Gay-Berne and electrostatic multipole based coarse-grain potential in implicit solvent

Johnny Wu,[1,a)] Xia Zhen,[1,a)] Hujun Shen,[2] Guohui Li,[2,b)] and Pengyu Ren[1,c)]
[1]*Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712-1062, USA*
[2]*Molecular Modeling and Design, State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhangshan Rd., Dalian 116023, People's Republic of China*

(Received 23 May 2011; accepted 26 September 2011; published online 20 October 2011)

A general, transferable coarse-grain (CG) framework based on the Gay-Berne potential and electrostatic point multipole expansion is presented for polypeptide simulations. The solvent effect is described by the Generalized Kirkwood theory. The CG model is calibrated using the results of all-atom simulations of model compounds in solution. Instead of matching the overall effective forces produced by atomic models, the fundamental intermolecular forces such as electrostatic, repulsion-dispersion, and solvation are represented explicitly at a CG level. We demonstrate that the CG alanine dipeptide model is able to reproduce quantitatively the conformational energy of all-atom force fields in both gas and solution phases, including the electrostatic and solvation components. Replica exchange molecular dynamics and microsecond dynamic simulations of polyalanine of 5 and 12 residues reveal that the CG polyalanines fold into "alpha helix" and "beta sheet" structures. The 5-residue polyalanine displays a substantial increase in the "beta strand" fraction relative to the 12-residue polyalanine. The detailed conformational distribution is compared with those reported from recent all-atom simulations and experiments. The results suggest that the new coarse-graining approach presented in this study has the potential to offer both accuracy and efficiency for biomolecular modeling. © *2011 American Institute of Physics*. [doi:10.1063/1.3651626]

I. INTRODUCTION

The ambition to understand molecular systems of increasing length and time scales drives the pursuit and development of coarse-grain (CG) computational models. It continues to be prohibitively expensive for all-atom molecular mechanics models to collect statistically converged measurements of molecular phenomena that involve large conformational rearrangements, such as protein folding, protein-protein interaction, and allosteric regulation.[1] Although there has been much development in the areas of enhanced sampling, the need to study the dynamics of large biomolecular systems over long time scales remains. Consequently, various coarse-graining strategies have been endeavored to model the systems of interest. Much effort has been made to develop coarse-grained models by matching the intermolecular interaction energy and force at the functional group or molecular level with all-atom simulations of specific systems. Klein and co-workers reported coarse-grained models of membrane lipids and proposed various coarse-graining strategies based on previous studies of polymer melts.[2,3] DeVane *et al*. have recently embarked on a method that employs the Lennard-Jones 9–6 and 12–4 forms to model nonbonded interactions of coarse-grain sites and have thus far validated the model on various amino acid side-chain analogs.[4] Hills *et al*. has demonstrated that a physics-based, isotropic site, solvent-free method is able to maintain the native structures of Trpzip, Trp-cage, and the open/close conformations of adenylate kinase[5] Moreover, the united-residue force field developed by Scheraga and co-workers has matured significantly and used to study the folding mechanism of specific domains of the staphylococcal protein A and the formin-binding protein.[6–13] Alternatively, sequence-based statistical potentials have been used as a coarse-grain approach to fold t-RNA, 5S, and 16S ribosomal RNA.[14,15]

In this work, a general coarse-grain model, consisting of rigid bodies of anisotropic Gay-Berne (GB) particles and point multipoles, has been developed. The Generalized Kirkwood method is applied to account for the solvation effects.[16] While the current CG model is constructed from atomic force fields as with other coarse-grained models, our focus is on representing the general components of intermolecular forces such as electrostatic and repulsion-dispersion at a CG level, rather than matching the overall effective forces produced by atomic models. The strategy is much similar to that of developing empirical atomic potential energy model from quantum mechanical principles. The resulting CG model is transferable and not limited to specific systems or environments. Another distinct feature is that the model adopts the common functional forms that are supersets of all-atom model, which will facilitate future multi-scale applications.

[a)]J. Wu and X. Zhen contributed equally to this work.
[b)]Author to whom correspondence should be addressed. Electronic mail: ghli@dicp.ac.cn. Tel: (0086) 411-8437-9593. Fax: (0086) 411-8467-5584.
[c)]Author to whom correspondence should be addressed. Electronic mail: pren@mail.utexas.edu. Tel: (512) 232-1832. Fax: (512)-471-0616.

II. MODEL AND METHODS

A. Gay-Berne potential

The coarse-grain repulsion-dispersion interactions are represented with anisotropic GB potentials. A full description of the Gay-Berne potential is available in our previous work[17,18] and in the supporting information as well. Based on Gaussian-overlap potential, the potential energy between two particles i and j has the form

$$U_{GB}(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \mathrm{r}_{ij}) = 4\varepsilon(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij}) \left[\left(\frac{d_w \sigma_0}{\mathrm{r}_{ij} - \sigma(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij}) + d_w \sigma_0} \right)^{12} - \left(\frac{d_w \sigma_0}{\mathrm{r}_{ij} - \sigma(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij}) + d_w \sigma_0} \right)^{6} \right], \tag{1}$$

where the range parameter $\sigma(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij})$ has the generalized form as

$$\sigma(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij}) = \sigma_0 \left[1 - \left\{ \frac{\chi\alpha^2(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{r}}_{ij})^2 + \chi\alpha^{-2}(\hat{\mathrm{u}}_j \cdot \hat{\mathrm{r}}_{ij})^2 - 2\chi^2(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{r}}_{ij})(\hat{\mathrm{u}}_j \cdot \hat{\mathrm{r}}_{ij})(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{u}}_j)}{1 - \chi^2(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{u}}_j)^2} \right\} \right]^{-\frac{1}{2}}, \tag{2}$$

and

$$\sigma_0 = \sqrt{d_i^2 + d_j^2}, \tag{3}$$

$$\chi = \left[\frac{(l_i^2 - d_i^2)(l_j^2 - d_j^2)}{(l_j^2 + d_i^2)(l_i^2 + d_j^2)} \right]^{1/2}, \tag{4}$$

$$\alpha^2 = \left[\frac{(l_i^2 - d_i^2)(l_j^2 + d_i^2)}{(l_j^2 - d_j^2)(l_i^2 + d_j^2)} \right]^{1/2}, \tag{5}$$

where l and d are the length and breadth of each particle, respectively.

The terms $\chi\alpha^2$, $\chi\alpha^{-2}$, and χ^2 can be calculated as

$$\chi\alpha^2 = \frac{l_i^2 - d_i^2}{l_i^2 + d_j^2}, \tag{6}$$

$$\chi\alpha^{-2} = \frac{l_j^2 - d_j^2}{l_j^2 + d_i^2}, \tag{7}$$

$$\chi^2 = \frac{(l_i^2 - d_i^2)(l_j^2 - d_j^2)}{(l_j^2 + d_i^2)(l_i^2 + d_j^2)}. \tag{8}$$

The total well-depth parameter is presented as

$$\varepsilon(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij}) = \varepsilon_0 \varepsilon_1^{\nu}(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j) \varepsilon_2^{\mu}(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij}). \tag{9}$$

The orientation-dependent strength terms are calculated in the following manner:

$$\varepsilon_1(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j) = [1 - \chi^2(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{u}}_j)^2]^{-1/2}, \tag{10}$$

$$\varepsilon_2(\hat{\mathrm{u}}_i, \hat{\mathrm{u}}_j, \hat{\mathrm{r}}_{ij}) = 1 - \left\{ \frac{\chi'\alpha'^2(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{r}}_{ij})^2 + \chi'\alpha'^{-2}(\hat{\mathrm{u}}_j \cdot \hat{\mathrm{r}}_{ij})^2 - 2\chi'^2(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{r}}_{ij})(\hat{\mathrm{u}}_j \cdot \hat{\mathrm{r}}_{ij})(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{u}}_j)}{1 - \chi'^2(\hat{\mathrm{u}}_i \cdot \hat{\mathrm{u}}_j)^2} \right\}, \tag{11}$$

where

$$\chi' = \left[\frac{\left(\varepsilon_{Si}^{1/\mu} - \varepsilon_{Ei}^{1/\mu}\right) \times \left(\varepsilon_{Sj}^{1/\mu} - \varepsilon_{Ej}^{1/\mu}\right)}{\left(\varepsilon_{Sj}^{1/\mu} + \varepsilon_{Ei}^{1/\mu}\right) \times \left(\varepsilon_{Si}^{1/\mu} + \varepsilon_{Ej}^{1/\mu}\right)} \right]^{1/2}, \tag{12}$$

$$\alpha'^2 = \left[\frac{\left(\varepsilon_{Si}^{1/\mu} - \varepsilon_{Ei}^{1/\mu}\right) \times \left(\varepsilon_{Sj}^{1/\mu} + \varepsilon_{Ei}^{1/\mu}\right)}{\left(\varepsilon_{Sj}^{1/\mu} - \varepsilon_{Ej}^{1/\mu}\right) \times \left(\varepsilon_{Si}^{1/\mu} + \varepsilon_{Ej}^{1/\mu}\right)} \right]^{1/2}. \tag{13}$$

The well depth of the cross configuration is denoted by ε_0, the well depth of the end-to-end/face-to-face configuration is presented as ε_E, and ε_S denotes the well depth of the side-by-side configuration.[19] Here we improved the accuracy of

the Gay-Berne model by separating the ratio of $\varepsilon_E/\varepsilon_S$ to two independent variables, ε_E and ε_S.

The new representations of χ' and α'^2 allow the consistent result for a pair of Gay-Berne particles of arbitrary types. Between unlike pairs, all ε_0 values and their ε_S and ε_E are specified explicitly or computed using a combining rule.[20] The d_W parameter describes the "softness" of the potential to allow better correlation with the all-atom energy profile. The parameters μ and ν were set to canonical values of 2.0 and 1.0, respectively. The current Gay-Berne potential with electrostatic multipole (GBEMP) model is implemented based on the TINKER molecular dynamics package.[21]

The terms χ'^2, $\chi'\alpha'^2$, and $\chi'\alpha'^{-2}$ were treated as inseparable and computed directly as

$$\chi'^2 = \frac{\left(\varepsilon_{Si}^{1/\mu} - \varepsilon_{Ei}^{1/\mu}\right) \times \left(\varepsilon_{Sj}^{1/\mu} - \varepsilon_{Ej}^{1/\mu}\right)}{\left(\varepsilon_{Sj}^{1/\mu} + \varepsilon_{Ei}^{1/\mu}\right) \times \left(\varepsilon_{Si}^{1/\mu} + \varepsilon_{Ej}^{1/\mu}\right)}, \tag{14}$$

$$\chi'\alpha'^2 = \frac{\left(\varepsilon_{Si}^{1/\mu} - \varepsilon_{Ei}^{1/\mu}\right)}{\left(\varepsilon_{Si}^{1/\mu} + \varepsilon_{Ej}^{1/\mu}\right)}, \tag{15}$$

$$\chi'\alpha'^{-2} = \frac{\left(\varepsilon_{Sj}^{1/\mu} - \varepsilon_{Ej}^{1/\mu}\right)}{\left(\varepsilon_{Sj}^{1/\mu} + \varepsilon_{Ei}^{1/\mu}\right)}. \tag{16}$$

Electrostatic potentials are represented with pairwise interactions of point multipole sites up to quadrupole. Each rigid body may contain zero or more off-center multipole sites where the local frame of the site is aligned with the principal axis of the rigid body. A complete description of electrostatic interactions of the GBEMP model are provided in previous work.[17,18]

III. RESULTS AND DISCUSSION

A. Benzene and methanol model

The improved Gay-Berne functional form has been validated on benzene and methanol molecules, which were represented by disk-like and rod-like particles, respectively. As with the previous studies,[17,18] the Gay-Berne parameters were derived by first fitting to the gas-phase homodimer intermolecular interaction energy and then refined in the liquid simulations. All-atom homodimer interactions energy for cross, end-end, face-face, and side-by-side configurations was obtained at various separations up to 12 Å apart. At each separation, the dimer interaction energy was calculated as a Boltzmann average over configurations generated by rotation about the primary axis of each Gay-Berne particle. Molecular electrostatic multipole (EMP) moments of benzene and methanol in liquid environments were obtained from atomic multipoles, including induced dipoles, given by the all-atom AMOEBA polarizable force field.[22,23]

In coarse-grained liquid simulations, the initial structures of benzene and methanol particles were created by mapping from all-atom structures. After rigid-body energy minimization, MD simulations of a box of ~300 molecules were performed with an NPT ensemble at 298 K and 1 atm. The periodic boundary condition was applied with a cutoff of 12 Å. Different time steps (up to 20 fs) were tested in the CG simulations.

A comparison of dimer interaction energies between the all-atom and GBEMP models shows that the new functions for combining the Gay-Berne well-depth parameters, ε_E and ε_S, produce a better agreement than the previous Gay-Berne function (see Table S1 of supplementary material).[49] The well-depth for benzene in the T shape configuration has increased to 0.91 Kcal/mol from 0.52 Kcal/mol (using the previous model) and more closely matches that of all-atom result (1.60 Kcal/mol) (see Figure S1).[49] Liquid simulations for benzene and methanol yield bulk properties, such as internal potential energy and density, which are in excellent agreement with the experimental values (error < 2%) (see Tables S2 and S3 of supplementary material).[49] More detailed comparison among CG and all-atom simulations, as well as experiment can be found in the supporting information. The GBEMP model is next extended to polyalanine peptides that consist of bonded coarse-grained particles.

B. Alanine model

In our CG model, a peptide is composed of covalently bonded rigid bodies, with Gay-Berne and/or electrostatic multipole sites. Bonding occurs between the Gay-Berne or EMP sites on different rigid bodies. Bond stretch energies adopt the fourth-order Taylor expansion of the Morse potential. Bond angle bend energies utilize a sixth-order potential. A three-term Fourier series expansion is calculated with the torsion energy. These valence functional forms are similar to those used by classical molecular mechanics potential such as MM3.[24] To use large time-step in MD simulations, the bond and angle terms can be restrained using rattle algorithm.[25]

In a previous work,[18] we have devised a general rigid-body representation containing an arbitrary number of off-centered Gay-Berne and multipole interaction sites that share the same local frame. Gay-Berne interactions are computed using orientation and site location vectors in Cartesian coordinates, relative to the local frame of the rigid body, as variables. Likewise, multipole interactions are computed via positions given by Cartesian coordinates relative to the local frame of

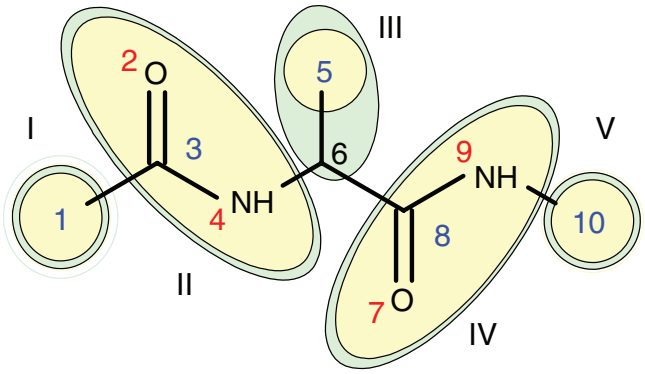

FIG. 1. Representation of dialanine GBEMP model. Ellipsoids encompass the rigid bodies (green) that contain Gay-Berne (blue) and multipole (red) interaction sites. The Gay-Berne particles are located at the center of the mass of the corresponding atoms.

the rigid body. The dialanine model consists of five rigid bodies (I through V) as depicted in Fig. 1. Gay-Berne parameters of amide and methyl groups were obtained with the same procedure as described above, by fitting to AMOEBA atomic force field. As in Fig. 1, the rigid body that represents the amide group consists of one Gay-Berne particle and two EMP sites. Gay-Berne sites 1, 5, and 10 are spherical methyl groups while sites 3 and 8 are equivalent ellipsoid amide groups. Similarly, sites 2 and 7 share the same EMP type, as do sites 4 and 9. Site 6 is used to compute bonded interactions only. Bonds exist between sites (1, 3), (4, 6), (6, 8), and (9, 10). An example of an angle is composed of sites (1, 3, and 2) and a torsion angle is composed of sites (3, 4, 6, and 8). The 12-mer alanine model polymerizes rigid bodies II and III from Fig. 1 as a repeating unit 12 times, thus, requiring 27 rigid bodies. For each rigid body type, the coordinates of the corresponding atoms are recorded in the local frame, which allow us to map the coarse grain molecules back to all-atom structures. Note that although the Gay-Berne particle is symmetric about the primary axis, the rigid body is not necessarily symmetric due to the presence of off-center site and/or multipoles.

Solvation is represented implicitly and is composed of polar and nonpolar contributions. Polar solvation employs the Generalized Kirkwood (GK) method,[16] a multipolar extension of the Generalized Born approach[26,27] and is computed for all the multipole sites. The Grycuk effective radius[28] is used in the polar solvation calculations. Nonpolar solvation is evaluated for all Gay-Berne sites with the ACE surface area method[29] and Still method[26,30] to estimate the effective radius of each particle. All solvation methods as well as effective radii estimation methods are implemented in the TINKER 5 (Ref. 21) molecular modeling package and adapted to the current GBEMP suite. Particle radii used for effective radii estimation are taken from the maximum of the Gay-Berne *l* or *d* parameters. Rigid bodies with more than one multipole site, like the amide groups in Fig. 1 (II and IV), uniformly divide the Gay-Berne radius value among all sites.

Parameters for the alanine model were obtained for the non-bonded terms, such as Gay-Berne and electrostatic multipole potentials, as well as the bonded terms, such as bond stretching, angle bending, and torsion energies. Applying the same procedure used to parameterize benzene and methanol, Gay-Berne and EMP parameters for each rigid body in an alanine residue were fit to all-atom homodimer energy and monomer multipole (in solution environments), respectively. Bond stretch and angle bend parameters were parameterized via Boltzmann inversion with atomic configurations generated from molecular dynamics of alanine dipeptide using AMOEBA. Molecular dynamics were executed in an NVT ensemble with explicit solvent (209 water molecules) in a 19.7 Å box with a 1 fs time step at 298 K. Torsional energy parameters were fit to the all-atom conformational energy map generated with fixed-charge OPLSAA with Generalized Born Surface Area implicit solvation.[26,29] OPLSAA is chosen as it is a commonly used atomic force field and uses the similar torsional energy function as in the current coarse-grain model. Nonetheless, the torsional parameters will be refined in the future by comparing directly to experimental data.[31] As we discuss below, the torsional term only contributes to a fraction of the conformational energy along with the intramolecular non-bonded electrostatic and van der Waals (vdW) interactions.

C. Dialanine energy components from coarse-grain model

The conformational energy of dialanine as a function of backbone dihedral angles, ϕ and ψ, is investigated in solution and gas phases. Conformations are generated at 30° intervals starting at the origin of the energy map by minimization with restraints. Conformational energies for the GBEMP model in solution- and gas-phase are shown in Fig. 2, compared with corresponding energies from all-atom model using the OPLSAA field.[32] The energy surface of the GBEMP model is smoother than that of the all-atom model as a consequence of coarse-graining. Nonetheless, the overall features of the CG gas phase energy maps are in fair agreement with the corresponding map of the atomic OPLSAA force field. Moreover, solution phase energy maps are in excellent qualitative agreement between the GBEMP and atomic force field. The agreement between solution phase energy maps is better than that of the gas phase maps and is expected since both are designed to describe solution phase properties. This is encouraging as the CG torsional parameters were only fit to the OPLSAA energy in solution. In addition, the solution-phase minima for alpha-helix, beta-sheet, as well as the less stable left-handed alpha-helix conformations are well manifested in the energy map.

When compared to the gas-phase electrostatic energy (Figs. 3(b) and 3(e)), the solvation energy contribution (Figs. 3(c) and 3(f)) clearly compensates the electrostatic interactions in gas-phase. This observation, true for both all-atom (OPLSAA) and the current CG potentials, is consistent with the physical interpretation that when secondary structure forms, intramolecular hydrogen bonds replace the hydrogen bonds between peptide and surrounding water.

We further compared the energy components of the coarse-grained GBEMP model with OPLSAA. A decomposition of the non-bonded interactions indicates that steric interaction given by the Gay-Berne function in the GBEMP model resemble that given the atomic vdW interaction energy of the OPLSAA force field over the Ramachandran map (Figs. 3(a) and 3(d)), including the scale. Likewise, contour maps of the gas-phase electrostatic energy (Figs. 3(b) and 3(e)), as well as the implicit solvation energy (Figs. 3(c) and 3(f)), show good agreement between the coarse grain and the all-atom results. Although the overall scales are different, the two components seem to mostly cancel each other as discussed above. As a result, the total energy minimum at the alpha-helix conformation mostly arises from the vdW contribution (Figs. 3(a) and 3(d)). A comparison of the torsional energy contribution (supporting information) between the CG and all-atom models also expresses a consistent behavior. The gas-phase conformational energy captures the C5 local minimum well.[33] However, the C7eq and C7ax minima have drifted slightly from the all-atom conformations. This may be due to the torsional energy contributions since their parameters were fit to the condensed-phase energy map. However, as with other all-

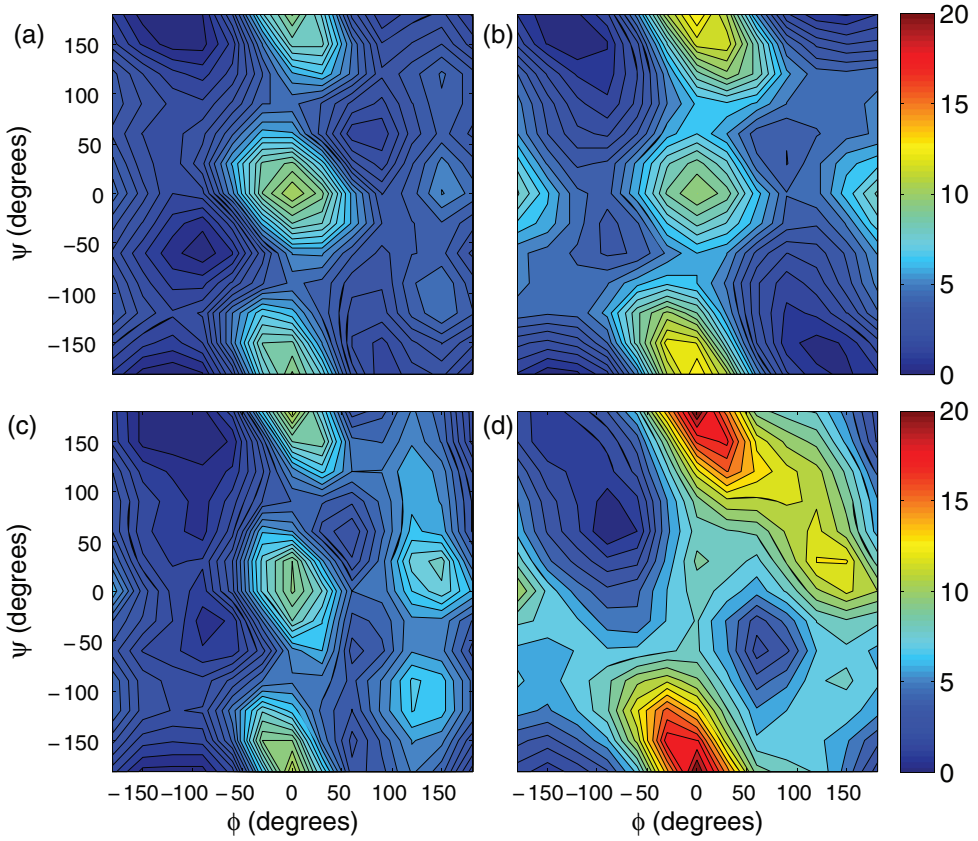

FIG. 2. Total conformational energy (kcal/mol) of alanine dipeptide: (a) CG model in solution, (b) CG model in gas-phase, (c) all-atom model (OPLSAA) in solution, (d) all-atom model (OPLSAA) in gas-phase.

atom fixed-charge models, transferability between gas- and solution-phase requires the inclusion of polarization effect.

D. Simulation of polyalanine

The conformation of polyalanine with various lengths has been investigated with both experimental and computational approaches.[31,34–47] To compare the GBEMP model with experiments and all-atom MD simulations, we investigated the blocked 5-mer polyalanine using GBEMP model in MD simulations. The aforementioned Generalized Kirkwood implicit solvent was utilized. The replica exchange molecular dynamics (REMD)[48] was performed to elucidate the conformational distribution of the 5-mer

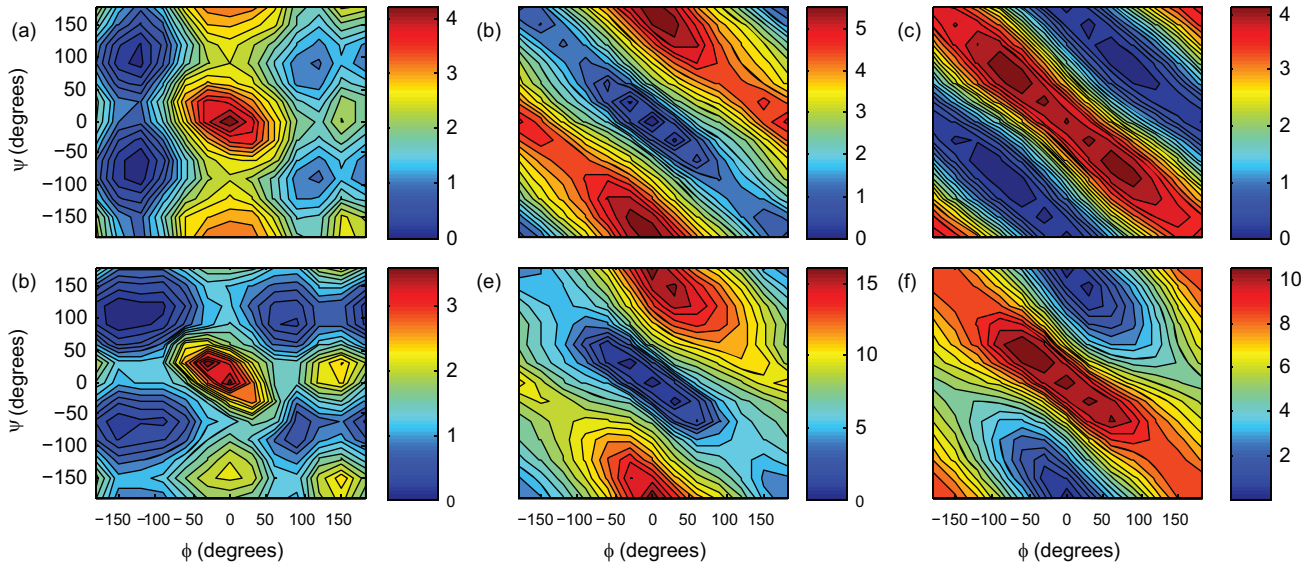

FIG. 3. Decomposition of alanine dipeptide energy (kcal/mol). Coarse-grain: (a) Gay-Berne energy, (b) gas-phase electrostatic energy, and (c) implicit solvation energy from GK/SA. All-atom: (d) vdW energy, (e) gas-phase electrostatic energy, and (f) implicit solvation energy from GB/SA.

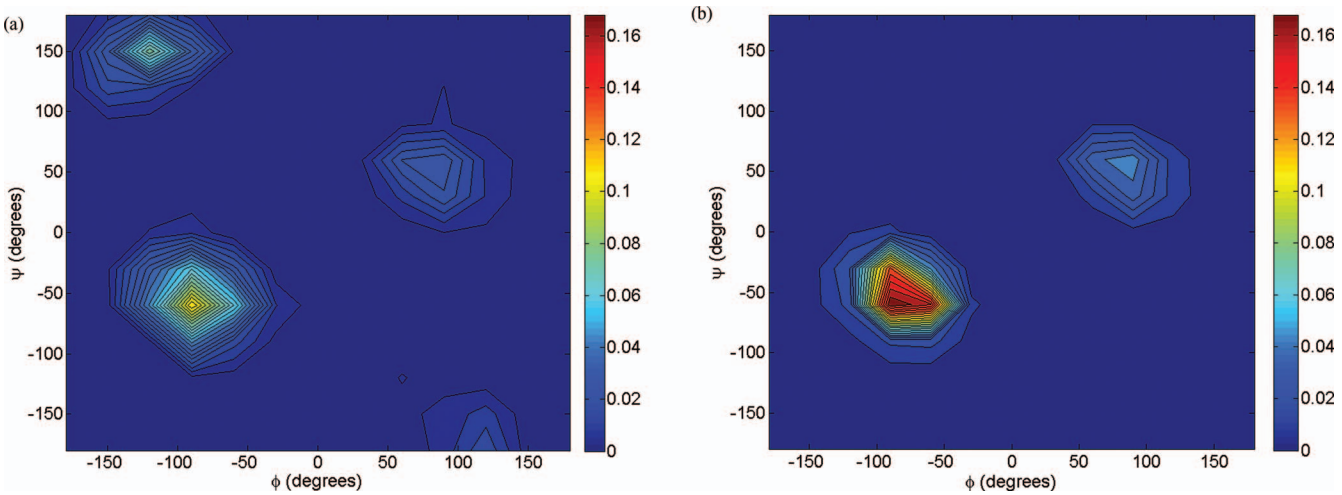

FIG. 4. Conformational distribution of 5-mer (a) and 12-mer (b) polyalanine from CG REMD simulations.

polyalanine. Thirty replicas were used between 298 and 800 K and the simulation time for each replica was 200 ns. The distribution of ϕ and ψ angles for all residues is shown in (Fig. 4(a)). Three dominant populations were observed: alpha-helix ($-160^\circ \leq \phi \leq -20^\circ$ and $-120^\circ \leq \psi \leq 50^\circ$), beta-strand ($-180^\circ \leq \phi \leq -90^\circ$ and $50^\circ \leq \psi \leq 240^\circ$; or $160^\circ \leq \phi \leq 180^\circ$ and $110^\circ \leq \psi \leq 180^\circ$), and left-handed helix ($20^\circ \leq \phi \leq 160^\circ$ and $-50^\circ \leq \psi \leq 120^\circ$). The 5-mer polyalanine conformations observed are comparable with all atom simulation results (Table I). Although circular dichroism (CD) spectroscopy and Fourier-transform infrared (FTIR) experiments reported somewhat less alpha-helix conformation,[47] the distributions sampled from MD simulations using all-atom force fields seem to be in qualitative agreement with what we obtained from the GBEMP simulations. Moreover, since the GBEMP model was developed based on interactions of all-atom force fields, it is reasonable for the model to behave consistently with all-atom simulation. Additionally, the population of full alpha-helices, in which ϕ and ψ angles of all five residues adopt the alpha-helical conformation, occurs at 4.62%, in comparison with 8% and 1% observed in all-atom simulations using CHARMM and Amber03 force field, respectively.[47]

To study the effects of chain-length, a 12-residue polyalanine system was simulated using REMD with 30 replicas and 500 ns for each replica. Residue-level fractions observed were 42%, 4.3%, and 21% for alpha-helix, beta-strand, and left-handed helix conformations, respectively. Although the beta-strand conformation exhibits a minima in the conformational energy landscape (Figs. 2(a) and 2(c)), a substantial (5-fold) decrease in the beta-strand fraction compared to the 5-mer polyalanine suggests that the hydrogen bonding scheme provided by the alpha-helix conformation stabilizes the 12-mer polyalanine. Additionally, simulated annealing MD simulations were performed to inspect the minimum-energy structure of the peptide after an initial rigid-body energy minimization. The systems were heated to 1000 K within the first 50 ps and then cooled linearly to less than 1 K over 60 ns. Five independent simulated annealing trials were performed and an example of the final structure is shown in Figure S5 of supplementary material.[49] The final polyalanine structures after simulated annealing all adopt the alpha-helical conformation at low temperatures (100 K, see Figure S3 of supplementary material).[49] A comparison of the RMSD between structures obtained from the simulated annealing trajectory and a canonical alpha-helix (see Figure S5 of supplementary material)[49] suggests that the accessible area of phase-space noticeably increases as the temperature rises above 500 K.

Furthermore, MD simulations of a few microseconds were performed at room temperature to verify the convergence of the conformational space determined by the GBEMP/REMD. These simulations started with different initial structures, including the extended conformation, alpha-helix, and partial alpha-helical and beta-strand conformations. The torsional distribution sampled from the GB-EMP MD simulation (6 μs for 12-mer and 2 μs for 5-mer) at 298 K is in agreement with the REMD conformational map and is provided in the supplementary material (Figure S6).

E. Computational efficiency of the Gay-Berne potential with electrostatic multipole model

The GBEMP model provides a great improvement in the performance of molecular modeling. Due to the reduction of

TABLE I. Per-residue fractions of 5-mer polyalanine from experiments and all-atom simulations.

Conformation	CD[a]	FTIR[a]	All-atom[a,b]	CHARMM 27/cmap[b]	OPLSAA/L[b]	GBEMP
Alpha-helix	$13 \pm 3\%$	$13 \pm 5\%$	4%–60% [a]	57.5%	32.8%	46%
Beta-strand	N/A	N/A	9.8%–55.5% [a]	19.8%	32.0%	28%

[a]Distributions from experiment and various force fields (see Ref. 47).
[b]Distributions of various force fields (see Ref. 31).

particle numbers and larger time-steps, the computational efficiency is enhanced by a factor of 50–800 compared to all-atom models tested with implicit and explicit solvent in this study (see Table S2 in the supplementary material).[49] Furthermore, the absence of high frequency motions, as required by all-atom models, allows time steps of up to 5 fs in MD simulations. Therefore, the CG model can achieve an improvement of about three orders of magnitude in the simulation speed and enable studies of large systems or extended simulation times from nanoseconds to microseconds.

IV. CONCLUSIONS

A unique coarse-grained GBEMP (Gay-Berne potential with electrostatic multipole) model has been developed based on the general physical principles of molecular interactions. In this CG potential, the fundamental components of intermolecular forces are represented explicitly: the van der Waals interaction is described by treating molecules as soft uniaxial ellipsoids interacting via a generalized anisotropic Gay-Berne function; the charge distribution is represented by off-center multipoles, including point charge, dipole, and quadrupole moments. The Generalized Kirkwood method and the ACE surface area method are used to calculate the polar and nonpolar solvation energy, respectively.[16,29] The coarse-grained GBEMP model has been implemented in the TINKER modeling package capable of rigid-body molecular dynamics simulation. The replica-exchange method is implemented to enhance sampling. The CG parameters are calibrated using all-atom force field (AMOEBA and OPLS-AA) and extension to other molecular systems is straightforward. Most importantly, there is no need for constant re-parameterization when applied to different environments. We tested the CG model on the alanine peptides of various lengths. The results show that the model and parameters can be directly transferred from gas phase to solution (with implicit solvent model), and from dialanine to polyalanine of different lengths. For the first time, we show that the individual energy components in the coarse-grained model, including vdW, electrostatics, solvation, and torsional energy contributions, match closely with those of all-atom force fields, in both gas-phase and solution. REMD and room-temperature MD simulations of 5-residue and 12-residue polyalanines predict reasonable alpha-helix and beta-sheet populations in comparison with all-atom simulations and experiments. Due to the reduction of particle numbers and larger time-steps, the computational efficiency is enhanced by a factor of up to 1000 compared with all-atom simulations. Further speedup is possible if the bonds and angles are restrained. The coarse-graining potential presented in this study can be extended to various biomolecular systems and even combined with all-atom potential in multiscale applications.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of General Medical Sciences (R01GM079686), an agency of the NIH, and the Robert A. Welch Foundation (F-1691). The HPC resources were provided by TeraGrid and Texas Advanced Computing Center at UT Austin (TG-MCB100057).

[1] D. Kern and E. R. P. Zuiderweg, Curr. Opin. Struct. Biol. **13**, 748 (2003).
[2] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, J. Phys. Chem. B **105**, 4464 (2001).
[3] S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, J. Phys. Condens. Matter **16**, R481 (2004).
[4] R. DeVane, M. L. Klein, C. C. Chiu, S. O. Nielsen, W. Shinoda, and P. B. Moore, J. Phys. Chem. B **114**, 6386 (2010).
[5] R. D. Hills, L. Y. Lu, and G. A. Voth, PLOS Comput. Biol. **6**(2010).
[6] M. Makowski, E. Sobolewski, C. Czaplewski, S. Oldziej, A. Liwo, and H. A. Scheraga, J. Phys. Chem. B **112**, 11385 (2008).
[7] G. G. Maisuradze, A. Liwo, S. Oldziej, and H. A. Scheraga, J. Am. Chem. Soc. **132**, 9444 (2010).
[8] A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga, J. Chem. Phys. **115**, 2323 (2001).
[9] C. Wu and J. E. Shea, Curr. Opin. Struct. Biol. **21**, 209 (2005).
[10] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Oldziej, K. Wachucik, and H. A. Scheraga, J. Phys. Chem. B **111**, 260 (2007).
[11] A. Liwo, C. Czaplewski, S. Oldziej, A.V. Rojas, R. Kazmierkiewicz, M. Makowski, R. K. Murarka, and H. A. Scheraga, in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, edited by G. Voth (CRC, Taylor & Francis Group, Farmington, CT, 2008), p. 107.
[12] C. Czaplewski, A. Liwo, M. Makowski, S. Oldziej, and H. A. Scheraga, in *Multiscale Approaches to Protein Modeling*, edited by A. Kolinski (Springer, New York, 2010).
[13] G. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC, Taylor & Francis Group, Farmington, CT, 2008).
[14] J. C. Wu, D. P. Gardner, S. Ozer, R. R. Gutell, and P. Y. Ren, J. Mol. Biol. **391**, 769 (2009).
[15] Z. Xia, D. P. Gardner, R. R. Gutell, and P. Ren, J. Phys. Chem. B **114**, 13497 (2010).
[16] M. J. Schnieders and J. W. Ponder, J. Chem. Theory Comput. **3**, 2083 (2007).
[17] P. A. Golubkov and P. Y. Ren, J. Chem. Phys. **125**, 64103 (2006).
[18] P. A. Golubkov, J. C. Wu, and P. Y. Ren, Phys. Chem. Chem. Phys. **10**, 2050 (2008).
[19] D. J. Cleaver, C. M. Care, M. P. Allen, and M. P. Neal, Phys. Rev. E **54**, 559 (1996).
[20] T. A. Halgren, J. Am. Chem. Soc. **114**, 7827 (1992).
[21] J. W. Ponder, TINKER – Software Tools for Molecular Design, Washington University Medical School, 2010; see http://dasher.wustl.edu/tinker.
[22] P. Ren, and J. W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).
[23] J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, J. Phys. Chem. B **114**, 2549 (2010).
[24] N. L. Allinger, Y. H. Yuh, and J. H. Lii, J. Am. Chem. Soc. **111**, 8551 (1989).
[25] H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).
[26] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[27] R. Constanciel and R. Contreras, Theor. Chim. Acta. **65**, 1 (1984).
[28] T. Grycuk, J. Chem. Phys. **119**, 4817 (2003).
[29] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[30] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, J. Phys. Chem. A **101**, 3005 (1997).
[31] R. B. Best, N. V. Buchete, and G. Hummer, Biophys. J. **95**, L7 (2008).
[32] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[33] T. Headgordon, M. Headgordon, M. J. Frisch, C. L. Brooks, and J. A. Pople, J. Am. Chem. Soc. **113**, 5989 (1991).
[34] P. Y. Chou and G. D. Fasman, Biochemistry **13**, 211 (1974).
[35] J. S. Richardson and D. C. Richardson, Science **240**, 1648 (1988).
[36] R. R. Hudgins, M. A. Ratner, and M. F. Jarrold, J. Am. Chem. Soc. **120**, 12974 (1998).
[37] Y. Levy, J. Jortner, and O. M. Becker, Proc. Natl. Acad. Sci. U. S. A. **98**, 2188 (2001).
[38] A. E. Counterman and D. E. Clemmer, J. Am. Chem. Soc. **123**, 1490 (2001).
[39] J. Nakano, S. Kuroki, I. Ando, T. Kameda, H. Kurosu, T. Ozaki, and A. Shoji, Biopolymers **54**, 81 (2000).

[40]H. D. Nguyen, A. J. Marchut, and C. K. Hall, Protein Sci. **13**, 2909 (2004).
[41]P. Soto, A. Baumketner, and J. E. Shea, J. Chem. Phys. **124**, 134904 (2006).
[42]J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, Biophys. J. **92**, 4289 (2007).
[43]J. W. Chu, S. Izvekov, and G. A. Voth, Mol. Simul. **32**, 211 (2006).
[44]M. F. Jarrold, Phys. Chem. Chem. Phys. **9**, 1659 (2007).
[45]J. Graf, P. H. Nguyen, G. Stock, and H. Schwalbe, J. Am. Chem. Soc. **129**, 1179 (2007).
[46]F. Albrieux, F. Calvo, F. Chirot, A. Vorobyev, Y. O. Tsybin, V. Lepere, R. Antoine, J. Lemoine, and P. Dugourd, J. Phys. Chem. A **114**, 6888 (2010).
[47]W. A. Hegefeld, S. E. Chen, K. Y. DeLeon, K. Kuczera, and G. S. Jas, J. Phys. Chem. A **114**, 12391 (2010).
[48]E. S. Penev, S. Lampoudi, and J. E. Shea, Comput. Phys. Commun. **180**, 2013 (2009).
[49]See supplementary material at http://dx.doi.org/10.1063/1.3651626 for following details: Table S1: Gay-Berne parameters of benzene, methanol, and water GBEMP models; Table S2: MD simulation results for benzene; Table S3: MD simulation results for methanol; Table S4: Bond, bond-angle, torsional, and multiple parameters for GBEMP model. Table S5: Comparison of computational efficiency of GBEMP model with all-atom simulations. Figure S1: comparison of homodimer interaction energy give by the Gay-Berne model and all-atom model for benzene and methanol; Figure S2: *phi* and *psi* torsion angle distribution of 12-mer polyalanine at temperature of 800 K to 900 K in the simulated annealing simulation; Figure S3: *phi* and *psi* torsion angle distribution of 12-mer polyalanine at temperature of 1 K to 100 K in the simulated annealing simulation; Figure S4: contribution of torsional energy to total conformational energy of dialanine GBEMP model and All-atom OPLSAA; Figure S5: Simulated annealing MD simulations to inspect the minimum-energy structure of the peptide after an initial rigid-body energy minimization. (a) A final snapshot of polyalanine from the 60-ns simulated annealing simulations using GBEMP potential. (b) Heavy-atom RMSD of the 12-residue polyalanine from 5 simulated annealing simulations; Figure S6: conformational distributions of 5-mer and 12-mer polyalanine from CG simulations at 298 K.