JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

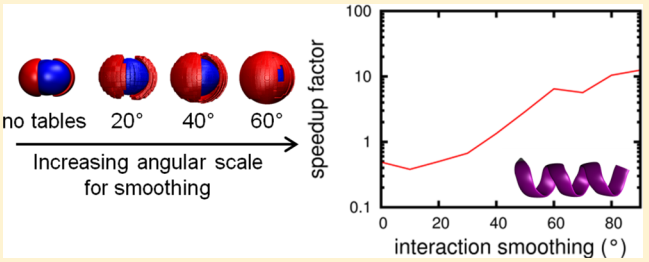
# Tunable Coarse Graining for Monte Carlo Simulations of Proteins via Smoothed Energy Tables: Direct and Exchange Simulations

Justin Spiriti and Daniel M. Zuckerman*

Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, United States

Supporting Information

**ABSTRACT:** Many commonly used coarse-grained models for proteins are based on simplified interaction sites and consequently may suffer from significant limitations, such as the inability to properly model protein secondary structure without the addition of restraints. Recent work on a benzene fluid (Lettieri, S.; Zuckerman, D. M. *J. Comput. Chem.* **2012**, *33*, 268−275) suggested an alternative strategy of tabulating and smoothing fully atomistic orientation-dependent interactions among rigid molecules or fragments. Here we report our initial efforts to apply this approach to the polar and covalent interactions intrinsic to polypeptides. We divide proteins into nearly rigid fragments, construct distance and orientation-dependent tables of the atomistic interaction energies between those fragments, and apply potential energy smoothing techniques to those tables. The amount of smoothing can be adjusted to give coarse-grained models that range from the underlying atomistic force field all the way to a bead-like coarse-grained model. For a moderate amount of smoothing, the method is able to preserve about 70−90% of the $\alpha$-helical structure while providing a factor of 3−10 improvement in sampling per unit computation time (depending on how sampling is measured). For a greater amount of smoothing, multiple folding−unfolding transitions of the peptide were observed, along with a factor of 10−100 improvement in sampling per unit computation time, although the time spent in the unfolded state was increased compared with less smoothed simulations. For a $\beta$ hairpin, secondary structure is also preserved, albeit for a narrower range of the smoothing parameter and, consequently, for a more modest improvement in sampling. We have also applied the new method in a "resolution exchange" setting, in which each replica runs a Monte Carlo simulation with a different degree of smoothing. We obtain exchange rates that compare favorably to our previous efforts at resolution exchange (Lyman, E.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2006**, *2*, 656−666).




## 1. INTRODUCTION

In an effort to simulate biological systems on larger spatial scales and longer time scales than can be achieved with all-atom simulation, the first coarse-grained models for proteins were constructed by Levitt and Warshel.[1,2] Since then, many coarse-grained force fields for biomolecules have been constructed according to a wide range of approaches.[3−37] Each of these approaches has its own advantages and limitations.

A large class of models explicitly uses experimental structural information. The simplest types, Gō and elastic network models, typically include only geometric features of the native state[7−11] or of a preselected set of structures.[12−16] The full Protein Data Bank can also be used to parametrize more chemically realistic models, typically based on multiatom beads, which may be restricted to a lattice[17−19] or not.[20,21] More elaborate geometries can also be used.[22−25,28,29] Presaging some of the present work, Ma et al. divided the amino acids into small "blocks" and constructed a knowledge-based packing potential dependent on the relative orientations of these blocks using orientational bins.[26,27] These force fields assume that the system under study is similar to those in the Protein Data Bank, which may not be true for all systems of biological interest.

It is also possible to parametrize a coarse-grained model directly from physicochemical data. A prominent example is MARTINI, which was originally developed for lipids[30,31] and subsequently extended to proteins.[32] For example, the Lennard-Jones parameters in MARTINI were derived primarily from the partitioning free energies of model compounds between aqueous and organic phases, while bonded parameters are derived from surveys of structures from the Protein Data Bank and comparisons with atomistic MD simulations.[31] However, the backbone particles of the MARTINI force field have different dihedral parameters depending on the type of secondary structure. Consequently, MARTINI cannot be used to model conformational changes that include changes in the secondary structure, such as protein folding or amyloid aggregation,[32] although efforts have been made to improve the force field for amyloid peptides.[33] With structured proteins, the inclusion of an elastic network model can improve the agreement with atomistic simulations,[38] at the price of limiting some large-scale motions. The PRIMO model[36] is constructed in a generally similar way. It is free from secondary structure

**Received:** July 15, 2014
**Published:** October 9, 2014

   dx.doi.org/10.1021/ct500622z | *J. Chem. Theory Comput.* 2014, 10, 5161−5177

biases but at the cost of a substantially finer resolution, with each particle representing approximately two to three atoms.

Another approach to coarse-graining involves systematically optimizing the potential energy function to maximize the agreement between forces or correlations determined in atomistic simulations and those determined in coarse-grained simulations. The most prominent example of this approach is the multiscale coarse-graining (MS-CG) method.[34,35] In principle, this method provides the coarse-grained potential energy function that comes closest to formally integrating out the degrees of freedom not included in the coarse-grained representation, given a specific functional form.[4,5] However, the resulting force fields are not strictly transferable across systems or thermodynamic states, so it may be necessary to repeat the parametrization process. Potential sensitivity to sampling and the choice of matching observables have been noted.[5,6,39]

The increase in speed that makes it possible to simulate larger systems for longer times using coarse-grained models comes primarily from three sources: the decrease in the number of particles that comes from replacing many atoms by a single coarse-grained "bead", the increased mass of coarse-grained particles (which enables a longer time step in MD simulations), and the smoothing of the energy landscape that results from averaging out fine-grained degrees of freedom.[3] There is a trade-off between this increase in speed and accuracy; a lower-resolution model offers more speedup at the cost of a reduction in accuracy. Since the parametrization strategy, as noted above, will dictate which observables will be modeled accurately, there is no single coarse-grained model or resolution that is optimal for all applications. Consequently, it would be helpful if, for example, the resolution of a coarse-grained model could be adjusted for each application, instead of being fixed.

Here we apply a highly flexible alternative strategy suggested by recent work employing tabulated distance and orientation-dependent interactions applied to a fluid of rigid benzene molecules.[37] In that work, tables were constructed of the interaction energy between two molecules of benzene as a function of the Cartesian coordinates of the net displacement and the two absolute orientations of the benzene molecules. The tables were used to calculate the energies in a Monte Carlo simulation. The radial distribution functions of the benzene were found to be very similar to those obtained in a Monte Carlo simulation conducted without the use of tables. Furthermore, the potential energy function represented by the tables could be smoothed by taking averages over neighboring energy values in the tables; by changing the number of cells averaged, the amount of smoothing could be varied. This smoothing was found to increase the diffusivity of the benzene molecules in the simulation by as much as 4.7 times compared with the original simulation without tables, without significant changes to the radial distribution functions. Combined with the computational saving obtained by tabulating the energies, this resulted in a net speedup of approximately 114 times.

In this paper, we extend the tabulation and smoothing strategy to construct coarse-grained models for proteins. We divide the 20 possible amino acids into rigid fragments and construct interaction tables between each pair of possible fragment types as a function of their relative displacement and orientation. We have also developed potential energy smoothing techniques appropriate to these tables. By adjusting the degree of smoothing, we can effectively vary the resolution of the model continuously, from a united-atom force field all the way to a MARTINI-like coarse-grained model in which each fragment is effectively replaced by a spherical particle. We note that potential energy smoothing has been applied in the past to proteins in the context of global optimization of the structure.[40−42]

The new strategy has a number of potential advantages. Because it is based on an atomistic force field rather than on experimental data or molecular dynamics simulations of specific systems, it is expected to have fewer implicit biases and be more transferable than other coarse-grained models, and should also be more suitable for multiscale modeling. The ability to control the effective resolution of the model through adjustment of the degree of smoothing allows selection of the best trade-off between atomistic accuracy and coarse-grained speed for a particular system, rather than being restricted to the decisions that were made when the coarse-grained model was first constructed. Furthermore, the smoothing technique has been constructed in such a way as to preserve the generalized second virial coefficient resulting from interactions between two fragments, implying that, to the extent that average properties of the system depend on interactions between pairs of fragments, they are kept unchanged by the coarse-graining procedure.

In addition, since the interaction tables are orientation dependent, they incorporate detailed information on the shape and charge distribution of each fragment. As will be shown below, the resulting model is able to maintain the structure of an $\alpha$-helix without additional restraints and performs reasonably for maintaining the structure of a $\beta$-sheet as well. By keeping track of the orientation of each fragment as well as its position, it is straightforward to generate atomistic structures during the simulation, allowing the use of atomistic bond, angle, and dihedral parameters, as well as further simplifying multiscale modeling.

Potential smoothing appears to be a novel way to implement the "resolution exchange" (ResEx) idea.[43,44] The ResEx idea, as pursued here, involves constructing a replica exchange simulation in which some replicas experience a coarse-grained model, while others simulate according to an atomistic model. In this way, the simulation benefits from the increased sampling provided by the coarse-grained model, while still providing canonical sampling according to an atomistic model. Earlier ResEx efforts were hampered by an inability to tune intermediate levels of resolution that were not simple mixtures of all-atom and coarse-grained models.[45] It is worth noting that, unlike temperature-based replica exchange, which is expected to sample unfolded space to a significant degree, Hamiltonian replica exchange and ResEx in principle can maintain sampling of folded configurations at all levels.[46]

We demonstrate preliminary results for this new method on two small systems: an $\alpha$-helix composed of 12 leucine residues (Leu12) and the $\beta$-hairpin consisting of residues 41−56 from the B1 domain of streptococcal protein G, otherwise known as the GB1 hairpin.[47] The first section of the paper describes how the amino acids are divided into fragments, how the tables are constructed, and how smoothing is accomplished. Particular attention will be given to the differences between the method presented here and our previous work on benzene.[37] Results from test simulations of this method on the two peptides are then described, along with additional tests of our method coupled with Hamiltonian replica exchange. The analysis of these simulations focuses on the ability of the method to preserve secondary structure and on the increase of sampling

that has been obtained. Finally, we discuss the strengths and weaknesses of the new method and suggest possible strategies for improvement.

## 2. THEORY AND METHODS

Our basic strategy of tabulating interactions and smoothing the tables follows previous work on benzene,[37] but details of our implementation differ significantly. Not only do the present coordinates and smoothing strategy differ, but treatment of polypeptides requires accounting for covalent interactions among the molecular fragments.

**2.1. Choice of Underlying Atomistic Force Field.** The orientation-dependent interaction energy tables used in this work were based on the CHARMM19 united-atom force field,[48] as implemented in TINKER.[49,50] Although an all-atom force field could have been used, a more elaborate fragmentation scheme would have been required to handle the $H^{\alpha}$ and $C^{\alpha}$ atoms while allowing the $\phi$ and $\psi$ bonds to remain rotatable. Also, given the goal of coarse-graining, united atom interactions seemed sufficient.

**2.2. Orientation-Dependent Interaction Energy Tables for Proteins.** The construction of the model for nonbonded fragments followed several steps. In overview, proteins were first divided into fragments that could be treated as rigid throughout the coarse of the simulation, similar to the way in which our previous work used rigid benzene molecules.[37] This ensured that tables of this interaction energy could be constructed in terms of only the relative position and orientation of the fragments, without any internal conformational coordinates. Second, a coordinate system was chosen and a discrete grid defined, in order to represent the relative displacement and orientation within a table. Here we chose to use spherical coordinates and Euler angles to construct the table, rather than the Cartesian coordinates and orientation library used previously. Once these choices had been made, the tables themselves could be constructed by calculating the interaction energy (sum of van der Waals and electrostatic energies) for each possible relative position and orientation for every pair of possible fragment types. (Covalently bonded fragments are discussed below.)

The 20 amino acids that generally make up proteins were divided into fragments according to the scheme shown in Figure 1, in which united-atom representations of the 20 amino acids out of which proteins are made are divided into 32 possible types of fragments. (A detailed list of the fragment types and their uses in representing the amino acids is shown in Table S1 in the Supporting Information.) Each of these fragments is approximately rigid in that it contains no rotatable bonds among its heavy atoms. A reference geometry for each of the 32 fragment types was constructed by energy-minimizing one of the amino acids containing the fragment in vacuum, then orienting the fragment so that the principal moments of inertia aligned with the coordinate axes. The configuration space sampled in our model is then represented by the position $\mathbf{r}_i$ of the center of mass and the orientation relative to this reference geometry (expressed as a quaternion $\mathbf{q}_i$[51−53]) for each fragment $i$. Since most peptide bonds are in the *trans* configuration, the reference geometries for peptide bond fragments have this configuration as well.

For each pair of fragment types, a six-dimensional table of interaction energies was constructed in terms of the relative displacement, expressed by spherical coordinates $(r, \theta, \phi)$ and the relative orientation expressed in terms of Euler angles $(\phi', \theta', \psi')$. In constructing the table, a regularly spaced grid was used for all the angular coordinates $(\theta, \phi, \phi', \theta', \psi')$. This grid was set up so that the resolutions for the angular coordinates $\Delta\theta = \Delta\phi$, and likewise the resolution for the orientational coordinates $\Delta\phi' = \Delta\theta' = \Delta\psi'$. An exponential grid was used for the radial coordinate $r$, so that the table had a higher resolution for close interactions between fragments, where the interaction energy varies more quickly with interfragment distance. For this exponential grid, the equations relating the radial grid index $n$ to the radius $r_n$ were as follows:

$$r_{n+1} = fr_n \quad \text{where} \quad f = 1 + \frac{\Delta r}{r_0}$$
$$r_n = f^n r_0$$
$$n = \text{round}\left(\frac{\log r_n/r_0}{\log f}\right) \tag{1}$$

The minimum radius $r_0$ was chosen to be 2 Å, and the maximum radius was chosen to be 12 Å for all tables; the maximum radius also served as a fragment based interaction energy cutoff, so that interactions between fragments more than 12 Å apart were not counted. During the subsequent simulations, the interaction energy between two fragments was determined by looking up the energy from the cell in the table corresponding to spherical coordinates and Euler angles closest to the relative displacement and orientation of the two fragments. (In the event two fragments were closer together than the minimum radius, an infinite positive energy was

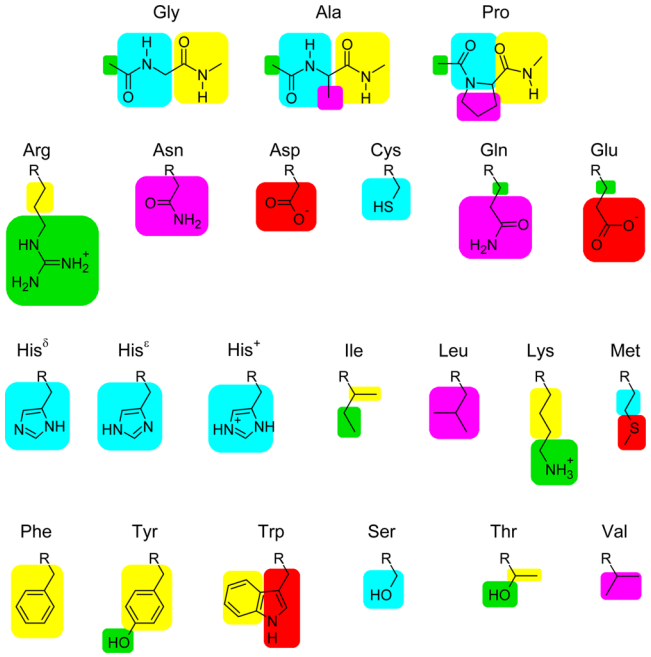



**Figure 1.** Fragmentation scheme for the 20 amino acids, including three tautomers of histidine. The fragments are shown as colored rounded rectangles underneath the atoms assigned to each fragment. The colors serve to distinguish fragments within each residue but do not correspond directly to fragment types. The division of the peptide backbone into fragments is shown for glycine, alanine, and proline. For the other amino acids, the division is the same as that for alanine. A detailed list of the fragment types and the atoms included in each is given in Table S1 in the Supporting Information.


dx.doi.org/10.1021/ct500622z | J. Chem. Theory Comput. 2014, 10, 5161−5177

Table 1. Resolution, Number, and Size of Tables Needed for Test Simulations

| system | | Leu12 | Leu12 (backbone tables) | GB1 hairpin | full set |
|---|---|---|---|---|---|
| polar | radial | 0.1−0.6 Å | 0.1−0.6 Å | 0.1−0.6 Å | 0.1−0.6 Å |
| | angular | 10° | 10° | 10° | 10° |
| | orientational | 15° | 15° | 30° | 30° |
| nonpolar | radial | 0.2−1.2 Å | 0.2−1.2 Å | 0.2−1.2 Å | 0.2−1.2 Å |
| | angular | 20° | 20° | 20° | 20° |
| | orientational | 30° | 30° | 30° | 30° |
| interaction tables | | 8 | 8 | 109 | 528 |
| backbone tables | | 0 | 2 | 0 | 12 |
| total size (GB) | | 1.33 | 1.36 | 2.79 | 15.04 |

returned so that all such configurations were rejected in our Monte Carlo simulations.)

Thus, the resolution of an interaction energy table can be expressed in terms of the minimum radial resolution, $\Delta r$, the angular resolution, $\Delta\theta = \Delta\phi$, and the orientational resolution, $\Delta\phi' = \Delta\theta' = \Delta\psi'$. The sizes and resolutions of the tables used in this work are shown in Table 1. In order to better capture the highly directional nature of hydrogen bonding, while at the same time limiting memory usage, a distinction was made between polar fragments (those capable of hydrogen bonding) and nonpolar fragments; higher resolution tables were used when both fragments were polar. The total size of all 528 tables needed to represent all possible interactions between the 32 different types of fragments, using the same resolutions as was used for the GB1 hairpin, comes out to approximately 15.1 GB.

The interaction energy used in this work was based fundamentally on the CHARMM19 force field,[48] and was constructed by replacing most of the van der Waals and electrostatic terms, by tables containing interaction energies also derived ultimately from the force field. In addition, some of the covalent terms involving the backbone may also be replaced by terms derived from tables. This results in the following form for the overall potential energy function.

$$U_{\text{total}} = U_{\text{interaction}} + U_{\text{backbone}} + U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}} + U_{\text{vdW}} + U_{\text{elec}} \quad (2)$$

In this equation, the first two terms are obtained by summing energies obtained from tables. $U_{\text{interaction}}$ comes from the previously described tables of the interaction energy between fragments, while $U_{\text{backbone}}$ comes from a separate set of tables giving the interaction energy between directly bonded fragments that form part of the peptide backbone. These "backbone tables" are described in more detail in section 2.5.

The remaining six terms have the same forms and parameters as in the force field, except that they do not include terms already counted in the tabulated interactions. The last two terms are needed in order to include certain van der Waals and electrostatic interactions that could not be included in the interaction tables because the fragments contained atoms separated by three or fewer covalent bonds. In order to incorporate a crude model for the effect of aqueous solvation, a distance dependent dielectric was used in which $\varepsilon = 2r$, and the charges of atoms in the side chains of charged amino acids were scaled by a factor of 0.6 relative to their original values. (Because the interaction energy is only tabulated in terms of the relative position and orientation of the fragments, it must be a pairwise interaction. Consequently, more sophisticated implicit solvent models, such as the generalized Born model,[54] could not be used, since the Born radii would depend on the positions and orientations of other fragments besides the two under consideration.)

The CHARMM19 force field has special rules for handling the van der Waals and electrostatic interactions involving atoms that are separated by three or fewer covalent bonds.[48] The interaction energies included in the tables were calcuated assuming that no such relationships existed between the fragments under consideration. Consequently, these tables could not be used for any pair of fragments that shared at least one pair of atoms separated by three or fewer covalent bonds. These interactions were instead calculated directly without the use of tables and comprise the $U_{\text{vdW}}$ and $U_{\text{elec}}$ terms in eq 2.

**2.3. Comparison to Previous Tabulation Strategy.** This structure for the interaction energy table outlined here has several advantages over the energy tables described in our previous work on benzene.[37] The fragments in the simulations described here are able to adopt any possible orientation with respect to the simulation frame of reference, whereas those in our previous work were restricted to a library of 10 orientations randomly chosen at the beginning of each simulation. In addition, because the tabulation is performed based on the relative orientation rather than the absolute orientation, interaction energy values are not duplicated over the many possible pairs of absolute orientations corresponding to the same relative orientation. This makes the tables significantly smaller; one interaction table for the GB1 hairpin is at most approximately 80 MB in size, whereas previously a table of similar resolution would have occupied about 1 GB in storage space.[37] The use of spherical coordinates and an exponential grid for the radial coordinate allow the table to have much higher resolution for closely interacting fragments, at the expense of coarser resolution for fragments that are further apart. This is helpful because the interaction energy is a much more rapidly varying function of relative position and orientation for closely interacting fragments, particularly when the fragments are irregularly shaped or strongly directionally dependent interactions such as hydrogen bonds are present. As discussed below, the spherical coordinates also aid in smoothing the table only in the angular and orientational directions, which allows smoothing to be applied without allowing a "collapse" of the protein through loss of the repulsive interactions between fragments.

The main disadvantage of this new approach is that significantly more calculation is required for each table lookup, since in order to look up the interaction energy between fragments $i$ and $j$, the program must calculate the relative orientation $\mathbf{q}_j\mathbf{q}_i^{-1}$ and convert this to Euler angles, as well as the relative displacement $\mathbf{q}_i^{-1}(\mathbf{r}_j - \mathbf{r}_i)\mathbf{q}_i$, which must be converted to spherical coordinates. This increase in calculation, coupled with the relatively small size of the fragments used here, means

that the implementation described here does not offer any significant computational speedup over calculating the interaction energies exactly during the simulation; any acceleration in the simulation time scales is due to the application of smoothing techniques to the tables. In the Discussion, we will point out avenues for addressing this computational cost.

*2.3.1. Overview of Tabulation.* In sum, proteins were divided into rigid fragments according to the scheme shown in Figure 1, and interaction energy tables were constructed in terms of the relative displacement and orientation between each possible pair of fragments. These energy tables embody the combined van der Waals and electrostatic terms of the CHARMM19 force field. While the precise resolution varied with interfragment distance and angular position, the tables had an effective resolution of on the order of 0.5 Å in the most important regions of configurational space. Thus, they can be considered to be nearly atomistic in resolution. These tables were used in place of calculating the corresponding van der Waals and electrostatic interactions during the simulations.

The use of finite-resolution interaction energy tables is similar in some ways to the use of finite resolution lattices to construct coarse-grained models.[17−19] Although in our models the fragments are not restricted to a lattice of absolute positions or orientations, the finite resolution of the tables and the use of the closest cell introduce similar possible errors to the use of a lattice. However, the resolution of our tables is finer than the 1.3−1.7 Å spacing of lattice models. Likewise, a 30° orientational resolution offers 864 possible relative orientations, about an order of magnitude more than the 56−90 possible basis vectors for $\alpha$-carbon traces available in these models.[19]

**2.4. Smoothing of the Interaction Tables.** As was done previously for benzene,[37] potential energy smoothing techniques were applied to the "raw" interaction energy tables in order to construct a coarse-grained model from them. These potential energy smoothing techniques are intended to reduce free energy barriers and improve sampling without changing the overall features of the free energy landscape, in much the same way that coarse-grained models reduce the frustration and thereby increase the rate at which conformational changes take place. The details of the smoothing algorithm differ due to the new structure for the table, however. Rather than averaging over a small cube of nearest neighbors in Cartesian space, the energies were averaged over table cells with the same radial coordinate but all possible angular and orientational coordinates, using a Gaussian-like kernel to ensure that nearest neighbor cells receive the greatest weight in the average. The kernels are the solutions to the diffusion equation in spherical[55] or orientational space. (Averaging in the radial direction was found to reduce the steric repulsion between fragments, leading to collapsed structures in the resulting simulations.) The width of this Gaussian can be used to adjust the extent to which nearest neighbor cells contribute to the average and consequently the degree of smoothing. It is analogous to the time allowed for diffusion; a wider Gaussian creates a smoother, more coarse-grained potential. In the limit of an infinitely wide Gaussian, all variation of the interaction energy with angular position or orientation is eliminated, leaving a spherically symmetric potential, similar to bead-based potentials such as MARTINI. As was done previously, Boltzmann averaging was used in order to prevent high, positive interaction energies from dominating the average and increasing the roughness of the landscape. The use of Boltzmann averaging also has the mathematical property that the generalized second virial coefficient for the interaction between two isolated fragments is left unchanged by smoothing.

Translational smoothing was carried out first, followed by orientational smoothing, although the effect is mathematically the same as if both types of smoothing were carried out at the same time. The smoothing procedure is controlled by three parameters: two angular scales, $\gamma_0$ and $\chi_0$, which control the degree of translational and orientational smoothing, respectively, as well as a smoothing temperature, $T_s$, with $\beta = 1/(k_B T_s)$.

The translationally smoothed interaction energy, $\tilde{U}(r_i, \theta_i, \phi_i, \phi_i', \theta_i', \psi_i')$, was computed from the unsmoothed interaction energy, $U(r_i, \theta_i, \phi_i, \phi_i', \theta_i', \psi_i')$ according to

$$\exp[-\beta\tilde{U}(r_i, \theta_i, \phi_i, \phi_i', \theta_i', \psi_i')] = \frac{\sum_{\theta_j,\phi_j} w_{\mathrm{trans}}(\gamma_{ij})\Delta V_j \exp[-\beta U(r_i, \theta_j, \phi_j, \phi_i', \theta_i', \psi_i')]}{\sum_{\theta_j,\phi_j} w_{\mathrm{trans}}(\gamma_{ij})\Delta V_j} \tag{3}$$

where the sum is taken over all cells in the table having the same radial coordinate $r$ and orientational coordinates $(\phi', \theta', \psi')$ as the cell to be smoothed but potentially different angular coordinates $(\theta, \phi)$. In this equation, $\Delta V_j = r_i^3(f-1)\sin\theta_j \sin\theta_i' \Delta\theta\Delta\phi\Delta\phi'\Delta\theta'\Delta\psi'$ is the configurational volume of cell $j$, $\gamma_{ij} = \cos\theta_i\cos\theta_j + \sin\theta_i\sin\theta_j\cos(\phi_i - \phi_j)$ is the angular spherical distance between spherical coordinates $(\theta_i, \phi_i)$ and $(\theta_j, \phi_j)$, and the translational smoothing kernel $w_{\mathrm{trans}}(\gamma_{ij})$ is given by[55]

$$w_{\mathrm{trans}}(\gamma_{ij}) = \sum_{l=0}^{l_{\max}} \frac{2l+1}{2} P_l(\cos\gamma_{ij}) \exp\left[-\frac{l(l+1)\gamma_0^2}{4}\right] \tag{4}$$

where $\gamma_0$ is the angular scale for translational smoothing, $P_l(x)$ is the $l$th order Legendre polynomial, and $l_{\max}$ was set at 1000.

Similarly, the fully smoothed interaction energy (with both translational and orientational smoothing) $\tilde{\tilde{U}}(r_i, \theta_i, \phi_i, \phi_i', \theta_i', \psi_i')$ was computed from the translationally smoothed interaction energy $\tilde{U}(r_i, \theta_i, \phi_i, \phi_i', \theta_i', \psi_i')$ according to

$$\exp[-\beta\tilde{\tilde{U}}(r_i, \theta_i, \phi_i, \phi_i', \theta_i', \psi_i')] = \frac{\sum_{\phi_j',\theta_j',\psi_j'} w_{\mathrm{orient}}(\chi_{ij})\Delta V_j \exp[-\beta\tilde{U}(r_i, \theta_i, \phi_i, \phi_j', \theta_j', \psi_j')]}{\sum_{\phi_j',\theta_j',\psi_j'} w_{\mathrm{orient}}(\chi_{ij})\Delta V_j} \tag{5}$$

Here the sum is taken over all cells in the table having the same translational coordinates $(r, \theta, \phi)$ as the cell to be smoothed, but potentially different orientational coordinates $(\phi', \theta', \psi')$. In this equation $\chi_{ij}$ is the overall angle of rotation needed to bring a fragment from the orientation given by $(\phi_i', \theta_i', \psi_i')$ to $(\phi_j', \theta_j', \psi_j')$ and is given by $\cos(\chi_{ij}/2) = \mathrm{Re}\,\mathbf{q}_j\mathbf{q}_i^{-1}$, where $\mathbf{q}_i$ and $\mathbf{q}_j$ are the corresponding quaternions.

The orientational kernel, $w_{\mathrm{orient}}(\chi_{ij})$, is likewise a solution to the diffusion equation in SO(3) (the group of rotations of three-dimensional space); a derivation is given in the Supporting Information. (To our knowledge, no solution to this problem has previously been published in the literature.) It is given by


dx.doi.org/10.1021/ct500622z | J. Chem. Theory Comput. 2014, 10, 5161−5177

$$w_{\text{orient}}(\chi_{ij}) = \sum_{j=0}^{j_{\max}} \frac{2j+1}{8\pi^2} \exp\left[-\frac{j(j+1)\chi_0^2}{4}\right] \frac{\sin\left(j+\frac{1}{2}\right)\chi_{ij}}{\sin\frac{1}{2}\chi_{ij}} \tag{6}$$

where $\chi_0$ is the angular scale for orientational smoothing and $j_{\max}$ was set at 1000. In order to speed up the smoothing process, for both translational and orientational smoothing, tables were constructed of $w_{\text{trans}}(\gamma)$ or $w_{\text{orient}}(\chi)$ as a function of $\cos\gamma$ or $\cos\chi$. These tables had $\sim 10^5$ entries. In addition, a cell $j$ was not counted in the average if its weight $w_{\text{trans}}(\gamma_{ij})$ or $w_{\text{orient}}(\chi_{ij})$ was less than $1 \times 10^{-6}$.

Because the smoothing kernels $w_{\text{trans}}$ and $w_{\text{orient}}$ are solutions to the diffusion equation in their respective spaces, this smoothing procedure can be interpreted as allowing the Boltzmann probability, $\exp(-\beta U)$, to diffuse along the angular and orientational coordinates. The angular scales $\gamma_0$ and $\chi_0$ control the extent of this diffusion and therefore the degree of translational or orientational smoothing. For $\gamma_0 = \chi_0 = 0$, no smoothing takes place and the interaction energy is equivalent to that given by the atomistic force field, with relatively small errors that come from the finite resolution of the table. In the limit as $\gamma_0 \to \infty$ and $\chi_0 \to \infty$, all variation in the interaction energy with the angular coordinates $(\theta, \phi, \phi', \theta', \psi')$ is eliminated, and the interaction potential becomes a spherically symmetric potential, dependent only on the interfragment distance $r$. Thus, by adjusting the values of $\gamma_0$ and $\chi_0$, and constructing tables accordingly, we can obtain a continuous range of protein force fields, ranging from a united-atom force field all the way to a MARTINI-like coarse-grained model with spherical "beads" centered on the center of mass of each fragment. Figure 2b shows this transformation, and a plot of the kernel functions $w_{\text{trans}}(\gamma)$ and $w_{\text{orient}}(\chi)$ for $\gamma_0 = 60°$ and $\chi_0 = 60°$ is shown in Figure 2a.

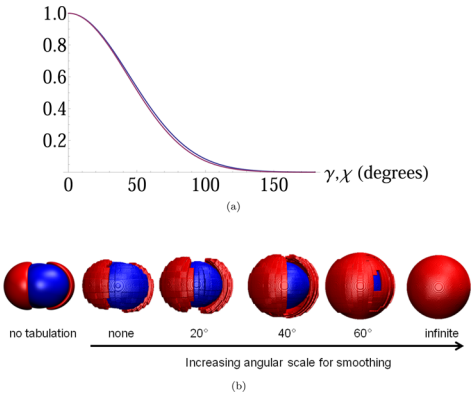



**Figure 2.** Smoothing kernels and effect of smoothing on interaction potentials. (a) Plots of the smoothing kernels $w_{\text{trans}}(\gamma_{ij})$ and $w_{\text{orient}}(\chi_{ij})$ given in eqs 4 and 6 for smoothing scale $\gamma_0 = \chi_0 = 60°$. (b) Contour plots of interaction energy between peptide-bond fragments of identical orientation as a function of angular scale of smoothing. Blue contour represents interaction energy of +1.0 kcal/mol; red contour represents interaction energy of −1.0 kcal/mol.

It should be noted that a smoothing angle of 60° constitutes a large amount of smoothing and largely removes the directional dependence of hydrogen bonding or nonspherical steric interactions. This can be seen in Figure 2a, which shows the effect of smoothing on the interaction energy of two peptide identically oriented peptide fragments. The two minima of the interaction energy between two peptide fragments visible in the exact potential energy surface, corresponding to two possible hydrogen-bonding configurations, have largely become one nearly spherically symmetric minimum in the interaction energy surface corresponding to a 60° smoothed table.

In addition, it can be demonstrated mathematically that the smoothing procedure leaves the generalized second virial coefficient for the interaction between two fragments unaltered at the smoothing temperature:

$$\begin{aligned} B_2 &= -\frac{1}{2}\int\{\exp(-\beta\tilde{\tilde{U}}(r,\theta,\phi,\phi',\theta',\psi')) - 1\}\,d\Omega \\ &\approx -\frac{1}{2}\sum_i V_i\{\exp(-\beta\tilde{\tilde{U}}(r,\theta,\phi,\phi',\theta',\psi')) - 1\} \\ &= -\frac{1}{2}\sum_i V_i\{\exp(-\beta U(r,\theta,\phi,\phi',\theta',\psi')) - 1\} \end{aligned} \tag{7}$$

This implies that the contribution to the partition function from pairwise interactions between fragments is left unchanged by the smoothing procedure outlined here and more generally ensures that the overall strength of the interaction between each pair of fragments is left approximately the same. A proof of this theorem is provided in the Supporting Information.

To summarize, coarse-grained models were created by smoothing the tabulated potential energy in the angular and orientational coordinates through Boltzmann averaging. It is expected that this smoothing will result in an overall free energy surface that is smoother, thus reducing the effective resolution of model and reproducing the ability of coarse-grained models to provide improved sampling relative to atomistic models. By adjusting the width of the Gaussian-like kernel for this smoothing, a range of coarse-grained models can be created, extending from a nearly atomistic model when no smoothing is used to a model with one to two spherical beads per side chain when infinite smoothing is used. The smoothing procedure ensures that the overall strength of interactions between fragments remains roughly the same by maintaining the second virial coefficient of the interaction.

**2.5. Backbone Tables and Smoothing.** Another source of barriers to conformational change comes from steric interactions along the protein backbone. The interaction energy tables described above cannot be used for the interactions between peptide fragments, because they do not include the corresponding bond, angle, or dihedral terms, nor do they take into account the special van der Waals or electrostatic interactions for atoms in 1−2, 1−3, 1−4 relationships. Consequently, a separate set of tables was constructed for pairs of covalently bonded peptide-like fragments, which included the necessary bond, angle, and dihedral terms, as well as its own smoothing technique. These tables included the necessary bond, angle, and dihedral terms, as well as the van der Waals and electrostatic interactions.

These "backbone tables" contained the interaction energy between two adjacent peptide fragments as a function of the two Ramachandran angles $\phi$ and $\psi$ and the N−$C^\alpha$−C bond angle (which will be designated $\alpha$ in the following equations). The tables had a resolution of 1° in $\phi$ and $\psi$, as well as a 1° resolution in $\alpha$, which ranged from 100° to 130°. The tables were used to calculate $U_{\text{backbone}}$ in eq 2, and the corresponding terms were omitted from the other terms in the equation.

These backbone tables could also be smoothed in the $\phi$ and $\psi$ directions. The smoothed backbone energy $\tilde{U}(\phi, \psi, \alpha)$ was


dx.doi.org/10.1021/ct500622z | J. Chem. Theory Comput. 2014, 10, 5161−5177

computed from the unsmoothed interaction energy $U(\phi, \psi, \alpha)$ according to

$$\exp[-\beta\tilde{U}(\phi_i, \psi_i, \alpha_i)] = \frac{\sum_{\phi_j,\psi_j} w_{bb}(\phi_i, \psi_i, \phi_j, \psi_j)\Delta\phi_j\Delta\psi_j \exp[-\beta U(\phi_j, \psi_j, \alpha_j)]}{\sum_{\phi_j,\psi_j} w_{bb}(\phi_i, \psi_i, \phi_j, \psi_j)\Delta\phi_j\Delta\psi_j} \quad (8)$$

where the smoothing kernel $w_{bb}(\phi_i, \psi_i, \phi_j, \psi_j)$ is given by

$$w_{bb}(\phi_i, \psi_i, \phi_j, \psi_j) = \exp\left[-\frac{(\phi_i - \phi_j)^2 + (\psi_i - \psi_j)^2}{\sigma_0^2}\right] \quad (9)$$

where $\sigma_0$ is the angular scale for backbone smoothing in Ramachandran space. No smoothing was carried out in the $\alpha$ direction.

In sum, in addition to the interaction tables described above, which include only van der Waals and electrostatic interactions, our model also includes backbone tables. These tables include both covalent and noncovalent interactions along the backbone. By smoothing these tables, one can lower the free energy barriers for backbone dihedral transitions, which is expected to enhance sampling beyond what can be achieved with the smoothing of the interaction tables alone.

**2.6. Simulation Protocol and Systems.** In order to test the new method, complete sets of interaction tables covering all 32 types of fragments included in our fragmentation scheme were constructed at smoothing levels in 5° increments from no smoothing to 30° and in 10° increments from 40° to 90°. In each case, the angular scale of smoothing along the angular coordinates was kept equal to that along the orientational coordinates so that the impact of smoothing on the simulations could be studied relative to a single "smoothing scale".

Using the tables, Monte Carlo simulations[56] were performed in the potential given by eq 2 on two test systems: an $\alpha$-helix having the sequence Ac-$L_{12}$-NMe (Leu12) and a $\beta$-hairpin consisting of residues 41−56 from the B1 domain of streptococcal protein G (the GB1 hairpin). Simulations of Leu12 were performed with smoothing levels in 10° increments, while simulations of the GB1 hairpin additionally used the tables smoothed at 5° increments. (Although simulations of the GB1 hairpin were performed up to 90° smoothing, results are only reported up to 30° smoothing because the simulations beyond this failed to converge, for reasons discussed below in section 3.2.) These simulations were then analyzed to determine the stability of the structure and the extent of sampling.

The initial coordinates for Leu12 in the CHARMM19 force field[48] were constructed using equilibrium bond lengths, bond angles, and dihedral angles using the internal coordinate facility in CHARMM,[57,58] including a $\phi$ angle of −79° and a $\psi$ angle of −39° for each amino acid. The initial coordinates for the GB1 hairpin were taken from residues 41−56 of the structure as determined by NMR spectroscopy (PDB code 2GB1),[47] with acetyl and *N*-methyl groups on the N- and C-termini, respectively (sequence Ac-GEWTYDDATKTFTVTE-NMe), and any missing coordinates were also filled in using the CHARMM internal coordinate facility. In either case, the structure was then minimized in CHARMM under the same conditions analogous to those that would be used for the subsequent Monte Carlo, except that a switching function between 12 and 14 Å for both electrostatic and van der Waals interactions was used in place of a fragment-based cutoff at 12 Å.

Each system was then divided into rigid fragments according to the fragmentation scheme shown in Figure 1. The starting centers, $\mathbf{r}_i$, and orientations, $\mathbf{q}_i$, of each rigid fragment were determined by minimizing the mass-weighted RMSD between the reference geometry of each fragment and the corresponding atoms in the minimized structure of the system according to the algorithm given by Coutsias et al.[59] All fragments were successfully fitted with a mass weighted RMSD of no more than 0.1 Å.

During the subsequent Monte Carlo simulations, trial moves were performed first on the fragment centers and orientations; then, after each trial move, the Cartesian coordinates of each united atom in the simulation were recalculated based on the new center and orientation of each fragment. Consequently, no subsequent RMSD fits needed to be performed during the simulation proper. For each fragment pair whose centers were within a 12 Å cutoff, the interaction energy was determined by calculating the relative displacement and orientation, converting this to spherical coordinates and Euler angles, and looking up the energy corresponding to the closest cell in the appropriate table. The resulting interaction energies were summed to determine the overall interaction energy term, $U_{interaction}$. The remaining terms in eq 2 were calculated from the Cartesian coordinates according to the force field. The moves were then accepted or rejected based on the Metropolis criterion.[56]

The trial moves used included backbone rotations, side chain rotations, and backrub moves;[60] detailed descriptions of each move and information on the distribution of move sizes are shown in Table S2 in the Supporting Information. The Cartesian coordinates of each atom were also written to the trajectory; consequently, the simulation produced structures at near atomistic resolution despite being in some respects a coarse-grained simulation. Each system was simulated at 300 K. Control Monte Carlo simulations without tabulation were also conducted, in which the van der Waals and electrostatic interactions between fragments were calculated directly from the atom positions during the simulation, rather than using the table. The lengths of all Monte Carlo simulations are shown in Table 2.

The generation of smoothed tables and subsequent use of these tables for Monte Carlo simulations were implemented in an in-house code written in C++. (Some code was adapted from the variable-resolution library-based Monte Carlo program previously developed in our laboratory.[61]) Computer times used for the determination of speedup factors were obtained by performing the above-described simulations using this code on a single processor. A speedup factor for energy calculations was obtained as the ratio of CPU time per trial move between a simulation with without tables and a simulation with tables with or without smoothing.

For reference, all-atom MD simulations of Leu12 and the GB1 hairpin were also performed, starting from the minimized structures, using the CHARMM22 force field with CMAP corrections.[62,63] These simulations also used a distance dependent dielectric and a switching function between 12 and 14 Å for both electrostatic and van der Waals interactions. They were run at 300 K using Langevin dynamics, a time step of 2 fs, and SHAKE[64] to maintain constant length for bonds involving hydrogen. The Leu12 simulation was run for 100 ns, and the GB1 hairpin simulation was run for 50 ns.

**Table 2. Monte Carlo Simulations Used for Testing of the New Coarse-Grained Model[a]**

| system | interaction table smoothing | backbone table smoothing | length ($10^9$ trial moves) |
|---|---|---|---|
| Leu12 | no tables | no tables | 5.0 |
| | 0–90° in 10° increments | no tables | 5.0 |
| | 0–90° in 10° increments | 10–30° in 5° increments | 4.0 |
| | replica exchange | no tables | 1.0 |
| GB1 hairpin | no tables | no tables | 5.0 |
| | 0–30° in 5° increments, 40–90° in 10° increments | no tables | 5.0 |
| | 0–90° in 10° increments | 10–30° in 5° increments | 3.0 |
| | replica exchange | no tables | 4.0 |
| Met-enkephalin | replica exchange | no tables | 5.0 |

[a]For replica exchange simulations, the lengths given are per replica, and exchange attempts were made every $10^4$ trial moves per replica. Schedules of replicas for the replica exchange simulations are shown in Table 3.

**2.7. Analysis.** To determine whether and how much the structural properties of Leu12 and the GB1 hairpin changed as a result of being simulated by means of this new method, the resulting trajectories were analyzed by constructing histograms of the backbone RMSD (omitting end residues) and by assigning the secondary structure of each residue using STRIDE,[65] from which the average fraction of $\alpha$-helical or $\beta$-sheet structure could be calculated.

In addition, sampling was assessed by the rate of Ramachandran transitions and by the rate of high-RMSD excursions. Each residue in each frame was assigned to one of the basins shown in Table S3 in the Supporting Information according to its Ramachandran $\phi$ and $\psi$ angles, and the rate of transitions between these basins per $10^6$ trial moves was computed. End residues were excluded from this analysis to prevent their flexibility from dominating the calculated rates. A sampling speedup factor was calculated as the ratio of the transition rate between basins for the simulation without tables and the corresponding rate for a simulation using tables with or without smoothing.

High-RMSD excursions were defined as excursions of the backbone RMSD above a threshold of 2.5 Å for Leu12 and 2.0 Å for the GB1 hairpin, followed by a return to lower values. Thresholds were determined by visual inspection of backbone RMSD distributions shown below.

**2.8. Hamiltonian/Resolution Exchange Simulations.** By adjusting the angular scale of smoothing, it is possible using this method to produce a continuous range of force fields ranging in resolution from a united-atom force field to a coarse-grained one. Since the configurational space in each force field is the same (the positions and orientations of the fragments), it is relatively easy to construct a Hamiltonian replica exchange simulation in which each replica uses a different level of smoothing, similar to previously proposed "resolution exchange" simulations in which different replicas are simulated at different coarse-grained resolutions.[43,44]

Leu12, the GB1 hairpin, and Met-enkephalin were all simulated in this way; a total of six replicas were used for Met-enkephalin, eight replicas were used for Leu12, and 10 replicas were used for the GB1 hairpin. (The structure of Met-enkephalin from PDB code 1PLW[66] was used without minimization.) In each case, one replica simulated the system without tabulation, calculating van der Waals and eletrostatic interactions directly from the atom positions. It was found necessary also to have an additional replica in which the potential energy was a 50/50 mixture of tabulated and directly calculated potential energy functions:

$$U_{50/50} = 0.5U_{\text{nontabulated}} + 0.5U_{\text{tabulated}} \quad (10)$$

The remaining replicas used tables with various levels of smoothing; schedules showing the amount of smoothing used for each replica are shown in Table 3. The temperature of each replica was 300 K. A trial exchange between one randomly chosen replica $i$ and its neighbor $j = i + 1$ was performed after every $10^4$ trial Monte Carlo moves of other kinds. This exchange was accepted or rejected with probability $p$ according

**Table 3. Schedules of Replicas and Average Exchange Probabilities for Hamiltonian Replica Exchange Simulations[a]**

| system | replica | Hamiltonian used | exchange probability (%) |
|---|---|---|---|
| Met-enkephalin | 1 | no tabulation (exact force field) | 19.6 |
| | 2 | tabulation with no smoothing | 39.3 |
| | 3 | tabulation with 10° smoothing | 53.1 |
| | 4 | tabulation with 15° smoothing | 59.1 |
| | 5 | tabulation with 20° smoothing | 17.5 |
| | 6 | tabulation with 40° smoothing | |
| Leu12 | 1 | no tabulation (exact force field) | 14.1 |
| | 2 | 50/50 mixture (eq 10) | 10.9 |
| | 3 | tabulation with no smoothing | 23.6 |
| | 4 | tabulation with 10° smoothing | 16.4 |
| | 5 | tabulation with 15° smoothing | 23.1 |
| | 6 | tabulation with 20° smoothing | 2.7 |
| | 7 | tabulation with 40° smoothing | 17.9 |
| | 8 | tabulation with 60° smoothing | |
| GB1 hairpin | 1 | no tabulation (exact force field) | 3.5 |
| | 2 | 50/50 mixture (eq 10) | 2.6 |
| | 3 | tabulation with no smoothing | 4.2 |
| | 4 | tabulation with 5° smoothing | 1.9 |
| | 5 | tabulation with 10° smoothing | 7.8 |
| | 6 | tabulation with 15° smoothing | 6.7 |
| | 7 | tabulation with 20° smoothing | 1.5 |
| | 8 | tabulation with 25° smoothing | |

[a]Average exchange probabilities are between the replica listed and the succeeding replica.

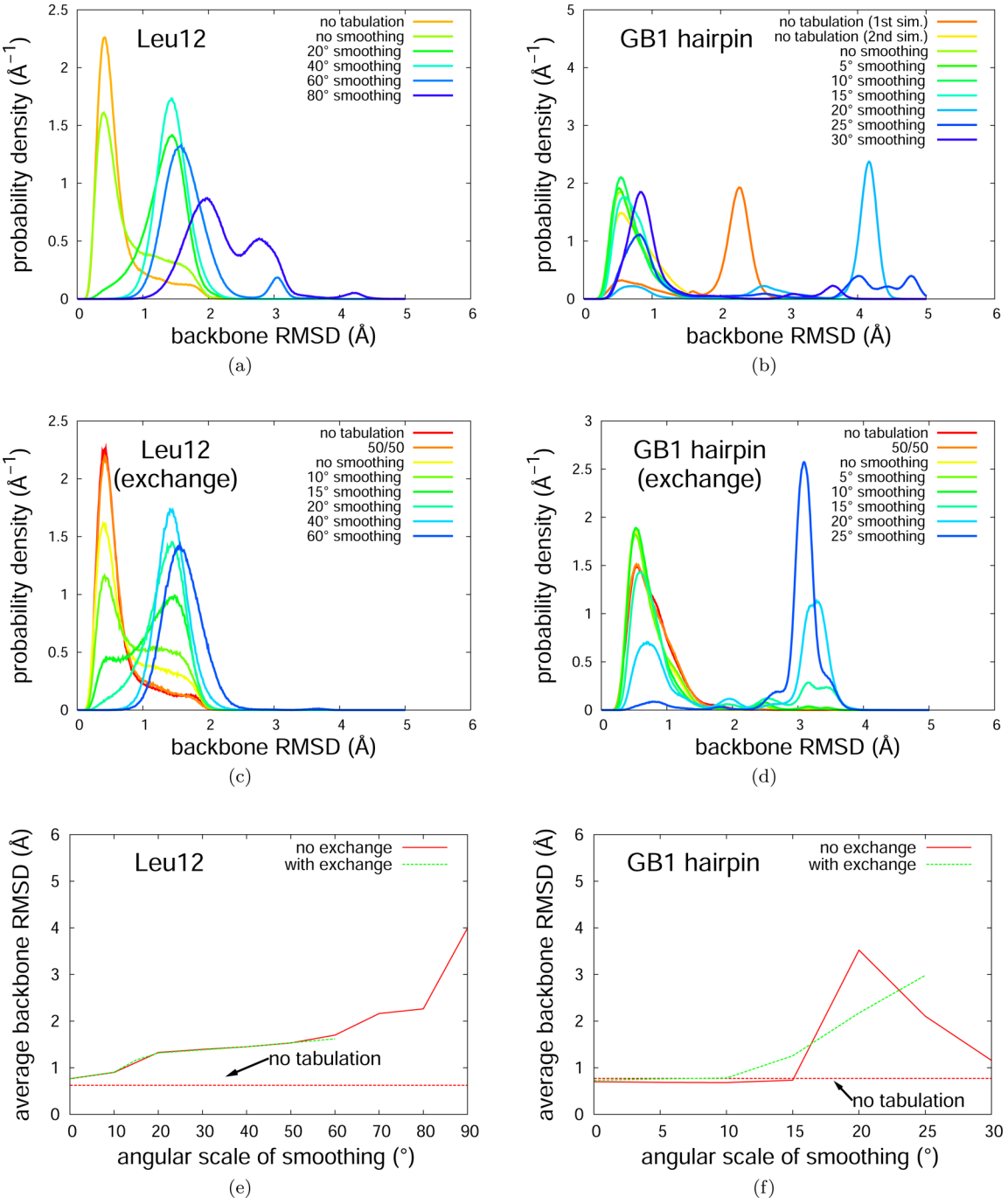



**Figure 3.** Backbone RMSD distributions in the various ensembles. (a, b) Backbone RMSD distribution in simulations without exchange of (a) Leu12 and (b) GB1 hairpin. (c, d) Backbone RMSD distribution in Hamiltonian replica exchange simulations of (c) Leu12 and (d) the GB1 hairpin. (e, f) Average backbone RMSD as a function of smoothing scale for simulations with and without exchange, for (e) Leu12 and (f) the GB1 hairpin.

to the detailed balance criterion for Hamiltonian replica exchange:

$$p = \min\left\{1, \frac{\exp[-\beta U_i(\mathbf{r}_j, \mathbf{q}_j)]\exp[-\beta U_j(\mathbf{r}_i, \mathbf{q}_i)]}{\exp[-\beta U_i(\mathbf{r}_i, \mathbf{q}_i)]\exp[-\beta U_j(\mathbf{r}_j, \mathbf{q}_j)]}\right\} \quad (11)$$

where in this equation $\beta = 1/(k_B T_i)$ with $T$ being the temperature of each replica, $\mathbf{r}_i$, $\mathbf{q}_i$, $\mathbf{r}_j$, and $\mathbf{q}_j$ are the fragment centers and orientations for the configuration from replicas $i$ and $j$, respectively, and $U_i$ and $U_j$ represent the energy functions for replicas $i$ and $j$. The average exchange probabilities were calculated for each replica. The total length of each simulation is shown in Table 2.

## 3. RESULTS

Using the coarse-grained model described here, we performed test Monte Carlo simulations of the $\alpha$-helical peptide Leu12 and the GB1 hairpin, which has primarily $\beta$-sheet structure. We have carried out a variety of analyses in order to characterize the ability of the new method to preserve the structure of the systems and speed up simulations.

**3.1. Effect of Finite-Resolution Tables on Structure.** Just as a careful choice of lattice was needed to obtain


dx.doi.org/10.1021/ct500622z | J. Chem. Theory Comput. 2014, 10, 5161−5177

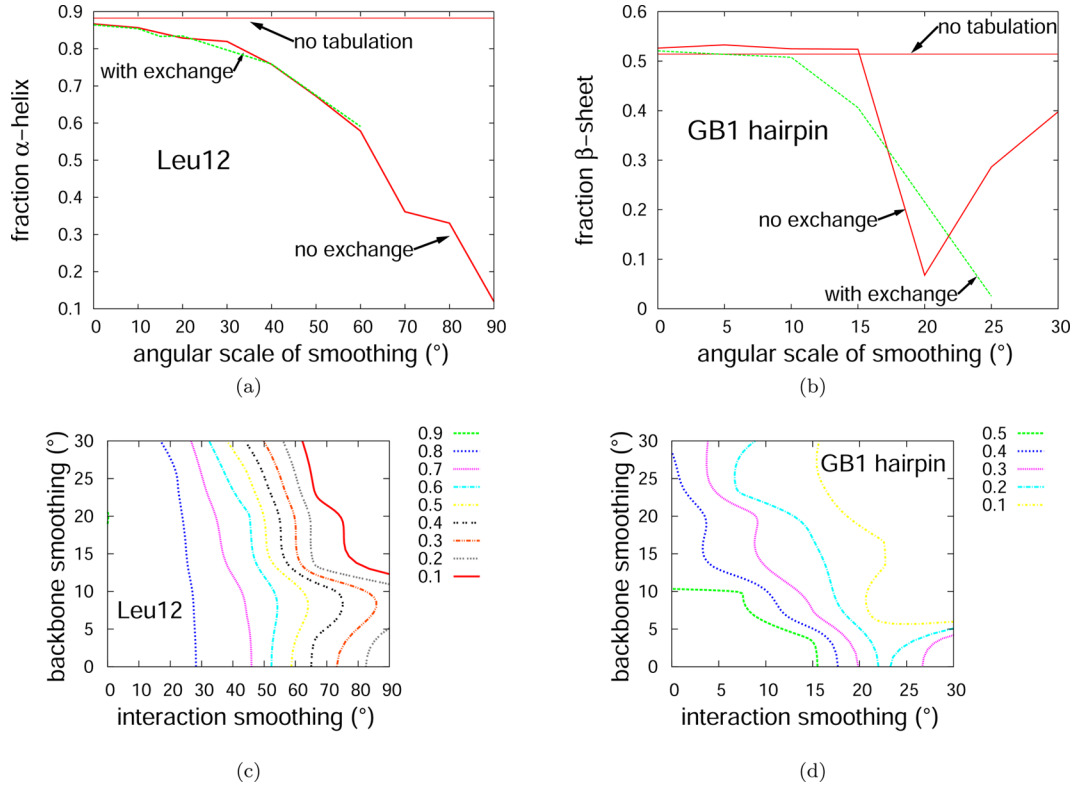



**Figure 4.** Secondary structure in tabulated and smoothed simulations. (a) Average fraction of $\alpha$-helical structure in Leu12 as determined either by regions of the Ramachandran plot or using STRIDE.[65] Thin horizontal lines indicate result for simulation without tables. (b) Average fraction of $\beta$-sheet structure in GB1 hairpin as determined by STRIDE. (c, d) Contour plot of average fraction of (c) $\alpha$-helical structure in Leu12 and (d) $\beta$-sheet structure in the GB1 hairpin, both as a function of interaction smoothing and covalent smoothing.

secondary structures in lattice-based coarse-grained models,[17] it is possible for the finite resolution of our tables to destabilize secondary structure in our models. To investigate this, we compared the average $\alpha$-helical or $\beta$-sheet fraction and the distribution of backbone RMSD in simulations with and without tables. The results are shown in Figures 3 and 4. For Leu12, both the fraction of $\alpha$-helical structure and the distribution of backbone RMSDs did not change when tables were used.

For the GB1 hairpin, the distribution of backbone RMSDs in the first simulation *without* tables showed a peak at about 2 Å that was not present when tables were used (Figure 3). This was due to a conformational transition that took place in the simulation without tables (Figure S2a in the Supporting Information). A repetition of this simulation under identical conditions but with different random number seeds did not reproduce this conformational transition (Figure 3 and Figure S2b in the Supporting Information). In addition, none of the other simulations of the GB1 hairpin reported here produced any conformations closer than approximately 1.2 Å in backbone RMSD to the conformation of the hairpin reached by this transition. Therefore, it appears that this conformational transition was an extremely rare event on the time scale of our simulations, although it may be important to the ensemble as a whole. The second simulation without tables had a backbone RMSD distribution and fraction of $\beta$-sheet structure that corresponded much more closely to the simulation using tables.

Thus, it appears that the limited resolution of the tables does not appear by itself to hinder the representation of the secondary structure of proteins. This is true despite the highly directional nature of the backbone hydrogen bonds within protein secondary structures.

**3.2. Effect of Smoothing on Structure.** The histograms of backbone RMSD and the secondary structure analyses also make it possible to assess the effect of smoothing on the conformations sampled by each simulation. In both systems, the quality of the structure decreases with increasing angular scale of smoothing, with increasing average RMSD deviation from the starting structure and decreasing fraction of $\alpha$-helical or $\beta$-sheet structure. For Leu12, the decrease in quality is small for smoothing scales up to about 40−50° (backbone RMSD deviations of 1−2 Å), but much greater at higher levels of smoothing. The use of backbone smoothing likewise decreases the secondary structure, and the effects of backbone and interaction smoothing appear to be roughly additive (Figure 4c). At higher levels of smoothing, unfolded conformations appear with significant deviations from the starting structure, although even heavily smoothed simulations return to structures near the original starting structure (Figures S1e,f in the Supporting Information).

The GB1 hairpin is more sensitive to the smoothing procedure. The fraction of $\beta$-sheet structure in the GB1 hairpin declines more rapidly with increasing smoothing scale than does the fraction of $\alpha$-helical structure in Leu12, and the data for the GB1 hairpin is also much noisier in standard simulations (Figure 3). Some simulations of the GB1 hairpin moved away from the original structure over the course of the simulation, toward more compact structures in which the $\beta$-sheet structure was lost. This can also be seen in the distribution of backbone RMSD and its evolution over the course of the simulations (Figure 3b and Figure S2 in the Supporting Information). This


dx.doi.org/10.1021/ct500622z | J. Chem. Theory Comput. 2014, 10, 5161−5177

Table 4. Relative Computational and Sampling Speedup of Simulations As a Function of the Use of Tables and Degree of Smoothing[a]

| | smoothing scale (deg) | | relative to MC without tables | | | relative to all-atom MD | | |
|---|---|---|---|---|---|---|---|---|
| system | interaction | backbone | energy | sampling | overall | energy | sampling | overall |
| Leu12 | no tables | none | 1.00 | 1.00 | 1.00 | 33.38 | 0.84 | 27.96 |
| | 0 | none | 0.77 | 0.64 | 0.49 | 25.75 | 0.54 | 13.83 |
| | 10 | none | 0.77 | 0.49 | 0.38 | 25.64 | 0.41 | 10.62 |
| | 20 | none | 0.78 | 0.65 | 0.50 | 25.93 | 0.54 | 14.05 |
| | 30 | none | 0.75 | 0.89 | 0.67 | 25.20 | 0.75 | 18.84 |
| | 40 | none | 0.77 | 1.74 | 1.34 | 25.80 | 1.46 | 37.56 |
| | 50 | none | 0.80 | 3.61 | 2.90 | 26.84 | 3.03 | 81.23 |
| | 60 | none | 0.77 | 8.36 | 6.46 | 25.80 | 7.00 | 180.66 |
| | 70 | none | 0.73 | 7.81 | 5.67 | 24.23 | 6.54 | 158.50 |
| | 80 | none | 0.74 | 14.06 | 10.44 | 24.80 | 11.78 | 292.07 |
| | 90 | none | 0.74 | 16.90 | 12.50 | 24.70 | 14.16 | 349.69 |
| | 10 | 10 | 1.00 | 0.62 | 0.62 | 33.35 | 0.52 | 17.26 |
| | 20 | 15 | 0.95 | 0.88 | 0.84 | 31.87 | 0.74 | 23.53 |
| | 30 | 20 | 0.93 | 1.73 | 1.61 | 30.97 | 1.45 | 44.97 |
| | 40 | 25 | 0.89 | 6.90 | 6.14 | 29.74 | 5.78 | 171.87 |
| | 50 | 30 | 0.84 | 29.10 | 24.38 | 27.97 | 24.38 | 681.87 |
| GB1 hairpin | no tables | none | 1.00 | 1.00 | 1.00 | 9.60 | 0.75 | 7.22 |
| | 0 | none | 1.10 | 0.46 | 0.51 | 10.58 | 0.35 | 3.70 |
| | 5 | none | 1.17 | 0.53 | 0.62 | 11.23 | 0.40 | 4.48 |
| | 10 | none | 1.10 | 0.82 | 0.90 | 10.60 | 0.62 | 6.53 |
| | 15 | none | 1.18 | 0.69 | 0.81 | 11.34 | 0.52 | 5.84 |
| | 20 | none | 0.97 | 0.71 | 0.68 | 9.27 | 0.53 | 4.93 |
| | 25 | none | 1.09 | 1.09 | 1.20 | 10.51 | 0.82 | 8.64 |
| | 30 | none | 1.08 | 0.86 | 0.94 | 10.41 | 0.65 | 6.77 |
| | 10 | 10 | 1.19 | 0.92 | 1.10 | 11.46 | 0.69 | 7.96 |
| | 20 | 15 | 1.04 | 1.90 | 1.97 | 9.95 | 1.43 | 14.22 |
| | 30 | 20 | 0.96 | 5.75 | 5.51 | 9.21 | 4.32 | 39.80 |
| | 40 | 25 | 0.94 | 3.42 | 3.21 | 9.02 | 2.57 | 23.22 |
| | 50 | 30 | 0.94 | 8.39 | 7.90 | 9.05 | 6.31 | 57.09 |

[a]Energy calculation speedup figures are based on total CPU time per MC or MD step. Sampling speedup figures compare the rate of backbone dihedral transitions in Ramachandran space as defined in Table S3, Supporting Information.

sensitivity is likely due to approximations made in this first implementation of tabulation and smoothing for peptides. Also, Figures 3 and 4 show that for the GB1 hairpin, Hamiltonian exchange simulations exhibit more physically intuitive, and perhaps better sampled, behavior. These data are discussed further below.

To further diagnose the structural changes in the GB1 hairpin, a temperature replica exchange Monte Carlo simulation[67] of the hairpin was performed without tables using eight replicas at temperatures of 300, 356, 423, 503, 597, 709, 842, and 1000 K, following a similar protocol to the resolution exchange simulations discussed above. Histograms of the backbone RMSD for the trajectory from each replica were prepared and are shown in Figure S3a in the Supporting Information. All replicas, including ones at low temperatures, showed conformations substantially different from the starting configuration. The replica at 300 K showed distorted, compact conformations with similar features to those seen in the smoothed, tabulated simulations of the hairpin. For example, in both simulations, glutamate and aspartate side chains moved toward the center of the hairpin in order to form additional hydrogen bonding or salt bridge interactions (Figure S3b,c in the Supporting Information). This strongly suggests that the distorted conformations represent a free energy well even when tables are not used and the potential consists of the force field and simple distance-dependent dielectric. Thus, they are not a direct result of the effect of smoothing on the free energy surface; rather, the smoothing serves to reduce the time scale needed to reach this free energy well in the simulations. In the Discussion, we propose possible improved solvation methods that could be combined with the tables in order to overcome this problem. We also examine other aspects of smoothing that contribute to the distortion of the hairpin and discuss possible ways of correcting for them.

### 3.3. Speedup in Energy Computation and Sampling.

Another important test for this new method is whether it offers enough increase in speed to allow simulations of larger systems for longer times than would be possible using more conventional techniques. This speedup could come from two possible sources: replacement of multiple energy-term calculations with a single look-up and smoothing of the energy landscape. The increase in sampling was quantified in two ways: using the rate of backbone transitions in Ramachandran space and the rate of "excursions" to high backbone RMSD conformations. Table 4 shows how sources of speedup contribute to the overall increase in sampling per unit computation time, in comparison with both Monte Carlo simulation without tables and all-atom MD simulation as described previously in section 2.6.

It turns out that no significant computational savings were achieved in the energy calculation, relative to Monte Carlo simulation of the underlying *united*-atom force field without


dx.doi.org/10.1021/ct500622z | J. Chem. Theory Comput. 2014, 10, 5161−5177

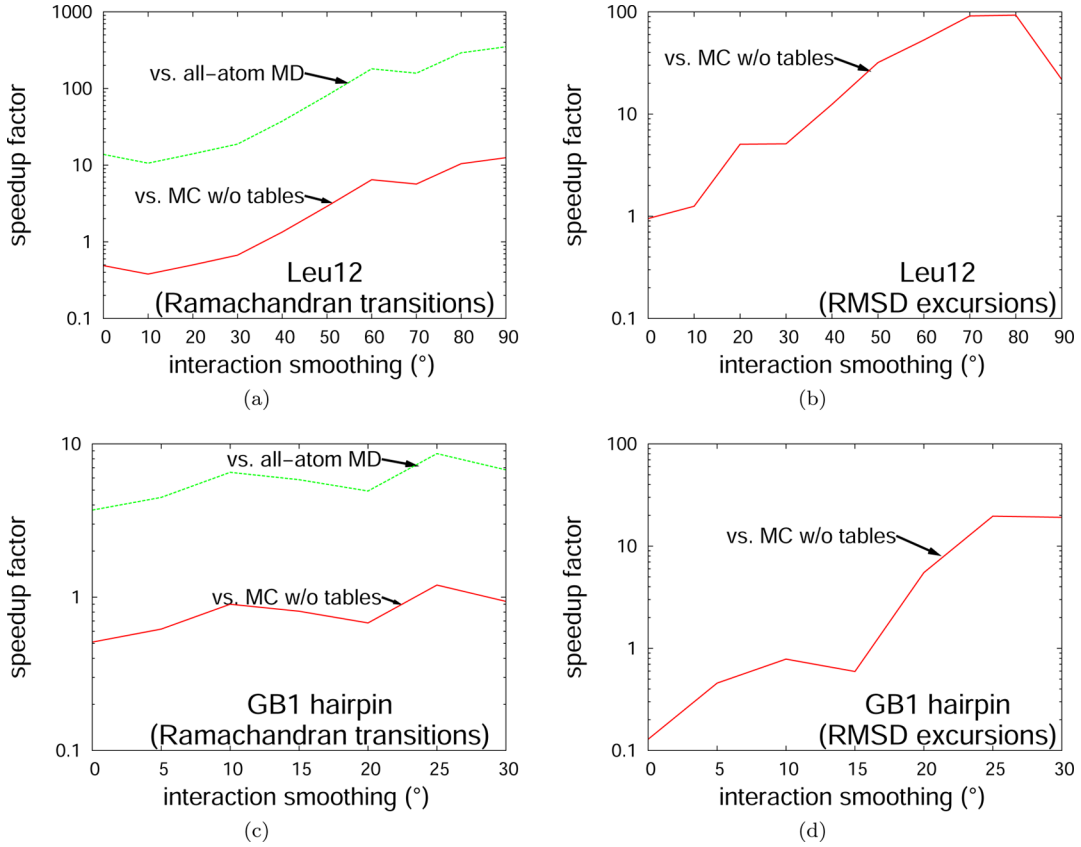



**Figure 5.** Overall speedup in sampling per unit CPU time as a function of the degree of interaction smoothing (a, b) for Leu12 or (c, d) for the GB1 hairpin relative to MC simulations without tabulation (red) or MD simulations (green). Sampling assessed based on rate of transitions between (a, c) Ramachandran plot regions or (b, d) excursions to high RMSD (>2.5 Å for Leu12, > 2.0 Å for GB1 hairpin).

tables. One reason for this is the small size of the fragments (an average of approximately four united atoms), which means that only relatively few atomistic interactions can be replaced by a table lookup. A detailed profiling analysis was conducted as described in the Supporting Information to determine where CPU time is being spent during each table lookup and in the simulation as a whole. The profiling data, shown in Table S4 in the Supporting Information, demonstrates that in a tabulated simulation the CPU time is distributed among several tasks. About 20−30% of the time is spent on calculating the nontabulated terms, and the remaining 60−70% of the CPU time is spent on table lookups and their supporting calculations. The most time-consuming tasks within the table lookup process are the computation of spherical coordinates and Euler angles, which together take about 30−35% of the total time in a simulation. By contrast, although in modern computers the main memory is much slower than the CPU (and in the case of table lookups the cache memory may not help much), the actual lookup from a table appears to be taking less than 5% of the total CPU time.

A much greater speedup in energy calculations is obtained relative to all-atom MD simulations, although this is the result of several factors not related to the use of tables. These include the increased number of particles in an all-atom model, compared with the united-atom model used here as a basis for the new method, and some computational efficiencies related to the use of Monte Carlo, such as not calculating the forces on atoms.

Although the use of tables did not reduce the amount of time needed to calculate the energy, the smoothing procedure for the tables does result in a smoother free energy surface, as indicated by increased conformational sampling in the simulations. We quantify this increased sampling by comparing the rates of Ramachandran transitions and RMSD excursions in smoothed versus unsmoothed simulations. For Leu12, both of these rates increase with the angular scale of smoothing, and moderate levels of interaction smoothing (about 30−60°) are able to counteract this effect (Figure 5). Depending on the amount of smoothing applied, up to a 1 order of magnitude increase in the rate of Ramachandran transitions and up to a 2 orders of magnitude increase in high-RMSD excursions could be obtained for both systems, relative to Monte Carlo simulations without tabulation. With more moderate levels of interaction smoothing (about 40−50°), which do not distort the structure of the system as significantly, an approximately 2.5-fold increase in the rate of Ramachandran transitions and a 10-fold increase in high-RMSD excursions may be possible. For the GB1 hairpin, the rate of Ramachandran transitions does not increase with increasing smoothing, but the rate of RMSD excursions does. A 5−10-fold increase in the rate of RMSD excursions is possible with moderate amounts of smoothing (about 20−30°), although even these appear to introduce significant deviations in the free energy surface as described above. The use of tables alone, without smoothing, decreases both of these sampling rates compared with simulations without tabulation. This appears to be due to the finite resolution of the tables, which causes discontinuous jumps in energy that increase the roughness of the energy surface.

In a Monte Carlo simulation, the sampling rate also depends on the types and sizes of trial moves used. All Monte Carlo


dx.doi.org/10.1021/ct500622z | J. Chem. Theory Comput. 2014, 10, 5161−5177

simulations used the same types of trial moves, which are described in detail in Table S2 in the Supporting Information and which allowed sampling of both backbone and side chain conformations. The acceptance rate of the moves was lower in tabulated simulations without smoothing than in simulations without tables, but increased with increasing smoothing. A range of possible move sizes were tested to determine the size that results in optimum sampling. To keep the number of possible combinations from becoming overwhelming, in this first study the maximum sizes of backbone dihedral rotations and backrub moves were restricted to be the same, and the maximum size of side chain dihedral rotations and fractions of the various move types were not varied. For both Leu12 and the GB1 hairpin the dependence of the transition rate on step size turned out to be relatively weak (data not shown). Since for the GB1 hairpin a maximum move size of about 20° appeared to give slightly better sampling than other move sizes, trajectories with this move size were selected for all of the previously described studies on the relationship between smoothing scale, structure preservation, and sampling. Further optimization of the types, sizes, and mixture of Monte Carlo moves to be used with this method will be the subject of future work.

**3.4. Hamiltonian/Resolution Exchange Simulations.** The first attempts at "resolution exchange" for a nontrivial peptide involved trying to make exchanges between all-atom and united-atom representations of Met-enkephalin.[44] Direct exchange between the two models resulted in a very poor exchange rate due to the lack of overlap between the ensembles sampled; consequently models of intermediate resolution were needed in order to bridge the gap. The models of ref 44 were constructed using "incremental coarsening" in which successive replicas converted one amino acid at a time from the all-atom to the united-atom representation. In this way, exchange rates of 2−18% were obtained over a total of six replicas. This approach is not uniform over the whole system, however, and since every bead must map onto a whole number of atoms, there are a limited number of discrete choices for this mapping.

The method described here allows for another approach to constructing models at intermediate resolution, since the angular scale of interaction smoothing can be adjusted to give models ranging from a united atom model all the way to a much more coarse-grained model. Consequently, an arbitrary number of replicas with intermediate levels of smoothing can be inserted, until the exchange rate between adjacent replicas is adequate to obtain good sampling.

This method was tried with Met-enkephalin, Leu12, and the GB1 hairpin (see section 2.8). For all three systems, replica 1 was made to simulate the protein without tables. For Leu12 and the GB1 hairpin, it was found necessary to insert an additional replica that used a 50/50 mixture of potentials with and without tables (eq 10) to obtain adequate exchange rates to the fully tabulated replica. The remaining replicas only used increasing levels of smoothing. The resulting exchange rates are shown in Table 3. We obtain exchange rates between 18% and 59% for Met-enkephalin, higher than those obtained previously.[44] In addition, we obtain exchange rates for Leu12 and the GB1 hairpin similar to those obtained previously for Met-enkephalin, even though both Leu12 and the GB1 hairpin are larger. Note that the maximum smoothing level for the GB1 hairpin was smaller than that for Leu12 because further smoothing was found to significantly decrease the overlap in configurational space (see Figure S2 in the Supporting Information). That said, increased exchange rates between replicas do not necessarily imply better sampling, and in the case of the GB1 hairpin, the unphysical structural distortions in heavily smoothed simulations may not have been helpful for enhancing sampling in the replicas using less or no smoothing.

The trajectories obtained from the replica exchange simulations were also analyzed for secondary structure using STRIDE,[65] and the fractions of secondary structure are shown in Figure 4, alongside the corresponding fractions from the trajectories not using exchange. The fractions of $\alpha$-helical structure were very similar for Leu12, whereas the replica exchange simulations from the GB1 hairpin did not experience as much deviation from their starting structures as did the simulations using smoothing without exchange. This may have been simply because the replica exchange simulations were not long enough for the hairpin to drift away from its original conformation as it did in the simulations without exchange.

## 4. DISCUSSION

The present report describes a generalization of an energy tabulation and smoothing scheme to peptides. Using all semirigid fragments found in proteins, we redesigned and extended the orientation-dependent tables from our preliminary studies on benzene[37] to accommodate polypeptides. This new application yielded promising results, but certainly there is room for improvement. We tested the new method on $\alpha$-helical and $\beta$-hairpin peptides. We found that for the $\alpha$-helix, there are intermediate levels of smoothing that allow a significant increase in sampling with only a small destabilization of the secondary structure, while for the $\beta$-hairpin it is more difficult to improve sampling without destabilizing the structure. We stress that the present implementation represents a first effort for proteins with the new strategy, and a number of significant improvements should be possible, as discussed below.

**4.1. Future Applications.** The strengths of the new coarse-grained model suggest a number of potential uses for addressing biophysical problems where existing coarse-grained models are insufficient. One possibility is to use orientation-dependent tables in a setting in which it makes sense to assume much larger rigid "fragments" than the ones used here. For example, tables of the type described here could be used to represent protein−protein aggregation, with a fragment representing a whole protein or domain. Such a model could then be used to study macromolecular crowding effects[68,69] or the assembly of subunits to form icosahedral viral capsids.[70,71] This would have the advantage of allowing a single table lookup to replace the calculation of a much larger number of interatomic interactions, thus speeding up energy calculations by a much greater factor than was obtained here.

Some multiscale procedures beyond exchange simulation could be facilitated by the present approach. Because the simulation keeps track of both the position and orientation of the rigid fragments, it is relatively straightforward to map back and forth from coarse-grained to atomistic coordinates. Furthermore, since the coarse-grained model is derived from an atomistic force field, it should incorporate similar physical assumptions, and thus it should be possible to combine coarse-grained and atomistic approaches without the complexities involved in other attempts.[72−74] Conseqeuntly, a mixed-resolution model could be constructed simply by evaluating the interaction energy for fragment pairs exactly when they are both inside an atomistic region, and using a smoothed table when one of the fragments is outside this region. Such a model

could be useful for drug docking; we previously obtained promising results with a mixed-resolution model based on combining an atomistic force field with a much simpler Gō potential.[61]

Another useful characteristic of our model is its tunable nature. By adjusting the amount of smoothing applied to the tables, it is possible to construct models of different effective resolution automatically. This makes the model potentially suitable for situations in which it is not obvious at the outset what resolution is most appropriate for capturing the essential interactions in a system without expending excessive computational effort. Furthermore, as demonstrated by the resolution exchange simulations presented here, the use of a tunable model represents a major step toward making the resolution exchange idea more practical. As pointed out earlier, although an improvement in exchange rates was observed compared with previous attempts at resolution exchange, achieving an actual enhancement in sampling may require more work. We expect that the improvements outlined below will make the smoothed ensembles more similar to the unsmoothed ones and thereby improve the effectiveness of resolution exchange as a sampling method beyond what has been demonstrated here.

Since the tables used by our method incorporate orientation-dependent interactions such as hydrogen bonds with sufficient resolution to resolve protein secondary structures, the model has the inherent ability to simulate secondary structure transitions without incorporating additional biases toward particular secondary structures. In addition, the approach does not incorporate information described by experimental structures or simulations, so it should be possible to use it in situations where these might introduce inappropriate biases. Although the coarse-grained model here is preliminary and requires improvement, it shows significant potential for future applications.

**4.2. Measurement of Sampling.** The effective speedup demonstrated here depends entirely on a comparison of the conformational sampling obtained in simulations with and without smoothing. However, the assessment of conformational sampling in biomolecular simulations is a complex problem and cannot easily be reduced to a single measurement or simple algorithm.[45,75−80] Here the rates of Ramachandran angle transitions and of RMSD excursions are used as measurements of the overall rates of sampling in our simulations. The rate of Ramachandran transitions measures the sampling of relatively small, local conformational fluctuations, while the rate of high RMSD excursions shows the effect on the sampling of more global conformational fluctuations. These processes were chosen because enough events of each type could be observed in each simulation to permit a reasonably accurate estimate of the rate. Many biomolecular simulation studies focus on larger conformational changes than observed here (such as protein folding), and thus the ability of our models to increase the rate of these larger conformational changes would be of more relevance. However, except for the structural distortion of the GB1 hairpin (which was apparently irreversible), we were not able to observe these kinds of conformational changes in our simulations. Nevertheless, the increase in rate of small-scale conformational fluctuations indicates that the overall free energy surface is smoother, which in turn suggests that the rate of larger-scale conformational changes should also be enhanced.

Another problem in comparing sampling rates across simulations is that, because of the smoothing, the simulations are not all sampling the same free energy surface. If there is a significant difference between free energy surfaces, transitions sampled on the smoothed surface may not be representative of corresponding transitions on the unsmoothed surface. Consequently, an increased rate of transitions may not really represent an improvement in sampling that would be relevant to the unsmoothed system, and a comparison of these rates may therefore not be meaningful. This is potentially a problem for all coarse-grained models; because of the physical approximations they make, the increased sampling they provide may not be relevant for the all-atom system. For Leu12, it appears that the deviations in the free energy surface introduced by our smoothing techniques are modest, since even heavily smoothed simulations still primarily sample the starting $\alpha$-helical basin. This suggests that it is still meaningful to use rates of Ramachandran transitions or RMSD excursions to compare sampling across Leu12 simulations with various levels of smoothing, since the nature of the conformational fluctuations are similar across the different simulations. For the GB1 hairpin, the situation is less clear. The effective speedup achieved by our measures is modest (a factor of 10 at most). The smoothed simulations sample different conformations from those without smoothing, so even this may not be a valid indication of actual increase in sampling. However, the ability of the smoothed simulations to reach a conformation (albeit a distorted one) that was not reached except through a temperature replica exchange simulation suggests that the smoothing may be effective in improving sampling in ways that are not captured by the measures adopted here.

**4.3. Improving the Treatment of Solvation.** The fact that distorted conformations of the GB1 hairpin occurred in the temperature replica exchange simulation strongly suggests that the simple distance-dependent dielectric solvation is a major reason for the unphysical conformations of the hairpin at high smoothing and that it should therefore be improved.

The manner in which solvation effects are incorporated into other coarse-grained models varies from one model to another. Many other coarse-grained models (particularly knowledge-based models) do not have a specific term in the potential energy devoted to representing solvation effects. Instead, solvation effects are reflected in the experimental structures used to create the model.[20−26] While it might be possible in principle to tabulate such a potential in order to incorporate solvation effects, this would defeat one of the purposes of the research presented here, which was to create a coarse-grained model that is free from biases introduced by incorporating experimental structural information.

Here, the initial choice of a distance dependent dielectric was made because more sophisticated solvation models such as generalized Born (GB) models[54,81] are not pairwise. In these models, the contribution of pairs of atoms to the solvation free energy is expressed in terms of their charges, the distance between them, and the Born radii. These Born radii depend on the positions of atoms in all the other fragments besides the two being considered and are therefore not known during the initial construction of a table of the type described here. Consequently, it will be necessary to redesign the interaction energy tables in order to incorporate a GB solvation energy.

Several strategies for implementing GB-type solvation may be possible within the tabulation framework by building on previously proposed approximate schemes. The calculation would require two stages: estimation of Born radii and then calculation of the overall energy. First, following ideas used in

the PRIMO coarse-grained model,[36] for example, fragment-averaged radii could be estimated for each fragment. Alternatively, estimates of the radii could be obtained using a new (second) set of tables to calculate configuration-dependent components of the radii based on fragment pair relative positioning. Once Born radii are obtained, the overall energy of a given configuration could be obtained from a modified set of energy tables based on both rigid-body coordinates (as in our current approach) and additionally on a parameter representing the product of fragment-averaged radii; the latter parameter is needed for the Still formula[81] and would be obtained from the initial estimates of the Born radii. This strategy is similar to an approach used for computational protein design.[82,83]

Thus, although the tables considered in the present work are only applicable to pairwise potentials, it may be possible to apply orientation-dependent tables of a more complex design to more accurate solvation potentials, which would be expected to significantly improve the conformational dynamics of the GB1 hairpin. The construction and testing of tables that can accommodate more accurate solvation models will be the focus of future work. It is noteworthy that, although improved solvation will necessarily increase the computational cost for our approach, it is likely the tabulation platform can accommodate an approximate GB scheme that is cheaper than existing implementations. If so, the net result would be that the speedup of the tabulation approach would improve relative to traditional calculations.

**4.4. Other Possible Improvements.** Given that this is an initial implementation of a nontraditional approach, there are several avenues for improvement beyond incorporating a better solvation model. The smoothing procedure employed here does not preserve the shape of fragments or the packing of amino acid side chains. In the limit of infinite smoothing, the fragments become spherical particles that are generally smaller than the original fragments in all directions as a result of the Boltzmann averaging. This proved to be a particular problem for the indole ring in the side chain of tryptophan. In early versions of our fragmentation scheme, this ring was a single fragment: smoothing caused this ring to become a small particle whose size was approximately equal to the van der Waals thickness of the ring. Consequently, other parts of the peptide could overlap with the indole ring, causing distorted conformations. This problem was overcome by splitting the indole ring into two smaller fragments. Similar distortions likely affected other nonspherical fragments, such as those containing aromatic rings. Since fragments of these types were present in the GB1 hairpin, but not in Leu12, this effect would have impacted the GB1 hairpin simulations more than those of Leu12. A more general way of overcoming this problem (without making the fragments smaller) is to divide each fragment into smaller "subfragments" comprising fewer atoms, calculating the pairwise interaction energy between each pair of subfragments, applying eqs 3 and 5 to these subfragment interactions, then adding the results to obtain the full interaction energy in a single table. In this way, it may be possible to separately smooth the interactions between subfragments, possibly allowing for better shape preservation without reducing the size of the fragments.

There are several other possible improvements that could be made to the model described here. Another weakness of the smoothing formalism used here is the assumption that each pair of fragments in the system to be simulated can assume any relative position and orientation, as if the fragments were isolated. However, many of these relative positions may not be accessible when the two fragments are embedded in a larger molecule. It should be possible to modify the smoothing process to better incorporate this information, which could be obtained from short simulations of model compounds.

The finite resolution of the table appears to cause a slowdown in sampling when tables with only small amounts of smoothing are used. Consequently a minimum of about 30−40° of smoothing (in the present scheme) is required to overcome this slowdown and obtain a net gain. Because the tables are six-dimensional, it is diffciult to improve their resolution without a substantial increase in the memory requirements. Interpolation techniques may help here, although they will necessitate additional computation. In addition, only a limited effort was made here to optimize the Monte Carlo moves; further optimization may also yield improvements in sampling.

Other potential improvements focus on trying to obtain a greater speedup than was obtained here, particularly in the energy calculation, or on reducing the size of the tables in memory. At best, the use of tables to calculate the interfragment energy took only marginally less time than calculating it directly, and in some cases the use of tables took longer than an equivalent direct energy calculation (Table 4). This was not the case for the tables used in the preliminary studies on benzene[37] because those tables were constructed in terms of absolute orientations, which were restricted to a predefined library. It should be possible to adopt a similar approach here. A library of orientations would be constructed, and for each fragment, the index of the closest orientation in the library would be found. A separate "division table" would then be used to obtain the index corresponding to a relative orientation from the indices corresponding to two absolute orientations. This would obviate the need to repeatedly convert relative orientations to Euler angles (which is responsible for approximately 20% of the total cost of a simulation). Some discretization error would be introduced but presumably this would be modest for sufficiently smoothed models of primary interest.

It may also be possible to reduce the size of the tables in memory by approximately 40% while maintaining their effective resolution by indexing the tables according to $\cos\theta$ or $\cos\theta'$, rather than $\theta$ or $\theta'$. This would eliminate the factors $\sin\theta \sin\theta'$ from the volume of a table cell, eliminating the waste of memory that occurs because the table is denser where $\theta$ or $\theta'$ is close to 0 or $\pi$. In addition, this would produce a modest computational savings (perhaps about 10% of the total time) because it would no longer be necessary to calculate the arccosine of $\cos\theta$ or $\cos\theta'$ during each table lookup.

## 5. CONCLUSIONS

We present here a first effort toward a tunable coarse-grained force field for proteins by extending our previous work on tabulation for simulations of fluid benzene.[37] We construct our force field by dividing proteins into rigid fragments, precomputing orientation-dependent interaction energy tables for those fragments, and applying smoothing techniques to the tables. The degree of smoothing can be adjusted to create coarse-grained models with an effective resolution that can be varied from a united-atom force field to a bead-like model. We have tested this approach on an $\alpha$-helix and a $\beta$-hairpin, and found that it can give improvement in sampling while preserving secondary structure. The approach has also been

tested as part of the resolution exchange method, producing an improvement in exchange rates over previous attempts.[44] In essence, resolution exchange is recast as Hamiltonian exchange since all degrees of freedom are retained in all replicas.

The initial results are encouraging, but important challenges remain. Thus, we have also proposed several improvements to the appproach, regarding solvation model, smoothing strategy, and computation speed, that could facilitate future applications. Because this model has been constructed differently from other coarse-grained models in the literature, it may prove useful for simulations in which other models may fail, including significant secondary structure changes as well as exchange simulations and mixed-resolution models. The tabulation approach could be particularly powerful at larger scales for encoding protein−protein interactions in cellular and viral-capsid contexts.

## ■ ASSOCIATED CONTENT

### Supporting Information

A derivation of the orientational smoothing kernel (eq 6), a proof of the theorem that the generalized second virial coefficient for interaction between fragments is left unchanged by the smoothing procedure, supplementary tables on the methods used, supplementary profiling data, and supplementary figures on the evolution of backbone RMSD in the simulations and on the temperature replica exchange simulation of the GB1 hairpin. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: ddmmzz@pitt.edu.

### Funding

This work was supported by NIH Grant No. P41-GM103712 and NSF Grant Nos. MCB-1119091 and CNS-1229064.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Dr. Sundar Raman Subramanian and Rohith Palli for assistance with the variable resolution library-based Monte Carlo code. We also thank Dr. Ernesto Suarez, Rory Donovan and Josh Adelman for helpful discussions and Dr. David Koes for technical assistance. We also thank the Department of Computational and Systems Biology and the Center for Simulation and Modeling at the University of Pittsburgh for computer time.

## ■ REFERENCES

(1) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694−698.
(2) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59−107.
(3) Takada, S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−137.
(4) Saunders, M. G.; Voth, G. A. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.
(5) Noid, W. G. *J. Chem. Phys.* **2013**, *139*, No. 090901.
(6) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423−12430.
(7) Taketomi, H.; Ueda, Y.; Gō, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 445−59.
(8) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531−1548.
(9) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(10) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181.
(11) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.
(12) Zuckerman, D. M. *J. Phys. Chem. B* **2004**, *108*, 5127−5137.
(13) Best, R. B.; Chen, Y.-G.; Hummer, G. *Structure* **2005**, *13*, 1755−1763.
(14) Okazaki, K.; Koga, N.; Takada, S.; Onuchic, J. N.; G.Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844−11849.
(15) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3860−3871.
(16) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807−822.
(17) Skolnick, J.; Kolinski, A. *Science* **1990**, *250*, 1121−1125.
(18) Kolinski, A.; Skolnick, J. *Proteins* **1994**, *18*, 338−352.
(19) Kolinski, A.; Skolnick, J. *Polymer* **2004**, *45*, 511−524.
(20) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 945−950.
(21) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534−552.
(22) Chebaro, Y.; Pasquali, S.; Derreumaux, P. *J. Phys. Chem. B* **2012**, *116*, 8741−8752.
(23) Liwo, A.; He, Y.; Scheraga, H. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16890−16901.
(24) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.
(25) Buchete, N.-V.; Straub, J. E.; Thirumalai, D. *Protein Sci.* **2004**, *13*, 862−874.
(26) Lu, M.; Dousis, A. D.; Ma, J. *J. Mol. Biol.* **2008**, *376*, 288−301.
(27) Ma, J. *Acc. Chem. Res.* **2009**, *42*, 1087−1096.
(28) Golubkhov, P. A.; Ren, P. *J. Chem. Phys.* **2006**, *125*, No. 046103.
(29) Wu, J.; Zhen, X.; Shen, H.; Li, G.; Ren, P. *J. Chem. Phys.* **2011**, *135*, No. 155104.
(30) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(31) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.
(32) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(33) Seo, M.; Rauscher, S.; Pomès, R.; Tieleman, D. P. *J. Chem. Theory Comput.* **2012**, *8*, 1774−1785.
(34) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116−4127.
(35) 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*, No. 244115.
(36) Kar, P.; Gopal, S. M.; Cheng, Y.-M.; Pretheus, A.; Feig, M. *J. Chem. Theory Comput.* **2013**, *9*, 3769−3788.
(37) Lettieri, S.; Zuckerman, D. M. *J. Comput. Chem.* **2012**, *33*, 268−275.
(38) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. *J. Chem. Theory Comput.* **2009**, *5*, 2531−2543.
(39) Larini, L.; Shea, J.-E. *J. Phys. Chem. B* **2012**, *116*, 8337−8349.
(40) Pappu, R. V.; Hart, R. K.; Ponder, J. W. *J. Phys. Chem. B* **1998**, *102*, 9725−9742.
(41) Pappu, R. V.; Marshall, G. R.; Ponder, J. W. *Nat. Struct. Biol.* **1999**, *6*, 50−55.
(42) Hart, R. K.; Pappu, R. V.; Ponder, J. W. *J. Comput. Chem.* **2000**, *21*, 531−552.
(43) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. *Phys. Rev. Lett.* **2006**, *96*, No. 028105.
(44) Lyman, E.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2006**, *2*, 656−666.
(45) Zuckerman, D. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; Taylor & Francis: Boca Raton, FL, 2009; Chapter 12, pp 171−184.
(46) Zuckerman, D. M. *Annu. Rev. Biophys.* **2011**, *40*, 41−62.
(47) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T. *Science* **1991**, *253*, 657−661.
(48) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(49) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016−1024.

(50) Kendrot, C. E.; Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1991**, *12*, 402−409.
(51) Hamilton, W. R. *Proc. R. Ir. Acad.* **1843**, *2*, 424−434.
(52) Hamilton, W. R. *Proc. R. Ir. Acad.* **1847**, *3*, 1−16.
(53) Karney, C. F. F. *J. Mol. Graphics Modell.* **2007**, *25*, 595−604.
(54) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(55) Bülow, T. Spherical Diffusion for Surface Smoothing and Denoising. Technical Report MS-CIS-01-38, University of Pennsylvania: Philadelphia, 2001.
(56) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(57) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(58) Brooks, B.; et al. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(59) Coutsias, E. A.; Seok, C.; Dill, K. A. *J. Comput. Chem.* **2004**, *25*, 1849−1857.
(60) Betancourt, M. R. *J. Chem. Phys.* **2011**, *134*, No. 014104.
(61) Mamonov, A. B.; Lettieri, S.; Ding, Y.; Sarver, J. L.; Palli, R.; Cunningham, T. F.; Saxena, S.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2012**, *8*, 2921−2929.
(62) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(63) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(64) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(65) Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566−579.
(66) Marcotte, I.; Separovic, F.; Auger, M.; Gagné, S. M. *Biophys. J.* **2004**, *86*, 1587−1600.
(67) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1608.
(68) Elcock, A. H. *Curr. Opin. Struct. Biol.* **2010**, *20*, 196−206.
(69) McGuffee, S. R.; Elcock, A. H. *PLOS Comp. Biol.* **2010**, *6*, No. e1000694.
(70) Nguyen, H. D.; Reddy, V. S.; Brooks, C. L. *Nano Lett.* **2007**, *7*, 338−344.
(71) Nguyen, H. D.; Reddy, V. S.; Brooks, C. L. *J. Am. Chem. Soc.* **2009**, *131*, 2606−2614.
(72) Rzepiela, A. J.; Louhivuori, M.; Peter, C.; Marrink, S. J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10437−10448.
(73) Sokkar, P.; Choi, S. M.; Rhee, Y. M. *J. Chem. Theory Comput.* **2013**, *9*, 3728−3739.
(74) Wassenaar, T. A.; Ingolfsson, H. I.; Priess, M.; Marrink, S. J.; Schaefer, L. V. *J. Phys. Chem. B* **2013**, *117*, 3516−3530.
(75) Lyman, E.; Zuckerman, D. M. *Biophys. J.* **2006**, *91*, 164−172.
(76) Grossfield, A.; Feller, S. E.; Pitman, M. C. *Proteins* **2007**, *67*, 31−40.
(77) Zhang, X.; Bhatt, D.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2010**, *6*, 3048−3057.
(78) Genheden, S.; Ryde, U. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8662−8677.
(79) Romo, T. D.; Grossfield, A. *Biophys. J.* **2014**, *106*, 1553−1554.
(80) Hansen, N.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2014**, *10*, 2632−2647.
(81) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(82) Archontis, G.; Simonson, T. *J. Phys. Chem. B* **2005**, *109*, 22667−22673.
(83) Gaillard, T.; Simonson, T. *J. Comput. Chem.* **2014**, *35*, 1371−1387.