THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

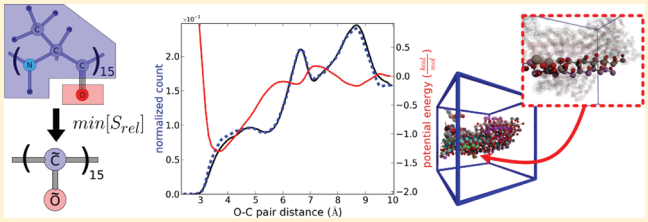
A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly

Scott P. Carmichael and M. Scott Shell*

Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, California

Supporting Information

ABSTRACT: Peptide self-assembly plays a role in a number of diseases, in pharmaceutical degradation, and in emerging biomaterials. Here, we aim to develop an accurate molecular-scale picture of this process using a multiscale computational approach. Recently, Shell (Shell, M. S. *J. Chem. Phys.* **2008**, *129*, 144108-7) developed a coarse-graining methodology that is based on a thermodynamic quantity called the relative entropy, a measure of how different two molecular ensembles behave. By minimizing the relative entropy between a coarse-grained peptide system and a reference all-atom system, with respect to the coarse-grained model's force field parameters, an optimized coarse-grained model can be obtained. We have reformulated this methodology using a trajectory-reweighting and perturbation strategy that enables complex coarse-grained models with at least hundreds of parameters to be optimized efficiently. This new algorithm allows for complex peptide systems to be coarse-grained into much simpler models that nonetheless recapitulate many correct features of detailed all-atom ones. In particular, we present results for a polyalanine case study, with attention to both individual peptide folding and large-scale fibril assembly.



INTRODUCTION

Self-assembled peptide systems are a versatile new class of materials that have diverse applicability in industry and medicine. For example, they are capable of templating the growth of metal nanowires,[1,2] forming three-dimensional cell-culture scaffolds,[3,4] and providing in vivo matrixes for axon regeneration[5] or wound healing.[6] These remarkable properties are due to their innate biocompatibility, sequence-specific self-assembled phase behavior, and ease of synthesis. While many model peptide systems have been identified that have interesting properties,[2,7,8] the vast peptide sequence space affords a huge potential for new systems that have yet to be explored. Simulations provide a natural means of exploring this space by investigating the nature of the sequence−structure relationship, but traditional all-atom (AA) simulations (using a force field such as AMBER[9]) cannot practically explore the large system sizes and time scales required to capture long-ranged structure formation. Therefore, there is a need for methods that can extract computationally fast coarse-grained (CG) peptide models (that retain sequence specific information) from accurate AA models, for eventual deployment in larger scale simulations.

Here we pursue a "bottom-up" strategy for modeling the self-assembly of peptide systems. To accomplish this, we transform accurate but compute-intensive AA peptide models (and reference simulations of them) into much faster CG ones, using a novel information-theoretic coarse-graining strategy. Many copies of these CG models are then used to simulate the self-assembly process at much larger scales. Importantly, our coarse-graining methodology generates CG models that retain the information encoded in the target peptide's amino acid sequence by coupling to AA simulations, and are therefore in principle capable of predicting differences in peptide self-assembly behavior that arise from specific sequence mutations.

The rich phase behavior that peptides exhibit is currently an area of intense research.[10−13] A heavily studied aggregate phase is amyloid fibrils. Implicated in the pathology of over 30 diseases,[14] amyloid fibrils are ordered structures composed of peptide strands that consist of very long, stacked β-sheet elements.[15] Experimental studies have indicated fibrils form via a nucleation-elongation pathway[16] accompanied by subsequent secondary nucleation,[17] and several simulations have further corroborated the nucleation-elongation finding.[18−21] There have also been efforts to discover non-natural, de novo designed amyloid-forming sequences,[22−29] and a mutation study by Chiti et al.[30] has uncovered a strong correlation between changes in average peptide properties due to mutations (i.e., net charge, hydrophobicity, β-sheet propensity, etc.) and amyloid aggregation rate. The molecular origin of this effect is currently unclear but is very much of the type of problem that motivates the need for accurate CG peptide models.

While both AA and CG simulations have been used to study fibril formation, the high computational cost of atomistic

Special Issue: Macromolecular Systems Understood through Multi-scale and Enhanced Sampling Techniques

Received: November 29, 2011
Revised: January 27, 2012
Published: February 2, 2012

dx.doi.org/10.1021/jp2114994 | *J. Phys. Chem. B* 2012, 116, 8383−8393

simulations has limited their applicability to studies of the stabilities of large preformed structures[31] or to the early behavior of small oligomer systems.[32] On the other hand, "toy model" CG simulations have been used to study the entire fibril formation process. These CG models for peptides have recently been reviewed by Shea[33] and Tozzini.[34] Here we highlight a few of them. Caflisch and co-workers[35] proposed a "toy" model with a variable oligomerization propensity, and used it to study the dynamics of fibril formation. Shea and co-workers[36,37] studied the effects of β-sheet propensity and temperature on the structure of fibrillar peptide aggregates using a three bead per amino acid model. Hall and co-workers[38] used a four bead per amino acid model, and applied it to map out aggregation phase diagrams for several peptides. While these CG models are highly informative about general properties of fibril formation, they are not generally capable of accurately representing the physics of any single peptide sequence, because they lack a direct connection to more fundamental interatomic interactions.

A basic challenge is finding unique and optimal ways of representing detailed AA physics with CG models. The problem has historically been circumvented by tuning CG models to reproduce basic bulk trends of phase behavior or native structures.[30,33,39−41] Such parametrization tasks are inevitably under-constrained (i.e., a few experimental parameters rationalize a large number of model variables) and likely do not extrapolate well beyond the original behaviors subject to matching. Further, these kind of simple, hand-tuned CG models do not connect back to realistic atomic-scale interactions and, importantly, the sequence-specific features that they naturally encode. Generally, these kinds of models are "toy" in the sense that they do not really represent a particular sequence but are rather used to probe generic self-assembly behavior.[42] Capturing the sequence landscape remains a crucial target in this field, as it enables an understanding of associated pathologies and the development of rational design techniques for creating new peptide materials.

Determining an effective CG force field from detailed, accurate AA models is a nontrivial inverse problem. A natural approach is to tune the CG model so as to match specific static or all-ensemble properties determined from equilibrium simulations of small but representative AA systems. One prominent technique is the method proposed by Voth and co-workers[43,44] that matches interatomic forces. For simple systems, this approach has had success in generating CG potentials for AA models of phospholipids[45,46] and biomolecules,[47−49] including peptides.[43,46,50] A second major technique has been iterative Boltzmann inversion, which parametrizes CG models to match specific probability distribution functions (e.g., pair correlations) of the target AA system.[51] This too has been widely used with great success for simple polymer and lipid systems;[52−54] however, as the models become more complex, the optimization becomes difficult to converge due to the highly correlated nature of intermolecular interactions.[51]

Our approach is to parametrize a CG model so that its configurational ensemble (the distinct set of probabilities for all configurational microstates) matches that of the AA system as closely as possible. Such ensemble matching results in a further matching of all of the system properties that are a function of the configurational ensembles (e.g., free energy differences, native and non-native structures). Since the AA and CG ensembles are characterized by distribution functions, the distance between them can be quantified in informatic terms, using the relative entropy or Kullback−Leibler divergence, as proposed by Shell.[55] In this sense, the relative entropy effectively measures the amount of information lost due to coarse-graining by quantifying the overlap between two configurational ensembles (AA and CG). When the relative entropy is minimized with respect to the CG model's force field, the two ensembles have maximum overlap, in an information-theoretic sense. Hence, relative entropy minimization provides a natural way of linking AA information to CG force fields. Shell and co-workers have previously used this approach to extract optimized, very simple CG water models that capture both bulk properties[56] and hydrophobic interactions,[57] and they have also shown that coarse-graining errors can be predicted and systematically eliminated using relative entropy ideas.[58]

The work we describe here extends this approach to many-parameter CG models, such as those useful for describing peptides. Our efforts focus on developing CG models for a test-case polyalanine peptide $(Ala)_{15}$. We compare the single peptide properties (structural distributions, averages, and behavior in temperature space) of these CG models directly to AA ones. Then, by deploying the CG models in large-scale simulations, we observe whether they exhibit behavior similar to that expected for amyloid-beta structural motifs. We also develop and compare several different peptide models of varying detail, a particularly novel aspect of this work. Ultimately, the results of our efforts suggest a new approach to studying large-scale peptide solution behavior, using a bottom-up strategy starting with AA physics and reaching larger scales in a manner guided by the relative entropy.

METHODS

Coarse-Graining by Relative Entropy Minimization. Shell and co-workers[55,56,58,59] developed a systematic multiscale coarse graining approach that relies upon minimizing the relative entropy, given by

$$S_{\mathrm{rel}} = \sum_{\nu} p_{\mathrm{AA}}(\nu) \ln\left(\frac{p_{\mathrm{AA}}(\nu)}{p_{\mathrm{CG}}(M[\nu])}\right) + S_{\mathrm{map}} \tag{1}$$

where $p_X(\nu)$ is the equilibrium ensemble probability of a particular configuration using model X. $M[\nu]$ is a mapping operator for turning atomistic to CG coordinates, and it stems from how one chooses to partition the AA into CG sites.[44,60] The summation is over the set of all atomistic configurations ν, and for the classical systems we investigate here, it is actually an integral in configuration space. S_{map} is the so-called mapping entropy that measures, on average, how many AA configurations map to the same CG one;[55,59] it is rigorously independent of the CG force field. A mathematical result of the probability distributions being normalized is that the relative entropy has a lower bound of zero, attained when the two ensembles overlap perfectly.

When Boltzmann configurational probabilities are substituted into eq 1, one obtains the canonical expression for the relative entropy

$$S_{\mathrm{rel}} = \beta\langle U_{\mathrm{CG}} - U_{\mathrm{AA}}\rangle_{\mathrm{AA}} - \beta(A_{\mathrm{CG}} - A_{\mathrm{AA}}) + S_{\mathrm{map}} \tag{2}$$

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

Here, $\beta = 1/k_BT$, U is the total potential energy, A is the configurational free energy, and the averaging is performed in the target AA ensemble:

$$\langle U_{CG} - U_{AA}\rangle_{AA} = \sum_{\nu} p_{AA}[U_{CG}(M[\nu]) - U_{AA}(\nu)] \tag{3}$$

where the mapping function M converts an AA configuration ν to a CG one, as before.

A general coarse-graining strategy based on minimizing the relative entropy requires careful consideration. For CG models of complex biomolecules like peptides, there are often many parameters describing the CG force field that must be determined, and thus optimizing the coarse model results in a search in a very high dimensional parameter space. For example, a single bead per amino acid model of polyalanine that uses modest splines to represent the pair, angle, and torsional potentials contains on the order of 100 potential parameters. An exhaustive search of this parameter space is impractical. Moreover, directly evaluating the relative entropy is difficult due to its dependence on the Helmholtz free energies (A_{CG} and A_{AA}) of the two systems, although computationally intensive free energy techniques could be employed in principle.[61] Fortunately, it is much easier to compute the derivatives of the relative entropy in parameter space, which can guide one toward the optimum location via steepest descent and related procedures. The first derivatives are given by

$$\frac{\partial S_{rel}}{\partial \lambda} = \beta\left\langle \frac{\partial U_{CG}}{\partial \lambda}\right\rangle_{AA} - \beta\left\langle \frac{\partial U_{CG}}{\partial \lambda}\right\rangle_{CG} \tag{4}$$

where λ is a vector composed of the set of all CG potential parameters λ_i (e.g., Lennard-Jones coefficients, spline knots, etc.). The second derivative Hessian matrix is given by

$$H_{ij} = \beta\left\langle \frac{\partial^2 U_{CG}}{\partial\lambda_i\partial\lambda_j}\right\rangle_{AA} - \beta\left\langle \frac{\partial^2 U_{CG}}{\partial\lambda_i\partial\lambda_j}\right\rangle_{CG} + \beta^2\left\langle \frac{\partial U_{CG}}{\partial\lambda_i}\frac{\partial U_{CG}}{\partial\lambda_j}\right\rangle_{CG} - \beta^2\left\langle \frac{\partial U_{CG}}{\partial\lambda_i}\right\rangle_{CG}\left\langle \frac{\partial U_{CG}}{\partial\lambda_j}\right\rangle_{CG} \tag{5}$$

Notice that if U is linear in λ (e.g., $U = \lambda\mathbf{F}(\nu) + ...$) then, at the relative entropy minimum, $\langle\mathbf{F}(\nu)\rangle_{AA} = \langle\mathbf{F}(\nu)\rangle_{CG}$; thus, optimal models will ensure that the averages of $\mathbf{F}$ will be identical. Moreover, for such functionalities, the curvature of the relative entropy in parameter space for CG models built entirely from such linear basis functions will always be zero or positive and thus admit only a single minimum relative entropy basin.

For CG force fields where pair, angle, or torsional interactions are modeled by splines, the minimum condition is taken with respect to the spline knots. For suitably flexible splines with many knots, this condition can be approximated by a functional derivative. If, for example, a particular component spline interaction is denoted by $u(\xi)$, then the minimum condition is given by $\langle\delta U_{CG}/\delta u(\xi)\rangle_{AA} = \langle\delta U_{CG}/\delta u(\xi)\rangle_{CG}$, which after evaluation of the functional derivative (involving delta functions) gives $\mathcal{P}_{AA}(\xi) = \mathcal{P}_{CG}(\xi)$, where $\mathcal{P}_X(\xi)$ are the probability distributions of an arbitrary structural coordinate ξ. The implication is that including a spline energy term in the CG model that is a function of a coarse structural coordinate ξ will result, after relative entropy minimization, in the CG distribution of ξ more or less exactly matching that of the reference AA system. Clearly, this establishes a direct connection with Boltmann inversion methods and a detailed discussion is found in ref 59.

We use two standard numerical techniques to minimize the relative entropy with respect to the model parameters (λ), steepest descent and Newton−Raphson minimization. These approaches use the calculated derivatives above to step iteratively through parameter space toward the relative entropy minimum. The Newton−Raphson iteration scheme yields the following equation

$$\lambda^{k+1} = \lambda^k - \chi_{NR}H^{-1}\cdot\frac{\partial S_{rel}}{\partial\lambda} \tag{6}$$

where χ_{NR} is a step-size parameter used to stabilize the algorithm's convergence. Each iteration step, k, through parameter space brings the CG force field closer to the optimum. The Newton−Raphson approach can only be used when the Hessian (eq 5) is positive definite (i.e., the current parameter set $\{\lambda_i\}$ is in a basin of positive curvature of S_{rel} in the parameter space). Otherwise, we take a steepest-descent step instead, in which case H in eq 6 is replaced by the identity matrix and a different step size χ_{SD} is used.

Determining the values for the relative entropy derivatives requires performing ensemble averages of the derivatives of the CG potential, in accordance with eqs 4 and 5. Evaluating these averages in the AA ensemble requires two steps. Finely discretized histograms of relevant structural coordinates (e.g., bond distances, bond angles, torsion angles, and pair distances) are generated from a reference AA trajectory. A given parameter in the CG potential corresponds directly to one of these structural coordinates; e.g., a spline knot in a pair potential affects the corresponding pair distribution. Thus, the relevant ensemble average can be found quickly from the sum of the product of the corresponding normalized histogram with the current derivative of the CG potential.[56]

The CG averages could be evaluated by performing a trial CG simulation at each iteration step and explicitly evaluating the terms in eqs 4 and 5, but this approach is computationally expensive and, more importantly, introduces a stochastic element in the minimization procedure. This was the approach originally developed by Shell and co-workers.[55,56] To ameliorate the stochastic nature of that strategy, we propose a new reweighting-based procedure. In it, a single initial reference CG simulation is performed using an initial "guess" parameter set (λ^0) for the potential, and a trajectory is saved from it that can be parsed to evaluate averages of the relative entropy derivatives. Using those averages, a Newton−Raphson iteration step is performed and a new parameter set (λ) is obtained. Frames from the reference trajectory are then reweighted to reflect the changed ensemble probabilities of the new parameter set, and the reweighted trajectory is used for the CG averages in the next iteration, and so on. Such reweighting strategies are well established in simulations;[62] the first derivative for the relative entropy, for example, becomes

$$\frac{\partial S_{rel}}{\partial\lambda} = \beta\left\langle \frac{\partial U_{CG}}{\partial\lambda}\right\rangle_{AA} - \beta\frac{\left\langle \frac{\partial U_{CG}}{\partial\lambda}w\right\rangle_{CG,\lambda^0}}{\langle w\rangle_{CG,\lambda^0}} \tag{7}$$

where $w \equiv \exp(\beta(U_{CG,\lambda^0} - U_{CG,\lambda})) = \exp(\beta\Delta U_{CG})$ and the superscript 0 indicates the original parameters. A related expression exists for reweighted versions of the second derivative (eq 5). By employing the reweighting formalism,

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

we are actually minimizing the difference between the relative entropy calculated using parameter set λ and that calculated using the initial guess parameter set λ^0

$$\Delta S_{\mathrm{rel}} = S_{\mathrm{rel},\lambda} - S_{\mathrm{rel},\lambda^0}$$

$$= \beta\langle U_{\mathrm{CG},\lambda} - U_{\mathrm{CG},\lambda^0}\rangle_{\mathrm{AA}} - \beta(A_{\mathrm{CG},\lambda} - A_{\mathrm{CG},\lambda^0}) \quad (8)$$

$$= -\beta\langle \Delta U_{\mathrm{CG}}\rangle_{\mathrm{AA}} + \ln\langle w\rangle_{\mathrm{CG},\lambda^0} \quad (9)$$

where we have expressed the Helmholtz free energy difference using the classic Zwanzig perturbation expression in the last line. This reweighting approach has several key advantages. It becomes much easier to evaluate ΔS_{rel}, which relies only upon ensemble averages, than for the absolute relative entropy, which may require special free energy techniques. Further, this objective function attains a deterministic value, since the CG trajectory is created only once, for the initial parameter set. Finally, ΔS_{rel} shares the same minimum in parameter space as the relative entropy, which is trivial to show, since $S_{\mathrm{rel},\lambda^0}$ is constant during minimization.

One challenge in the reweighting approach is the classic problem of averaging exponentials. Namely, if the current parameter set λ is too different from the guess trajectory one λ^0, statistical errors in the reweighted averages in eq 7 become large owing to the large weights given to rarely sampled configurations. When these errors would become too big, we solve this problem by occasionally creating a new guess trajectory from the current parameter set and restarting the minimization procedure. To judge whether a new trajectory is required, the algorithm monitors the effective fractional number of frames from the reference trajectory that contribute to reweighting, $f_{\mathrm{RW}} = \exp(-\sum p_i \log p_i)/n$, where n is the total number of frames and p_i gives the probability weight of frame i in reweighting averages, i.e., $p_i = w_i/\sum_j w_j$.

Progression of the Reweighting-Based S_{rel} Coarse-Graining Algorithm. In practice, our implementation of the algorithm described above consists of several steps. As input, it takes an AA *target* trajectory, a CG mapping function which describes the geometry of the *model* (i.e., how AA configurations are turned into CG geometries), and a set of adjustable potential functions which define the interactions (force field) within the model system (the potential functions are initialized with "guess" parameters). After S_{rel} minimization, the algorithm then returns a set of optimized CG force field parameters.

The sequence of events that take place within the algorithm is as follows. The CG mapping function is used to project the atomic coordinates of each frame in the *target* AA trajectory onto representative CG sites. This CG representation of the AA system yields histograms of all of the structural coordinates relevant to the potential energy functions of the CG force field, which are later used to calculate the AA ensemble averages found in eqs 4 and 5. The algorithm then initializes a CG simulation using an initial guess set of parameters. The ultimate optimized model should be invariant to these if the CG force field is linear in all of its parameters, as there should be a single relative entropy minimum in this case.[55,59]

The algorithm next performs a constant-temperature (Langevin thermostat) molecular dynamics simulation of the current CG model. The production period for this run is used to generate an initial CG trajectory that serves to later calculate the relative entropy Jacobian and Hessian (eqs 4 and 5). These multidimensional derivatives are then used in either a steepest-descent or a Newton−Raphson iteration scheme to update the potential parameter set, and therefore step through parameter space toward the global minimum, as depicted in Figure 1. The

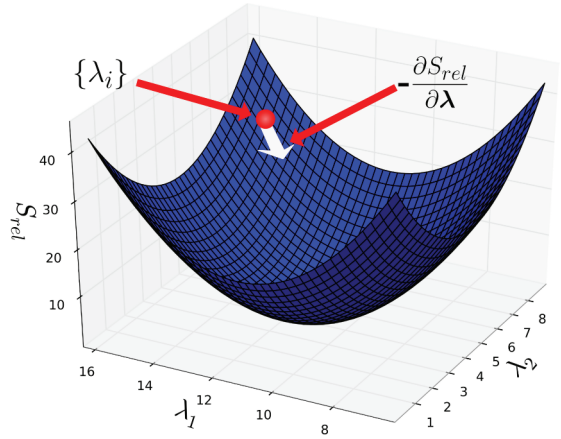


Figure 1. Schematic of the coarse-graining strategy. The current CG model parameter set is guided toward the relative entropy minimum by the gradient and curvature in parameter space.

updated parameters at each iteration are used to reweight CG trajectory frames and averages in accordance with the reweighting strategy of eq 7. The process is repeated until the parameters no longer change by more than 10^{-8} fractional value with each iteration. Periodically, a new CG trajectory is generated and minimization restarted in accordance with the procedure outlined above.

Several techniques are employed to ensure the stability of the minimization procedure. Many of these involve manipulating the step sizes that are taken through parameter space. First, each parameter has a maximum allowable change per iteration step, typically 50% of their initial guess value. These are predefined during the initial setup of the CG model, and prevent "runaway" parameter changes that may occur due to random statistical fluctuations. Second, the step size parameters χ_{NR} and χ_{SD} are dynamically adjusted so that the second-order forward and backward finite-difference changes in S_{rel} differ between 1 and 20%. As part of this, backtracking of a step is applied if it is found that the relative entropy changes outside of this range. Finally, when taking Newton−Raphson steps, statistical fluctuations in the Hessian entries can become magnified upon its inversion (eq 6), leading to inaccuracies in the direction (through parameter space) of the step. To mediate these effects, we use Hessian conditioning according to the well-established method of diagonal increments.[63]

It is worthwhile to note that the reformulated relative entropy minimization method (eq 7) enables the application of conjugate gradient approaches as well, since it admits a unique value of the objective function ΔS_{rel}. Our tests show that the conjugate gradient approach performs with about the same computational efficiency and speed of convergence as the Newton−Raphson version.

RESULTS AND DISCUSSION

Algorithm Validation. To validate the numerical functionality of the S_{rel} minimization algorithm, our initial efforts focus on performing a test run that consists of recovering known

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

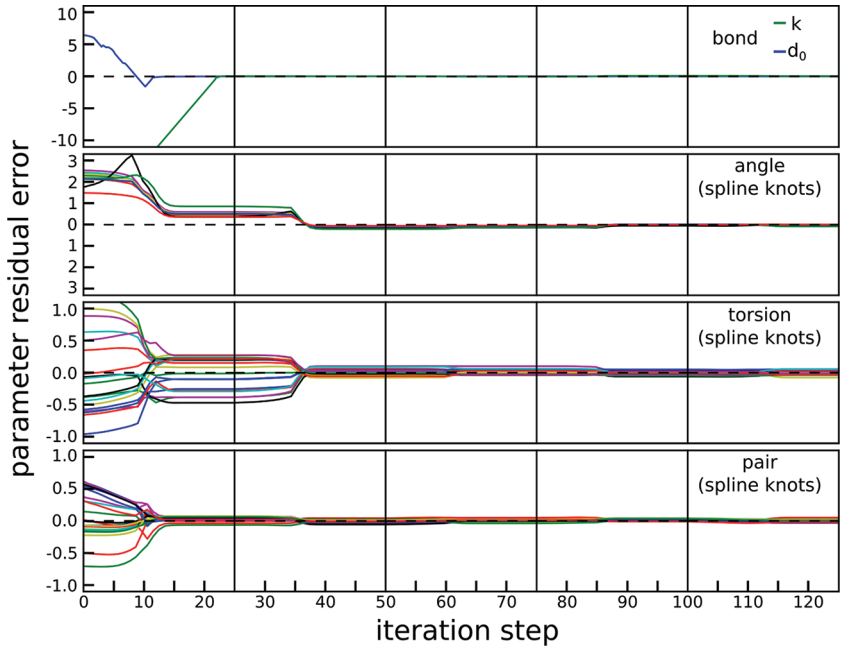


Figure 2. Convergence behavior of all free potential parameters in a test single bead polymer. Each colored line indicates the value of the residual of a free parameter for the given potential about its true value. Tighter grouping about the black dashed zero line indicates better convergence. The units of the parameter residual errors are as follows: bond force constant, k (kcal/mol Å^2), bond equilibrium distance d_o (Å), spline knots for angle, torsion, and pair (kcal/mol).

"seed" parameters from a constructed, already coarse target system. While this is not a test of coarse-graining, it does address issues of numerical stability and fundamental issues associated with relative entropy minimization. This toy model is a 15-bead linear homopolymer with bond stretching, angle bending, torsional, and nonbonded pair interactions. Cubic splines are used for all of these potentials except for bond stretching, which employs a simple harmonic spring functionality. The 82 parameters in the model $\{\lambda_i\}$ include the equilibrium distance and force constant of the bonded interaction, and the spline knots for each of the remaining potential terms. The seed potentials employed are the ones that are optimized for a single bead CG model of $(Ala)_{15}$ and are depicted in the Figure S1 of the Supporting Information. We use an equilibrium molecular dynamics trajectory of this model as the *target* for relative entropy minimization, and pick an initial *model* CG system with potential parameters all initialized at zero (except for the bond distance potential, which requires a finite equilibrium bond distance and force constant).

We initially choose to optimize the model potentials sequentially instead of simultaneously. This means that we minimize the relative entropy for only a subset of parameters, in stages, while the others are held fixed, to ensure stability and a smooth approach to the minimum. We optimize the potentials first in order of bond, angle, torsion, and pair; subsequently, we optimize in a simultaneous manner where all of the potential functions are further adjusted in parallel. To examine the convergence of the system during successive optimizations, we repeat the simultaneous optimization four times. Importantly, at least one simultaneous optimization is required to capture cross-correlation effects of different potential terms on each other.

The optimization routine in this test recovers the seed parameters within high numerical tolerance (e.g., ≤0.5% error). Figure 2 demonstrates the decay of the residuals between the model system's potential parameters and their respective target ("seed") values as the optimization progresses. This encouraging result shows that the optimization algorithm locates the correct minimum, and converges without inconsistencies or instabilities. The sequential optimizations are represented in the first block of each series (i.e., iteration steps less than 25), while the final four blocks give the four consecutive simultaneous optimizations. With each successive optimization, the parameters converge more closely to the seed values, but the system is within 99.5% of the correct solution by the end of the initial simultaneous optimization (iteration step 50).

This test optimization took ∼10 h to complete on a single 2.44 Ghz processor. A similar validation was performed using four simultaneous optimizations only (forgoing the initial sequential optimizations) and was found to converge, with similar accuracy, within 7 h. The speed of these optimizations is much faster than previous efforts by Shell and co-workers with a much simpler model[56] that took ∼2 weeks on an identical computer. In those previous efforts, the CG system had only 5 force field parameters (versus 82 here). This dramatic improvement in computational efficiency and speed is likely due to our use of the reformulated relative entropy minimization algorithm based on trajectory reweighting and full inversion of the Hessian.

Coarse Models of $(Ala)_{15}$. We proceed to generate CG models of a "real" peptide, polyalanine. This simple system minimizes the number of bead types and hence potentials. Moreover, alanine rich peptides (which are model amyloid systems) have been extensively studied both experimentally and in simulations.[12,19,38,43,46,64−71] We develop CG models for several length polyalanines, namely, $(Ala)_5$, $(Ala)_{10}$, and $(Ala)_{15}$. However, we focus our analysis on the $(Ala)_{15}$ system, which in AA simulations forms transient α-helices and β-hairpins as monomers and β-sheets as small oligomers, and therefore

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

presents a challenging balance of secondary structure behavior for our CG models to reproduce.

To generate the reference AA target trajectory of $(Ala)_{15}$, we use replica-exchange molecular dynamics (REMD). We employ the AA force field AMBER ff96[9] and an implicit water model based on the generalized born/solvent accessible surface area model.[72,73] This combination has been previously shown to capture the correct secondary structures of both α-helical and β-peptides.[74−76] These simulations use an Andersen thermostat, and 20 replicas spanning temperatures from 270 to 600 K, spaced exponentially. We perform 50 ns of REMD simulation before using a further 50 ns for trajectory generation (frames taken at 1 ps intervals). We only use trajectory data from the 330 K replica that is approximately at the simulation folding temperature. It is important for the reference trajectory to sample as much configuration space as possible in order to generate accurate statistics for the relative entropy minimization algorithm. The folding temperature is in some sense the ideal coarse-graining temperature, as the peptide will be continuously transitioning between folded and unfolded states.

We then perform our relative entropy optimization procedure using three separate model geometries that represent each amino acid residue using one, two, and three beads, respectively, where the CG beads are mapped to the centroid positions of the atom groups shown in Figure 3. The models

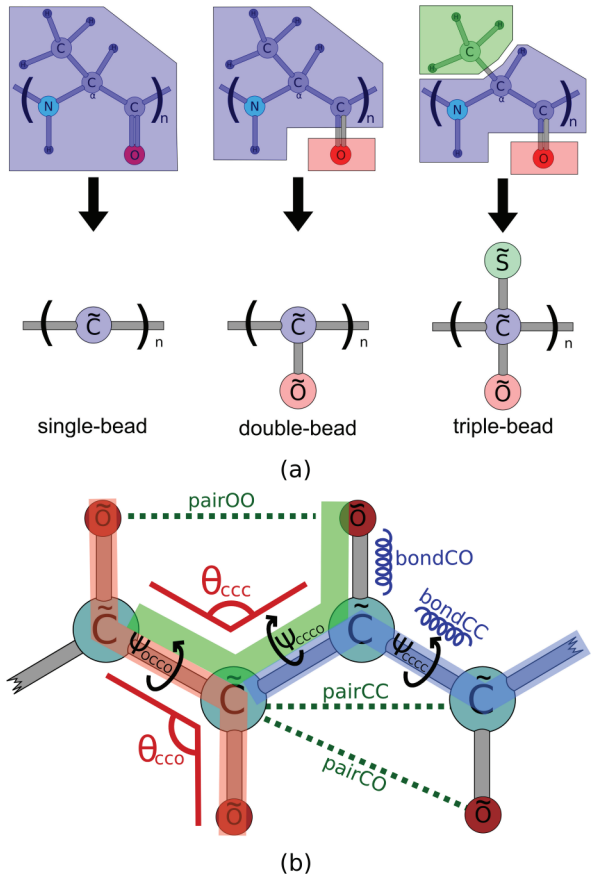


Figure 3. (a) Several CG model geometries for $(Ala)_n$. (b) All of the distinct potentials that can be applied to a two bead per amino acid CG model of this peptide.

contain 82, 224, and 444 force field parameters, respectively. We choose to represent the carbonyl oxygen as an independent site in the double and triple bead models in order to allow hydrogen-bonding-like interactions to be expressible; likewise, the single bead model should not be capable of stabilizing an α-helix through nonbonded interactions.

Algorithm Convergence. Using a single 2.44 GHz processor, the optimization runs reach stable convergence within 27, 138, and 648 h for the single, double, and triple bead models, respectively. For each CG model, all potentials were optimized simultaneously. The convergence times appear to scale superlinearly with the number of parameters in the model. It is important to note that the initial potential parameters in each case are zeroed, except for the bond parameters and a repulsive core made of a linear ramp in nonbonded pair potentials that ensures sterics. Better initial guesses—based for example on inverted Boltzmann distributions—would require significantly less convergence time. One contributor to the run times is the length of the simulations of the CG models. In particular, we find that increased CG model equilibration periods are necessary for more detailed models to avoid significant oscillations in parameter convergence (described more below). REMD simulations to speed equilibration or use of graphical processing units would likely substantially decrease the reference CG model run times.

One method of assessing the stability of the algorithm is to inspect the behavior of the model parameters as a function of iteration step. The representative set of such data shown in Figure 4 provides a visual record of model parameters as the

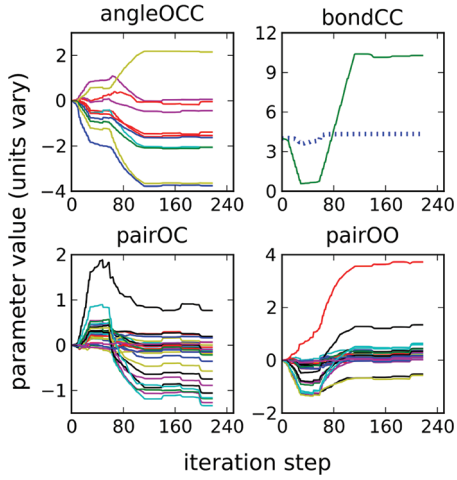


Figure 4. The evolution of model parameters for a double bead model of $(Ala)_{15}$. The units of the parameter values (y-axis) are kcal/mol for all of the panels except for "bondCC", which has the following units: green solid line. kcal/mol Å^2; blue dashed line, Å.

optimization routine guides them toward their converged values. Relatively smooth convergence behavior indicates good stability in this case. In particular in the graph labeled "bondCC", we observe that the bond force constant parameter (green solid line) initially moves away from its optimum value, before reversing direction later in the optimization and converging to a higher value than its initial. This behavior can be attributed to the relative entropy choosing an optimum path through parameter space. The path first passes through a lower bond force constant that permits the system to sample a wider range of bond lengths, at which point the equilibrium bond distance can then be optimized before the force constant then returns to a high value.

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

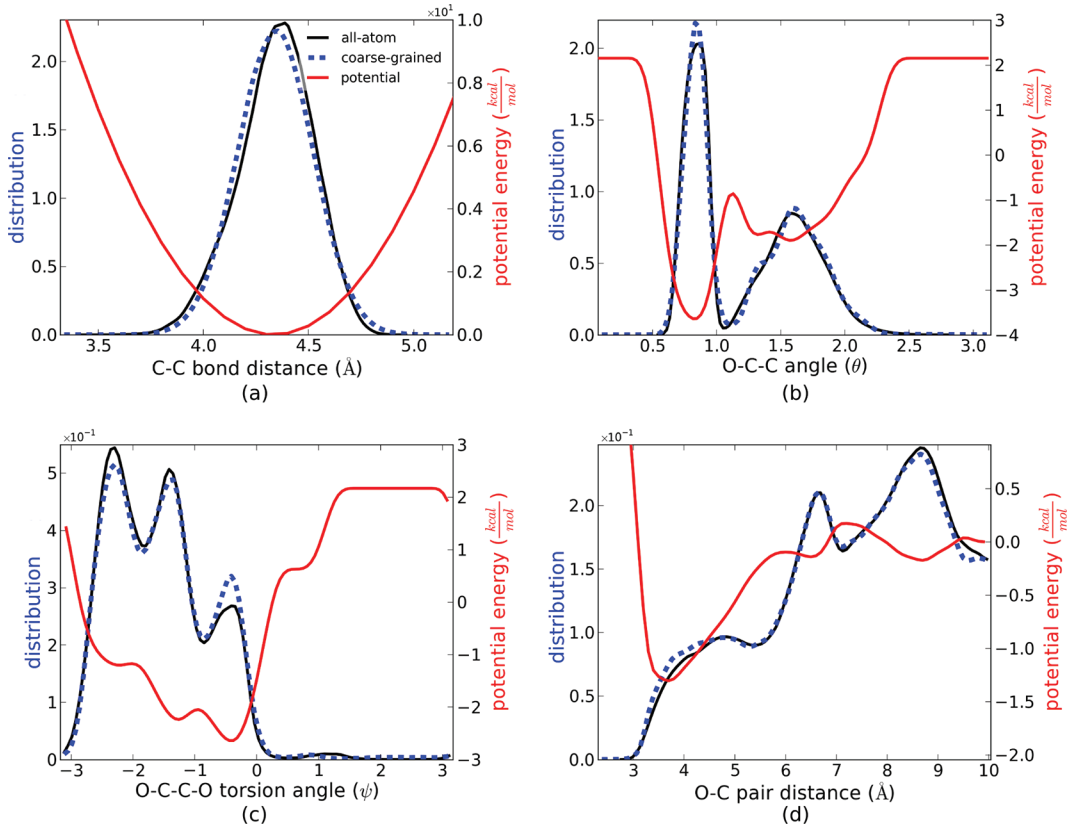


Figure 5. Representative structural correlations of a double bead per amino acid CG model of $(Ala)_{15}$ and a corresponding AA model. Shown in red (right axes) are the optimized CG potentials.

While there appear to be random fluctuations in the data near the ends of each graph, these fluctuations appear to occur about the optimum parameter values, and each successive optimization round likely brings the mean error (relative to the values consistent with a potential of mean force expressed in the CG potential basis set) of each potential's parameters increasingly closer to zero. Similar behavior is observed in the model validation convergence data (Figure 2).

Evaluating the Quality of Coarse-Grained Peptide Models. We employ a variety of analyses to evaluate how accurately the CG models reproduce ("replicate") AA behaviors. Although we analyze all three CG models, for brevity, we only provide a detailed account of the double bead model, while providing more selected analyses for the others. (Detailed results for all models, including potential parameters, can be found in Figures S1−S3 in the Supporting Information.)

Structural Correlations. By the theoretical considerations described in the Methods section, we expect to achieve near perfect replication of the bond angle, torsion angle, and nonbonded pair distance distributions, since each of these has a corresponding highly tunable spline function in the CG force field. That is, the distribution of these angle and distance coordinates should be equivalent in two different ensembles: (1) the reference AA system, in which these coordinates are projected, coarse observables of configurations under the action of the AA force field; and (2) the relative-entropy optimized CG model, in which these coordinates are under the action of the CG force field. Simulations of the AA and optimal CG systems give normalized histograms of these structural correlations for comparison. Figure 5 shows a representative sample (4 of 10 such histograms) for the two bead per amino acid $(Ala)_{15}$ model. When all of the data are considered, including those for metrics not shown, the histograms agree within 6%, evaluated as the cumulative residual error of all histogram bins. These results are quite encouraging and further indicate that our algorithm is properly implementing the relative entropy theory, as any systematic breakdown would result in disagreement between the distributions.

Macromolecular Characteristics. Further tests involve examining behaviors that the CG force field does not influence directly. We examine two such macromolecular attributes, the radius of gyration and the root-mean-squared (rmsd) deviation from an α-helical conformation. Both depend on the influence of all of the model's potentials acting in concert, and neither was explicitly or directly trained to AA distributions, so to speak. It is important to note that, for a fair comparison, the AA distributions must be extracted from coarse coordinates; i.e., these are constructed by first projecting each AA trajectory frame onto the CG geometry and then evaluating the metric (although the actual distributions do not vary noticeably among the three different CG models). Figure 7a shows that our model correctly predicts the upper and lower bounds of the radius of gyration, which seems relatively impressive. In between these values, it captures the detailed form of the distribution only qualitatively, however. The distributions are likely strongly affected by the temperature at which the optimization was performed (330 K). Lower temperature optimizations will probably emphasize folded states and thus shift which parts of the distribution match.

The results for the rmsd from an ideal α-helical structure show that the CG model does not sample configurations with an rmsd below ~4 Å, and therefore does not appear to fold into

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

a full α-helix but rather attains only states with partial helicity. It is possible that this could be attributed to a lack of adequate sampling of phase space, since in the AA system the full α-helix structure is only rarely sampled. It may also stem from the inability of such a coarse model to capture the cooperativity required to form a stable, full helix. For the rmsd region spanning 4−10 Å, the cumulative residual error is 30% of the total area under the AA curve, although the qualitative shape of the CG distribution is in excellent agreement for this range. The model correctly predicts the location of the peak, for example. This peak, which involves condensed structures with at most partial helicity, is consistent with the radius of gyration distribution, which is shifted to lower (globular) values for the CG model.

Model Transferability. We also address the transferability of the CG model to conditions beyond those at which the coarse-graining is performed, namely, different temperatures. Good temperature transferability in particular would be suggestive that we are capturing realistic coarse interactions that will scale to larger simulations of self-assembly. Achieving this is particularly challenging due to the balance of enthalpic and entropic interactions contributing to folding behavior that shifts dramatically with temperature and peptide chain length.

Figure 6 shows the computed folding curves (radius of gyration vs temperature) for the (projected) AA and CG

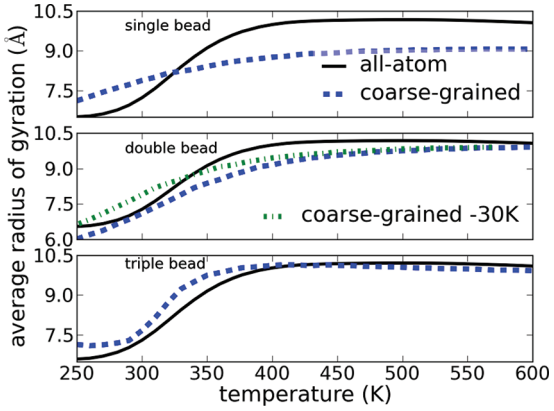


Figure 6. Comparison of folding curves for the single, double, and triple bead per amino acid CG models.

models. The double bead model is capable of qualitatively reproducing the folding behavior of the AA system. The folding temperatures agree to within 30 K, which might suggest that the CG balance of entropic and enthalpic driving forces is realistic. The double bead model, though, systematically underpredicts the radius of gyration (Rg), which is consistent with the fact that its value at the folding temperature, where it was parametrized, is off by ~0.5 Å. If the double bead CG Rg is uniformly shifted left to match the AA one at this temperature, the overall agreement appears much better. Even so, it is obvious that folding is less cooperative in the double bead model, as indicated by the shallow slope of its folding curve. This is not terribly surprising given that folding cooperatively is likely contingent on the detailed many-atom interactions that are only roughly approximated in this heavily coarse-grained model.

Comparison of Different Levels of Model Detail. Analyses of the kind described above are also useful for evaluating CG models with different levels of detail, i.e., different numbers of pseudoatoms and degrees of freedom. Ultimately, one desires models at some optimal level of detail that balances efficiency and accuracy. It is tempting to consider the (minimized) value of the relative entropy itself for each of these models, which would provide something of a single equilibrium "scoring metric" or fitness. Indeed, past work suggests that many CG model errors are often tightly correlated with S_{rel}.[58] However, calculation of the absolute relative entropy remains an expensive numerical task due to its dependence on system free energies. Instead, here we simply examine structural metrics as before.

Figure 6 compares the folding curves of the single, double, and triple bead (per amino acid) CG models to that of the original all atom system. Notably, the slope of each curve near its folding temperature increases with increasing model detail, suggesting that more cooperatively folding models emerge as more detail is incorporated. Interestingly, the triple bead model appears to well-reproduce the folding curve at both the low and high temperature extremes; in between, it is slightly *more* cooperative than the original AA system.

We also evaluate the radius of gyration distributions for these models (Figure 7a). All of the CG models have the same

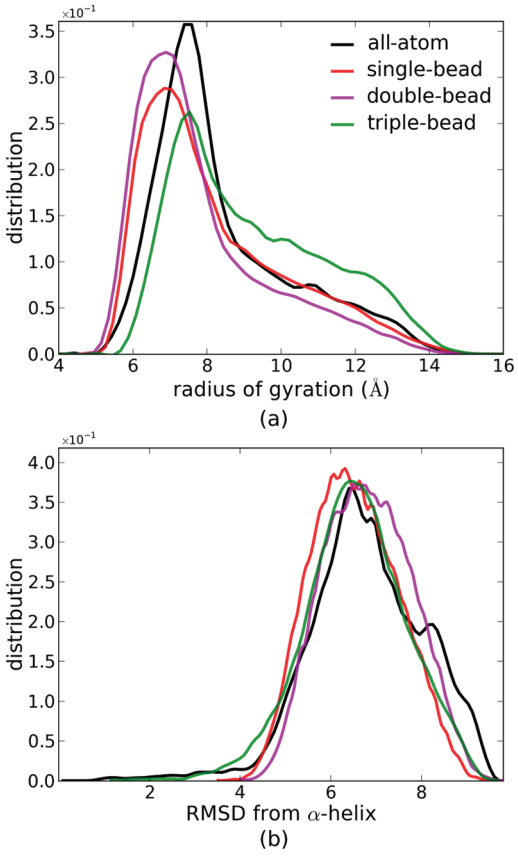


Figure 7. (a) Comparison of the radius of gyration distributions for the AA, single, double, and triple bead models of $(Ala)_{15}$ at 330 K. (b) A comparison of the rmsd from α-helix distributions for the AA, single, double, and triple bead models of $(Ala)_{15}$ at 330 K. For both graphs, a single reference all-atom distribution is shown, as it does not change appreciably with different projections (single, double, triple bead) onto the coarse models.

qualitative form as the original AA one, and their quantitative accuracy can be evaluated using the cumulative residual error of the distributions (relative to the AA one). Interestingly, the

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

single bead model has the lowest residual error (25.0%), while the double and triple bead models have slightly higher residuals (38.5 and 38.0%, respectively). Of course, there is no guarantee that models with more detail and lower relative entropy values should recapitulate this property better, as Rg is just one reduced structural metric characterizing the entire ensemble distribution. For the same reason, one should not consider models better in reproducing Rg as also better in replicating other properties.

Both the single and double bead models' Rg distribution peaks occur at a lower Rg than the all atom system's (~0.75 Å or ~10% relative error). The location of the triple bead model's peak is much closer to the all atom one (~0.11 Å or ~1.5% relative error). These results may in fact suggest that the single and double bead systems populate low Rg conformations that differ in structural character compared to those populated by the AA system, while the triple bead system has similar low Rg structures but rebalances their populations as compared to higher ones.

Analysis of each system's distribution of rmsd values from an α-helical structure shed some light on this result (Figure 7a). While all three models qualitatively capture the peak of the rmsd distributions, only the triple bead model is able to form full α-helical structures (indicated by the rmsd population <2 Å). Therefore, it is likely that the single and double bead models transition between partially helical/random coil configurations (low Rg) and extended unfolded structures, while the triple bead model transitions between a full α-helix (Rg of ~7.2 Å) and an extended unfolded structure. The effectively increased folding cooperativity of the triple bead model therefore likely minimizes forays into "off pathway" partial helical/random coil configurations.

Perhaps the most informative mode of analysis to compare the structural propensities of these models is free energy surfaces of helical propensity and Rg. As shown in Figure 8, all

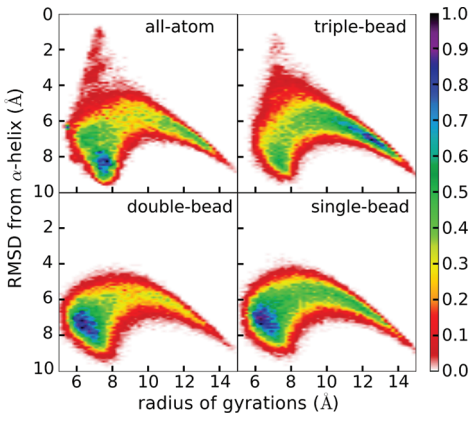


Figure 8. Free energy surfaces spanning the dimensions of Rg and helical propensity for the four models considered at 330 K.

three CG models capture the general shape of the AA free energy surface. Notably, the topology of the free energy basins that these CG models explore is substantially different than that of a freely jointed chain (see Figure S5 in the Supporting Information). Of the three CG models, only the triple bead one forms α-helical structures (basin centered near 1.0 rmsd and 7.2 Rg) and β-hairpin structure (basin centered near 8.5 rmsd and 7.8 Rg). This result is clearly manifested in the free energy surface for the triple bead model near its folding temperature of 320 K (see Figure S4 in the Supporting Information).

Large-Scale Simulations. As a second test of this new multiscale approach, we examine how the CG models perform in self-assembly simulations. As a preliminary study along these lines, we inserted 25 copies of a double bead model for $(Ala)_{15}$ into a periodic simulation box at 300 K (physiological temperature) and allowed them to aggregate in a molecular dynamics simulation. We chose a fairly high number density (0.0035 sites/Å^3) in order to decrease the mean free path of each peptide, and thus speed the aggregation process. We found that CG peptides quickly formed an amorphous phase, followed by internal reorganization to a β-like structure. In this structure, individual peptides are aligned in a coplanar arrangement with pleats propagating along their backbones and hydrogen bonding sites located between the peptides. Figure 9 shows a snapshot of this simulation at later times.

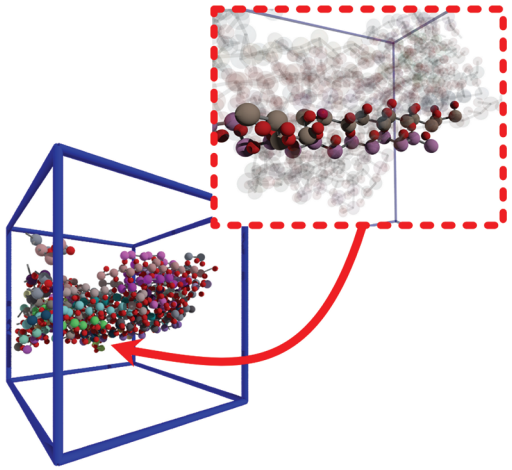


Figure 9. Twenty-five double bead models of $(Ala)_{15}$ assemble into a β-sheet-like structure. The inset highlights two of the peptides in a β-like structure.

There is a clear pattern of coarse "hydrogen bonds" formed between strands, due to the oxygen atom that is maintained separately from the remaining atoms in each coarse site. There is also a clear β-pleat along the assembled sheets.

Figure 10a shows a map of residue−residue contacts in a smaller five-peptide system, from one time-point late in the simulation. A contact is defined when two nonbonded CG sites are within 5 Å of separation distance. There are many signatures of β-like parallel and antiparallel chain alignments between different peptides, manifesting as contact groups along one of the diagonals. While other more detailed polyalanine CG models form β motifs,[19,38,70] it is particularly remarkable that this very simple double bead model can aggregate to such ordered structures. It is even more surprising when one considers that the many secondary structure motifs in the reference AA simulation are helical in nature. These stabilizing interpeptide contacts form rapidly, as there is a marked increase in the total number of interpeptide contacts with simulation time step (Figure 10b). While the dynamics of the CG model do not incorporate viscous effects due to solvent or the removed degrees of freedom and while the exact time scale of a simulation time step is not yet clear, it is obvious that the coarse model enables simulation of what would be very long time scale events had an AA approach been used. We leave a more

dx.doi.org/10.1021/jp2114994 | J. Phys. Chem. B 2012, 116, 8383−8393

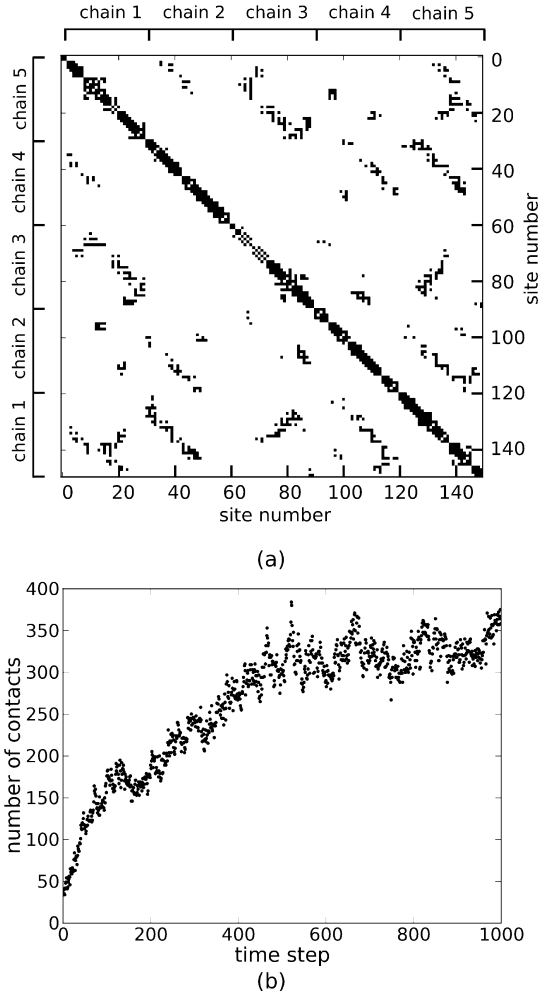


Figure 10. (a) Contact map depicting peptide chains that are in contact (defined as sites separated by ≤5 Å). (b) Time series of the total number of contacts in a simulation.

detailed examination of the aggregation properties of these models for a future study.

CONCLUSIONS

In this work, we extended the relative entropy coarse-graining methodology originally developed by Shell and co-workers,[55,56,59] using a trajectory reweighting approach. This new numerical coarse-graining approach affords the development of complex CG models with high dimensional parameter spaces. Importantly, any equilibrated AA simulation can be used for this procedure, including those with different force fields or explicit solvent models. We validated this method first by quantitatively recovering known seed potential parameters from a simple homopolymer model. Subsequently, we developed several coarse models of polyalanine varying in detail. We found that the coarse-graining procedure exhibited stable convergence behavior, and required times that scaled with the level of model detail. We further showed that the coarse polyalanine models all qualitatively and in many cases quantitatively reproduced AA system properties, including distance and angle correlation functions, radius of gyration and helicity distributions, and folding curves. Our analysis found an increase in folding cooperativity and ability to capture full α-helical structures with increasing model detail.

We also performed a preliminary large-scale simulation containing many copies of a polyalanine model optimized with two beads per amino acid, one backbone, and one hydrogen-bonding site. These molecular dynamics simulations revealed aggregation and arrangement into a highly ordered β-like structure. Impressively, the β character emerges implicitly and naturally through the coarse-graining procedure, despite the brute simplicity of the model and the fact that the reference AA runs used for coarse-graining show also helical structure. These features suggest that models developed using this technique may prove quite accurate and thus valuable to molecular mechanics studies of many biomolecular systems. The approach may be applied to any peptide sequence, which may lead to models that give insights into sequence-specific assembly mechanisms and driving forces. Moreover, this coarse-graining approach seems generally applicable to many complex molecules, including nucleic acids, synthetic polymers, or small molecules. For heterogeneous systems beyond the simple homopolymer case here, a straightforward extension of our coarse-graining approach would be to use simulations of associating dimers as an all-atom reference (rather than a single peptide chain).

ASSOCIATED CONTENT

Supporting Information

Coarse-grained potential parameters for the single, double, and triple bead models of $(Ala)_{15}$. Additional free energy surfaces for the AA and CG models, as well as a freely jointed chain model. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: shell@engineering.ucsb.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We greatly appreciate the support of the National Science Foundation (Award No. CBET-0845074).

REFERENCES

(1) Scheibel, T.; Parthasarathy, R.; Sawicki, G.; Lin, X.; Jaeger, H.; Lindquist, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4527−4532.
(2) Reches, M.; Gazit, E. *Science* **2003**, *300*, 625−627.
(3) Zhang, S.; Holmes, T. C.; DiPersio, C. M.; Hynes, R. O.; Su, X.; Rich, A. *Biomaterials* **1995**, *16*, 1385−1393.
(4) Holmes, T. C.; de Lacalle, S.; Su, X.; Liu, G.; Rich, A.; Zhang, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6728−6733.
(5) Ellis-Behnke, R. G.; Liang, Y.; You, S.; Tay, D. K. C.; Zhang, S.; So, K.; Schneider, G. E. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5054−5059.
(6) Meng, H.; Chen, L.; Ye, Z.; Wang, S.; Zhao, X. *J. Biomed. Mater. Res., Part B* **2009**, *89B*, 379−391.
(7) Tycko, R. *Annu. Rev. Phys. Chem.* **2001**, *52*, 575−606.
(8) Yan, X.; Zhu, P.; Li, J. *Chem. Soc. Rev.* **2010**, *39*, 1877.
(9) Gunsteren, W. F. v. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; Springer: 1997; Vol. 3, pp 83−96.
(10) Auer, S.; Kashchiev, D. *Phys. Rev. Lett.* **2010**, *104*, 168105.
(11) Reches, M.; Gazit, E. *Curr. Nanosci.* **2006**, *2*, 105−111.
(12) Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G. *Biophys. J.* **2009**, *96*, 5020−5029.

(13) Zhang, S.; Altman, M. *React. Funct. Polym.* **1999**, *41*, 91−102.
(14) Jahn, T. R.; Makin, O. S.; Morris, K. L.; Marshall, K. E.; Tian, P.; Sikorski, P.; Serpell, L. C. *J. Mol. Biol.* **2010**, *395*, 717−727.
(15) Török, M.; Milton, S.; Kayed, R.; Wu, P.; McIntire, T.; Glabe, C. G.; Langen, R. *J. Biol. Chem.* **2002**, *277*, 40810−40815.
(16) Lomakin, A.; Chung, D. S.; Benedek, G. B.; Kirschner, D. A.; Teplow, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1125−1129.
(17) Knowles, T. P. J; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. *Science* **2009**, *326*, 1533−1537.
(18) Zhang, J.; Muthukumar, M. *J. Chem. Phys.* **2009**, *130*, 035102.
(19) Nguyen, H. D.; Hall, C. K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16180−16185.
(20) Nguyen, H. D.; Hall, C. K. *J. Biol. Chem.* **2005**, *280*, 9074−9082.
(21) Pellarin, R.; Caflisch, A. *J. Mol. Biol.* **2006**, *360*, 882−892.
(22) Sipe, J. D.; Cohen, A. S. *J. Struct. Biol.* **2000**, *130*, 88−98.
(23) Guijarro, J. I.; Sunde, M.; Jones, J. A.; Campbell, I. D.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4224−4228.
(24) Chiti, F.; Webster, P.; Taddei, N.; Clark, A.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3590−3594.
(25) West, M. W.; Wang, W.; Patterson, J.; Mancias, J. D.; Beasley, J. R.; Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11211−11216.
(26) López de la Paz, M.; Serrano, L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 87−92.
(27) López de la Paz, M.; Goldie, K.; Zurdo, J.; Lacroix, E.; Dobson, C. M.; Hoenger, A.; Serrano, L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16052−16057.
(28) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.
(29) Jahn, T. R.; Radford, S. E. *Arch. Biochem. Biophys.* **2008**, *469*, 100−117.
(30) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. *Nature* **2003**, *424*, 805−808.
(31) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 111−116.
(32) Irbäck, A.; Mitternacht, S. *Proteins* **2008**, *71*, 207−214.
(33) Wu, C.; Shea, J. *Curr. Opin. Struct. Biol.* **2011**, *21*, 209−220.
(34) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.
(35) Pellarin, R.; Guarnera, E.; Caflisch, A. *J. Mol. Biol.* **2007**, *374*, 917−924.
(36) Bellesia, G.; Shea, J. *J. Chem. Phys.* **2009**, *130*, 145103.
(37) Bellesia, G.; Shea, J. *J. Chem. Phys.* **2009**, *131*, 111102.
(38) Nguyen, H. D.; Hall, C. K. *Biophys. J.* **2004**, *87*, 4122−4134.
(39) Go, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183−210.
(40) Caflisch, A. *Curr. Opin. Chem. Biol.* **2006**, *10*, 437−444.
(41) Friedman, R.; Pellarin, R.; Caflisch, A. *J. Mol. Biol.* **2009**, *387*, 407−415.
(42) Thirumalai, D.; Klimov, D.; Dima, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 146−159.
(43) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. *Biophys. J.* **2007**, *92*, 4289−4303.
(44) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.
(45) *Coarse-graining of condensed phase and biomolecular systems*; Voth, G. A., Ed.; CRC Press: 2009; p 85.
(46) Thorpe, I. F.; Zhou, J.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 13079−13090.
(47) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045−15048.
(48) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192−198.
(49) Izvekov, S.; Chung, P. W.; Rice, B. M. *J. Chem. Phys.* **2010**, *133*, 064109.
(50) Thorpe, I. F.; Goldenberg, D. P.; Voth, G. A. *J. Phys. Chem. B* **2011**, *115*, 11911−11926.
(51) Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(52) Praprotnik, M.; Site, L. D.; Kremer, K. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545−571.
(53) Lyubartsev, A. P. *Eur. Biophys. J.* **2005**, *35*, 53−61.
(54) Sun, Q.; Faller, R. *J. Chem. Theory Comput.* **2006**, *2*, 607−615.
(55) Shell, M. S. *J. Chem. Phys.* **2008**, *129*, 144108-7−.
(56) Chaimovich, A.; Shell, M. S. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1901−1915.
(57) Hammer, M. U.; Anderson, T. H.; Chaimovich, A.; Shell, M. S.; Israelachvili, J. *Faraday Discuss.* **2010**, *146*, 299−308.
(58) Chaimovich, A.; Shell, M. S. *Phys. Rev. E* **2010**, *81*, 060104.
(59) Chaimovich, A.; Shell, M. S. *J. Chem. Phys.* **2011**, *134*, 094112.
(60) Noid, W. G.; Chu, J.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
(61) *Free energy calculations: theory and applications in chemistry and biology*; Chipot, C., Pohorille, A., Eds.; Springer: 2007; pp 92−112.
(62) *Understanding Molecular Simulation: From Algorithms to Applications*, 1st ed.; Frenkel, D., Smit, B., Eds.; Academic Press, Inc.: 1996; pp 183−189.
(63) Piegorsch, W. W.; Casella, G. *SIAM Rev.* **1989**, *31*, 428−434.
(64) Soto, P.; Baumketner, A.; Shea, J. *J. Chem. Phys.* **2006**, *124*, 134904.
(65) Forood, B.; Perezpaya, E.; Houghten, R. A.; Blondelle, S. E. *Biochem. Biophys. Res. Commun.* **1995**, *211*, 7−13.
(66) Rousseau, F.; Schymkowitz, J.; Serrano, L. *Curr. Opin. Struct. Biol.* **2006**, *16*, 118−126.
(67) Measey, T. J.; Smith, K. B.; Decatur, S. M.; Zhao, L.; Yang, G.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2009**, *131*, 18218−18219.
(68) Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954−965.
(69) Ravikumar, B. *Hum. Mol. Genet.* **2002**, *11*, 1107−1117.
(70) Nguyen, H. D.; Marchut, A. J.; Hall, C. K. *Protein Sci.* **2004**, *13*, 2909−2924.
(71) Friedman, R.; Pellarin, R.; Caflisch, A. *J. Phys. Chem. Lett.* **2010**, *1*, 471−474.
(72) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(73) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383−394.
(74) Shell, M. S.; Ritterson, R.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 6878−6886.
(75) Lin, E.; Shell, M. S. *J. Chem. Theory Comput.* **2009**, *5*, 2062−2073.
(76) Gee, J.; Shell, M. S. *J. Chem. Phys.* **2011**, *134*, 064112.