JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

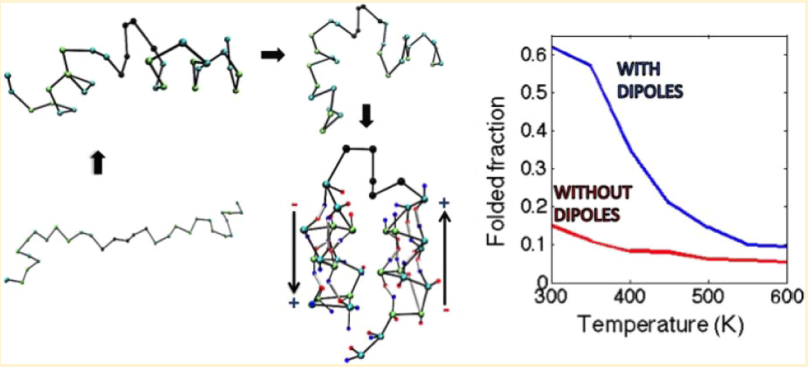
Open Access on 05/09/2015

Role of Backbone Dipole Interactions in the Formation of Secondary and Supersecondary Structures of Proteins

Sai J. Ganesan and S. Matysiak*

Fischell Department of Bioengineering, University of Maryland, College Park, Maryland 20742, United States

Supporting Information



ABSTRACT: We present a generic solvated coarse-grained protein model that can be used to characterize the driving forces behind protein folding. Each amino acid is coarse-grained with two beads, a backbone, and a side chain. Although the backbone beads are modeled as polar entities, side chains are hydrophobic, polar, or charged, thus allowing the exploration of how sequence patterning determines a protein fold. The change in orientation of the atoms of the coarse-grained unit is captured by the addition of two oppositely charged dummy particles inside the backbone coarse-grained bead. These two dummy charges represent a dipole that can fluctuate, thus introducing structural polarization into the coarse-grained model. Realistic α/β content is achieved *de novo* without any biases in the force field toward a particular secondary structure. The dipoles created by the dummy particles interact with each other and drive the protein models to fold into unique structures depending on the amino acid patterning and presence of capping residues. We have also characterized the role of dipole−dipole and dipole−charge interactions in shaping the secondary and supersecondary structure of proteins. Formation of helix bundles and β-strands are also discussed.

INTRODUCTION

Proteins are marginally stable and rely on cooperative effects to keep them in their folded structure.[1−3] A balance between charge interactions, hydrogen bonding, and hydrophobicity is responsible for the formation of stable native folds. A simple description of a hydrogen bond can be based on an electrostatic dipole−dipole interaction. The formation of a protein backbone's hydrogen bonds leads to a particular orientation of dipoles in various types of secondary structural elements.[4] For example, it is well established that the CO and NH dipoles in an α-helix are electrostatically aligned and almost parallel to the axis of the helix.[5] Furthermore, studies have shown that the helical macrodipole can be exploited to stabilize or destroy a helical moiety by charge capping at the chain ends.[6−8] Placement of a positive amino acid at the C-terminus or negative at the N-terminus enhances helicity by a charge−macrodipole interaction.

Protein design studies have revealed that it is possible to design helix-bundle proteins and β-strands considering mainly the pattern of polar and nonpolar amino acids. Statistical analysis of folded structures has shown that nonpolar amino acids appear in the protein sequence, every three to four amino acids in α-helical structures, and every two amino acids in solvent exposed β-sheets.[9,10]

Molecular simulations have been extremely useful in providing a molecular understanding of experimental observations. Coarse-grained (CG) models allow us to understand which interactions are essential and which ones can be approximated. In addition, by reducing the level of resolution, CG models decrease the computational requirements compared to all-atom simulations and smooths out the free-energy landscape, facilitating the sampling from one conformation to another. Simple models where residues are represented by a few beads have been very valuable in advancing our understanding of the protein folding process.[3,10−12] But their main drawback is that the parameters are normally tuned for a particular system and are not transferable. Therefore, they have to be recalibrated according to the system of interest. On the other hand, there exist intermediate resolution models where

Received: December 18, 2013
Published: May 9, 2014

the backbone is modeled with fine resolution and the side chain coarse-grained. Upon inclusion of a hydrogen bonding interaction, these models are able to fold into helical structures without the addition of biases toward a native fold but fail to stabilize β-sheet structures.[13,14] The inclusion of explicit dipole−dipole interactions in these types of models has been shown to stabilize β-sheets.[1,15,16] However, all the above-mentioned models renormalize the role of the solvent through effective short-range, inter-residue interactions. Thus, they are not appropriate for studying the effect of the environment in protein folding because explicit solvent is needed. Also, the importance of the role of dipole interactions is evident from the work done by Scheraga et al. His group has shown that an optimization of the electrostatic interaction, by aligning the dipole of each backbone plane to a protein's electric field, is enough to fold small peptides using Monte Carlo simulations without solvation.[17−19]

The recently developed water-explicit MARTINI CG force field[20] has been parametrized using water/oil and water/vacuum partitioning coefficients. By doing so, the model parameters are transferable and do not depend on a particular system. This CG model captures many lipid membrane and transmembrane protein properties. However, it fails to capture changes in secondary structure and thus it cannot be used to study protein folding.

In this work, we present a generic water-explicit CG model of proteins that can be used to characterize the driving forces behind protein folding without the addition of biasing potentials such as dihedral potentials or dependencies in the bending potential with secondary structure propensities. The model has roots in the MARTINI CG force field and thus it has the potential to be extended to model membrane protein folding. In the current implementation, each amino acid is coarse-grained with two beads, backbone, and side chain. The change in orientation of the atoms underlying the backbone coarse-grained unit is captured by a flexible dipole that is created by oppositely charged dummy particles inside each backbone coarse-grained bead. These dipoles interact with each other through Coulombic potentials and introduce structural polarization into the model. Here we explore if just the addition of effective dipole interactions, in addition to Lennard-Jones interactions, is sufficient to produce secondary and super-secondary folds depending on sequence patterning. We have evaluated the performance of the model by folding protein structures on the basis of the sequence of amino acids. A molecular picture of how charge-capping stabilizes helical structures is also provided.

We use the sequence patterning of helices (-HHPPHH-) and sheets (-HPHPHP-)[9] to fold into distinct secondary structures. We then explore the role of helix capping by terminating helices with charged residues. Helix and sheet bundles are designed on the basis of the above *de novo* patterns with the inclusion of turning regions. The paper is organized as follows: a detailed description of the CG model is presented in the next section. The Results and Discussion are subdivided into two parts. In the first section we look at charge dipole interactions of capped helices. In the next section, we discuss the formation of protein supersecondary structures, namely helix bundles and sheet-strands and compare the results to the model without dipoles, thus drawing attention to the relevance of the backbone dipolar interactions in determining protein structure.

METHODS

Overview. An amino acid is modeled by two beads, a polar backbone bead (BB) that embeds a dipole, and a side chain bead (SC). The SC beads are broadly classified into hydrophobic (H), polar (P), and positively and negatively charged (C+/C−), as depicted in Figure 1a. The protein model

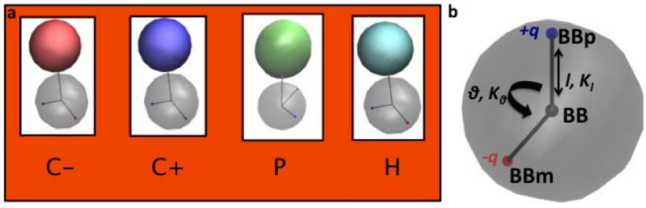


Figure 1. (a) Bead types (left to right). C− is the negatively charged residue, C+ is the positively charged residue, and P and H are the polar and hydrophobic residues, respectively. Gray represents the polarizable backbone (BB) bead and red, blue, green, and cyan represent the side-chain (SC) beads. (b) Polarizable BB bead. VdW radius of the BB bead encloses dummy particles, BBm (negatively charged), and BBp (positively charged). The five tunable parameters (l, q, θ, k_l, k_θ) are depicted.

explores the role of backbone dipoles in driving secondary and supersecondary structure formation. Therefore, the beads P and BB are treated differently. The distinction between H and P beads are purely pairwise interactions. The CG protein model is combined with the recently developed polarizable coarse-grained water model.[21] The spherical backbone coarse-grained bead consists of three interaction sites, the center bead BB and two dipole particles, BBm and BBp, similar to the polarizable water model and depicted in Figure 1b. The main site, the center of the BB bead, interacts with other CG beads through a pairwise Lennard-Jones potential. Dipole particles BBm and BBp are harmonically bound to the central particle BB (equilibrium distance (l), force constant (k_l)), and carry a positive and negative charge of equal magnitude (q), respectively. These dipole particles interact with other particles via electrostatic interactions. A harmonic angle potential (equilibrium angle (θ) and angular force constant (k_θ)) is used to control the rotation of BBm and BBp particles. Because the location of the dipole particles are not fixed, the model is polarizable; i.e., the dipole orientation and moment of the backbone bead are dependent on the electric field of the surrounding environment. The rationale for using a polarizable backbone dipole as opposed to a fixed dipole is to make the dipoles environment sensitive, thus enabling future studies of the structural changes induced by backbone dipoles in different dielectric environments. To avoid overpolarization, a small repulsive core is added to the dipolar particles, as commonly done in polarizable all-atom force fields. Because each main coarse-grained site is covalently bonded to its nearest neighbors by a harmonic bond potential, and adjacent bonds are connected by a harmonic angle potential, all the 1−2 and 1−3 nonbonded interactions are excluded from the main coarse-grained sites and the corresponding embedded dipole particles. In addition, the nonbonded interactions between dipole particles inside the backbone CG bead are excluded as well. The mass of the whole bead (72 amu) is distributed equally among the three particles (24 amu each) in backbone beads.

Force Field Parameters. The force field consists of bonded parameters (harmonic bonded and angular potential;

dihedral potential is not included) and nonbonded parameters (12-6 Lennard-Jones (LJ) potential and Coulombic potential). The dipole moment distribution of the BB bead is parametrized to match that of a peptide bond (3.5 D).[22] It is worth mentioning that with dipole moments of BB less than 3.5 D, helices are not formed. There are five parameters for the backbone CG bead that can be tuned to obtain the desired average dipole moment (l, q, θ, k_l, k_θ). For the backbone CG bead, the following set of parameters yields an average dipole moment of 3.5 D: l = 0.14 nm, q = 0.34, θ = 0°, k_l = 5000 kJ/mol, and k_θ = 7.2 kJ/mol. Each BB bead is covalently bound to the adjacent BB bead by a bond distance of 3.85 Å and k_l = 5000 kJ/mol, distance parametrized from the average Cα−Cα bond distance observed from a 12mer polyalanine atomistic simulation with GROMOS force field.[23] Other bonded interactions and LJ interaction strengths are borrowed from Marrink et al.[20] The SC−BB harmonic bond distance is chosen as 0.4 nm with k_l = 5000 kJ/mol on the basis of a typical two-bead residue in the MARTINI force field. The BB−BB−BB harmonic angle is defined by θ = 105° and k_θ = 75 kJ/mol,[24] and the SC−BB−BB harmonic angle by θ = 100° and k_θ = 75 kJ/mol.[20]

The interaction strength ε of the LJ potential for the different nonbonded interactions is listed in Table 1. An effective size of

Table 1. Interaction Strength ε of LJ Interactions (kJ/mol)[a]

bead (bead type)	BB (P5)	H (C1)	P (P4)	C+ (Qd)	C− (Qa)	W (POL)
BB (P5)	5	2	5	4	4	4.5
H (C1)	2	3.75	2	2.3	2.3	0.2
P (P4)	5	2	5	4	4	5
C+ (Qd)	4	2.3	4	3.5	4	5
C- (Qa)	4	2.3	4	4	3.5	5
W (POL)	4.5	0.2	5	5	5	4

[a]In parentheses is the corresponding MARTINI force-field bead type.

σ = 0.47 nm is assumed for all main CG interaction pairs in the LJ potential. With the interaction strengths taken from the MARTINI force field, sheets are not formed. It has been observed that hydrophobic interactions play a dominant role in stabilizing β-sheet conformations.[25] The most hydrophobic bead type (a.k.a. C1 bead in the MARTINI force field) is not hydrophobic enough to stabilize *de novo* sheet formation in the current model. A 12 mer chain of backbone beads only folds into a helical structure (results not shown). If hydrophobic side chains have low hydrophobicity, then the conformational ensemble is dominated by backbone properties that have an inherent preference over helical conformations. To correct for this, the ε between H beads and water (C1-POL) has been decreased from 2 to 0.2 kJ/mol, and between two H beads (C1−C1) increased from 3.5 to 3.75 kJ/mol, where sheet formation is observed. In addition, the charge SC (C+/C−) backbone interaction strength is decreased from 5 to 4 kJ/mol to observed helix capping effects, as described in the Results and Discussion.

Simulation Setup. All simulations were carried out using the GROMACS package[26] version 4.5.5 and visualized on VMD.[27] All system sizes are reported in the Supporting Information (Table S1). A Nose−Hoover thermostat (time constant = 1 ps) and Parinello−Rahman barostat (time constant = 1 ps, isothermal compressibility = 3×10^{-5} (kJ mol^{-1} nm^{-3})$^{-1}$) were used to keep the temperature and pressure constant at 300 K and 1 bar. Only for the construction of melting curves were *NVT* simulations performed instead of *NPT* from 300 to 600 K. A time step of 5 fs was used in all the simulations, and the neighbor list was updated every 10 steps. The long-range electrostatic interactions with periodic boundary conditions (*xyz*) were calculated by the particle-mesh Ewald method.[28] A global dielectric constant of 2.5 was used. The LINCS algorithm[29] was used to constraint the bonds of the water molecules (between the central CG site and the dipole particles). The starting conformation for all simulations was an extended random coil. For each system, the simulation data was collected over a 320 ns *NPT* run, and the last 300 ns was analyzed. A speedup of 6 is observed vis-a-vis a fully atomistic system. However, because coarse-graining smooths the energy landscape, more transitions from folded to unfolded states are observed in comparison with atomistic simulations.

Electrostatic Field. The alignment of the backbone dipoles with respect to the protein electric field is determined from the angle θ_i between the local electrostatic field, $E_i(r_i)$, and the ith dipole moment (μ_i) using the following equation:

$$\theta_i = \arccos[\mu_i(r_i)\cdot E_i(r_i)/|\mu_i(r_i)||E_i(r_i)|] \quad (1)$$

where

$$\mu_i(r_i) = q_{\mathrm{BBm}}(r_{\mathrm{BBm}_i} - r_i) + q_{\mathrm{BBp}}(r_{\mathrm{BBp}_i} - r_i) \quad (2)$$

The quantities r_{BBm_i}, r_{BBp_i}, q_{BBm}, and q_{BBp} represent position vectors and charge of the two dummy particles and r_i is the position vector:

$$r_i = (|q_{\mathrm{BBm}}|r_{\mathrm{BBm}_i} + |q_{\mathrm{BBp}}|r_{\mathrm{BBp}_i})/(|q_{\mathrm{BBm}}| + |q_{\mathrm{BBm}}|) \quad (3)$$

The electric field due to the protein molecule is computed using Coulomb's law:

$$E_i(r_i) = \sum_j{}' q_j(r_i - r_j)/(|r_i - r_j|^3) \quad (4)$$

where the summation runs over all other charged sites in the structure, except the ones belonging to i, $i \pm 1$, and $i \pm 2$, as these interactions are excluded in the model and do not contribute to the local electric field of the considered dipole i.

RESULTS AND DISCUSSION

Molecular dynamics simulation of different test systems, representing fundamental secondary and supersecondary motifs were performed. Different patterns of polar and nonpolar amino acids, with and without the presence of capping charge residues were used to test the capability of our model to predict different types of motifs, and explore folding landscapes.

Role of Helix Capping. In agreement with previous findings,[9,30−32] a 12mer sequence of -HHPP- folds into an α-helix. Figure 2a depicts the potential mean force (PMF) of the helical system, using two reaction coordinates, the average absolute dihedral angle between backbone beads (50° corresponds to a perfect helix; a comparison of α dihedral angles of CG helix and sheet ensembles with those of PDB structures is shown in Figure S1, Supporting Information) and the i to i + 3 distance of backbone beads, H1[33] (5.5 Å for a perfect helix). The two minima correspond to the folded helical state and the unfolded ensemble. Figure 2b depicts a typical folded conformation of the 300 ns trajectory. Similarly, a 12mer sequence of -HPHP- folds into a β-sheet. Refer to Figure S2 (Supporting Information) for PMF of the sheet system using

dx.doi.org/10.1021/ct401087a | J. Chem. Theory Comput. 2014, 10, 2569−2576

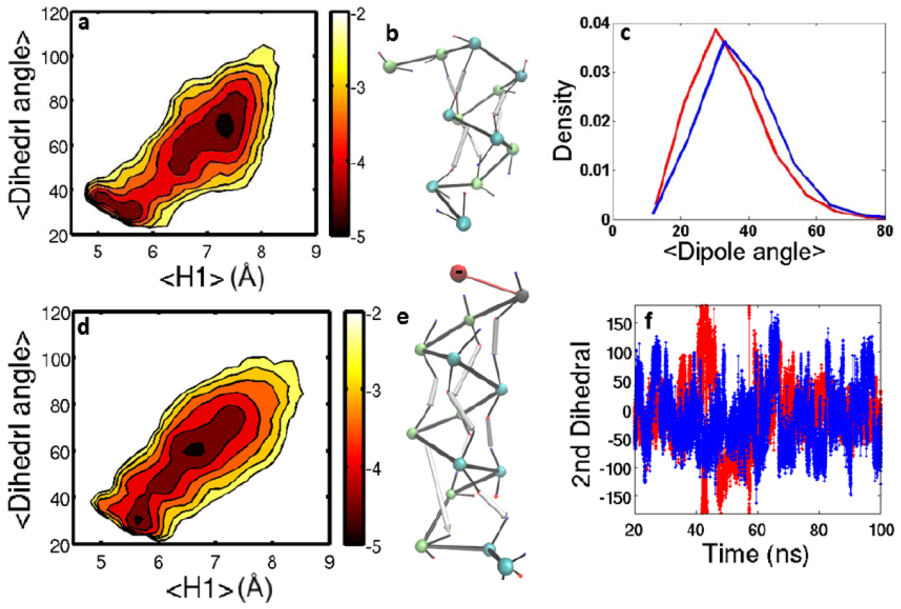


Figure 2. (a) Potential of mean force plot of a noncapped helical polymer at 300 K, as a function of the mean absolute backbone dihedral and H1 parameter. Each contour level marks a kT unit. (b) Representative conformation of the helical peptide in the folded ensemble. Only backbone beads are shown (green for polar and cyan for hydrophobic residues). The arrows indicate dipole orientation. (c) Probability distribution of angles between oriented dipoles in the folded ensemble, of the capped system (red) and noncapped helical polymer (blue). (d) Potential of mean force plot of a capped helical protein at 300 K, as a function of mean absolute backbone dihedral and H1 parameter. (e) Representative conformation of the capped helical peptide in the folded ensemble. Only backbone beads are shown (green for polar and cyan for hydrophobic residues), except for the negatively charged residue where the side chain is shown in red. (f) Comparison of the time evolution of the second dihedral angle (counting from the charged end) between the capped system (red) and noncapped system (blue).

two reaction coordinates, pair contacts (Q) and end-to-end distance (L_c).

To explore charge capping effects, we introduced at one end of the 12-mer peptide a negatively charged amino acid (C−). The PMF plot for the capped system is shown in Figure 2d and displays two distinct minima that correspond to a fully folded helical state and a partially unfolded state that has some helical content. It is evident from a comparison of Figure 2a,d, that capping one end with a charge residue leads to an increase in helicity in the folded ensemble because the minima of the folded state is shifted toward H1 = 5.5 Å. Also, upon helix capping the unfolded state becomes partially folded with a shift of H1 distances to lower values indicating more helical content.

Because the dipoles in the model act as pseudo hydrogen bonds, we also characterize the microdipole angles present in the folded helical structures. A dipole vector is defined from the negatively charged dipole bead of BB i (BBm) to the positively charged dipole bead of BB i+4 (BBp), as depicted by the arrows in Figure 2b,e. In Figure 2c, the effect of charge capping is evident from the probability distribution of dipole angles, which shows better alignment of the dipole angles of the capped system over the noncapped system. The hydrogen bonds in helices of crystal structures have the CO and NH dipole aligned, and almost parallel to the axis of the helix. Therefore, the enhancement of dipole alignment with capping signals an increase in helicity.

The time evolution of the second dihedral angle (counting from the charged end) depicted in Figure 2f clearly shows that the charged end contains a helical turn, which is stable throughout the whole trajectory, even in the partially unfolded state that helical turn is seen to exist, most of the time. On the other hand, the second dihedral angle of the noncapped system displays more coil-like conformations, as evident from the large fluctuations from 50°. As shown in Figure 2e, the negatively charged SC bead of C-residue (red bead) groups the positively charged dipole particles of bonded neighboring backbone beads and creates a helical turn. It is worth mentioning that similar results are obtained when a positively charged amino acid is used at one end. In that case, the dipole particles of bonded neighboring BB beads reorient pointing the negatively charged dipole particles toward the positively charged SC bead, creating a helical turn (results not shown). Our model mimics the fact that the N-terminus of a helix has amide groups that lack hydrogen bonds and the C-terminus has carbonyl groups that also lack hydrogen bonds; therefore, the four residues of the ends are unable to satisfy the classical α-helix hydrogen bonding interactions without proper capping, thus causing helix destabilization.[34,35] Our results indicate that the enhancement of helicity by charge capping is due to a charge−dipole interaction. The dipole alignment caused by charge capping promotes the formation of subsequent dipoles, thus promoting helix stabilization. This observation is in agreement with experimental data, showing that positively charged amino acids have a preference to the C terminus and, similarly, negatively charged side chains to the N terminus.[36] However, it is important to note that our model cannot distinguish N- and C-termini, as the BB beads are identical. As mentioned in the Methods, the LJ interaction strength between the charge side-chain and backbone beads (Qa/Qd-P5 pair interaction of the MARTINI force field) had to be decreased, to allow for stabilization of the helical structure, as observed in Figure 2d. Without this change, the C− SC bead was interacting strongly with the backbone beads by LJ interactions through the main CG sites, which did not allow for the dipole particles of the bonded nearest neighbor BB beads to reorient toward the negative charge that creates a helical turn (results not shown).

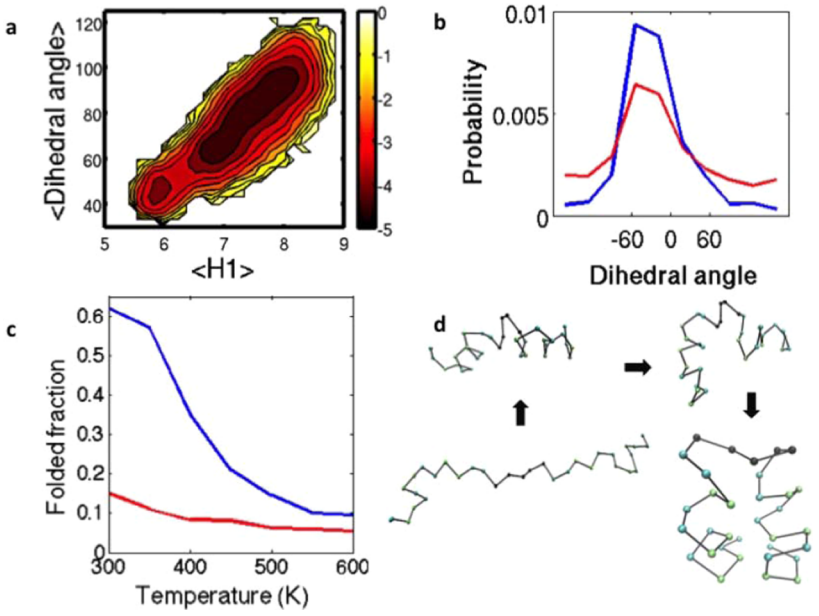


Figure 3. (a) Potential of mean force plot of a helix bundle at 300 K, as a function of the mean absolute backbone dihedral angle and the H1 parameter. Each contour level marks a kT unit. (b) Probability distribution of mean absolute dihedral angle in the folded ensemble of each of the two (red, blue) helices in the bundle. (c) Folding curve of a helix bundle with dipoles (blue) and without dipoles (red). (d) Stages in helix bundle formation, starting from a coil-like conformation. Only backbone beads are shown (green for polar and cyan for hydrophobic residues).

To further evaluate the effect of dipolar interactions in stabilizing secondary folds, we have computed the angle θ_i between the local electrostatic field, $E_i(r_i)$ and the ith dipole moment (μ_i) using eq 1. The folded ensembles of the 12-mer α-helix and 12-mer β-sheet are used for this calculation. An angle of 30.1 ± 8.7° is observed for the α-helix and 54.8 ± 12° for the β-sheet. These values agree very well with the reported values from Scheraga et al. on a nonredundant set of PDB structures with atomistic resolution.[17] The authors found that the angle made between the peptide bond dipole and the protein's local electrostatic field is approximately 35° for α-helices and 50° for β-sheets. Furthermore, the average dipole energy ($\langle -(\mu E) \rangle$) of the CG helix is 4.5 times higher than that of the CG sheet. This value is slightly greater than the one reported from atomistic PDB structures by Scheraga et al. (average dipole energies between helix and sheets was found to be between 2.64 and 3.24).[17] However, the same effect is observed; helices are more stabilized by the backbone dipoles than sheets. Also, helices are better aligned with the electric field than sheets.

Supersecondary Structures. *Helix Bundles.* Because the peptide with the repeating sequence of -HHPP- folds into a helix without any added bias, we have tested our model in predicting supersecondary structures such as helix bundles. Hydrophilic loops are very common in proteins, and statistical analysis of protein structures has shown that the most common loop lengths in helix bundles are between three and six amino acids.[37−39] Taking all this into account, we have designed turning regions (T) with five polar residues. We have tested the behavior of our model in *de novo* folding of a 33 residue peptide with the sequence $(HHPP)_{14}$-$(T)_5$-$(PPHH)_{14}$. As shown in Figure 3d, the chain folds into an α-helix bundle in a hierarchical manner where initially the peptide folds into two distinct helices on either end. This is followed by the collapse of the chain into the helix bundle. Similar results were obtained when a turning region of three, four, or six residues was used (results not shown). The initial folding of the helices is induced by the orientation of the backbone dipoles, which is a local effect. Therefore, helix and supersecondary structure formation is uncoupled. Clearly, the secondary structure is sufficiently stable in the absence of tertiary interactions; thus, folding proceeds in a hierarchical manner as observed for several proteins, such as the engrailed homeodomain.[40,41] It is evident from the PMF (Figure 3a), using as reaction coordinates the average absolute dihedral angle between backbone beads and the distance H1, that there is a distinct folded and unfolded state. The unfolded state is very broad due to the fact that unfolded structures have different degrees of helical content. This is in agreement with experimental observations, that for some helix bundle systems such as α3D, the thermal unfolded state is biased toward local helical content.[42]

It is known that right- or left-handed helices are present in helix bundles[43] with more proportion in the nature of right-handed helices. For the sequence pattern explored here, we observe that the helix bundles have either right- or left-handed helices (Figure 3b). This is because of the minimalist description of the backbone, which does not have any chirality. By adding the microdipoles of each helix (arrows in Figure 4b) of the helix bundle, we have computed the macrodipole that exists in each helix. Figure 4 depicts the probability distribution of the angle between the two macrodipoles (big black arrows in Figure 4b) for folded conformations. The macrodipole angular distribution has a peak around 160°, which signals an almost antiparallel dipole orientation that provides a favorable electrostatic interaction energy for helix bundle formation. Known helix bundles of protein structures in the literature, exhibit the same antiparallel orientation.[44]

To further explore the effects of dipolar interactions in the folding of helix bundles, we have computed the folded fraction as a function of temperature in Figure 3c. A folded structure is considered if the average backbone dihedral of the whole protein structure is between 40° and 60° (Figure S1b

dx.doi.org/10.1021/ct401087a | J. Chem. Theory Comput. 2014, 10, 2569−2576

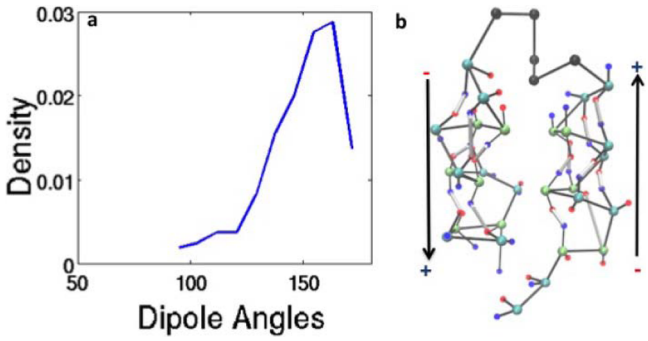


Figure 4. (a) Probability distribution of angles between macrodipole vectors of helices (taking the helix as a whole) in the helix bundles. (b) Representative structure of a helix bundle, white arrows are between i and i + 4 dummy particles. The black arrows represent the macrodipole vectors of the helices in the bundle.

(Supporting Information) and Figure 3a). The blue curve in Figure 3c is for our model, and the red one is without the inclusion of dipole particles. Clearly, without backbone dipoles, the folded fraction is almost constant at around 0.1 folded fraction at all temperatures. The effect of the addition of backbone dipoles is that the folding process becomes cooperative, achieving a folding fraction of 0.6 at 300 K, as evident from the sigmoidal shape of the blue curve. Furthermore, Figure S3a (Supporting Information) shows that without the inclusion of dipoles, the system (T = 300 K) does not fold into a helix bundle, a representative of a collapsed structure is shown as an inset of that figure. Also, the system without the dipole particles does not exhibit any helicity.

Clearly, the folding of helical bundles is influenced by both the effect of dipolar interactions and pairwise LJ interactions; i.e., it is a combination of the sequence patterning and electrostatic interactions. However, we observe that even though there are two major energetic contributions, the effect of dipoles is largely significant in the systems considered. To further elaborate, we characterized these two energetic quantities at 300 K as a function of the reaction coordinate H1. It is clear from Figure 5b that the effect of dipoles has a marginal effect on the LJ energetic contribution, which is lowest in the helical ensemble. It is also interesting to note that the helical ensemble is more stabilized by the Coulombic contribution (Figure 5d), as can be seen by the larger energetic difference between folded (H1 around 6 nm) and unfolded conformations. Also, the presence of dipoles increases cooperativity in folding the helical bundles, this is supported by both (a) the large drop in Figure 5d, seen between folded and unfolded ensemble and (b) the sigmoidal behavior displayed by the melting curves of the helix bundles with dipoles (Figure 3c).

Sheet Bundles. Finally, we tested the ability of the model in predicting β-strand bundles, solely on the basis of sequence patterning. Statistical analysis of β-sheet structures has shown that hydrophilic short turning regions of two amino acids are very common.[45−47] Therefore, to model sheet bundles, we decided to use two polar residues to create turns. The sequence pattern HPHPHTTHPHPHTTHPHPH folds without any biases into a three-strand antiparallel β-sheet, consisting of two hairpins, whose structure is reminiscent of the β3s protein.[48,49] This sequence pattern is common in solvent-exposed β-sheets.[9] The folding of the three-stranded β-sheet

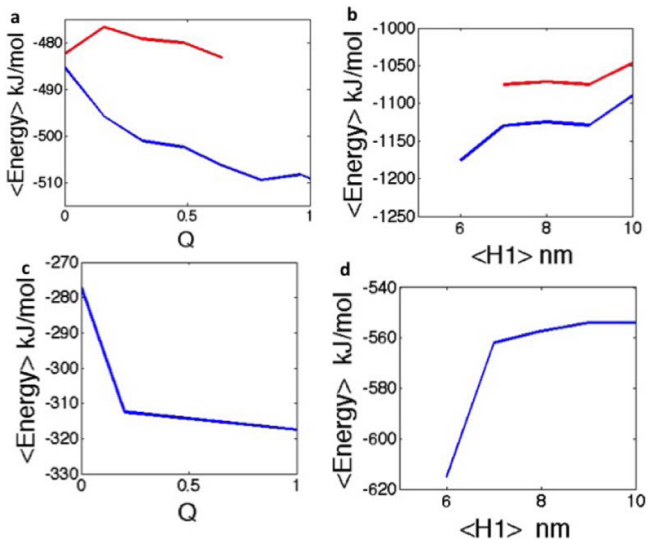


Figure 5. (a) LJ energetic contribution of sheet bundles with (blue) and without dipoles (red), as a function of sheet contacts (Q). (b) LJ energetic contribution of helix bundles with (blue) and without dipoles (red), as a function of H1. (c) Coulombic energetic contribution of sheet bundles with dipoles, as a function of Q. (d) Coulombic energetic contribution of helix bundles with dipoles, as a function of H1. All represented ensembles obtained at 300 K.

occurs through a hierarchical order of hairpin formation (Figure 6c). One hairpin is formed first with the second hairpin forming at the last stage of folding, in agreement with all-atom simulations of β3s.[48] It is worth noticing the matching of hydrophobic and polar residues in the folded structure. Figure 6a shows the PMF as a function of the fraction of sheet contacts (Q) and end to end distance (L_c). Q is defined as the fraction of sheet pairs formed in a bundle. Two residues on neighboring strands within 4−7 Å are defined as a pair. As shown in the PMF (Figure 6a), this protein exhibits a two-state folded behavior, characterized by the folded and unfolded state. It is noteworthy that short sequences with a single turning region (HPHPHTTHPHPH) fold into a β-hairpin and four-stranded β-sheets can be folded by adding a third turning region and an extra-HP- sequence (results not shown).

The melting curve of the sheet bundle exhibits a sigmoidal behavior characteristic of cooperative processes, as shown by the blue curve of Figure 6b. A folded structure is defined as structures with $Q > 0.75$. A control run without the inclusion of dipole particles was also performed using the sequence HPHPHTTHPHPHTTHPHPH. As depicted by the red curve of Figure 6b, without backbone dipoles the folded fraction remains almost constant at around 0.1. It is also important to note that the control run at T = 300 K does not fold into a β-sheet nor exhibit any secondary structure, as shown in Figure S3b (Supporting Information). To further characterize the driving forces behind the folding of sheet-bundles, we have characterized the energetic contributions (LJ and Coulombic) as a function of Q at 300 K. It is evident from Figure 5a that the presence of backbone dipoles has an effect on the LJ contribution of the protein, which is minimum in the fully folded state. Figure 5c depicts the changes in Coulombic energy as a function of Q. By comparing Figure 5c and Figure 5d (helix bundles), one can say that Coulombic energy plays a larger and a more cooperative role in helix bundles. Therefore, in this model the secondary and supersecondary structures are

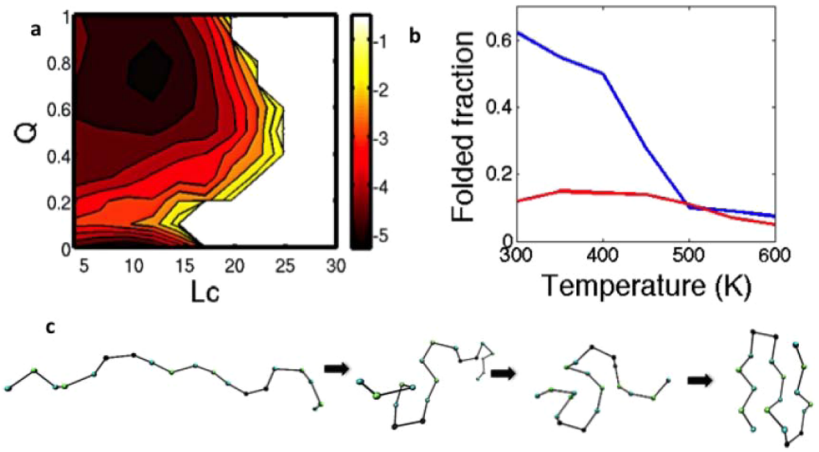


Figure 6. (a) Potential of mean force plot of a sheet bundle at 300 K, as a function of the fraction of sheet contacts (Q) and end-to-end distance (Lc). Each contour level marks a kT unit. (b) Folding curve of a sheet bundle with dipoles (blue) and without dipoles (red). (c) Stages in sheet bundle formation, starting from a coil-like conformation. Only backbone beads are shown (green for polar and cyan for hydrophobic residues).

driven by the presence of the dipoles and the sequence patterning; however, the effect of the added dipoles is more significant. The explored sequences do not fold into helix or sheet bundles in the absence of backbone dipoles.

CONCLUSIONS

In conclusion, the presented model introduces polarization into the protein backbone coarse-grained beads, which allows for the *de novo* folding of secondary structure and supersecondary structure assemblies based solely on the primary sequence. We have shown that our model is capable of folding helix bundles and β-sheet strands. Because of the minimalist description of the backbone, the helices fold with either right- or left-handedness. The chirality can be fixed by introducing *ad hoc* a quadratic term in the potential involving only quadruplets of successive backbone beads.[50] We have also shown that secondary structure formation is driven by backbone dipole interactions and sequence patterning. The charge−dipole interactions are found to be crucial in stabilizing helical folds. The presence of backbone dipoles increases the cooperativity of the folding process. Interestingly, helical conformations are more stabilized by dipolar interactions than β-strands. Our model is currently composed of four bead types for the side chain; however, it is possible to extend the model with more degrees of polarity and hydrophobicity (by varying LJ parameters), thus mimicking more amino acids.

Our model provides a step forward towards the development of a transferable coarse-grained force field for proteins without biases towards secondary structure. In addition, because some of the model parameters were borrowed from the MARTINI coarse-grained force field, our new model has the potential to be extended to model membrane protein folding. Also, because dipole interactions are influenced by the dielectric environment, we expect the proposed model to be sensitive to the nature of the environment (low or high dielectric).

Future work will focus on exploring more complex folds like $\beta-\alpha-\beta$ and protein aggregation. Work along these lines is currently being pursued in the lab.

ASSOCIATED CONTENT

Supporting Information

Table of system sizes. Figures of probability distribution of α dihedral angles and of the potential of mean force for a beta hairpin and for the helix and sheet bundle sequences without the inclusion of backbone dipoles. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*S. Matysiak: e-mail, matysiak@umd.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported in part by the UMD-NCI Cancer Partnership for Cancer Technology and by the National Science Foundation through XSEDE resources provided by The Texas Advanced Computing Center (TACC) under grant numbers TG-MCB110075 and TG-MCB120045. We also thank Sudi Jawahery for the initial parametrization of the polarizable backbone bead.

REFERENCES

(1) Bereau, T.; Deserno, M. *J. Chem. Phys.* **2009**, *130*, 235106.
(2) Matysiak, S.; Das, P. *Phys. Rev. Lett.* **2013**, *111*, 058103.
(3) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(4) Hol, W. G. J.; Halie, L. M.; Sander, C. *Nature* **1981**, *294*, 532−536.
(5) Sheridan, R. P.; Allen, L. C. *Biophys. Chem.* **1980**, *11*, 133−136.
(6) Aqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *88*, 2026−2030.
(7) Sali, D.; Bycroft, M.; Fersht, A. R. *Nature* **1988**, *335*, 740−743.
(8) Oommachen, S.; Ren, J.; McCallum, M. *J. Phys. Chem. B* **2008**, *112*, 5702−5709.
(9) West, M. W.; Hecht, M. E. *Protein Sci.* **1995**, *4*, 2032−2039.
(10) Bellesia, G.; Jewett, A. I.; Shea, J.-E. *Protein Sci.* **2009**, *19*, 141−154.
(11) Matysiak, S.; Clementi, C. *Arch. Biochem. Biophys.* **2008**, *468*, 29−33.
(12) Head-Gordon, T.; Brown, S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 160−167.
(13) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. *J. Chem. Phys.* **1999**, *110*, 11616.
(14) Irback, A.; Sjunnesson, F.; Wallin, S. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 13614−13618.
(15) Mu, Y.; Gao, Y. Q. *J. Chem. Phys.* **2007**, *127*, 105102.

(16) Chen, N.-Y.; Su, Z.-Y.; Mou, C.-Y. *Phys. Rev. Lett.* **2006**, *96*, 078103.
(17) Ripoll, D. R.; Vila, J. A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7559−7564.
(18) Piela, L.; Scheraga, H. A. *Biopolymers* **1987**, *26*, S33−S58.
(19) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283−1303.
(20) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(21) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, e1000810.
(22) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman Publishers: New York, 1992.
(23) Scott, W. R.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(24) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 10712−10717.
(25) Narayanan, C.; Dias, C. L. *J. Chem. Phys.* **2013**, *139*, 115103.
(26) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(27) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(28) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(29) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 33−38.
(30) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680−1685.
(31) Wei, Y.; Liu, T.; Sazinsky, S. L.; Moffet, D. A.; Pelczer, I.; Hecht, M. H. *Protein Sci.* **2003**, *12*, 92−102.
(32) Roy, S.; Ratnaswamy, G.; Boice, J. A.; Fairman, R.; McLendon, G.; Hecht, M. H. *J. Am. Chem. Soc.* **1997**, *119*, 5302−5306.
(33) Kemp, J. P.; Chen, Z. Y. *Phys. Rev. Lett.* **1998**, *81*, 3880.
(34) Aurora, R.; Rose, G. *Protein Sci.* **1998**, *7*, 21.
(35) Forood, B.; Feliciano, E. J.; Nambiar, K. P. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 838−842.
(36) Doig, A. J.; Baldwin, R. L. *Protein Sci.* **1995**, *4*, 1325−1336.
(37) Wilmot, C.; Thornton, J. *J. Mol. Biol.* **1988**, *203*, 221−232.
(38) Efimov, A. V. *Curr. Opin. Struct. Biol.* **1993**, *3*, 379−384.
(39) West, M. W.; Wang, W.; Patterson, J.; Mancias, J. D.; Beasley, J. R.; Hecht, M. H. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11211−11216.
(40) Clarke, N. D.; Kissinger, C. R.; Desjarlais, J.; Gilliland, G. L.; Pabo, C. O. *Protein Sci.* **1994**, *3*, 1779−1787.
(41) Daggett, V.; Fersht, A. R. *Trends Biochem. Sci.* **2003**, *28*, 18−25.
(42) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C.-Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 15486−15491.
(43) Cole, B. J.; Bystroff, C. *Protein Sci.* **2009**, *18*, 1602−1608.
(44) Sheridan, R. P.; Levy, R. M.; Salemme, F. R. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 4545−4549.
(45) Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4236−4237.
(46) Alba, E. D.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 854−865.
(47) Gellman, S. H. *Curr. Opin. Struct. Biol.* **1998**, *2*, 717−725.
(48) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. *J. Phys. Chem. B* **2008**, *112*, 8701−8714.
(49) Carr, J. M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 8760−8769.
(50) Bellesia, G.; Fedorov, M. V.; Kuznetsov, Y. A.; Timoshenko, E. G. *J. Chem. Phys.* **2005**, *122*, 134901.