous environment. *Phys. Chem. Chem. Phys.* **2019**, *21*, 8559–8568.
- (59) Sahoo, A.; Matysiak, S. Effects of applied surface-tension on membrane-assisted A β aggregation. *Phys. Chem. Chem. Phys.* **2021**, *23*, 20627–20633.
- (60) 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.
- (61) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (62) 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.
- (63) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (64) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.
- (65) Gallivan, J. P.; Dougherty, D. A. Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459–9464.
- (66) Zondlo, N. J. Aromatic–Proline Interactions: Electronically Tunable CH/ π Interactions. *Acc. Chem. Res.* **2013**, *46*, 1039–1049.
- (67) Baker, E. G.; Williams, C.; Hudson, K. L.; Bartlett, G. J.; Heal, J. W.; Porter Goff, K. L.; Sessions, R. B.; Crump, M. P.; Woolfson, D. N. Engineering protein stability with atomic precision in a monomeric miniprotein. *Nat. Chem. Biol.* **2017**, *13*, 764–770.
- (68) Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M. S.; Chakrabarti, P. CH– π hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12648–12683.
- (69) Huang, J.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.* **2014**, *5*, 3144–3150.
- (70) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 774–786.
- (71) Yap, E.; Fawzi, N. L.; Head-Gordon, T. A coarse-grained α -carbon protein model with anisotropic hydrogen-bonding. *Proteins* **2008**, *70*, 626–638.
- (72) 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.

- (73) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (74) Posch, H. A.; Hoover, W. G.; Vesely, F. J. Canonical dynamics of the Nosé oscillator: Stability, order, and chaos. *Phys. Rev. A* **1986**, *33*, 4253–4265.
- (75) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (76) Paschek, D.; Hempel, S.; García, A. E. Computing the stability diagram of the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754–17759.
- (77) Huang, J. L.; Noss, M. E.; Schmidt, K. M.; Murray, L.; Bunagan, M. R. The effect of neat ionic liquid on the folding of short peptides. *Chem. Commun.* **2011**, *47*, 8007–8009.
- (78) Day, R.; Paschek, D.; Garcia, A. E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1889–1899.
- (79) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β -hairpins. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578–5583.
- (80) Du, D.; Zhu, Y.; Huang, C.-Y.; Gai, F. Understanding the key factors that control the rate of β -hairpin folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15915–15920.
- (81) Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H. Folding dynamics and pathways of the trp-cage miniproteins. *Biochemistry* **2014**, *53*, 6011–6021.
- (82) Paschek, D.; Nymeyer, H.; García, A. E. Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water. *J. Struct. Biol.* **2007**, *157*, 524–533.
- (83) Wafer, L. N. R.; Streicher, W. W.; Makhatadze, G. I. Thermodynamics of the Trp-cage Miniprotein Unfolding in Urea. *Proteins* **2010**, *78*, 1376–1381.
- (84) Markiewicz, B. N.; Yang, L.; Culik, R. M.; Gao, Y. Q.; Gai, F. How Quickly Can a β -Hairpin Fold from Its Transition State? *J. Phys. Chem. B* **2014**, *118*, 3317–3325.
- (85) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17845–17850.
- (86) Reiner, A.; Henklein, P.; Kiefhaber, T. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4955–4960.
- (87) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules* **1983**, *22*, 2577–2637.
- (88) Latshaw, D. C.; Cheon, M.; Hall, C. K. Effects of Macromolecular Crowding on Amyloid β (16–22) Aggregation Using Coarse-Grained Simulations. *J. Phys. Chem. B* **2014**, *118*, 13513–13526.
- (89) Wang, Y.; Bunce, S. J.; Radford, S. E.; Wilson, A. J.; Auer, S.; Hall, C. K. Thermodynamic phase diagram of amyloid- β (16–22) peptide. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 2091–2096.
- (90) Lin, D.; Luo, Y.; Wu, S.; Ma, Q.; Wei, G.; Yang, X. Investigation of the Aggregation Process of Amyloid- β -(16–22) Peptides and the Dissolution of Intermediate Aggregates. *Langmuir* **2014**, *30*, 3170–3175.
- (91) Samantray, S.; Yin, F.; Kav, B.; Strodel, B. Different Force Fields Give Rise to Different Amyloid Aggregation Pathways in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 6462–6475.
- (92) Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. Amyloid fibril formation by A β 16–22, a seven-residue fragment of the Alzheimer's β -amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* **2000**, *39*, 13748–13759.
- (93) Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. Amyloid Fibril Formation by A β 16–22, a Seven-Residue Fragment of the Alzheimer's β -Amyloid Peptide, and Structural Characterization by Solid State NMR. *Biochemistry* **2000**, *39*, 13748–13759.
- (94) Nguyen, P. H.; Li, M. S.; Derreumaux, P. Effects of all-atom force fields on amyloid oligomerization: Replica exchange molecular dynamics simulations of the A β 16–22 dimer and trimer. *Phys. Chem. Chem. Phys.* **2011**, *13*, 9778–9788.
- (95) Arnarez, C.; Uusitalo, J. J.; Masman, M. F.; Ingólfsson, H. I.; de Jong, D. H.; Melo, M. N.; Periolo, X.; de Vries, A. H.; Marrink, S. J. Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. *J. Chem. Theory Comput.* **2015**, *11*, 260–275.
- (96) Ganesan, S. J.; Xu, H.; Matysiak, S. Influence of Monovalent Cation Size on Nanodomain Formation in Anionic–Zwitterionic Mixed Bilayers. *J. Phys. Chem. B* **2017**, *121*, 787–799.
- (97) Sahoo, A.; Matysiak, S. Microscopic Picture of Calcium-Assisted Lipid Demixing and Membrane Remodeling Using Multi-scale Simulations. *J. Phys. Chem. B* **2020**, *124*, 7327–7335.

Recommended by ACS

Ribonucleic Acid Folding Prediction Based on Iterative Multiscale Simulation

Dinglin Zhang, Guohui Li, *et al.*

OCTOBER 19, 2022

THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS

READ 

Folding Coarse-Grained Oligomer Models with PyRosetta

Theodore L. Fobe, Michael R. Shirts, *et al.*

SEPTEMBER 30, 2022

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

Developing Bonded Potentials for a Coarse-Grained Model of Intrinsically Disordered Proteins

Azamat Rizuan, Jeetain Mittal, *et al.*

SEPTEMBER 06, 2022

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Enhanced Conformational Sampling with an Adaptive Coarse-Grained Elastic Network Model Using Short-Time All-Atom Molecular Dynamics

Ryo Kanada, Yasushi Okuno, *et al.*

MARCH 24, 2022

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

Get More Suggestions >