JCTC Journal of Chemical Theory and Computation

Protein Backbone Dynamics Simulations Using Coarse-Grained Bonded Potentials and Simplified Hydrogen Bonds

Tap Ha-Duong*

Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement Université d'Evry-Val-d'Essonne Rue du Pere André Jarlan, 91025 Evry Cedex, France

Received August 5, 2009

Abstract: A new set of bonded potentials is introduced to model the flexibility of coarse-grained polypeptide chains. Based on a statistical analysis of known structures, the bonded potentials are sequence-dependent, and the secondary-structure propensity of each amino acid is partially reflected in the S_i—B_i—B_{i+1}—B_{i+2} pseudotorsion angle, where S_i and B_i denote the side-chain and backbone beads, respectively. To stabilize the secondary structures during simulations, the bonded force field must be balanced by a simplified model of the protein hydrogen bonds, based on dipole−dipole interactions. Tested on eight polypeptides with sequence lengths ranging from 17 to 98, using 200-ns molecular dynamics simulations, the coarse-grained model yields trajectories with RMSDs ranging from 3 to 8 Å from the experimental conformations. The less-structured regions of the simulated proteins exhibit the largest-amplitude movements.

1. Introduction

It is widely accepted that the conformational dynamics of backbone proteins plays an important role in their biological functions.[1] To mention just one example among many, the opening and reclosing motion of the two flaps that protect the active site of the HIV-1 protease is one of the key steps of its enzymatic mechanism.[2,3] The allosteric effect in proteins, which regulates their biological activities, is also well-known to involve conformational rearrangements of their backbones[4] and/or alterations of their dynamic properties.[5] These backbone motions can be probed by experimental techniques, particularly NMR spectroscopy, which can measure the angular mobility of the N—H bonds.[6] Protein structural changes can also be examined by theoretical methods, such as classical molecular dynamics simulations.[7] Nevertheless, both NMR and computational approaches generally meet difficulties in studying the broad-amplitude and long-time-scale movements of large proteins. Hence, the development of novel methodologies to investigate the functional internal motions of large biomolecular systems is still a very active research field.

Coarse-grained (CG) models of polymers,[8] particularly proteins, are now very popular, as the reduction in the number of particles enhances the exploration of phase space,[9] accelerates computer calculations, and provides insight into biological processes occurring on up to a microsecond time scale.[10,11] Among such simplified models, those at the residue level, which describe each amino acid with one or a few beads, succeed in combining computational efficiency with realistic descriptions of protein structural details. Thus, since the pioneer work of Levitt in 1976,[12] a large number of CG protein force fields have been developed, mainly to tackle the protein folding issue, but also to simulate the conformational dynamics of large proteins.[13,14] CG protein models have also been applied to the protein−protein docking problem, as bead softness can implicitly account for side-chain local flexibility and improve the predictions of matching interfaces between quasirigid proteins.[15]

The reliability of reduced protein models depends on a fine balance between the various force-field terms, and their interpretative strength relies on a simple formulation and a clear separation of the different physical driving forces. As in classical atomic models, CG protein force fields usually have a nonbonded (or long-range) contribution, which includes van der Waals and electrostatic interactions, and a

* E-mail: thaduong@univ-evry.fr.

10.1021/ct900408s
Published on Web 02/04/2010

bonded (or short-range) contribution, which determines the local geometry and flexibility of the polypeptide chains.[12,16−24] Whereas a physical basis can guide the building of nonbonded potentials between protein coarse grains,[25−29] the empirical parametrization of the bonded terms is not straightforward, as their ability to reproduce protein secondary and tertiary structures depends on the details of the nonbonded interactions, particularly the hydrogen bonds.

The flexibility of CG proteins can be efficiently modeled using elastic network models, which replace all of the interactions between pairs of beads that are separated by a distance lower than a cutoff parameter with quadratic potentials. Despite their simplicity and ease of implementation in molecular modeling programs, these one-parameter models can capture the essential features of the functional low-frequency large deformations of proteins.[30−34] One drawback of most elastic network models, however, is the absence of any explicit reference to the sequence of proteins, which hinders the study of the influence of mutations on protein dynamic behaviors. In addition, these models cannot describe large anharmonic motions and possible binding-induced structural changes of the polypeptide chains. To study these latter phenomena, one can hardly avoid the development of bonded potentials.

Effective bonded force fields for CG flexible proteins can be divided into two families that differ in terms of the level of resolution. In the high-level group, the peptide backbone is generally described with three united atoms, one for the nitrogen and its hydrogen atom, another for the α-carbon and its hydrogen atom, and the third for the carbonyl carbon and its oxygen atom.[9,22,23,35,36] In these descriptions, the backbone dihedral angles are similarly defined as in atomic models and can be directly calibrated to reproduce the Ramachandran energy landscapes. In addition, the two backbone united atoms NH and CO allow for the natural introduction of the hydrogen-bond interactions that stabilize secondary structures.[22,35,37] In the low-resolution models, the amino-acid backbone is represented with a single bead. In that case, the residue propensity to form secondary structures has to be implicitly encoded in the backbone pseudobending and/or pseudotorsion potentials,[18,24,31,38] and the hydrogen-bond stabilizing effect has to be introduced through empirical potentials[3,17,39] or electrostatic interactions.[16,38] These one-bead backbone models significantly reduce the number of local minima in the conformational space and generate overall less-frustrated energy landscapes.[9] However, in addition to the difficulty of calibrating the balance between the various energetic terms, most of these models require more or less preliminary information about native secondary or tertiary structures to simulate the protein dynamic conformation, such as in refs 3 and 24. Except for a few studies including those carried out by Scheraga, Liwo, and co-workers[40,41] and the recent one by Majek and Elber,[38] the conformational stability and dynamics of CG polypeptidic chains over long trajectories has seldom been examined using off-lattice unbiased one-bead backbone models.

This article presents an effort to build a general empirical bonded force field of CG proteins that allows their conformational changes to be studied by means of molecular dynamics (MD) simulations. This model of polypeptide flexibility is the natural continuation of the CG nonbonded potential that was recently derived from an all-atom force field by Basdevant et al.[29] The CG bonded potentials are completed with a simplified model of hydrogen bonds, formulated in terms of dipolar interactions and not biased toward any particular protein conformation or secondary structure. The CG protein model does not yet include a consistent description of solvation, especially to account for hydrophobic effects. Using instead a distance-dependent dielectric function as a crude model of hydration, the aim of this work is to bring out a minimal set of CG physical potentials that can reproduce the dynamic stability of proteins with MD simulations. This study therefore primarily focuses on the equilibrium structural properties of CG proteins and compares them with experimental observations, principally those provided by NMR spectroscopy.

2. Methods

2.1. Description of the Amino Acids. In the CG protein model presented herein, each amino acid is described with one bead for the backbone atoms and one or two beads for the side chain, depending on its size (Figure 1).[29] Each bead is located at the geometric center of the heavy atoms that it represents. As in classical all-atom force fields, it is assumed that the nonbonded contribution can be separated and expressed as sums of pairwise nonpolar and Coulombic energy functions depending on the grain-to-grain distances. In the present study, all of the beads are neutral, except for those corresponding to the two terminal backbone residues ($Nter^+$ and $Cter^-$) and the extremities of the charged side chains (ARG^+, ASP^-, GLU^-, and LYS^+). In the following discussion, the backbone grains are denoted B_i, and the side-chain grains are denoted S_i and S_i^*. The coarse-grained nonpolar potentials were obtained by numerical integration of the mean Lennard-Jones forces between pairs of amino acids. To calculate these latter quantities, 20 all-atom MD simulations for all possible homologue pairs of amino acids were performed in vacuo, using fully flexible molecules and a vanishing charge in order to capture the purely nonelectrostatic interaction. Then, to obtain a numerically tractable expression for the nonpolar energies, we fit all of the computed potentials of mean force with a unique mathematical function. The function that was found to best fit all 29 potentials, regardless of the size and softness of the coarse grains, consists of a repulsive part in r^{-6} and a Gaussian attractive part

$$V_{\mathrm{vdW}}(r_{ij}) = \varepsilon_{ij}\left\{\left(\frac{\lambda_{ij}}{r_{ij}}\right)^6 - \exp\left[-\left(\frac{r_{ij}}{\sigma_{ij}}\right)^2\right]\right\} \qquad (1)$$

The values of the parameters ε_{ii}, λ_{ii}, and σ_{ii} for the self-van der Waals interactions, as well as comparisons between CG energies and averaged all-atom Lennard-Jones potentials, are given in ref 29. To include cross-interactions and keep the number of parameters as small as possible, the empirical Lorentz−Berthelot mixing rules were applied: $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$, $\lambda_{ij} = (\lambda_{ii} + \lambda_{jj})/2$, and $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$.

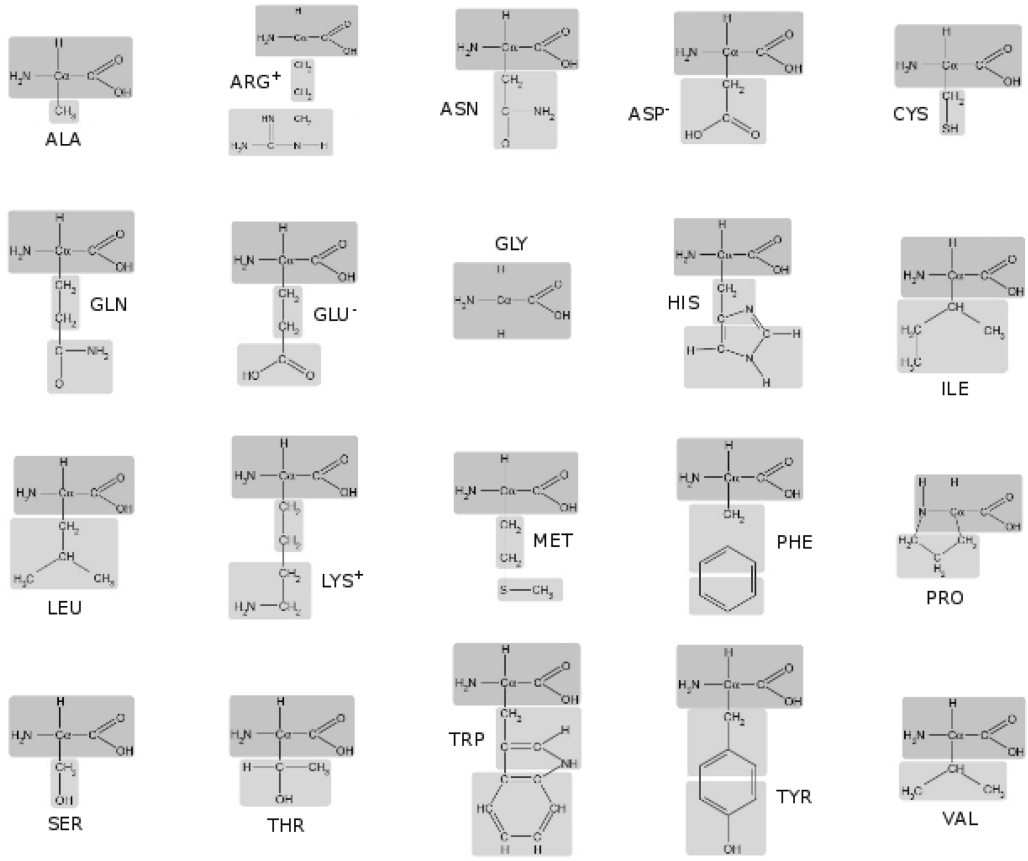

Figure 1. Presentation of the coarse-grained amino acids. Each gray rectangle represents a bead, whose position is located at the geometric center of the heavy atoms that form the coarse grain.

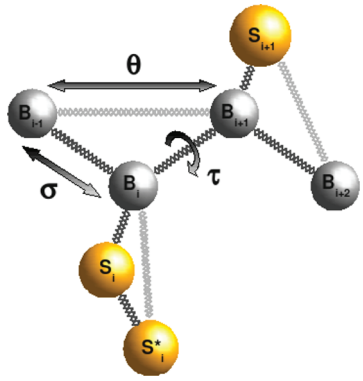

Figure 2. Schematic representation of the CG polypeptide model with its pseudobonds, pseudobends, and pseudotorsions. The backbone beads are denoted B_i, and the side chain beads are denoted S_i and S_i^*. Dark gray springs describe pseudobonds between the coarse grains, whereas light gray springs represent distance-dependent potentials used to account for the bending conformations.

2.2. Determination of the Bonded Potentials. The bonded potentials presented here are subdivided into three contributions (Figure 2), namely, a bond term depending on the length σ of all of the pseudobonds, a bending term expressed as a function of the distance θ between grains separated by two successive bonds, and a torsion term for all dihedral angles: $V_{\text{bonded}} = \sum V_{\text{bon}}(\sigma) + \sum V_{\text{ben}}(\theta) + \sum V_{\text{tor}}(\tau)$. To estimate these energy functions, a statistical analysis was first performed on a nonredundant set of 550 experimental structures of proteins (listed in Table 1 of the Supporting Information). Then, the knowledge-based potentials were extracted from the normalized probability distribution functions P, using Boltzmann inversion procedures: $V = -\gamma k_B T \ln(P)$, where k_B is the Boltzmann constant and T is the temperature. This approach, introduced by Miyazawa and Jernigan,[42] has been employed by many groups mainly to build short-range and long-range energy potentials that enable the discrimination of protein native folds from decoy structures.[20,43−46] In this study, this method was used to extract only the profiles of empirical bonded energy functions that govern the local conformation of the CG protein backbone, as the nonbonded contributions were previously determined from an all-atom protein model.[29] The empirical factors γ are weight scaling parameters introduced to balance the bonded energies against the nonbonded potentials that were determined using another approach. Preliminary trials on small peptides showed that, to generate stable MD trajectories, the bonded potentials needed to be stiffened using γ factors higher than 1. The tests revealed that values of $\gamma_{\text{bon}} = 4$, $\gamma_{\text{ben}} = 8$, and $\gamma_{\text{tor}} = 6$ for the bond, bending, and torsion potentials, respectively, can yield satisfactory results. Finally, the bonded potentials were fitted with tractable mathematical expressions to be implemented in an MD algorithm.

The database of 550 proteins includes both bound and unbound structures from the Protein Data Bank,[47] having less than 50% sequence similarity, refined to a crystallographic resolution lower than 3 Å, with no missing or unresolved heavy atoms. This set represents about 158000

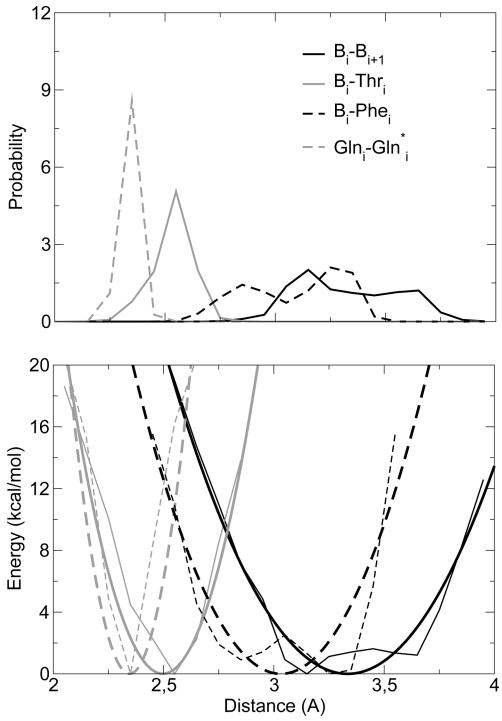

Figure 3. Top: Probability distribution functions for various pseudobond types. Bottom: Associated potentials (thin lines) and fitting energy functions (thick lines).

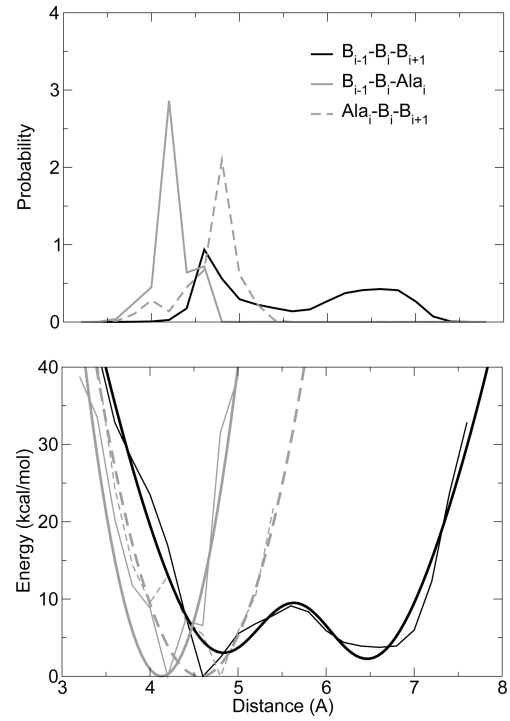

Figure 4. Top: Probability distribution functions for various pseudobend types. Bottom: Associated potentials (thin lines) and fitting energy functions (thick lines).

amino acids, of which 37% are considered in α-helices, 34% in β-strands, and 29% in nonstructured coils.

2.3. Bond, Bending, and Torsion Energy Functions. Figure 3 shows the probability distribution functions for the B_i—B_{i+1} and three others bond types. Most of the bond probability distributions have a peak shape around a single equilibrium value, whereas a few of them present a bimodal profile, including some virtual bonds B_i—S_i involving large side chains (Phe, Trp, Tyr) and especially the B_i—B_{i+1} bond distribution whose two peaks are associated with the helical and extended conformations of the backbone. Nevertheless, the length values between the two maxima are also significantly populated, and Boltzmann inversion yields a bond potential whose the second minimum is not very pronounced. Therefore, a single-well quadratic function was assumed to be a reasonable approximation of the B_i—B_{i+1} energy, as well as of all the other bond potentials (Figure 3)

$$V_{\text{bon}}(\sigma) = K_0(\sigma - \sigma_0)^2 \tag{2}$$

The energy functions associated with the bending angles were replaced with Urey−Bradley-like potentials depending on the distance between beads separated by two successive bonds (Figure 2). Preliminary works using bending-angle-dependent probability distributions generated a B_{i-1}—B_i—B_{i+1} energy profile whose second minimum associated with the β-structures was not very well-defined. For this reason, distance-dependent potentials are used instead of the angle-dependent ones, in order to better account for the two preferential B_{i-1}—B_i—B_{i+1} bending conformations. Because the side chains have a nonsymmetrical orientation relative to the direction of the protein backbone, it is important to differentiate between the B_{i-1}—B_i—S_i and S_i—B_i—B_{i+1} bending types, as illustrated in Figure 4. The bending probability distribution functions generally have a single-peak profile, except for eight bends that have a clear bimodal shape (B_{i-1}—B_i—B_{i+1}, B_{i-1}—B_i—His_i, B_{i-1}—B_i—Phe_i, B_{i-1}—B_i—Trp_i, B_{i-1}—B_i—Tyr_i, Asp_i—B_i—B_{i+1}, B_i—Phe_i—Phe^*_i, and B_i—Tyr_i—Tyr^*_i). Therefore, all of the bending potentials were fitted with quadratic functions, except the previously mentioned ones, for which double-well functions better fit the potentials (Figure 4). These functions were built by adding a harmonic term and a Gaussian term

$$V_{\text{ben}}(\theta) = K_1(\theta - \theta_1)^2 + K_2 \exp\left[-\left(\frac{\theta - \theta_2}{\theta_3}\right)^2\right] + K_3 \tag{3}$$

It should be noted that the two minima of the B_{i-1}—B_i—B_{i+1} generic energy function (Figure 4) are characteristic of α-helix and β-coil secondary structures, and that the tendency of the B_{i-1}—B_i—B_{i+1} triplet to adopt an α- or β-coil local conformation is indirectly modulated by the two other bends B_{i-1}—B_i—S_i and S_i—B_i—B_{i+1}. Furthermore, the bimodal profiles of the B_{i-1}—B_i—B_{i+1} probability distribution and potential were found by both Levitt[12] and Scheraga and co-workers[48,49] to be correlated with those of the adjacent backbone torsions. Nevertheless, it appears quite difficult to estimate how strong this correlation is and the extent to which it is a causal correlation rather than being due to a third physical factor, such as hydrogen bonding or other interactions. Thus, it is not clear whether this correlation has to be explicitly included in the CG bonded force field. The first version of the CG model presented herein neglects this correlation and assumes that the backbone internal coordinates are all independent degrees of freedom. Neglect of this

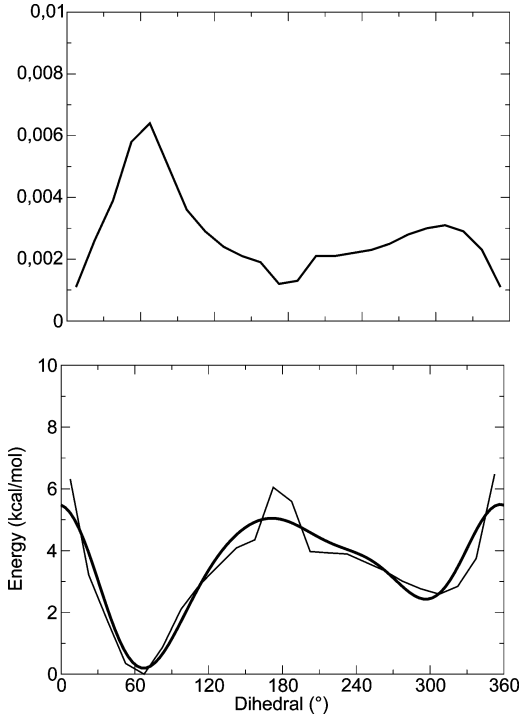

Figure 5. Top: Probability distribution function for the B_{i-1}—B_i—B_{i+1}—B_{i+2} pseudotorsion type. Bottom: Associated potential (thin line) and fitting energy function (thick line).

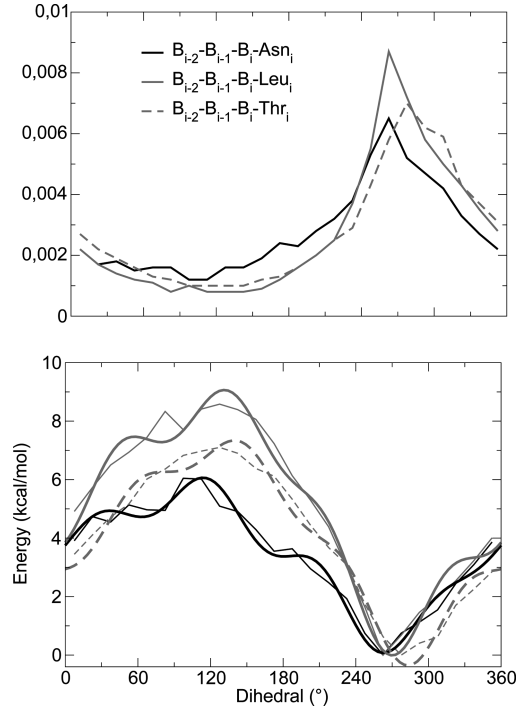

Figure 6. Top: Probability distribution functions for three B_{i-2}—B_{i-1}—B_i—S_i torsion types. Bottom: Associated potentials (thin lines) and fitting energy functions (thick lines).

correlation might have led to the need to stiffen the CG bonded potentials using the weight scaling parameters γ introduced in the previous section to generate stable MD trajectories. At present, further developments are being conducted to study the infuence of an explicit implementation of this correlation on the performance of the force field.

The pseudotorsion potentials are critical in determining a close description of the protein backbone secondary-structure propensity and possible large-amplitude dynamics. In the proposed model, the conformation of each backbone dihedral angle is locally determined by three potential types: the B_{i-1}—B_i—B_{i+1}—B_{i+2} torsion, which is sequence-independent, and the two distinct torsions B_{i-2}—B_{i-1}—B_i—S_i and S_i—B_i—B_{i+1}—B_{i+2}, which involve the side-chain grains. This approach differs significantly from other CG pseudotorsion models, which introduce only one potential for the B_{i-1}—B_i—B_{i+1}—B_{i+2} dihedral angle but whose the shape and parameters depend on the nature of the amino acids S_i and S_{i+1}.[16,18,20,38] The presented model allows for the consideration of only $1 + (2 \times 19)$, rather than 20×20, energy functions.

The torsion B_{i-1}—B_i—B_{i+1}—B_{i+2} probability distribution function and its associated potential (Figure 5) have general shapes similar to those found in several previous studies,[18,20] with two preferential conformations. However, it should be emphasized that, whereas the first minimum (around 70°) is clearly associated with α-helical structures, the second one (around 300°) predominantly corresponds to unstructured coils and widely overlaps the β-conformations, which have backbone torsion mean values around 190°. Hence, by itself, the B_{i-1}—B_i—B_{i+1}—B_{i+2} potential cannot yield stable β-strand structures, and these latter need to be stabilized by other local interactions. Compared to previous similar studies,[18,20] the difference in the position of the second minimum of the B_{i-1}—B_i—B_{i+1}—B_{i+2} potential arises slightly from the location of the B_i grains, which is at the geometric center of the backbone groups and not the Cα atoms. [9] It also certainly comes from differences in the number and proportion of α/β-coils in the protein databases that were used to extract the potentials. This is probably a significant drawback of knowledge-based approaches, but despite this limitation, they allow empirical energy functions to be generated easily and provide instructive insights into protein local short-range interactions.

For all of the amino acids except proline, the B_{i-2}—B_{i-1}—B_i—S_i probability and energy functions have similar profiles with a single preferential conformation between 240° and 300° (Figure 6). It should be noticed here that all of these torsion energies allow for the maintenance of the local chirality of the backbone beads B_i (with the exclusion of the Gly, N*ter*, and C*ter* residues). Indeed, for a given B_{i-2}—B_{i-1}—B_i—B_{i+1} backbone dihedral conformation, the position of the side chain S_i relative to the three beads B_{i-1}, B_i, and B_{i+1} is energetically determined by the single minimum of the torsion potential B_{i-2}—B_{i-1}—B_i—S_i (in addition to the bond length B_i—S_i and the two bending potentials B_{i-1}—B_i—S_i and S_i—B_i—B_{i+1}). In contrast, the S_i—B_i—B_{i+1}—B_{i+2} torsion types present more various behaviors that clearly depend on the nature of the amino acid. Their torsion potentials generally have two minima, one between 60° and 90° associated with β-strands and coils and a second around 210° that is characteristic of α-helices (Figure 7). One can distinguish three tendencies among these energy functions: For amino acids such as Ala, Arg, Glu, Gln, Lys, and Met, the minimum associated with the β-strand and coil conformations is clearly less deep than the α-related minimum. On the contrary, for residues such as Asn, Asp,

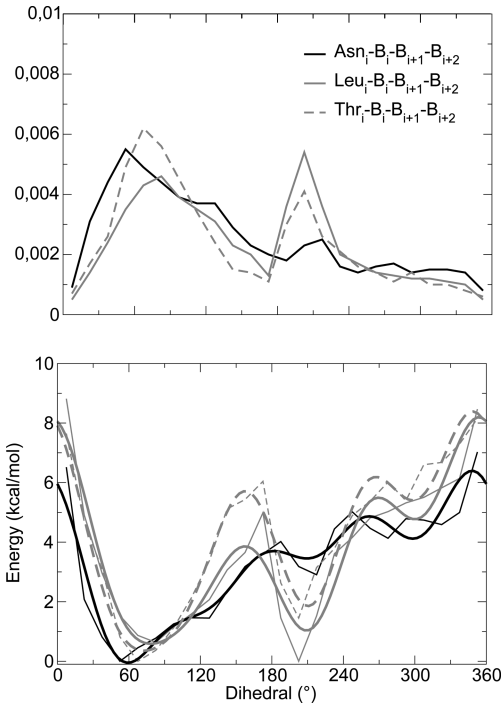

Figure 7. Top: Probability distribution functions for three S_i—B_i—B_{i+1}—B_{i+2} torsion types. Bottom: Associated potentials (thin lines) and fitting energy functions (thick lines).

Cys, Pro, Ser, and Thr, the well associated with the α structures is unfavored relative to the other one. In the third group, including the His, Ile, Leu, Phe, Trp, Tyr, and Val amino acids, the two minima coexist to a similar extent. Although these tendencies can hardly be quantitatively correlated to existing secondary-structure propensity scales, these observations reflect the overall residue probability to form secondary motifs previously computed by Chou and Fasman.[50] The chosen mathematical function to fit all torsion potentials is a fourth-order polynomial of sines, which is a good compromise between simplicity of the analytical form and accuracy of the fit

$$V_{\text{tor}}(\tau) = \sum_{n=1}^{4} A_n \sin^n(\tau - \tau_n) \qquad (4)$$

All of the parameters for the pseudobond, bending, and torsion potentials are given in Tables 2−4 of the Supporting Information.

2.4. Coarse-Grained Model of Hydrogen Bonds. The hydrogen bonds within proteins are particularly known to stabilize the protein secondary structures. These interactions, which are mainly electrostatic in nature, occur not only between amino acids separated by four or few residues, as in α-helices, but also between amino acids that could be very distant in the sequence, as in β-sheets. Following the idea of Liwo et al.,[16] a simplified model of protein hydrogen bonds can be introduced through dipole−dipole interactions between all pairs of backbone grains. Nevertheless, instead of placing a dipolar vector at the center of each backbone bead and calculating orientation-dependent interactions, a positively charged extra site H_i is attached to each grain B_i (which now carries an negative charge) in order to reproduce

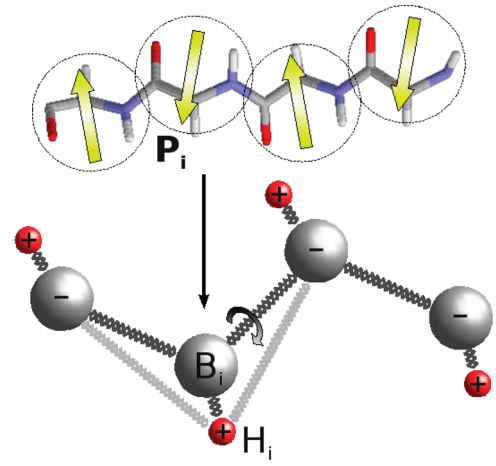

Figure 8. Modeling the protein backbone dipole moment, orientation, and fluctuations using classical Drude-like oscillators. Each particle H_i is bonded to the backbone bead B_i. Assuming that the two particles carry charges of ±0.5*e*, the length of the B_i—H_i bond is equal to $2P_i$, where P_i is the backbone dipole moment. Its orientation relative to the two neighboring backbone beads B_{i-1} and B_{i+1} is determined by the two distances B_{i-1}—H_i and H_i—B_{i+1}.

the dipole of the amino acid (Figure 8). The addition of the particle H_i follows in spirit the classical Drude oscillator model introduced in recent empirical force fields to describe the atomic polarizability.[51,52] In the present model, the auxiliary particles H_i are used to account for both the permanent dipole and the induced polarization of backbone beads.

To determine the backbone dipole moment and orientation relative to the neighboring residues, a statistical analysis was performed on the previous set of experimental structures of all-atom proteins, using the atomic partial charges from the second-generation Amber force field.[53] Assuming arbitrarily that the two grains B_i and H_i carry charges of ±0.5*e*, the analysis yields the probability distribution functions for the B_i—H_i bond length, as well as the B_{i-1}—B_i—H_i and H_i—B_i—B_{i+1} pseudobends (Figure 9). Then, applying the Boltzmann inversion generates the potentials that allow the moment, orientation, and fluctuations of the backbone dipoles to be modeled. As seen in Figure 9, the B_i—H_i bond potential can be satisfactorily fitted with a quadratic function, whereas the two Urey−Bradley-like potentials for the B_{i-1}—B_i—H_i and H_i—B_i—B_{i+1} bends are better fitted with two double-well functions (eq 3), which account for the two preferential α and extended β conformations. It should be noted that the averaged length of the B_i—H_i bonds is equal to 1.4 Å, which reflects a mean dipole moment of 0.7 eÅ for the backbone beads. This value is different from the dipole moment of the peptide bond NH—CO that links two successive Cα atoms (about 2.3 eÅ), because each backbone bead encompasses an amine group NH preceding a CαH and a carbonyl group CO succeeding it. Therefore, in contrast to the planar and quasirigid peptide group in which the bonds NH and CO dipoles are almost colinear and are added, the backbone bead dipole depends on the internal spatial distribution of the NH—CαH—CO atoms, and its moment is less strong and more variable than that of the peptide bond. Nevertheless,

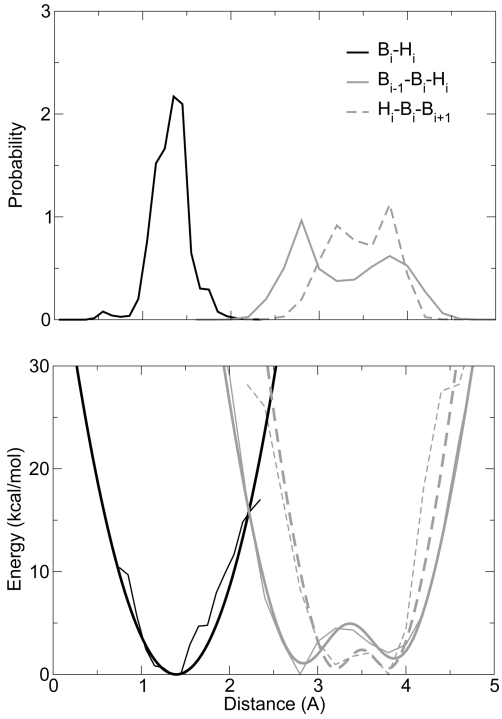

Figure 9. Top: Probability distribution functions for the B_i—H_i bond distance and the B_{i-1}—B_i—H_i and H_i—B_i—B_{i+1} pseudobends. Bottom: Associated potentials (thin lines) and fitting energy functions (thick lines).

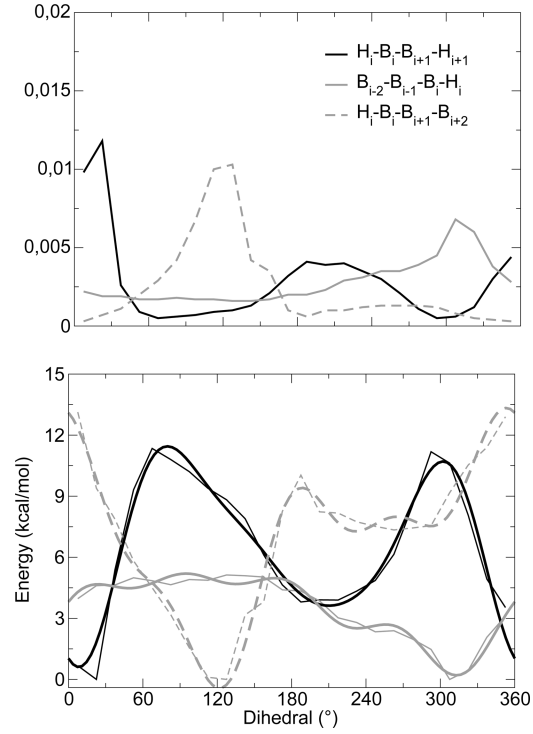

Figure 10. Top: Probability distribution functions for the H_i—B_i—B_{i+1}—H_{i+1}, B_{i-2}—B_{i-1}—B_i—H_i, and H_i—B_i—B_{i+1}—B_{i+2} pseudotorsions. Bottom: Associated potentials (thin lines) and fitting energy functions (thick lines).

the backbone bead dipole mainly arises from the carbonyl group dipole (which is twice as strong as the NH bond), so that, in extended β-strands, where the dipoles of the two NH and CO bonds seem to cancel out, the total backbone bead dipolar moment remains around 0.6 eÅ whereas it has a slightly higher value of around 0.8 eÅ in α-helices.

In most all-atom force fields, the 1−4 nonbonded interactions are treated differently from the others, because they are partially taken into account through dihedral potentials. Similar considerations for the CG model result in the introduction of three additional torsion potentials to account for the dipolar interactions of the B_i bead with its neighboring beads $B_{i\pm1}$ and $B_{i\pm2}$, which interact only through bonded potentials. Figure 10 displays the probability distributions and energy functions for these H_{i-1}—B_{i-1}—B_i—H_i, B_{i-2}—B_{i-1}—B_i—H_i, and H_i—B_i—B_{i+1}—B_{i+2} additional torsion types. Similarly to the B_{i-2}—B_{i-1}—B_i—S_i torsions, the B_{i-2}—B_{i-1}—B_i—H_i energy potential, which has a single minimum, allows for the maintenance of the "chiral" geometry of the substituents of the backbone beads B_i and prevents possible "flips" of the particles H_i. Finally, it can be noticed that this approach locates the dipole of the backbone grains at their geometric center in contrast to the model of Liwo et al.[16] which accounts for the peptide bond dipoles and then locates them in the middle of two successive Cα atoms. Thus, contrary to their dipolar model, the B_i—H_i dipole orientation is not explicitly correlated to the B_{i-2}—B_{i-1}—B_i—B_{i+1} dihedral conformation.[54] Nevertheless, the preferential orientation of the backbone dipoles relative to their neighboring residues is favored by the two torsion potentials H_i—B_i—B_{i+1}—B_{i+2} and H_{i-1}—B_{i-1}—B_i—H_i. All of the parameters for the potentials involving the auxiliary particle H_i are listed in Tables 2−4 of the Supporting Information.

3. Results

The CG bonded potentials described in the preceding section have been combined with the previously developed nonbonded van der Waals potentials (eq 1) in an MD algorithm to test their capability to yield stable trajectories of a few peptides and small proteins. In particular, an important question is the extent to which the flexible reduced protein model preserves the secondary structures and accounts for the loop mobility.

3.1. Simulation Computational Details. The MD simulations of the tested polypeptides were performed with the program Orac,[55] in the canonical *NVT* ensemble using the Nosé−Hoover thermostat[56] and an integration time step of 1 fs. All systems were first progressively heated from 100 to 300 K for 0.6 ns, then equilibrated at this latter temperature for an additional 1.2 ns, and finally simulated for 198.2 ns without any constraint. No cutoff distance was used, so that all nonbonded interactions were taken into account. The present study does not include any explicit hydration model. Nevertheless, to mimic the screening effect of the high-dielectric solvent, all Coulombic interactions were damped using a sigmoidal distance-dependent dielectric function.[57,58] The dielectric screening effect is theoretically dependent on the charge exposure,[59,60] but to keep the solvent model as simple as possible, only two different sigmoidal functions were used: one with a large slope for interactions involving charged side chains, which are generally exposed to the solvent, and one with a smaller slope for interactions between

Table 1. Polypeptides Whose Dynamic Structures Were Simulated with the CG Model

PDB code	no. of residues	nos. of atoms/grains	motif(s)	AA/CG CPU time ratio
1E0Q[68]	17	285/56	β	26.0
2FLY[69]	20	356/70	α	20.5
1VII[64]	36	596/123	α	16.1
1PPT[70]	36	581/119	α/β	19.0
1PGB[71]	56	855/183	α/β	16.7
1BPI[72]	58	892/190	α/β	15.8
1B75[73]	94	1533/319	α/β	17.5
1AJ3[65]	98	1596/334	α	16.7

the backbone beads B_i and H_i, which are overall more buried sites (see Figure 1 in the Supporting Information).

3.2. Equilibrium Structural Properties. Here are reported the results of the MD simulations for eight polypeptides, with sequence lengths ranging from 17 to 98, having various secondary-structure motifs. Table 1 lists the tested peptides and shows that the CG model allows the computational times to be reduced by a factor of about 15−25, relative to the AA description (without explicit solvent). All simulations were started from the protein experimental structure available in the Protein Data Bank,[47] without any addition of counterions. As shown in Figure 2 of the Supporting Information for four of the studied polypeptides, the time evolution of the total energies and their bonded contributions V_{bonded} are stable along the last nanoseconds of the trajectories, suggesting that simulations have reached some equilibrium states. Figure 11 displays the time evolutions of the protein root-mean-square deviations (RMSDs) from their initial conformation, calculated over their backbone beads. Overall, it is observed that most of the tested

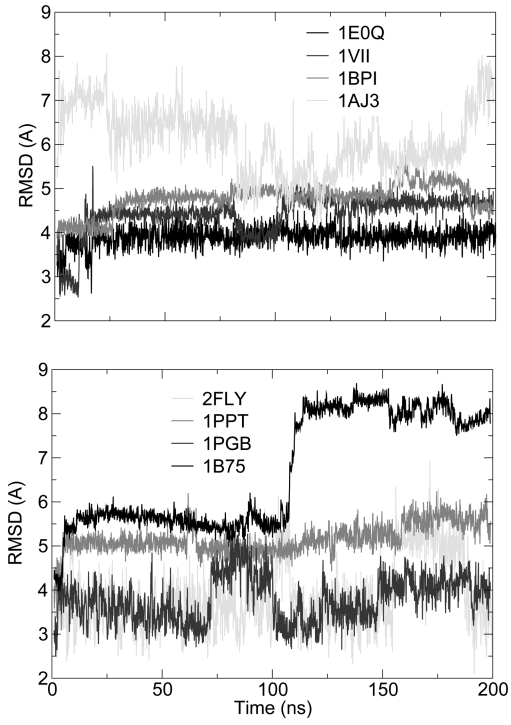

Figure 11. Time evolutions of the RMSD values (calculated over the backbone beads) relative to the initial conformations for the polypeptides (top) 1E0Q, 1VII, 1BPI, and 1AJ3 and (bottom) 2FLY, 1PPT, 1PGB, and 1B75.

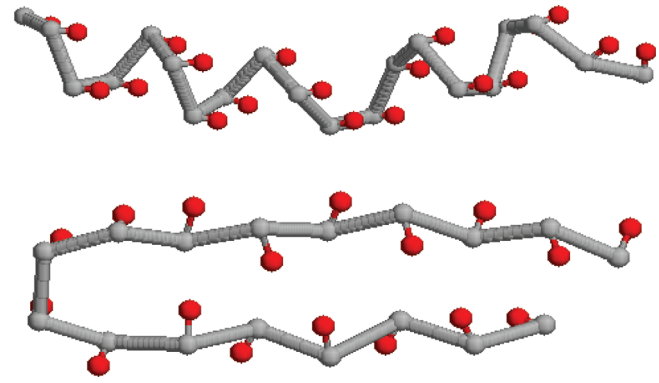

Figure 12. Ball-and-stick representations of the backbone of (top) the α-helix 2FLY and (bottom) the β-hairpin 1E0Q at the end of simulations. The B_i grains are displayed in gray, and the H_i particles are in red.

polypeptides reach stable conformations not too far from the native structures, with RMSDs ranging from 3 to 6 Å after 200 ns of simulation. In contrast, the protein 1B75 structure is quite stable until about 110 ns, after which its RMSD rapidly increases and stabilizes around 8 Å. The protein 1AJ3 seems to have reached some stable conformations between 25 and 75 ns and between 130 and 180 ns, but its RMSD increases at the end of the simulation to a rather high value between 7 and 8 Å. As shown in Figure 3 of the Supporting Information and as discussed below, the final RMSD values for these two proteins clearly indicate large deformations of their tertiary structures. Finally, it can be noticed that the CG bonded potentials do not prevent the proteins, such as 2FLY or 1PGB, to transitionally visit conformations other than the equilibrium one.

Comparisons of these first results with similar CG protein models were difficult, as most of the latter were designed to treat the folding problem and were not used to test the protein stability by MD simulations. Other residue-scale force fields, including the model of Tozzini and McCammon[61] and the MARTINI force field,[24] were developed to study the dynamics properties of large CG proteins over long MD trajectories.[3,62,63] However, as stressed by the authors, their CG force fields were biased to retain either the secondary or the native tertiary structures of proteins, making comparisons with the presented study very delicate. The dynamic conformations generated by their models are certainly closer to the experimental structures than those yielded by ours, as it is generally difficult to have a completely unbiased and accurate force field at the same time. Compared to the recent study of Majek and Elber,[38] which uses an unbiased one-bead backbone model to simulate protein conformational stability, the equilibrium RMSD values displayed in Figure 11 are close to the mean value (5 Å) of the RMSD yielded by their MD simulations. This suggests that the CG force field presented here, which is less elaborate than theirs but similarly models the bonded potentials and hydrogen bonding, allows for the generation of dynamic conformations of simplified proteins as far to the native structures as their simulations.

As illustrated in Figure 12 for the peptides 2FLY and 1E0Q (as well as in Figure 3 of the Supporting Information for the proteins 1PGB and 1AJ3), the polypeptide secondary

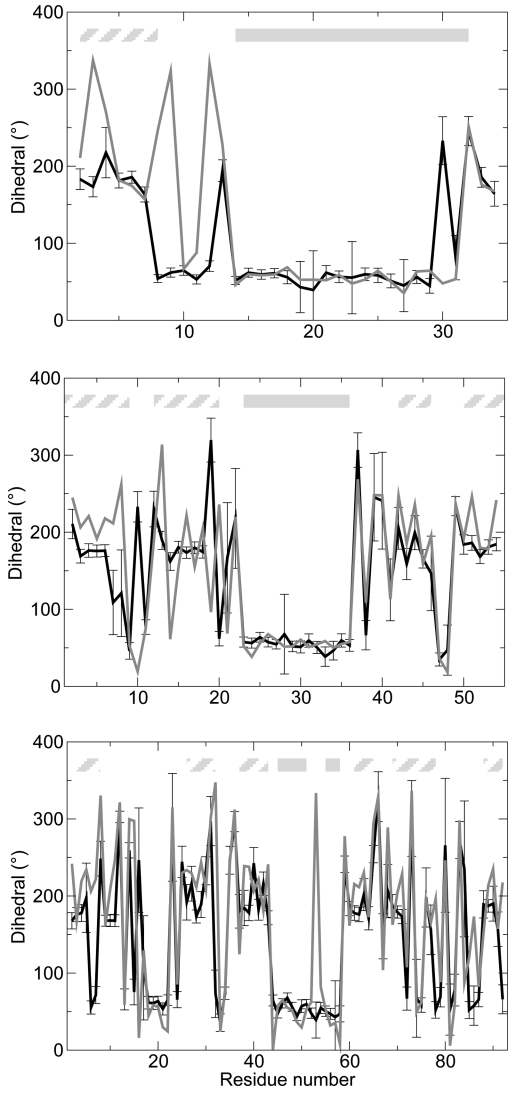

Figure 13. Residue-averaged values of the B_{i-1}—B_i—B_{i+1}—B_{i+2} torsions for the proteins (top) 1PPT, (middle) 1PGB, and (bottom) 1B75. Black and gray lines represent the theoretical predictions and experimental measurements, respectively. Horizontal plain and striped bars indicate the α-helix and β-strand positions, respectively, in the protein sequence.

structures are maintained during the simulations through the electrostatic interactions between backbone beads: In helical chains, the B_i—H_i bonds are mainly oriented along the helix axis, and the grains B_i strongly interact with the H_{i+3} and H_{i+4} particles. In β-sheets, the B_i—H_i dipoles are almost coplanar, perpendicular to the strand direction and alternatively point toward and from the neighboring strand. Without the simplified model of hydrogen bonds, the α-helix 2FLY bends rapidly, and a kink occurs in its middle, close to the Gly residue, and in the β-hairpin simulations, the two strands locally move apart, become highly curved, and dramatically lose their parallel orientation.

The values of the backbone torsions B_{i-1}—B_i—B_{i+1}—B_{i+2} provide a quantitative indicator of the stabilities of the secondary structures during the MD simulations.[18] Figure 13 displays the averaged values of these dihedral angles along the sequence of polypeptides 1PPT, 1PGB, and 1B75, which have various lengths and a common α/β motif, in order to

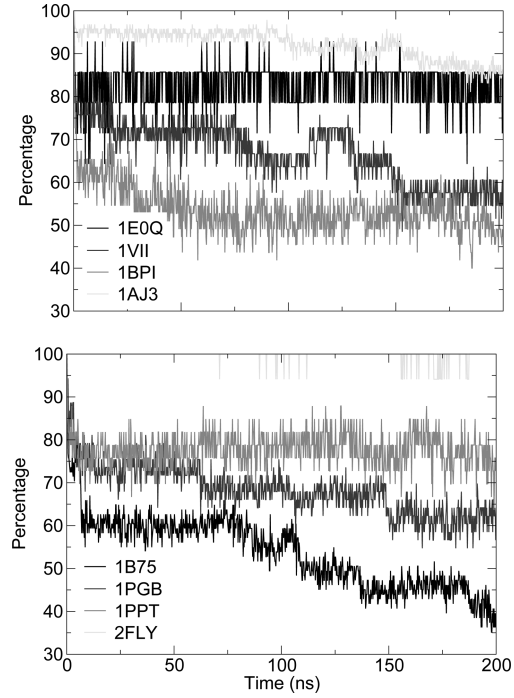

Figure 14. Time evolution of the percentage of native backbone torsions for the polypeptides (top) 1E0Q, 1VII, 1BPI, and 1AJ3 and (bottom) 2FLY, 1PPT, 1PGB, and 1B75.

display the different angle values that correspond to the backbone torsions. When compared with the values measured in the PDB structures, it can be observed that most of the backbone torsions keep a conformation close to the experimental one, especially in the helical and β-strand structures. Several of them undergo transitions, and such conformational changes mainly occur near or in the loop regions. In the case of 1PPT, an α → β transition of the residue 30 backbone torsions can be observed that slightly kinks the helix end, as well as a structural change into an α-conformation of the loop region 8−12, not observed by experiments. In the protein 1B75 simulation, several dihedral angles undergo β → α transitions, particularly in the first and fifth β-strands, revealing large conformational rearrangements consistent with the relatively high RMSD time evolution. It should be emphasized here that, whereas the residues in α-helices have dihedral values around 70°, which is the angle value of the first minimum in the B_{i-1}—B_i—B_{i+1}—B_{i+2} potential (Figure 5), the residues in β-strands have average torsions around 190°, which does not correspond to the second minimum of the backbone potential. These β conformations are, in fact, partially stabilized by the H_i—B_i—B_{i+1}—H_{i+1} torsions (Figure 10), as well as by the S_i—B_i—B_{i+1}—B_{i+2} potentials (Figure 7), which are sensitive to the nature of the amino acids. After many tests, no stable structure of proteins could be obtained without the H_i—B_i—B_{i+1}—H_{i+1} potential. Overall, the CG bonded potentials allow the backbone dihedral conformation of the proteins to be well conserved: As shown in Figure 14, which displays a plot of the time evolutions of the percentage of native backbone torsions (plus or minus 30°), most of the MD trajectories preserve more than 60% of the protein backbone dihedral angles in the experimental conformation. The two simulations of proteins 1BPI and 1B75

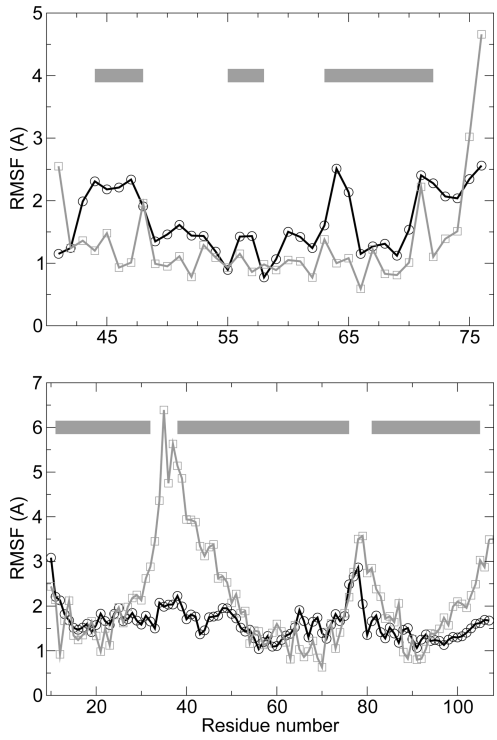

Figure 15. Comparison of the residue RMSFs relative to averaged conformations, computed from simulations (black lines) or provided by NMR experiments (gray lines), for the proteins (top) 1VII and (bottom) 1AJ3. Horizontal dark gray and light gray lines represent the α-helix positions.

are those that least conserve the backbone structures, with a percentage of torsion values close to experiments of around 50%.

If the secondary structures are relatively well conserved along the MD simulations, some protein tertiary structures are slightly deformed relative to the experimental structures. For example, the two last α-helices of the protein 1AJ3 are more curved and twisted at the end of the simulation than in the initial conformation (see Figure 3 of the Supporting Information). During the simulation of 1VII, the orientation of its first α-helix relative to the other two can change significantly. At the end of the 1PGB simulation, the α-helix is oriented almost parallel to the β-strand direction, whereas it diagonally crosses the sheets in the experimental structure. In the protein 1B75, the initial conformation in the β-barrel become flatter instead of remaining roughly cylindrical. All of these discrepancies could be due to the absence of counterions and of an electrostatic description of the neutral polar side chains. They probably also arise from the crude model of solvation, which cannot account for hydrophobic interactions within the protein cores or for the molecular nature and exclusion-volume effect of the solvent. Further corrections of these deficiencies in the CG protein model are at present under development.

3.3. Structural Fluctuations and Thermal Instability. One question addressed here is whether the simplified backbone model can reasonably simulate the conformational fluctuations of proteins, particularly their loops, which are expected to undergo the largest-amplitude movements. In Figure 15 are plotted the residue root-mean-square fluctuations (RMSFs) relative to the averaged structures for the 1VII and 1AJ3 proteins. These RMSFs are directly compared to those provided by NMR experiments that have solved the three-dimensional structures.[64,65] The RMSFs calculated for 1VII appear larger than the NMR values, especially for residues in the first α-helix and at the beginning of the third α-helix. This confirms that the CG protein tertiary structure is less compact than the experimental structure and thus does not restrain the movements of its secondary structures much. In contrast, the 1AJ3 loops undergo fewer large-amplitude motions than in NMR experiments, especially the loop connecting the first two α-helices. When examining the CG protein structure (Figure 3 of the Supporting Information), it can be noticed that the loops come close to and interact with the C-terminus and N-terminus instead of being fully solvated, which could explain their rather restrained motions. Nevertheless, it is overall observed that the residues in the secondary structures have the lowest RMSFs, whereas the loop regions have the largest fluctuations, no more nor less accurately than calculations provided by less detailed elastic network models.

A finer indicator of the protein backbone conformational fluctuations is provided by the measurement of the so-called S^2 order parameter with NMR spectroscopy.[66] This parameter reflects the angular mobility of the backbone N—H bonds and can be compared to the angular mobility of the B—H dipoles in the CG protein model. In simulations, it can be calculated as

$$S^2 = \frac{3}{2}\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\langle \mu_\alpha \mu_\beta \rangle^2 - \frac{1}{2} \qquad (5)$$

where μ_α and μ_β denote the three components, x, y, and z, of the normalized vector along the B—H bonds and where the brackets symbolize a time average over simulations.[67] An S^2 order parameter close to 1 indicates that the N—H or B—H vectors are quite constrained in space and that the H particle is probably involved in a hydrogen bond. If it is close to 0, these vectors are, on the contrary, free to move and rotate. The S^2 parameters calculated from the simulations are displayed in Figure 16 as a function of the residue numbers of the two proteins 1BPI and 1PGB. Overall, it is observed that the B—H bonds have small angular fluctuations for residues in α-helices and β-strands, whereas they are significantly more mobile for residues in loops. This corroborates the role of the electrostatic interactions between B—H polar groups in stabilizing the secondary structures, similarly to the hydrogen bonds between the N—H and C═O atoms. However, discrepancies between simulations and experiments can be noticed: In the 1BPI simulation, the two β-strands are significantly more mobile, whereas the loop between them has more restrained movements than observed by NMR spectroscopy, indicating a slight deformation of these secondary structures. The S^2 parameter profile of the protein 1PGB is in overall good agreement with experimental observations, except for the third β-strand, which appears to be more mobile, as well as the second loop before the α-helix, which undergoes larger-amplitude motion. Further investigations, especially into the stabilizing role of solvent, are being conducted to explain these discrepancies and

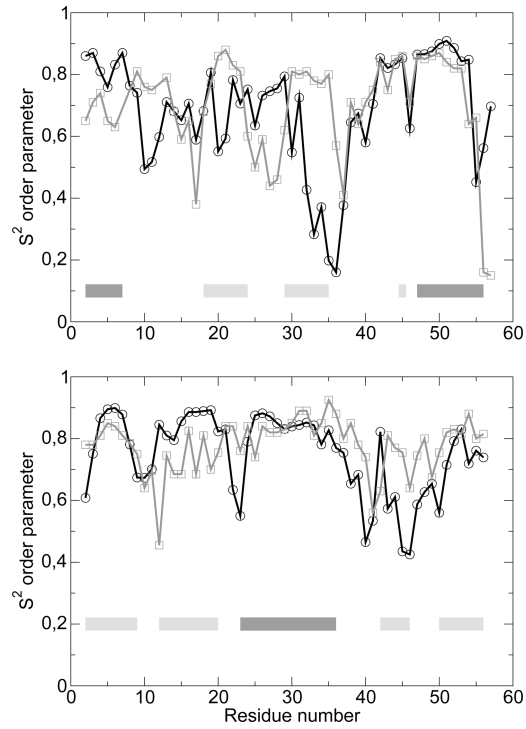

Figure 16. Comparison of the B—H order parameters (black line) with the N—H order parameters measured by NMR spectroscopy (gray line) for the proteins (top) 1BPI and (bottom) 1PGB. Horizontal dark gray and light gray lines represent α-helix and β-strand positions, respectively.

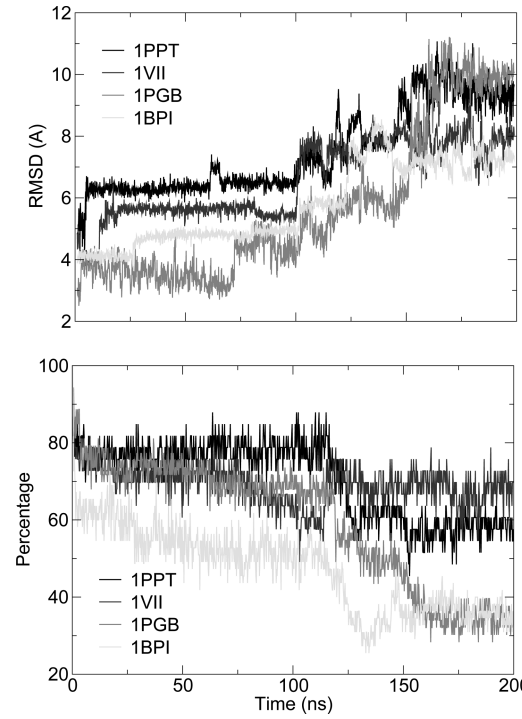

Figure 17. Time evolutions of the (top) RMSD relative to the initial conformation and (bottom) percentage of native backbone torsions for the polypeptides 1PPT, 1VII, 1PGB, and 1BPI. At 100 ns, the temperature was increased from 300 to 400 K, and at 150 ns, it was increased again to 500 K.

improve the capability of the CG model to yield stable and meaningful dynamic conformations of proteins.

As previously mentioned, the CG bonded potentials presented here possibly enable conformation changes of the simulated proteins, contrary to elastic network models, which can study only harmonic deformations around a single structure. To clarify the capability of CG proteins to explore non-native conformations, simulations at high temperatures were performed for the four medium-size proteins 1PPT, 1VII, 1PGB, and 1BPI. In practice, their previous MD trajectories at 300 K were continued at 400 K from 100 to 150 ns then at 500 K from 150 to 200 ns. Figure 17 shows the time evolutions of the RMSDs from the experimental structures and the percentages of native backbone torsions (±30°) for the four simulations. It can be observed that the RMSDs of the four proteins increase during the 50-ns trajectories at 400 K and, for the proteins 1PPT and 1PGB, continue to rise during the last 50 ns at 500 K to values ranging from 7 to 10 Å. This clearly indicates a denaturation of their native conformation upon heating. Nevertheless, regarding the percentages of native backbone torsions, the simulations of the two proteins 1PPT and 1VII maintain a rather high ratio of backbone torsions in native conformation (between 60% and 70%), suggesting that their tertiary structure is lost before their secondary structure. In contrast, the percentages of native backbone torsions of the two protein 1PGB and 1BPI rapidly decrease to low values between 30% and 40%. For these denatured proteins, the secondary structures are lost concomitantly to their tertiary conformation.

4. Conclusions

This work introduces a set of bonded potentials for modeling the backbone flexibility of proteins described with residue-scale coarse grains. The main feature of these potentials is that the local secondary-structure propensity of the amino acids seems to be encoded in the pseudotorsions S_i—B_i—B_{i+1}—B_{i+2}. Combined with the nonbonded van der Waals potential reported recently[29] and a simplified dipolar model of hydrogen bonds, the CG bonded potentials overall generate stable protein structures in the neighborhood of the experimental conformations. Despite some backbone torsion conformational changes, the protein secondary structures are quite well conserved along the 200-ns MD trajectories, but some tertiary structures deviate from the initial ones, especially the β-barrel protein 1B75. The amino acids in the loops connecting helices and strands are found to have the largest internal mobility. At present, further improvements of the model, particularly a better description of the polar residues and their interactions with high-dielectric solvents, are in development. It is believed that this CG protein model will provide, at a low computational cost, a reasonable dynamic picture of protein equilibrium structures and useful insights into the functional role of large protein conformational changes.

Acknowledgment. The author acknowledges Grant ANR-08-JCJC-0081-01 from the Agence Nationale de la Recherche.

Supporting Information Available: All of the parameters of the coarse-grained protein bonded potentials de-

scribed herein. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.

(2) Ishima, R.; Freedberg, D.; Wang, Y.; Louis, J.; Torchia, D. *Structure* **1999**, *7*, 1047.

(3) Trylska, J.; Tozzini, V.; Chang, C.; McCammon, J. *Biophys. J.* **2007**, *92*, 4179.

(4) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433.

(5) Popovych, N.; Sun, S.; Ebright, R.; Kalodimos, C. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831.

(6) Ishima, R.; Torchia, D. *Nat. Struct. Biol.* **2000**, *7*, 740.

(7) Karplus, M.; McCammon, J. *Nat. Struct. Biol.* **2002**, *9*, 646.

(8) Muller-Plathe, F. *Chem. Phys. Chem.* **2002**, *3*, 755.

(9) Zhou, J.; Thorpe, I.; Izvekov, S.; Voth, G. *Biophys. J.* **2007**, *92*, 4289.

(10) Klein, M.; Shinoda, W. *Science* **2008**, *321*, 798.

(11) Durrieu, M.; Bond, P.; Sansom, M.; Lavery, R.; Baaden, M. *Chem. Phys. Chem.* **2009**, *10*, 1548.

(12) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59.

(13) Kolinski, M.; Skolnick, J. *Polymer* **2004**, *45*, 511.

(14) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144.

(15) Bonvin, A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 194.

(16) Liwo, A.; Pincus, M.; Wawak, R.; Rackovsky, S.; Scheraga, H. *Protein Sci.* **1993**, *2*, 1715.

(17) Wallqvist, A.; Ullner, M. *Proteins* **1994**, *18*, 267.

(18) DeWitte, R.; Shakhnovich, E. *Protein Sci.* **1994**, *3*, 1570.

(19) Reva, B.; Finkelstein, A.; Sanner, M.; Olson, A. *Protein Eng.* **1997**, *10*, 865.

(20) Bahar, I.; Kaplan, M.; Jernigan, R. *Proteins* **1997**, *29*, 292.

(21) Haliloglu, T.; Bahar, I. *Proteins* **1998**, *31*, 271.

(22) Derreumaux, P. *J. Chem. Phys.* **1999**, *111*, 2301.

(23) Van Giessen, A.; Straub, J. *J. Chem. Theory Comput.* **2006**, *2*, 674.

(24) Monticelli, L.; Kandasamy, S.; Periole, X.; Larson, R.; Tieleman, D.; Marrink, S. *J. Chem. Theory Comput.* **2008**, *4*, 819.

(25) Gabdoulline, R.; Wade, R. *J. Phys. Chem.* **1996**, *100*, 3868.

(26) Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.

(27) Izvekov, S.; Voth, G. *J. Chem. Phys.* **2005**, *123*, 134105.

(28) Prampolini, G. *J. Chem. Theory Comput.* **2006**, *2*, 556.

(29) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Phys. Chem. B* **2007**, *111*, 9390.

(30) Tirion, M. *Phys. Rev. Lett.* **1996**, *77*, 1905.

(31) Bahar, I.; Atilgan, A.; Erman, B. *Fold. Des.* **1997**, *2*, 173.

(32) Hinsen, K. *Proteins* **1998**, *33*, 417.

(33) Tama, F.; Sanejouand, Y. *Protein Eng.* **2001**, *14*, 1.

(34) Micheletti, C.; Carloni, P.; Maritan, A. *Proteins* **2004**, *55*, 635.

(35) Klimov, D.; Betancourt, M.; Thirumalai, D. *Fold. Des.* **1998**, *3*, 481.

(36) Voegler Smith, A.; Hall, C. *Proteins* **2001**, *44*, 344.

(37) Takada, S.; Luthey-Schulten, Z.; Wolynes, P. *J. Chem. Phys.* **1999**, *110*, 11616.

(38) Majek, P.; Elber, R. *Proteins* **2009**, *76*, 822.

(39) Yap, E.; Fawzi, N.; Head-Gordon, T. *Proteins* **2008**, *70*, 626.

(40) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. *J. Phys. Chem. B* **2005**, *109*, 13798.

(41) Liwo, A.; Khalili, M.; Scheraga, H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362.

(42) Miyazawa, S.; Jernigan, R. *Macromolecules* **1985**, *18*, 534.

(43) Sippl, M. *J. Mol. Biol.* **1990**, *213*, 859.

(44) Bryant, S.; Lawrence, C. *Proteins* **1993**, *16*, 92.

(45) Skolnick, J.; Jaroszewski, L.; Kolinski, A.; Godzik, A. *Protein Sci.* **1997**, *6*, 676.

(46) Betancourt, M.; Thirumalai, D. *Protein Sci.* **1999**, *8*, 361.

(47) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *28*, 235.

(48) Kozlowska, U.; Liwo, A.; Scheraga, H. *J. Phys. Condens. Matter* **1997**, *19*, 285203.

(49) Liwo, A.; Pincus, M.; Wawak, R.; Rackovsky, S.; Oldziej, S.; Scheraga, H. *J. Comput. Chem.* **1997**, *18*, 874.

(50) Chou, P.; Fasman, G. *Biochemistry* **1974**, *13*, 211.

(51) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025.

(52) Vorobyov, I.; Anisimov, V.; MacKerell, A., Jr. *J. Phys. Chem. B* **2005**, *109*, 18988.

(53) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K., Jr.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(54) Liwo, A.; Oldziej, S.; Czaplewski, C.; Kozlowska, U.; Scheraga, H. *J. Phys. Chem. B* **2004**, *108*, 9421.

(55) Procacci, P.; Darden, T.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848.

(56) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.

(57) Mehler, E.; Eichele, G. *Biochemistry* **1984**, *23*, 3887.

(58) Hingerty, B.; Ferrell, T.; Turner, J. *Biopolymers* **1985**, *24*, 427.

(59) Sandberg, L.; Edholm, O. *Proteins* **1999**, *36*, 474.

(60) Mallik, B.; Masunov, A.; Lazaridis, T. *J. Comput. Chem.* **2002**, *23*, 1090.

(61) Tozzini, V.; McCammon, J. *Chem. Phys. Lett.* **2005**, *413*, 123.

(62) Shih, A.; Arkhipov, A.; Freddolino, P.; Schulten, K. *J. Phys. Chem. B* **2006**, *110*, 3674.

(63) Treptow, W.; Marrink, S.; Tarek, M. *J. Phys. Chem. B* **2008**, *112*, 3277.

(64) McKnight, C.; Matsudaira, P.; Kim, P. *Nat. Struct. Biol.* **1997**, *4*, 180.

(65) Pascual, J.; Pfuhl, M.; Walther, D.; Saraste, M.; Nilges, M. *J. Mol. Biol.* **1997**, *273*, 740.

(66) Barchi, J., Jr.; Grasberger, B.; Gronenborn, A.; Clore, G. *Protein Sci.* **1994**, *3*, 15.

(67) Smith, P.; van Schaik, R.; Szyperski, T.; Wuthrich, K.; van Gunsteren, W. *J. Mol. Biol.* **1995**, *246*, 356.

(68) Zerella, R.; Chen, P.; Evans, P.; Raine, A.; Williams, D. *Protein Sci.* **2000**, *9*, 2142.

(69) Lucyk, S.; Taha, H.; Yamamoto, H.; Miskolzie, M.; Kotovych, G. *Biopolymers* **2006**, *81*, 295.

(70) Blundell, T.; Pitts, J.; Tickle, I.; Wood, S.; Wu, C. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4175.

(71) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. *Biochemistry* **1994**, *33*, 4721.

(72) Parkin, S.; Rupp, B.; Hope, H. *Acta Crystallogr. D* **1996**, *52*, 18.

(73) Stoldt, M.; Wohnert, J.; Gorlach, M.; Brown, L. *Embo J.* **1998**, *17*, 6377.

CT900408S