THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

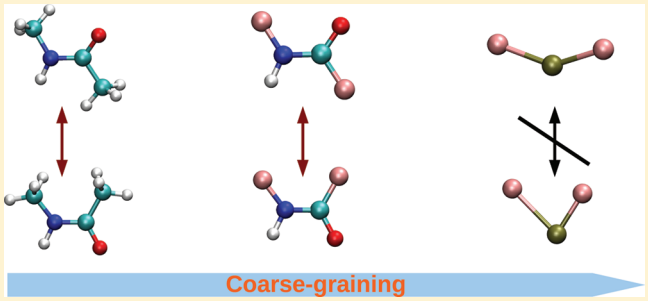
Coarse-Grained Modeling of Simple Molecules at Different Resolutions in the Absence of Good Sampling

Luca Larini and Joan-Emma Shea*

Department of Chemistry and Biochemistry and of Physics, University of California, Santa Barbara, Santa Barbara, California 93106, United States

Supporting Information

ABSTRACT: Many problems of interest in modern science originate from the complex network of interactions of different molecular structures, each possessing its own typical length and time scale of relevance. In such materials, nontrivial properties emerge from the different length and time scales involved that could not be predicted from the properties of each individual subunit taken alone. A solution to the formidable theoretical and computational issues raised by these systems involves coarse-graining, a procedure in which multiple atoms are grouped into a few interaction sites. The coarse-grained approach aims at constructing an effective Hamiltonian from available information about the system and then using this Hamiltonian to investigate the behavior of the system on the length and time scales of interest. In this paper, we aim at determining how far we can coarse-grain a system using only the commonly used pairwise, spherically symmetric potentials, as well as assessing the impact of poor initial sampling on the quality of the resulting coarse-grained model. Coarse-graining is performed following the multiscale coarse-graining (MS-CG) methodology, and we use as a model system the *N*-methylacetamide (NMA) molecule, a simple representation of a peptide bond, which can adopt two conformations, cis and trans. Our simulations reveal that as the coarse-graining becomes more aggressive multibody effects start to emerge and that the initial sampling of conformations can adversely bias the model in the case of heavy coarse-graining.



INTRODUCTION

Soft matter is a branch of condensed matter physics that deals with materials that can be easily deformed by an external field. A wide variety of materials, ranging from liquids, to polymers, to biological molecules, fall under this umbrella.[1] The typical energy scales associated with deformations are on the order of room temperature thermal fluctuations. At this scale, bonds are unlikely to break, and the molecular structure of the material does not degrade via chemical reactions. Rather, the material undergoes structural rearrangements.[2−4] The chemical stability that characterizes these deformations generally allows a disregarding of quantum effects, and the system can then be described by a simple classical force field. For instance, atomistic force fields used in molecular dynamics (MD) simulations such as AMBER,[5−11] CHARMM,[12−15] and OPLS-AA,[16−18] describe the interactions among atoms as classical potentials and then propagate motion through the integration of Newton's equations of motion.[19]

The approximations introduced in classical force fields allow the study of larger systems (on the order of nanometers) and time scales (on the order of nanoseconds) than would be permitted using quantum calculations (for which typical length and time scales are on the order of Angstroms and picoseconds, respectively).[20] Despite these simplifications, many systems of interest still lay beyond current available computational power. For example, protein folding can occur on a time scale of seconds,[21,22] and the slow relaxation typical of the glass transition can reach time scales of days.[23] Protein aggregation into macroscopic structures can reach dimensions of micrometers,[24−29] while a colloidal solution can contain particles as large as hundred of nanometers each.[30]

Further approximations are required to study these systems. One approach is to disregard atomistic details and focus on more general properties of the system, such as its excluded volume, connectivity, or shape. In these models, the chemical identity of each molecule is usually lost, leading to a general description of a class of materials. This scheme works very well, for instance, in the case of polymers, where many mechanical properties can be predicted on the basis of simple considerations about excluded volume and connectivity.[2]

While this approach is successful in a number of instances, there are cases where a general treatment that disregards chemical details is not appropriate. For instance, the binding of a drug to a protein is usually characterized by very specific

Special Issue: Macromolecular Systems Understood through Multiscale and Enhanced Sampling Techniques

Received: October 10, 2011
Revised: January 24, 2012
Published: January 31, 2012

   dx.doi.org/10.1021/jp2097263 | *J. Phys. Chem. B* 2012, 116, 8337−8349

interactions that are not easily captured by generic models. Since binding and associated conformational changes lie at the limit of what can be simulated in a reasonable time frame using fully atomistic force fields, it is still necessary to simplify the system under consideration. The most straightforward way to deal with this type of problem is to group single atoms into interacting sites and then build a coarse-grained (CG) force field based on its atomistic counterpart. In this manner, the number of interacting sites is reduced, allowing for more efficient simulations while at the same time retaining a sufficient level of specificity. Examples of these types of coarse-grained force fields include the UNRES[31,32] and MARTINI[33] force fields. These force fields offer an invaluable tool for the study of many phenomena, but they suffer from the same problem as atomistic force fields, namely, that the level of resolution is fixed beforehand. This may be undesirable in situations where a higher or lower level of resolution is required or when one wishes to use different levels of resolutions for different parts of the system. With this in mind, methodologies have been developed that allow the construction of a force field at the desired resolution starting from data obtained either from simulation or experiment, such as inverse Monte Carlo[34−38] and iterative Boltzmann inversion.[37,39] In the present work, we use one such methodology, the multiscale coarse-graining (MS-CG) method,[40−44] to investigate how far a molecule can be coarse-grained. The MS-CG methodology is particularly well-suited for this type of analysis as it allows us to build, in a reliable manner, force fields at different resolutions starting from atomistic simulation data. It has been shown[42] that the CG force field that can best reproduce the all-atom data is in general a force field including multibody interactions.[45−47] However, for practical purposes, multibody nonbonded potentials are usually difficult to implement and computationally inefficient. For these reasons, force fields are usually constructed using simpler functional forms, typically adopting two-body and spherically symmetric potentials for the nonbonded part of the potential. Using the atomistic data as a reference, it is possible to analyze how the approximations involved at different resolutions of the CG model affect the properties of the system.

The primary aim of this work is to build different coarse-grained force fields at different resolutions and to explore how far we can push the coarse-graining while still capturing the physical properties of the original system. We will show that some levels of resolution cannot properly be represented, whereas other levels can be optimally simulated, and that multibody effects start playing an important role as the system is progressively coarse-grained. We will discuss the importance of the configurations sampled in the atomistic simulations, which serve as the starting points in the MS-CG method. It is indeed possible that the system under examination undergoes major structural rearrangements during its evolution in time and that the time scale necessary to explore all conformations exceeds the available time window in the initial atomistic simulations. Hence, it is of critical importance to assess whether a CG force field developed under conditions of poor sampling is nonetheless capable of describing the process of interest and of capturing the conformations that were not visited in the atomistic simulations.

To investigate these questions, we turn to the *N*-methylacetamide molecule (NMA, see Figure 1 for the molecular structure) as a model system. This molecule offers the simplest possible representation of a peptide bond

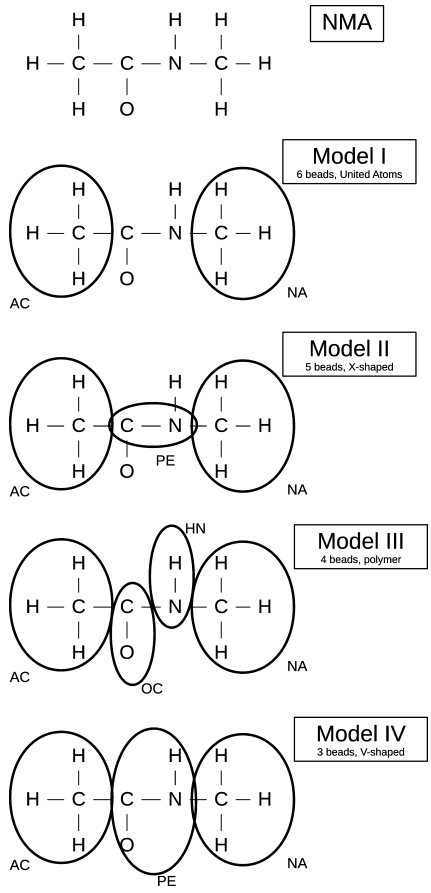


Figure 1. Structure of NMA and proposed CG models. (a) Fully atomistic, explicit hydrogen representation of the *N*-methylacetamide (NMA) molecule, our model system to be studied at different resolutions. (b) The highest resolution corresponds to a 6-bead united atom model, where nonpolar hydrogens have been coarse-grained. In addition, three different models for the peptide bond are taken into account. (c) 5-bead model: the hydrogen atom and the oxygen, which are involved in the formation of hydrogen bondings, are not coarse-grained. (d) 4-bead polymer-like model, where the carboxyl and the amine groups are mapped to two different sites. (e) The lowest resolution model, the 3-bead model, describes the peptide bond as one interacting site. The names used for the CG groups are explained in the text concerning each different level of resolution.

and can adopt two possible conformations via rotation around this bond: cis or trans. The transition between the cis and trans conformation can be fully sampled using standard molecular dynamics simulations with an all-atom representation of the molecule. We will exploit this feature to examine issues that may arise when coarse-grained models are constructed using atomistic data drawn only from the cis or the trans populations, rather than from the full simulation data. We will systematically coarse-grain the NMA from its explicit hydrogen, all-atom representation (top panel, Figure 1) to a united atom model, all the way to a very simple 3-bead model in which the peptide bond is described by one interacting site.

This article is structured as follows. The section Theoretical Background describes the theoretical framework underlying the coarse-graining procedure adopted in this work. The Methods section details how the coarse-graining and the simulations were performed. The Results section gives a detailed account of the issues connected with developing CG models at different levels of resolution. The section Discussion and Conclusions

dx.doi.org/10.1021/jp2097263 | *J. Phys. Chem. B* 2012, 116, 8337−8349

summarizes the findings of our work and outlines immediate applications as well as future developments.

THEORETICAL BACKGROUND

In this section, we provide a brief account of the theory behind the MS-CG methodology. We emphasize that although the discussion will be focused on the MS-CG method, many of the results and conclusions can be easily extended to other methodologies. The starting point of every MS-CG computation is the information available through molecular dynamics simulations data. The method is general, so any atomistic force field can be employed. In the rest of this work, the focus will be on simulations performed in the canonical ensemble, even though the problems analyzed apply to other ensembles as well.[48] The first step in any CG method consists of defining the resolution of the CG model. In this work multiple resolutions will be considered, to fully appreciate the impact that the removal of degrees of freedom could have on the system under examination. The different resolutions that we will consider are reported in Figure 1. The choice of a resolution is mathematically expressed through the definition of a mapping operator M, which maps the conformations available from atomistic simulations to their CG counterpart. If $\{\vec{r}_i\}$ is the set of coordinates that represents each conformation (namely, each snapshot of the MD trajectory), the equivalent CG conformation is described by the set of CG coordinates $\{\vec{R}_I\}$ defined as[42]

$$\vec{R}_I = M_I(\{\vec{r}_i\}) \tag{1}$$

In this equation, i is the ith atom in the atomistic resolution, whereas I is the Ith interaction site in the CG resolution. The function $M_I(\{\vec{r}_i\})$ defines which atoms belong to the Ith CG site and how the position of the Ith CG site has to be computed. An example of a mapping function is the function that maps a set of coordinates to their center of mass.

Once the resolution is decided, the second step consists of deciding which property (or properties) of the atomistic system should be correctly conserved in the CG model. The MS-CG methodology has been developed to correctly reproduce the force acting on each interaction site. The target forces can be computed from the atomistic trajectory according to

$$\vec{F}_I^{\mathrm{AA}} = \sum_{i \in I} \vec{f}_i \tag{2}$$

Thus, the total atomistic force $\vec{F}_I^{\mathrm{AA}}$ acting on the Ith CG site is due to the sum of the atomistic forces $\vec{f}_i$ acting on each single atom i that belongs to the CG site I.

Once the resolution is defined and the atomistic force acting on each interacting site is evaluated, the CG force field can be computed through a mean least-squares optimization of the $\vec{F}_I^{\mathrm{AA}}$. In particular, the residual χ of the following function is minimized

$$\chi^2 = \frac{1}{3N}\left\langle \sum_I^N |\vec{F}_I^{\mathrm{CG}} - \vec{F}_I^{\mathrm{AA}}|^2 \right\rangle \tag{3}$$

where $\vec{F}_I^{\mathrm{CG}}$ is the CG approximation to the $\vec{F}_I^{\mathrm{AA}}$ acting on the Ith CG site. The average is computed over all the available configurations (namely, every available snapshot from the atomistic trajectory) and over all the degrees of freedom of an ensemble of N CG sites.

It can be shown that a solution to this equation exists and is equivalent to[42,44]

$$U(\{\vec{R}_I\}) = -k_{\mathrm{B}}T \ln\left[\frac{Z_N}{z_n}\int \mathrm{d}\vec{r}^{\,n} \prod_{I=1}^{N} \delta(\vec{R}_I - M_I(\{\vec{r}_i\})) \times e^{-u(\vec{r}^{\,n})/k_{\mathrm{B}}T}\right] \tag{4}$$

where k_{B} is the Boltzmann's constant; T is the temperature; U and Z_n are the CG potential and partition function, respectively; and u and z_n are their atomistic counterparts. Although this equation is exact, it is of small practical applicability. In fact, it makes several assumptions regarding the quality of the data available from the initial atomistic trajectory. One of the most important assumptions is that the partition functions are available for both the atomistic and the CG system. This is usually not the case and sometimes the very reason why the system is coarse-grained, namely, to sample more configurations. Hence, it would be of great interest to check how important this assumption really is.

Another aspect worth noting is that eq 4 does *not* assume any specific partitioning (namely, resolution) of the system. However, we will show that under the approximation involved in the CG force field some resolutions give better agreement with the atomistic simulations than others.

Definition of the Problem. The section above showed that once a proper mapping is defined it is possible, at least in principle, to construct a CG model for the system under examination. However, for practical purposes, problems can arise associated with lack of a good initial sampling.

To formalize this problem, we start with a simple example of a two-state system that is characterized by two minima at the same energy separated by a barrier.

From a kinetic point of view, the barrier is what determines the transition time τ between the two states. The time τ tells us that for time scales smaller than τ the system is very likely trapped in one of the two minima. This is a typical problem in molecular dynamics simulations, where the time scales explored are often much smaller than the typical τ of the system under investigation.

On the other hand, from an equilibrium point of view, ensemble averages of a quantity σ are defined on the full free-energy surface by

$$\langle \sigma \rangle = \int \mathrm{d}x p(x)\sigma(x) \tag{5}$$

For practical purposes, however, the probability $p(x)$ of finding the system in the state x is significantly different from zero only in very few regions of the free-energy surface. From the point of view of statistical mechanics, $p(\vec{x}) = \exp(-\beta H(\vec{x}))/Z$ (canonical ensemble), where Z is

$$Z = \int e^{-\beta H(\vec{x})} \mathrm{d}\vec{x} \tag{6}$$

and $H(\vec{x})$ is the Hamiltonian of the system and $\vec{x}$ is a $6N$ dimensional variable describing position and momentum of a system of N particles. As noted above, this function is

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

significantly different from zero only around the minima of the free-energy surface, so that[49−52]

$$Z \simeq \sum_{\alpha} e^{-\beta\Omega_\alpha} \int_{X_\alpha} e^{-\beta\Delta_\alpha(\vec{x})} d\vec{x} \tag{7}$$

where X_α is the set of points belonging to the minimum α; Ω_α is the value of the Hamiltonian at the minimum α; and $\Delta_\alpha(\vec{x})$ is a non-negative function that evaluates the energy at the point $\vec{x}$ inside the basin α. The formalism above was taken from the theory developed by Stillinger and Weber for dealing with inherent structures of liquids, for example, supercooled liquid.[49−52] For the purposes of this paper, it suffices to say that eq 7 implies that the partition functions (and consequently the probability of each point $\vec{x}$) can be properly evaluated only if the system is able to explore all the available minima of the free-energy landscape.

In the case, commonly met when coarse-graining a system, that not all the minima α are visited, but only a subset α', the partition function Z' is biased toward the sampled region of space, so that

$$\gamma = Z'/Z = \frac{\sum_{\alpha'} e^{-\beta\Omega_{\alpha'}} \int_{X_{\alpha'}} e^{-\beta\Delta_{\alpha'}(\vec{x})} d\vec{x}}{\sum_{\alpha} e^{-\beta\Omega_\alpha} \int_{X_\alpha} e^{-\beta\Delta_\alpha(\vec{x})} d\vec{x}} \tag{8}$$

where γ is a measure of this distortion.

To clarify this point even further, we consider a two-state system, for example, a molecular species that can only populate its cis or trans conformation, such as NMA. In this case, eq 5 can be approximated as (one-dimensional case)

$$\langle\sigma\rangle \simeq \int_{x_1-\delta_1}^{x_1+\delta_1} dx p(x)\sigma(x) + \int_{x_2-\delta_2}^{x_2+\delta_2} dx p(x)\sigma(x) \tag{9}$$

where the distribution is significantly different in a region $2\delta_i$ around the ith minimum x_i. The extension of $2\delta_i$ depends on the thermal fluctuations of the system and the local properties of the minimum.

However, if only one minimum is sampled, the exact probability $p(x) = \exp(-\beta H(x))/Z$ is changed into $p'(x) = \exp(-\beta H(x))/Z'$, affecting the outcome of the average. For example, if $\langle\sigma\rangle$ is the average torsional angle of one molecule and the molecule is trapped in either a cis or trans conformation, its value would be different.

It should be noted that eq 7 allows us to split the problem of sampling the full free-energy surface into two distinct sampling problems. The first consists of sampling the full set of minima α of the system. The second issue is related to sufficiently sampling the conformations X_α around each minimum. We will refer to the first issue as incomplete sampling, whereas the second one will be referred to as sampling statistics.

In this paper, we will critically examine this assumption using the simple molecule *N*-methylacetamide (NMA) which is characterized by two possible conformations: cis and trans. Thus, this system can be described by eq 9.

In particular, we wish to assess whether a force field constructed from either the cis or the trans conformation is capable of capturing the other conformation. This will be achieved by restraining the dynamic of the system in one of the two basins and using the resulting force field to reproduce the properties of the other basin. Formally, the procedure consists of fixing the probability $e^{-\beta\Omega_\alpha}$ of being in one of the minima and constructing the CG force field from the sampling of the function $e^{-\beta\Delta_\alpha(\vec{x})}$ in this minimum. The latter is then used as the best approximation to the CG force field defined on the full free-energy landscape.

For this specific system, the two basins are not equivalent, with the trans conformation highly favored. Thus, it is surprising to notice that the force field obtained by the cis conformation usually performs better.

Practically, our approach mimics the situation commonly found for small peptides. These molecules are usually flexible enough to populate multiple conformations, but only a few of them are accessible experimentally.

As a final note, we want to mention that eq 7 is useful to define a "local transferability" of the CG force field. With this term we mean that a force field constructed for one of the minima α can be used for describing the properties of the other minima. As will be shown in the remaining part of the paper, not all the CG force fields developed obey this property. This property is however very important because based on it we can determine whether a limited or biased sampling can be safely used for exploring the free-energy landscape of the system under examination.

Basis Set Expansion and Force Field Assessment. The solution to eq 4 involves, in its more general form, multibody potentials to describe the interactions among the CG interacting sites.

To better appreciate the importance of this relationship, eq 4 can be recast as

$$\frac{e^{-U(\{\vec{R}_I\})/k_BT}}{Z_N} = \frac{1}{z_n} \int d\vec{r}^n \prod_{I=1}^{N} \delta(\vec{R}_I - M_I(\{\vec{r}_i\})) \times e^{-u(\vec{r}^n)/k_BT} \tag{10}$$

This formulation shows how the canonical Boltzmann's factor for the CG model relates to its atomistic counterpart. This relation also states that the probability of finding one conformation in the ensemble explored by the CG dynamics has to exactly match the same probability in the atomistic system.

From eq 10, it is possible to compute the CG *m*-particle density $\rho_N^{(m)}$ for a system of N independent particles, defined as[4]

$$\rho_N^{(m)}(\{\vec{R}_I\}) = \frac{N!}{Z_N(N-m)!} \int d\vec{R}^{(N-m)} e^{-U(\{\vec{R}_I\})/k_BT} \tag{11}$$

Using eq 10, this definition becomes

$$\rho_N^{(m)}(\{\vec{R}_I\}) = \frac{N!}{z_n(N-m)!} \int d\vec{R}^{(N-m)} \times \int d\vec{r}^n \prod_{I=1}^{N} \delta(\vec{R}_I - M_I(\{\vec{r}_i\})) e^{-u(\vec{r}^n)/k_BT} \tag{12}$$

At this point, it should be noted that the r.h.s. of this equation is the definition in atomistic coordinates of the *m*-particle density. This relationship means, for example, that when mapping particles to their centers of mass each *m*th order density profile should match between atomistic and CG resolution.

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

A straightforward consequence is that the m-particle distribution function $g_N^{(m)}$, defined as

$$g_N^{(m)}(\{\vec{R}_I\}) = \frac{\rho_N^{(m)}(\vec{R}_1, ..., \vec{R}_m)}{\prod_{I=1}^{m} \rho_N^{(1)}(\{\vec{R}_I\})} \quad (13)$$

is the same in both atomistic and CG resolution.

A particular instance of eq 13 is the radial distribution function $g(r)$ that is equivalent to $g_N^{(2)}$ for the case of an isotropic fluid.

The relations derived above show that a correct approximation of the multibody potential will lead to a correct reproduction of the structural properties of the system under examination. However, multibody potentials pose major computational issues, which is why they are usually approximated by simpler functional forms. One of the most common approximation consists in approximation of the nonbonded interaction with a pairwise additive potential.

Under this assumption, the system obeys the Yvon−Born−Green (YBG) hierarchy of equations that establishes an explicit relation between the forces acting on the system and its structural properties (CG case)

$$k_B T \frac{\partial \rho_N^{(m)}(\{\vec{R}_I\})}{\partial \vec{R}_1} - \sum_{J=2}^{m} \vec{F}_{1J} \rho_N^{(m)}(\{\vec{R}_I\}) - \int d\vec{R}_{m+1} \vec{F}_{1,m+1} \rho_N^{(m+1)}(\{\vec{R}_I\}) = 0 \quad (14)$$

A similar equation holds for the atomistic system. As demonstrated above, if the CG procedure is successful, this equation must hold also for the case when the force is from the CG model and the m-particle distribution function from atomistic simulation (and vice versa).

In conclusion, eq 12 and eq 14 show that the quality of the CG force field can generally be assessed by looking at its structural properties. In general, the radial distribution function $g(R)$ is enough to characterize the system under investigation, and it will be used in what follows to assess the quality of the CG force field generated by the MS-CG methodology.

METHODS

Samples. The sample used in this study consists of 1000 molecules of NMA. Specifically, each molecule was built linking together the groups ACE and NMA available in the OPLS-AA force field.[16−18] The NMA molecule can undergo a transition between a trans and cis conformation. To build a reliable force field that can correctly reproduce both species, we performed simulations (described in the next section) in which each molecule was not allowed to undergo an isomerization transition. To achieve this, the dihedral angle OCNH of the peptide bond was modified from the original OPLS-AA force field and kept fixed. The functional for this dihedral potential V_{OCNH} was changed into

$$V_{OCNH}(\phi) = k(1 + \cos(\phi - \phi_0)) \quad (15)$$

where ϕ is the dihedral angle and ϕ_0 its equilibrium value with $\phi_0 = 180$ for the cis conformation and $\phi_0 = 0$ for the trans conformation. An arbitrary value of $k = 300$ kJ/mol was used. No other changes to the original atomistic force field were done.

For the purpose of this work, other samples have been constructed with a fixed OCNH dihedral, at different concentrations of the cis and trans conformations. Specifically: 100% cis, 100% trans, 50% cis + 50% trans, 70% trans + 30% cis. The reasons for each specific choice will be detailed later in this work.

All-Atom Simulations. Molecular dynamics simulations were performed to sample relevant configurations for NMA. The GROMACS package[53−56] was used in association with the OPLS-AA force field.[16−18] To avoid numerical instabilities during the coarse-graining of the system, no constraints were applied to the bonded interactions. Electrostatics were computed using the particle mesh Ewald method[57,58] with a real space cutoff of 1.2 nm. A cutoff of 1.2 nm was also employed for the computation of the van der Waals interactions. Long-range dispersion corrections for both energy and pressure were used beyond this cutoff. All simulations were performed at constant temperature $T = 300$ K using the Nosé−Hoover thermostat.[59−61] The equations of motions were integrated using the leapfrog algorithm[62] with an integration step of 1 fs. To find the equilibrium volume, a preliminary simulation at constant pressure $P = 1$ bar was performed using the Berendsen barostat[63] with a coupling constant $\tau_p = 1$ ps. For this specific simulation, a Berendsen thermostat[63] was also employed rather than the Nosé−Hoover thermostat, as both the Berendsen barostat and thermostat are more efficient for equilibrating the system.[64] The equilibrium box edge was 5 nm long. A cubic box was employed as well as periodic boundary conditions. The equilibrium volume was evaluated for the case of flexible NMA. To equilibrate the initial configuration, a 3 ns long simulation was performed at 2000 K in the constant temperature and volume NVT ensemble. A further 3 ns equilibration at 300 K in the NVT ensemble was performed. Finally, data were collected in the NVT ensemble at 300 K for 10 ns, storing the configurations every 1 ps.

Coarse-Graining Procedure. The different coarse-grained models presented in this work have been parametrized using the MS-CG methodology.[40−42,65] This methodology performs a mean least-squares fitting of the forces acting on each CG site using data from molecular dynamics simulations. Assuming that a set of atomistic coordinates $\{\vec{r}_i\}$ is given, the coarse-graining sites are located according to a mapping operator M that transforms those coordinates into a set of CG coordinates $\{\vec{R}_I\}$. In the current work, the mapping operator is defined as the function that maps the atomistic coordinates into their centers of mass. The forces acting on each CG site are parametrized using the MS-CG algorithm as explained in the section Theoretical Background. The CG potential $\mathcal{V}$ has a general form of

$$\mathcal{V} = \mathcal{V}_b(R_{IJ}) + \mathcal{V}_a(\theta_{IJK}) + \mathcal{V}_d(\phi_{IJKL}) + \mathcal{V}_{nb}(R_{IJ}) \quad (16)$$

where the intramolecular interactions are represented by the chemical bond potential $\mathcal{V}_b(R_{IJ})$, the angular potential $\mathcal{V}_a(\theta_{IJK})$, and the dihedral potential $\mathcal{V}_d(\phi_{IJKL})$). The nonbonded interactions $\mathcal{V}_{nb}(R_{IJ})$ are assumed to be pairwise additive and spherically symmetric.[47,66−68] The current version of the MS-CG code allows the use of three-body nonbonded interactions,[43,45] but we found that their use in the present work does not lead to any relevant improvement.

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

The MS-CG code fits the atomistic forces with β-spline functions. For the actual simulations, the intramolecular interactions have been fitted to analytic forms. Specifically

$$\mathcal{V}_b(R_{IJ}) = k_b(R_{IJ} - R_0)^2 \quad (17)$$

$$\mathcal{V}_a(\theta_{IJK}) = k_a(\theta_{IJK} - \theta_0)^2 \quad (18)$$

$$\mathcal{V}_d(\phi_{IJKL}) = \frac{1}{2}k_1[1 + \cos(\phi_{IJKL})] + \frac{1}{2}k_2[1 - \cos(2\phi_{IJKL})] \quad (19)$$

where R_0 is the equilibrium bond length and θ_0 the equilibrium angle. The nonbonded interactions were tabulated with a binning of 0.001 nm.

Coarse-Graining Simulations. The CG simulations were performed using the molecular dynamics package LAMMPS.[69] The simulations were performed according to the following protocol. The initial configuration was taken from the last snapshot of the atomistic trajectory. The system was initially equilibrated in the NVT ensemble at 2000 K for 1 ns and then equilibrated for 1 ns at 300 K. Production runs were performed at 300 K for 10 ns. The temperature was kept constant using the Nosé–Hoover thermostat.[59−61] The equations of motion were integrated according to the methodology proposed by Tuckerman et al.[70] with a time step of 1 fs.

Assessment of the Force Field. The purpose of this paper is to understand the impact of a biased sampling of the atomistic simulation on the CG potential constructed from these data. The procedure adopted is as follow:

1. An atomistic molecular dynamic simulation is performed with molecules restrained in their cis or trans conformation.
2. A CG force field is developed from the all-cis or all-trans atomistic simulations.
3. The first test consists of establishing that the CG force field is able to reproduce its reference atomistic simulation. For example, the CG force field built from an all-cis sample should be able to reproduce the properties of the same atomistic system.
4. At this point, each CG force field is used to reproduce the property of the other system. For example, the all-cis CG force field is used to reproduce the properties of the all-trans atomistic system.

The quantity used to evaluate the quality of the CG force field is the radial distribution function. This choice has been explained in detail in the section "Basis Set Expansion and Force Field Assessment". The full set of radial distribution functions is reported in the Supporting Information for each resolution examined.

RESULTS

Model I: United Atom Resolution. The first model considered, shown in Figure 1, is very close to the original atomistic model. The only portions of the molecule that are coarse-grained are the two terminal methyl groups, now termed AC (for the ACE terminal group) and NA (for the NAC terminal group). Even though the two groups have the same chemical identity, they are better described by different effective potentials. This is a consequence of the fact that the two methyl groups are chemically bonded to groups with different chemical properties. This level of resolution, even though not very far from the underlying atomistic model, is a good test case for some of the approximations that will be used for the other models that we will study. The first important approximation is the lack of electrostatic interactions, with the latter now implicitly accounted for in the effective potential developed. To assess how this approximation affects the structural properties of liquid NMA, we report in Figure 2 both

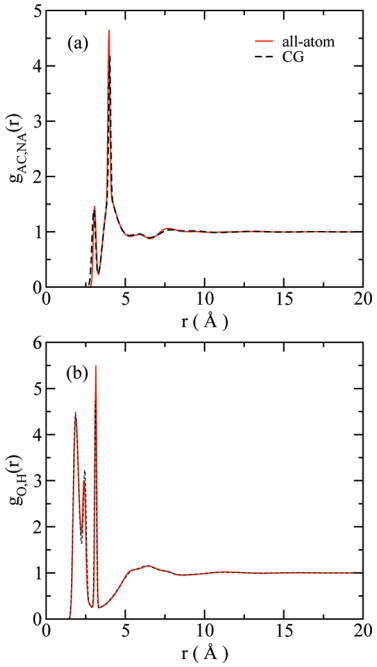


Figure 2. Radial distribution functions for model I. (a) The $g_{AC,NA}(r)$ reflects the internal degrees of freedom, (b) whereas the $g_{O,H}(r)$ between the oxygen and the hydrogen atom indicates how well the hydrogen bonds are reproduced. Only the situation for a fully flexible model is reported. All the RDFs for this model can be found in the Supporting Information.

the radial distribution function (RDF) $g_{AC,NA}(r)$ between the two terminal groups and $g_{O,H}(r)$ between the hydrogen and oxygen atoms. These two RDFs are particularly important because $g_{AC,NA}(r)$ is extremely sensitive to internal degrees of freedom (namely, to the correct parametrization of bonds, angles, and dihedral), whereas $g_{O,H}(r)$ accounts for the correct reproduction of the hydrogen bonding network.

The agreement between the coarse-grained model and the original atomistic model is excellent at this level of resolution. This shows that the removal of the explicit electrostatics does not considerably affect the structural properties under consideration. In addition, the CG potential takes into account the fact that in a condensed matter phase particles screen each other. In the specific case of NMA, a cutoff of 10 Å for the CG potential is used, even though a cutoff of 12 Å was used in the atomistic case for both the dispersion forces and the real space part of the electrostatic.

In addition to the computational efficiency of this model, another interesting aspect comes from the coarse-graining procedure itself. The first step in validating the CG force fields was to check whether there exists a force field that can reproduce at the same time both the cis and the trans conformation of NMA. To verify this, the MS-CG methodology was initially applied to the all-cis and the all-trans sample

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

independently. The two force fields obtained were then compared.

Even though both the all-cis and the all-trans samples converged to the same final force field (showing, that for this level of resolution, incomplete sampling in the original atomistic simulations does not hinder the construction of a reliable CG force field), surprisingly enough, the all-cis sample converged much faster and more reliably than its all-trans counterpart. In fact, unless the sampling statistics are extremely good, the force field obtained from the all-trans potential is unable to reproduce the correct radial distribution functions.

We have found that even at the other level of resolution considered the effective potential obtained from the all-cis sample works better and can correctly reproduce the all-trans case. This observation is actually surprising in light of the fact that the cis conformation is not the most populated conformation. In particular, the fact that the peptide bond is usually in a trans conformation was exploited by Pauling and Corey[71,72] to predict the secondary structure of proteins even before their structures were determined experimentally. The fact that the cis conformation is instead the most suitable one to parametrize force fields that can reproduce both cis and trans conformations leads to the more general implication that in a coarse-graining procedure the most populated structure is not necessarily the best choice for developing a CG force field.

On the other hand, it was *not* possible to obtain any working force field starting from a flexible NMA molecule, namely in the situation when the molecule is allowed to freely switch between the trans and cis conformation. The failure to converge to the same force field as in the all-trans/all-cis case can be attributed to two effects. The first is the fact that when a molecule switches between the two conformations it explores transient transition states. These transient states are usually poorly sampled leading to numerical instabilities. These instabilities are linked to the fact that inside the poorly sampled region the potential is varying in a discontinuous way, making the fitting with analytic functions problematic. A second source of error can be related to the cross interactions between cis and trans molecules. Indeed, an all-cis sample shows different structural properties than an all-trans sample (see the RDFs in the Supporting Information). This means that the local environment around a cis-NMA is different than around a trans-NMA. When the two species are mixed, it can happen that some of the cross-interactions are not well sampled, leading once again to numerical instabilities.

To better investigate this behavior, a 50:50 cis/trans mixture was studied. In this mixture the molecules cannot switch between the two conformations but are restrained to remain in their initial cis or trans conformation. Two CG force fields were built: In the first, the same CG potential is assumed for all molecules, irrespective of their conformational state. In the second, a different CG potential is built depending on whether the molecules are in the cis or trans state. In the first case, no working force field was obtained, whereas the second procedure led to a working force field, the same as obtained previously from the all-cis/all-trans case. This test shows that the instability in the flexible case is related to the sampling of the cross interactions between cis and trans molecules and not to the transient states explored when the molecule switches conformation. Transient states are, by definition, unlikely. As a consequence, it makes sense that their impact on the final force field is minimal. This property will be very useful for building lower-resolution models.

We have established that a CG force field does exist that can reproduce both the cis and trans conformation, but is it possible to build a flexible CG model? Since the same angular and bond potentials can be obtained from the all-trans and all-cis sample, the problem lies in how to treat the dihedral potential.

Our study with the 50:50 mixture shows that the transition states are not likely to introduce any particular instabilities. Furthermore, the internal degrees of freedom of the molecule, such as the dihedral angle, are probably decoupled from the nonbonded interactions. For these reasons, we took the dihedral angle potentials from the coarse-graining of the flexible atomistic sample. In this manner, we were able to obtain a flexible CG model. As can be seen in Figure 2, the radial distribution function is in excellent agreement with the underlying atomistic model.

Model II: 5-Bead Model. Model II has the same resolution as model I, with the only difference being that the carbon and nitrogen atoms are now considered as a unique interacting site. This new site is named PE (for "peptide bond"). We found that with this model it is not possible to find a force field capable of reproducing both the trans conformation and the cis conformation of the molecule. For this reason, we need to consider separate force field sites for the trans conformation (denoted with a prefix *t*, such as *t*O, *t*PE) and for the cis conformation (with prefix *c*, such as *c*O, *c*PE).

This force field was parametrized using a 50:50 mixture of nonflexible cis and trans conformation, and a different force field was built for the cis and trans conformations, plus the cross interactions. The only constraint was that the terminal groups AC and NA have the same parameters for the nonbonded part of the potential irrespective of whether they belong to a molecule with a trans or cis conformation (this restraint has a minimal impact on the CG force field). To assess the transferability of this CG force field, simulations for an all-cis and an all-trans sample were performed and compared to the atomistic case. Figure 3 shows a good agreement between the CG and the atomistic simulations. (A more detailed comparison is reported in the Supporting Information.) At equilibrium, the atomistic simulation has an average trans:cis ratio of 70:30. As stated above, the construction of a flexible CG model for this resolution is not possible. However, a comparison with the atomistic data is still possible, once a CG 70:30 mixture of nonflexible CG molecules is used for the analysis. This approximation correctly reproduces the radial distribution functions of the atomistic flexible model as shown in Figure 3 and Figure 4.

At this point, it is of interest to compare this model with the one presented in the previous section. The difference lies only in the fact that the C and N atoms involved in the peptide bond are now coarse-grained into the CG site PE. The effective potentials acting on the site PE in model II and the sites C and N in model I are compared in Figure 5. Except in the case when AC is compared, we report an average value of C and N together. This "average" potential is only correct if the interaction sites keep the same distance from both C and N at any time. Nonetheless, it gives a rough estimate of the effective potential expected when these atoms are coarse-grained into the PE site. With the exception of the case when AC is considered, we find reasonable agreement with the effective potentials computed. The AC site shows a peculiar behavior, with its effective potentials somewhat flatter between 4 and 6 Å and with two shallow minima around 4.2 and 5.6 Å. It is also interesting to check how the interactions between

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

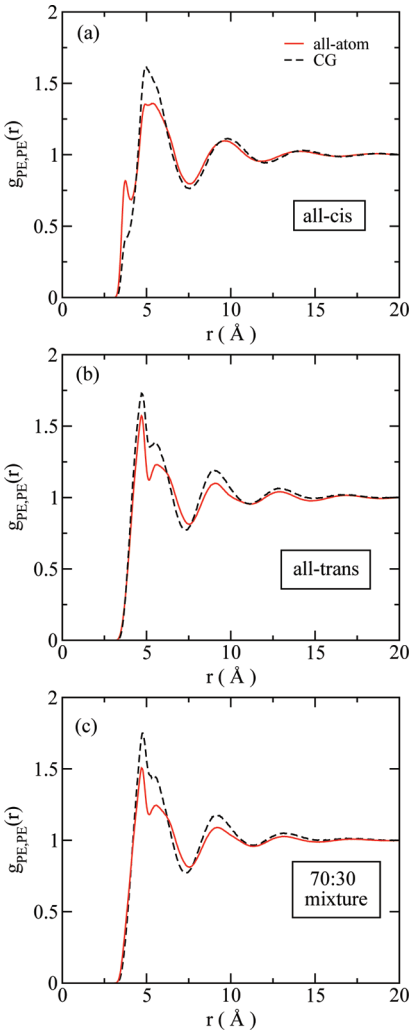


Figure 3. Comparison of the radial distribution function for the group PE in model II, in the case when (a) an all-cis, (b) an all-trans, or (c) a 70:30 mixture is considered. Notice that the mixture is composed of nonflexible CG molecules.

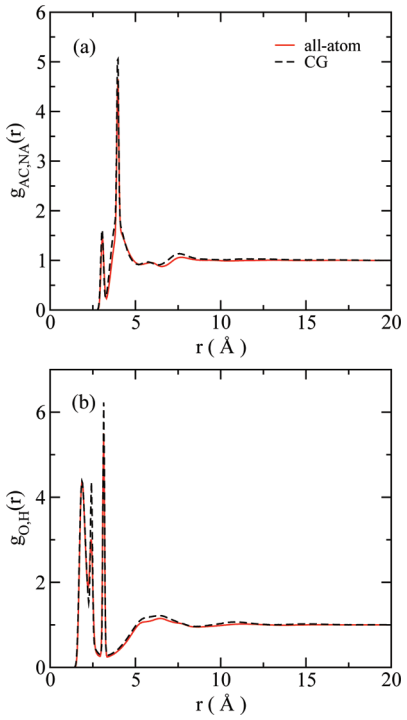


Figure 4. Radial distribution functions for model II. (a) The $g_{O,H}(r)$ shows that the hydrogen bond network is well described at such resolution. (b) The $g_{AC,NA}(r)$ is reported as well to assess the quality of the coarse-graining for the internal degrees of freedom. More RDFs are shown in the Supporting Information. It should be noticed that the comparison is between the atomistic NMA and a mixture 70:30 of trans and cis nonflexible CG molecules. See the text for further details.

O and H change, these interactions being critical for the formation of hydrogen bonds. As shown in Figure 5 they have mainly electrostatic interactions with the C and N atoms in model I, whereas in model II, the electrostatic interactions are considerably weakened, in particular for the H site. The O site seems to experience an average repulsion from the PE interaction sites, probably due to the dominance of the interaction with the N site. In summary, the main deviations from the atomistic case are probably electrostatic in nature. This would also explain why the *c*PE−*c*H interactions are poorly reproduced at short-range. To better understand this point, it is useful to recall that atomistic force fields are built by associating an excluded volume (dispersion forces) and a charge (electrostatic interactions) to each atom type. In the CG models proposed in this work, these interactions are replaced with one spherically symmetric effective potential. However, the most general description for a set of charges (such as atoms C and N) is given by the multipole expansion[73]

$$V(\vec{R}) = \frac{1}{4\pi\varepsilon_0}\sum_{l=0}^{\infty}\sqrt{\frac{4\pi}{2l+1}}\frac{1}{R^{l+1}}\mathcal{A}(\hat{R}, \{\vec{r}_i\}) \qquad (20)$$

where $\vec{R} = R\cdot\hat{R}$ is the location of the CG particle with modulus R and versor $\hat{R}$. $\mathcal{A}$ is the function that takes into account the angular dependence of the potential $V(\vec{R})$ and that contains information about the location $\{\vec{r}_i\}$ and the magnitude of the charges in the atomistic system. This formula is valid when $R > \max\{|\vec{r}_i|\}$.

The first term in the above expansion is the total charge of the CG site, followed by the total dipole, quadrupole, etc. As can be seen from this equation, the higher the order of the term in the expansion, the shorter its interaction length. Another aspect that becomes obvious from this expansion is that the highest order terms are usually not spherically symmetric.

Similar considerations can be applied to the specific case of the PE group as well. The effective interaction between the sites PE and O is on average repulsive, suggesting that the first term in the expansion of eq 20 is the dominant one. For these reasons a spherically symmetric potential can optimally reproduce this interaction. On the other hand, the PE−H interaction suffers from a strong averaging due to the charge compensation between the C and N sites. In this situation, the first term in the expansion of eq 20 is much smaller than in the previous case. For these reasons, the other terms in the expansion start becoming important. In this case, a spherically symmetric potential may not correctly reproduce the effective interaction, particularly at short distances. This would explain why the *t*PE−*t*H is less affected than the *c*PE−*c*H. Indeed, hydrogen atoms inside a trans conformation tend to be further away than in a cis conformation as shown by their radial

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

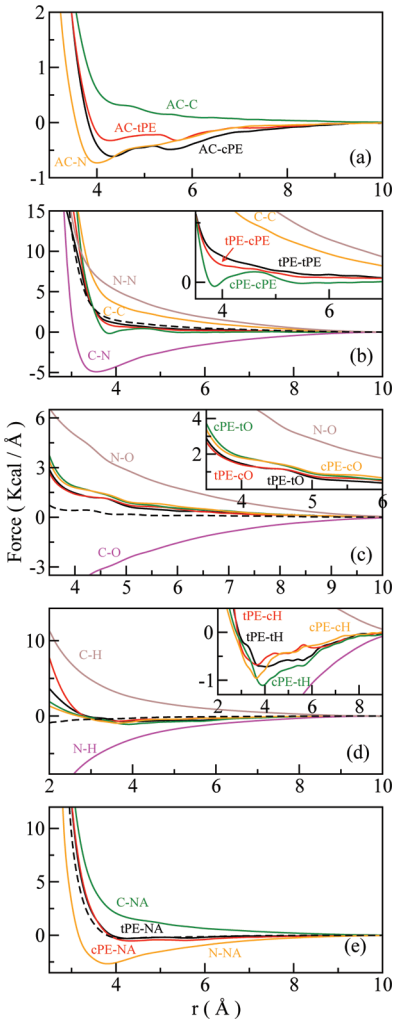


Figure 5. Comparison of the effective forces acting on the interaction sites C and N of model I and PE of model II. The average force (broken line) represents the interaction that should act on a CG site in the case that, on average, it is equidistant from both the C and N sites at any time. Each label refers to the line with the same color.

distribution function (see Supporting Information). In this way, short-range multipole moments do not have a relevant effect.

Model III: Linear Chain. This model resembles a typical polymer-like model for the protein. Interestingly, for this model, the all-trans sample was unable to yield a working force field, while the force field obtained from the all-cis conformation was able to correctly reproduce both the trans and the cis conformation but not the bonded interaction. For these reasons, the same nonbonded interactions were used for both the cis and trans conformation, whereas the bonded interactions were independently fitted (Figure 6). The major difference between the two lies in the parametrization of the angle OC−HN−NA that changes from 142° in the case of the cis conformation to 102° for the trans case (Figure 7). This model, as well as model IV, shows how multibody effects arise when developing coarse-grained models. In model II, their origin was attributed to electrostatic interactions. In the present case, the problem is due to the coupling between internal degrees of freedom. A commonly used assumption, which has

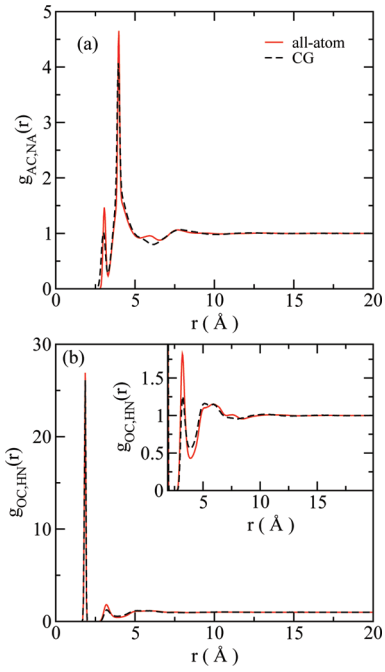


Figure 6. Radial distribution function for model III. (a) $g_{OC,HN}(r)$ shows that hydrogen bonds are well reproduced in the CG model. (b) $g_{AC,NA}(r)$ indicates that the internal degrees of freedom can be reproduced correctly as well. Notice that the figure refers to a mixture 70:30 of nonflexible CG molecules. See text for details.

been used in this work as well, consists of decomposing the bonded interaction into the form

$$\mathcal{V}_{\text{bonded}}(R_{IJ}, \theta_{IJK}, \phi_{IJKL}) = \mathcal{V}_{\text{b}}(R_{IJ}) + \mathcal{V}_{\text{a}}(\theta_{IJK}) + \mathcal{V}_{\text{d}}(\phi_{IJKL}) \quad (21)$$

This decomposition is however only an approximation. In real molecules, cross term correlations can appear. In this situation, the basis set expansion should be augmented including bond-angle, bond-dihedral, and angle-dihedral contributions. In the specific case of model III, there is a very strict association between the OC−HN−NA angle and the dihedral angle.

Model IV: 3-Bead Model. The last model to be considered consists of three interaction sites: the two terminal groups AC and NA and the peptide bond PE. As should be expected from the analysis of models II and III, it is not possible to construct a common force field for both the cis and trans conformation. For this reason, the same nonbonded interactions were used for the sites AC and NA, whereas a different force field was developed for the PE site, namely, *c*PE for the cis case and *t*PE for the trans conformation. The force field was parametrized using a 50:50 mixture of nonflexible cis and trans molecules, to obtain also the cross-interactions between cis and trans conformations. A CG mixture of 70:30 trans:cis was then validated against the atomistic simulations in the same way as for model II (Figure 8).

This model is a good example of the most extreme situation that is met when coarse-graining, namely, a strong association between bonded and nonbonded interactions. To understand this point better, it is useful to notice that a transition between cis and trans conformations can be modeled as a change in the valence angle in this CG model. This can be easily achieved using a form for the CG valence angle potential that allows

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

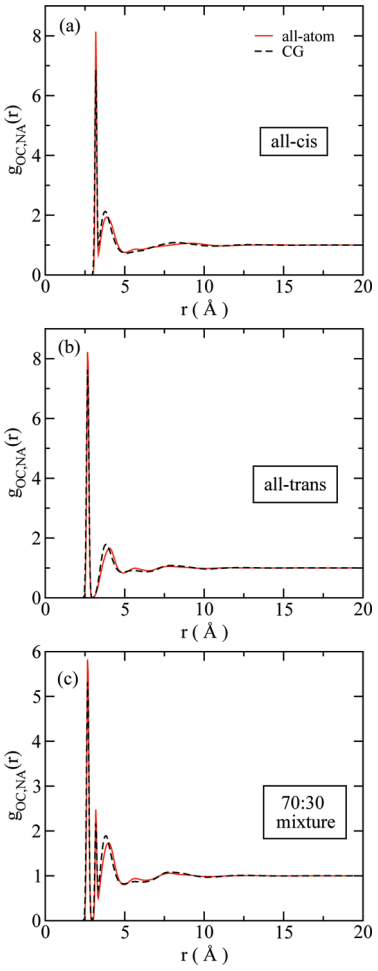


Figure 7. Impact of the angle OC−HN−NA in model III on the internal degrees of freedom for the case of (a) all-cis or (b) all-trans molecules or (c) when a mixture is used. Notice that the nonflexible CG mixture is plotted against the flexible atomistic model.

two minima. A transition between the two minima would correspond to a cis to trans conformation. Despite this simple solution, the nonbonded interactions are different for the two conformations, so that they should change as well during the transition, making the overall parametrization nontrivial.

In this situation, the only viable option is to use a more complex functional form for the nonbonded part of the potential, eventually including multibody interactions in the nonbonded part of the CG potential. However, the inclusion of multibody interactions usually leads to computationally inefficient force fields that may not be competitive against a higher resolution model with a simpler functional form. Although in the present version of the MS-CG code it is possible to parametrize 3-body nonbonded interactions,[45] their use did not lead to any appreciable improvement compared to (the more computationally efficient) 2-body approximation.

DISCUSSION AND CONCLUSIONS

In this work, multiple force fields were constructed and validated to represent the same NMA molecule at different resolutions.

In principle, when a system is coarse-grained, an exact form for the force field will include multibody interactions as defined by eq 4. However, multibody potentials are usually computationally inefficient, so that standard all-atom force fields (and related MD packages) limit themselves to the simpler pairwise, spherically symmetric potentials. In the case of all-atom force fields, this type of approximation is usually sufficient to correctly reproduce experimental observations. However, the validity of this assumption for CG models, in particular when the coarse-graining becomes very aggressive, has not been systematically assessed. For these reasons, it is of broad interest to investigate how far a system can be coarse-grained using functional forms that are commonly employed in standard fully atomistic simulations.

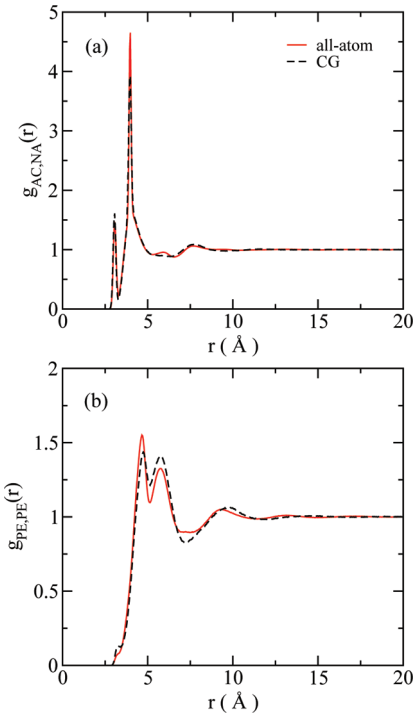


Figure 8. Radial distribution functions for model IV. The case shown is the 70:30 trans:cis nonflexible CG mixture.

In the context of the present work, we showed that we can construct a "united atom model" (model I) using a single short-range CG potential instead of the atomistic dispersion forces and full electrostatics. An established simulation protocol that employs short-range electrostatic interactions to speed up simulations is the reaction field method.[74−76] In this approach, the parameters for the interactions are usually taken from an existing atomistic force field and then properly smoothed to zero at the cutoff. This cutoff is usually longer than the real space cutoff used in association with PME. The simulations presented in this paper show that atomistic force fields (like the OPLS-AA force field here used) can be optimally coarse-grained to provide smooth and short-range potentials that could be coupled to the reaction field method.

As the atomistic system was further coarse-grained, the ability of reproducing subtle effects associated with the electrostatics degraded. This is clearly seen in model II, which differs from model I only in the coarse-graining of the C and N atoms involved in the peptide bond into one single site. As we explain in the section about model II, this seemingly minor modification invalidates the assumption that the effective force field is spherically symmetric, likely introducing the necessity of including dipoles.

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

In the case of model III, a polymer-like 4-bead model, there is a coupling among internal degrees of freedom. In particular, different valence angles are associated with different values for the dihedral angle. This kind of coupling is associated with an overall geometry change of the molecules. An incorrect treatment of this coupling can lead to improper packing of the molecules.

Models II and III are remarkable examples of how complex interactions can originate from atomistically simple functional forms for the potentials. Thus, it is not surprising that when the system is further coarse-grained both effects add up and a totally different potential had to be developed for the cis or trans conformation in model IV.

These latter models show that simple functional forms for the potentials may not be suitable to coarse-grain the system at very low resolutions. The reason for this is that as the topology becomes simpler the force field can become more and more complicated to match local properties of the system. A simple example is a system of two charges of opposite sign. If they are coarse-grained into a single CG site, the potential that better describes their interaction with the environment is a dipole (see also the discussion about model IV). More complex situations can arise when charged, hydrophilic, and hydrophobic regions of the system are coarse-grained together. It is clear from this example that there is an intrinsic connection between the way the system is partitioned and the complexity of the CG force field.

Despite these drawbacks, we have shown that models II, III, and IV can be successfully used to reproduce equilibrium properties. Although a flexible model could not be developed, we were able to construct a successful force field from a system composed of two nonflexible species (the cis and trans conformations), with a composition matching the equilibrium population of cis and trans conformations from atomistic simulations.

Another aspect that is central to this work is the question of the importance of incomplete sampling. This question becomes particularly pertinent when the system under examination cannot be explored through atomistic simulations because of the time and length scales involved. In this situation, the atomistic simulation can only access a limited number of possible conformations. It then becomes critical to assess whether a CG force field obtained from an incomplete atomistic sampling is suitable to explore the properties of the system at longer time and length scales or whether the inadequate sampling has introduced biases or artifacts. Our study shows that in some cases it is indeed possible to build an accurate force field from incomplete sampling. A requirement is that force fields generated from different samples (for instance an all-cis and all-trans sample) converge to the same common force field (as was the case in model I). The first important aspect is to identify N independent conformations of the system. For example, in the case of biomolecules, one structure, representative of the folded state, can be taken from the PDB database, and another, describing an unfolded state, can be generated, for instance, by heating the native structure to high temperatures. Once these conformations are available, two MD trajectories can be performed, and then two separate CG force fields can be developed. If the two resulting force fields are the same, then they can be used to study the system (in this example, the conformational change between the unfolded and folded state). If the two force fields fail to converge, a higher resolution model is required. These aspects of the parametrization were detailed and carefully assessed for model I presented in this work.

On the other hand, if one is not concerned about modeling of the transition state between two different conformations, a lower resolution model can be used, with different force fields for each conformation, so long as each conformation is populated according to its equilibrium distribution. The latter information can eventually be available from experimental data. In this situation, the advantage is a considerable reduction in the number of degrees of freedom.

For instance, consider the case of macromolecular crowding[77−79] (as in the inside of the cell), where the phenomena of interest happen in an environment crowded with other molecular species. While these species may not be directly involved in the process under investigation, their mere presence, at high concentrations, can have a considerable impact on the process of interest. In this kind of environment, a detailed characterization of these crowding agents is not needed, but these particles can nonetheless not be ignored. Thus, the process under examination can be modeled at a very high resolution, whereas the crowding agents can be aggressively coarse-grained (such as in model IV) to retain only their more basic features.

In conclusion, a central question that we sought to address in this work is defining what constitutes a good CG model and under which conditions a reliable force field can be constructed. The answer to this question is not unique and is tied to which property is deemed to be most relevant for the correct description of the phenomena of interest.

The two examples given above, namely, crowding versus NMA cis−trans dynamics, illustrate that different properties can become relevant at different resolutions, setting different boundaries to what can be considered the limitation of a good CG model. Ultimately, whether a CG model is optimal or not depends on the context. If the model has to be used to sample the relevant conformations of a molecule, then the CG model requires a higher level of detail (see model I in the present work). On the other hand, sometimes the molecules to be coarse-grained are not immediately relevant to the system under examination (for example, in the case of a buffer or crowding agent). In this latter situation, a more aggressive coarse-graining approach can be employed, eventually using a different force field for different conformations (see model IV in the present work).

Finally, this paper opens to the possibility to simulate complicated molecular systems, in particular proteins. The NMA is reputed a good model for the peptide bond,[80] and the various levels of resolutions explored in this paper allow for the study of proteins in a complex environment, namely, surrounded by crowders and buffer molecules, eventually using a different level of resolution for the different portions of the system. Connected to these aspects is the challenge to correctly capture cooperative effects that are at the heart of protein folding. In this spirit, Klimov and Thirumalai[81] have successfully defined a quantitative connection between resolution of the CG model and cooperativity and its impact on the reproduction of experimental data.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

For each model a table with the force field parameters and all the radial distribution functions. This material is available free of charge via the Internet at http://pubs.acs.org.

dx.doi.org/10.1021/jp2097263 | J. Phys. Chem. B 2012, 116, 8337−8349

AUTHOR INFORMATION

Corresponding Author

*E-mail: shea@chem.ucsb.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

L.L. acknowledges Vinod Krishna for helpful discussions about the theoretical aspects of this manuscript. Support from the National Science Foundation (MCB-1158577) and the David and Lucile Packard Foundation is gratefully acknowledged. This research was supported in part by the National Science Foundation through TeraGrid resources provided by the Texas Advanced Computing Center under grant number TG-MCA05S027. This research used ShaRCS, UC Shared Research Computing Services Cluster, which is technically supported by multiple UC IT divisions and managed by the University of California, Office of the President.

REFERENCES

(1) Chaikin, P. M.; Lubensky, T. C. *Principles of Condensed Matter Physics*; Cambridge University Press: Cambridge, 2000.
(2) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Oxford University Press: New York, 1986.
(3) de Gennes, P. G.; Prost, J. *The Physics of Liquid Crystals*; Oxford University Press: Oxford, 1995.
(4) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 3rd ed.; Academic Press: New York, 2006.
(5) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(6) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461−469.
(7) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(8) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(9) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950−1958.
(10) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(11) García, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782−2787.
(12) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(13) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(14) Feller, S. E.; MacKerell, A. D. *J. Phys. Chem. B* **2000**, *104*, 7510−7515.
(15) Foloppe, N.; MacKerell, A. D. Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(16) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(17) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827−4836.
(18) Kaminski, G. A.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(19) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: New York, 2002.
(20) Tozzini, V. *Acc. Chem. Res.* **2010**, *43*, 220−230.
(21) Shea, J.-E.; Brooks, C. L. III *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535.
(22) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. *Annu. Rev. Biophys.* **2010**, *39*, 159−183.
(23) Roland, C. M. *Macromolecules* **2010**, *43*, 7875−7890.
(24) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.
(25) Ganser-Pornillos, B. K.; Yeager, M.; Sundquist, W. I. *Curr. Opin. Struct. Biol.* **2008**, *18*, 203−217.
(26) Krishna, V.; Ayton, G. S.; Voth, G. A. *Biophys. J.* **2010**, *98*, 18−26.
(27) Straub, J. E.; Thirumalai, D. *Curr. Opin. Struct. Biol.* **2010**, *20*, 187−195.
(28) Straub, J. E.; Thirumalai, D. *Annu. Rev. Phys. Chem.* **2011**, *62*, 437−463.
(29) Wu, C.; Shea, J.-E. *Curr. Opin. Struct. Biol.* **2011**, *21*, 209−220.
(30) Ganapathy, R.; Buckley, M. R.; Gerbode, S. J.; Cohen, I. *Science* **2010**, *327*, 445−448.
(31) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13785−13797.
(32) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323−2347.
(33) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(34) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730−3737.
(35) Müller-Plathe, F. *ChemPhysChem* **2002**, *3*, 754−769.
(36) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869−1892.
(37) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211−3223.
(38) Savelyev, A.; Papoian, G. A. *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 20340−20345.
(39) Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(40) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.
(41) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
(42) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
(43) Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954−965.
(44) Krishna, V.; Larini, L. *J. Chem. Phys.* **2011**, *135*, 124103.
(45) Larini, L.; Lu, L.; Voth, G. A. *J. Chem. Phys.* **2010**, *132*, 164107.
(46) Izvekov, S. *J. Chem. Phys.* **2011**, *134*, 034104.
(47) Mullinax, J. W.; Noid, W. G. *Phys. Rev. Lett.* **2009**, *103*, 198104.
(48) Das, A.; Andersen, H. C. *J. Chem. Phys.* **2010**, *132*, 164106.
(49) Stillinger, F. H.; Weber, T. A. *Phys. Rev. A* **1982**, *25*, 978−989.
(50) Stillinger, F. H.; Weber, T. A. *J. Phys. Chem.* **1983**, *87*, 2833−2840.
(51) Sciortino, F.; Kob, W.; Tartaglia, P. *Phys. Rev. Lett.* **1999**, *83*, 3214−3217.
(52) Sciortino, F.; Kob, W.; Tartaglia, P. *J. Phys.: Condens. Matter* **2000**, *12*, 6525−6534.
(53) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(54) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; C.Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701−1719.
(55) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(56) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(57) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(58) Essman, U.; Perela, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(59) Nosé, S. *Mol. Phys.* **1984**, *52*, 255−268.
(60) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511−519.
(61) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(62) Verlet, L. *Phys. Rev.* **1967**, *159*, 98−103.

(63) Berendsen, H. J. C.; Postma, J. P. M; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(64) Hünenberger, P. *Adv. Polym. Sci.* **2005**, *173*, 105−149.
(65) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
(66) Izvekov, S.; Chung, P. W.; Rice, B. M. *J. Chem. Phys.* **2010**, *133*, 064109.
(67) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116−4127.
(68) Mullinax, J. W.; Noid, W. G. *J. Chem. Phys.* **2009**, *131*, 104110.
(69) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1−19.
(70) Tuckerman, M. E.; Alejandre, J.; López-Rendón, R.; Jochim, A. L.; Martyna, G. J. *J. Phys. A: Math. Gen.* **2006**, *39*, 5629−5652.
(71) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 729−740.
(72) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205−211.
(73) Jackson, J. D. *Classical Electrodynamics*; Wiley: New York, 1998.
(74) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451−5459.
(75) Barker, J. A.; Watts, R. O. *Mol. Phys.* **1973**, *26*, 789−792.
(76) Neumann, M. *Mol. Phys.* **1983**, *50*, 841−858.
(77) Zhou, H.-X.; Rivas, G.; Minton, A. P. *Annu. Rev. Biophys.* **2008**, *37*, 375−397.
(78) Dhar, A.; Samiotakis, A.; Ebbinghaus, S.; Nienhaus, L.; Homouz, D.; Gruebele, M.; Cheung, M. S. *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 17586−17591.
(79) Kudlay, A.; Cheung, M. S.; Thirumalai, D. *Phys. Rev. Lett.* **2009**, *102*, 118101.
(80) Whitfield, T. W.; Martyna, G. J.; Allison, S.; Bates, S. P.; Crain, J. *Chem. Phys. Lett.* **2005**, *414*, 210−214.
(81) Klimov, D.; Thirumalai, D. *Folding Des.* **1998**, *3*, 127−139.