JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

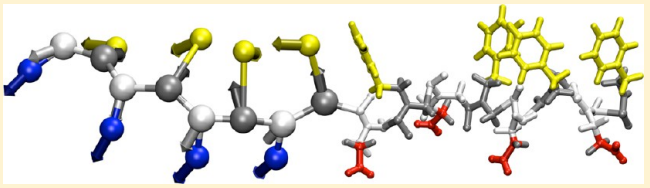
Modeling Sequence-Specific Polymers Using Anisotropic Coarse-Grained Sites Allows Quantitative Comparison with Experiment

Thomas K. Haxton,* Ranjan V. Mannige, Ronald N. Zuckermann, and Stephen Whitelam*

Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: Certain sequences of peptoid polymers (synthetic analogs of peptides) assemble into bilayer nanosheets via a nonequilibrium assembly pathway of adsorption, compression, and collapse at an air−water interface. As with other large-scale dynamic processes in biology and materials science, understanding the details of this supramolecular assembly process requires a modeling approach that captures behavior on a wide range of length and time scales, from those on which individual side chains fluctuate to those on which assemblies of polymers evolve. Here, we demonstrate that a new coarse-grained modeling approach is accurate and computationally efficient enough to do so. Our approach uses only a minimal number of coarse-grained sites but retains independently fluctuating orientational degrees of freedom for each site. These orientational degrees of freedom allow us to accurately parametrize both bonded and nonbonded interactions and to generate all-atom configurations with sufficient accuracy to perform atomic scattering calculations and to interface with all-atom simulations. We have used this approach to reproduce all available experimental X-ray scattering data (for stacked nanosheets and for peptoids adsorbed at air−water interfaces and in solution), in order to resolve the microscopic, real-space structures responsible for these Fourier-space features. By interfacing with all-atom simulations, we have also laid the foundation for future multiscale simulations of sequence-specific polymers that communicate in both directions across scales.



Many of the most important processes in molecular biology, including allostery,[1,2] enzyme catalysis,[3] molecular recognition,[4] protein homeostasis,[5] and nucleic acid metabolism,[6] involve the cooperative motion of large, precisely self-assembled[7] biomolecules. Engineering synthetic materials with similarly sophisticated functionality will require methods to relate the chemical sequence of large molecules (e.g., biomolecules or sequence-defined synthetic polymers) to their self-assembled form and function. These methods must span many orders of magnitude in time and space in order to describe atomically detailed interactions and correlated supramolecular motions, both of which contribute to materials assembly and function.

One class of nanomaterials that show promise as scaffolds for molecular recognition and catalysis[8] are peptoid nanosheets, solid bilayers that assemble from sequence-defined peptoid polymers[9] (positional isomers of peptides) due to a mechanical protocol that acts on many peptoids collectively (Figure 1).[10−13] Exposed to an air−water interface, amphiphilic peptoid polymers first adsorb from solution onto the interface, forming a structured monolayer. Subsequently compressing the monolayer past a certain pressure induces irreversible collapse into bilayer nanosheets 2.9 nm thick and up to 100 μm wide. Such a process is determined by mechanisms operating at multiple length scales: electrostatic interactions at the angstrom scale link side chains on neighboring polymers; amphiphilic patterning at the 1 nm scale allows for adsorption to the air−water interface; the motion of polymers on scales up to their full 10 nm length determines whether in-plane ordering occurs; and nanosheets can extend to scales of order 100 μm. The associated time scales range from picoseconds, for atomic and molecular fluctuations, to the seconds or minutes on which nanosheets are produced.

Developing a detailed, real-space picture of such a multiscale process requires a modeling approach able to account for mechanisms operating on a broad range of length and time scales. Studies of macromolecules such as nucleic acids and proteins have shown that coarse-grained modeling can, in principle, span scales efficiently, by representing explicitly only the most important molecular degrees of freedom, and representing implicitly other degrees of freedom via "effective" interactions.[14−24] However, reducing the number of degrees of freedom unavoidably discards information, so the resulting coarse-grained model cannot capture all aspects of the underlying all-atom system. For example, it has been shown generally that coarse-grained models parametrized to reproduce all-atom pair distribution functions cannot correctly reproduce thermodynamic properties such as energy and pressure, and vice versa.[25−27]

To mitigate this *representability* problem,[25−27] careful choices must be made in the two key aspects of a coarse-graining scheme: the choice of which degrees of freedom (or "sites") to be retained, and how interactions between sites should be parametrized. Most work has focused on the latter, resulting in

Received: September 6, 2014
Published: December 10, 2014

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

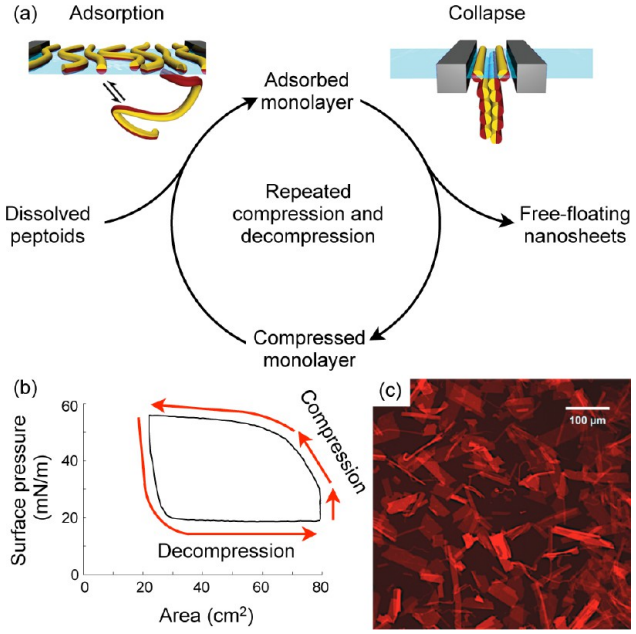


Figure 1. (a) Nanosheet production cycle. Exposed to an air−water interface, amphiphilic peptoid polymers adsorb from solution onto the air−water interface. Compressing the monolayer induces irreversible collapse into free-floating bilayer nanosheets. Decompressing the interface completes the cycle, allowing additional peptoids to adsorb. (b) Experimental plot of the surface pressure and surface area of the monolayer during one production cycle, illustrating the large hysteresis associated with the irreversible formation of nanosheets.[10−13] (c) Nile red fluorescence micrograph of a solution of free-floating nanosheets.[8]

the development of rigorous interaction-parametrization schemes to target particular features (typically distribution functions, forces, or energies) of related all-atom simulations[28−32] or to minimize the relative entropy between coarse-grained and all-atom ensembles.[33−36]

Fewer authors have investigated the choice of which degrees of freedom to retain. Coarse-grained models based on rigorous parametrization schemes have employed isotropic (spherically symmetric) interactions, for which these schemes are most tractable. However, it has been shown that the accuracy of isotropic coarse-grained models declines as the underlying all-atom system becomes more anisotropic, due to the fact that more information is lost when averaging spherically over anisotropic interactions than over isotropic ones.[37,38]

One way to improve the accuracy of a coarse-grained model with spherically symmetric sites is to increase the number of sites, a strategy often employed in protein modeling.[39−41] For example, by including between 3 and 8 sites per amino acid residue, the PRIMO protein model[41] can estimate all-atom configurations with such accuracy that all-atom configurations can be passed between PRIMO and atomistic models with 0.1 Å resolution,[42] allowing the model to seamlessly interface with all-atom force fields in multiscale simulations.[43] While such an approach clearly represents a protein in great detail, the large number of degrees of freedom represented means that simulations are only one order of magnitude faster than all-atom simulations.[41]

Here, we demonstrate an alternative strategy for combining accuracy and efficiency within coarse-grained modeling: retain only a minimal number of coarse-grained sites (two per monomer) but include fluctuating orientational degrees of freedom for each site. Coarse-grained modelers have recognized the importance of anisotropic interactions in biomolecules, having incorporated directional nonbonded interactions into models for proteins,[44−51] DNA,[52−55] and lipids.[56] Efforts in protein modeling have focused on capturing the directionality of backbone hydrogen bonding (absent in peptoids) by approximating the dipole−dipole interaction between peptide groups using only the positions of nearby alpha carbons on the backbone.[45−50] In addition, at least one model has accounted for the anisotropic shape of protein side chains by using ellipsoidal (but energetically isotropic) side chain sites,[44] and one model has accounted for dipole−dipole interactions between polar side chains.[51] Coarse-grained DNA models have focused more on directional nonbonded interactions, including separate base-pairing, base-stacking, and cross-stacking interactions depending on the relative orientation of interacting nucleobases.[52−55]

Motivated by the success of those strategies, and recognizing the importance of directional interactions and torsional conformations in peptoids, we created a model with directional interactions depending on *independently fluctuating* orientations, associating each site with both a position and an independent symmetry axis. As far as we know, the only previous uses of orientational degrees of freedom in coarse-grained biomolecule models are the protein model of Spiga, Alemani, Degiacomi, Cascella, and Dal Peraro, which includes rotating electric dipoles in polar side chains,[51] and the DNA model of Morriss-Andrews, Rottler, and Plotkin, which includes one soft orientational degree of freedom per base.[53] Relative to an isotropic model, including a symmetry axis increases the number of degrees of freedom per site from three to five but improves the accuracy of our model in two ways. First, it allows us to parametrize bonded and nonbonded interactions that incorporate atomic-level details like covalent-bond dihedral angle distributions and electric dipole interactions. Second, it allows us to estimate, or *backmap*, all-atom configurations with sufficient accuracy to perform detailed scattering calculations and interface with all-atom simulations.

We plan to use our model to investigate the dynamic, large-scale processes in the nanosheet production cycle (Figure 1): adsorption of solvated peptoid polymers to the air−water interface, ordering of the adsorbed monolayer, and collapse into a free-floating bilayer. In the current study, we establish that our model reproduces all known structural features of the equilibrium states involved in this production cycle, features measured by solution X-ray scattering of solvated polymers, grazing-incidence X-ray scattering of monolayers, and oriented X-ray scattering of stacks of bilayer nanosheets. In so doing, we provide a microscopic, real-space interpretation of these Fourier-space features. In addition, we show how our approach can interface with all-atom simulations with an accuracy comparable to computationally more demanding models that possess more degrees of freedom, laying the groundwork for efficient multiscale simulations[57−60] of sequence-specific polymers that communicate in both directions across scales.

MODEL AND METHODS

In this paper, we focus on the block-charge peptoid illustrated in Figure 2a−c, although we note that our coarse-graining scheme is transferable to different peptoid chemistries and is generalizable to proteins[61]. "Block-*n*" peptoids are (poly)-peptoids of the form $((\mathrm{Nae{-}Npe})_{n/4}{-}(\mathrm{Nce{-}Npe})_{n/4})$, built from two equal-length blocks, each consisting of nonpolar *N*-

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

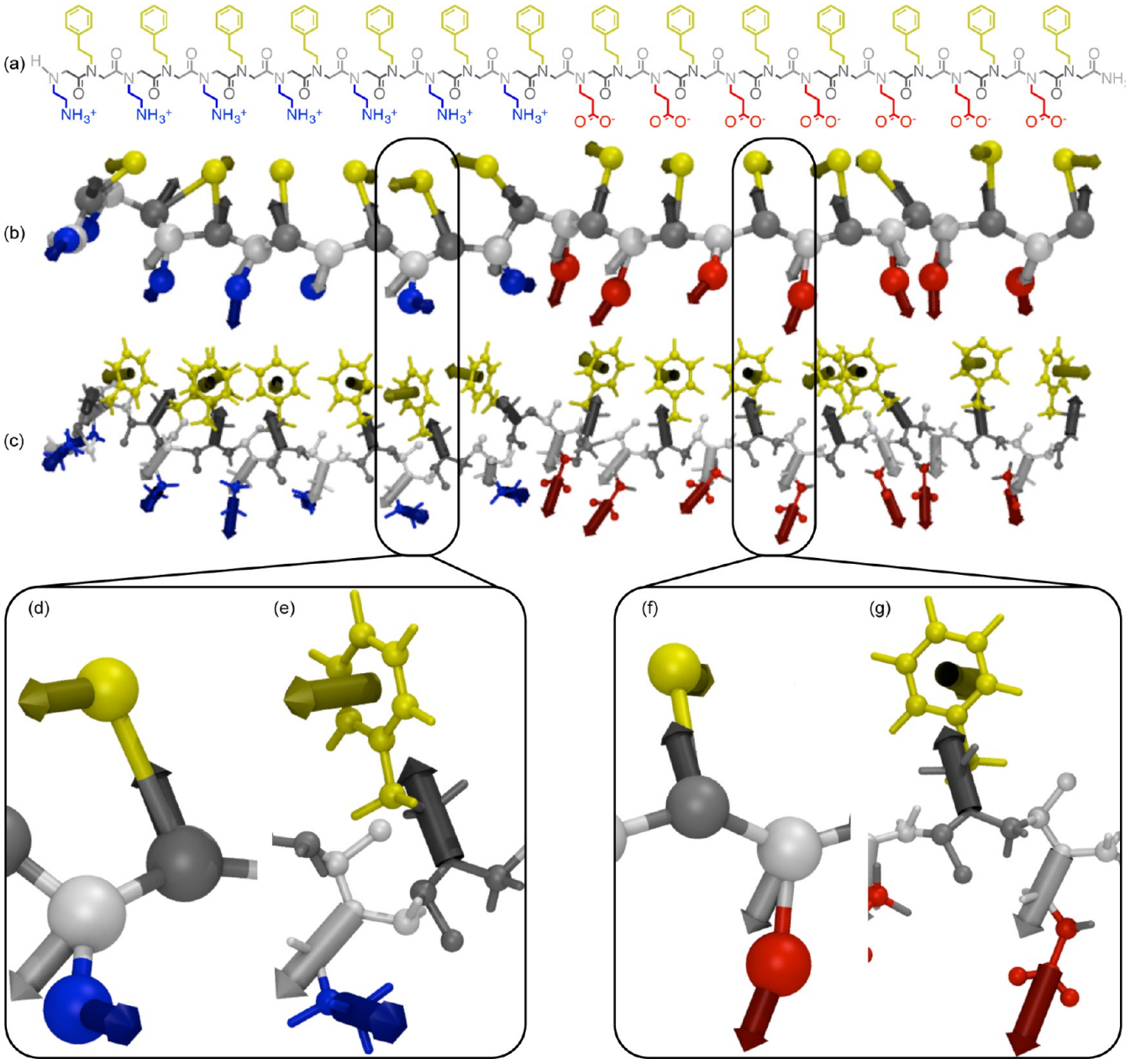


Figure 2. (a) Chemical diagram of a 28-monomer "block-28" peptoid polymer $(Nae\text{-}Npe)_7$−$(Nce\text{-}Npe)_7$. Blue and red parts of the diagram denote positively and negatively charged polar side chains, yellow parts denote nonpolar side chains, and grey parts denote the backbone. (b) Ball-and-stick representation of our coarse-grained peptoid model, taken from a single strand equilibrated within a monolayer. Blue, red, and yellow balls denote positions of the positively charged side chain sites, grey balls denote positions of the backbone sites, and arrows denote the symmetry axis of each site. (c) All-atom representation backmapped from the positions and (superposed) orientations of the coarse-grained sites. Atoms are colored according to the site with which they are associated. (d-e) Close-up view showing two monomers in the aminoethyl block in (d) coarse-grained and (e) all-atom representations. (f-g) Close-up view showing two monomers in the carboxyethyl bock in (f) coarse-grained and (g) all-atom representations.

(2-phenylethyl)glycine monomers alternating with charged monomers. The first block contains positively charged *N*-(2-aminoethyl)glycine monomers, and the second block contains negatively charged *N*-(2-carboxyethyl)glycine monomers. We investigated "block-*n*" peptoids of several lengths but present results primarily for block-28, for which the most extensive set of experimental data exists.

As illustrated in Figure 2b, we construct the *positions* of our coarse-grained model in the usual way, associating groups of atoms with coarse-grained sites and linking those sites with virtual or "coarse-grained" bonds.[14−24] We assign two coarse-grained sites per monomer, one for the backbone and one for the side chain. The resulting bonded network is a linear chain of *n* backbone sites, each branched with one side chain site, plus (optionally) an additional backbone site representing the relatively bulky carboxy-terminus. We associate the backbone sites with the backbone atoms of each monomer plus the first methylene bridge of the side chain. We associate the phenylethyl, aminoethyl, and carboxyethyl side chain sites with the remaining atoms of the side chains. With these associations, the backbone maps to the molecule *N*-

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

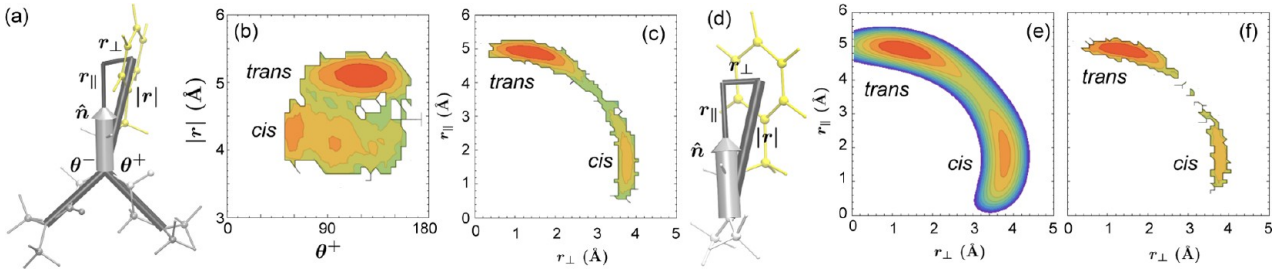


Figure 3. (a) Snapshot of a minimal peptoid dimer with three backbone nitrogens and one phenethylamine side chain, overlaid with virtual bonds between coarse-grained sites and an arrow denoting the orientation of the central backbone site. (b) Two-dimensional potential of mean force for two variables that would be treated separately in a typical coarse-grained model. (c) Two-dimensional potential of mean force for the variables treated collectively in our model to describe the position of the side chain site relative to the position and orientation of the backbone site. (d) Snapshot of the isolated peptoid monomer used to parametrize the side chain bonded interaction. (e) Coarse-grained potential of mean force for the relative position of the side chain site, parametrized to match the (f) all-atom potential of mean force, which agrees well with the analogous potential of mean force for the dimer (panel c).

methylacetamide and the side chain sites map to the "side chain analogs"[62] toluene, methylammonium, and acetate.

Our model distinguishes itself from previous coarse-grained models by defining each site $X_i = \{\vec{r}_i, \hat{n}_i\}$ by both a position $\vec{r}_i$ and a director $\hat{n}_i$ (arrows in Figure 2b) representing the dominant symmetry axis of the atoms associated with the site. For the backbone sites, defining $\vec{r}_i$ by the backbone N location and $\hat{n}_i$ by the first side chain ($N-C_\beta$) bond allows us to capture the softest covalent interactions (bond torsions) with effective interactions that account for the accumulated torsion along the backbone and side chain branches of the covalent network. For the phenylethyl side chain site, defining $\vec{r}_i$ by the center of the aromatic ring and $\hat{n}_i$ by its normal allows us to employ anisotropic nonbonded interactions that respect the aromatic ring's approximately axial symmetry. For the aminoethyl and carboxyethyl side chains, defining $\vec{r}_i$ by the center of mass of the associated atoms and $\hat{n}_i$ by the direction of the central bond ($C_\gamma-N_{amino}$ or $C_\gamma-C_{carboxy}$) aligns the site with the dipole moment of the associated atoms and leaves only hydrogens and the carboxy oxygens off the symmetry axis, allowing us to employ anisotropic nonbonded interactions capturing the dominant (monopolar and dipolar) electrostatic interactions among charged groups.

The principle of coarse-grained modeling lies in replacing a potentially exact many-body effective interaction between coarse-grained sites with a sum of few-body interactions that can be efficiently incorporated into molecular simulations.[14−24] We followed this principle by decomposing our effective interaction into bonded and nonbonded terms, as conventionally done, plus height-dependent solvation terms accounting for the preference of some molecular groups (such as the aromatic groups) to adsorb to the air−water interface. Height-dependent solvation terms are not common in coarse-grained models; instead, coarse-grained models used to study interfaces (such as the MARTINI lipid model[63]) typically employ explicit coarse-grained water particles. A height-dependent solvation interaction has been used previously to model surfactant adsorption onto a solid interface.[64] In future work, we plan to allow the shape of the air−water interface to fluctuate, including a fourth potential energy term to couple these fluctuations to the surface tension. We expect that including these fluctuations will be necessary to allow investigation of the mechanism by which the monolayer collapses into a bilayer.

We parametrized the potential energy function in two stages. First, we applied a "bottom-up" parametrization, performing direct Boltzmann inversion[65] to fit a set of effective potentials to all-atom potentials of mean force,

$$\frac{U^\alpha(\{R_i^\alpha\})}{k_B T} = -\ln\left(\int d\mathbf{r} \exp\left(-\frac{u(\mathbf{r})}{k_B T}\right) \prod_i \delta(\tilde{R}_i^\alpha(\mathbf{r}) - R_i^\alpha)\right) \quad (1)$$

where $\{R_i^\alpha\}$ is an orthogonal set of generalized coordinates describing a particular bonded, nonbonded, or solvation interaction U^α; $d\mathbf{r}$ are a set of atomic coordinates including water atoms; $\tilde{R}_i^\alpha(\mathbf{r})$ is the mapping from atomic coordinates to generalized coordinates; and $u(\mathbf{r})$ is an all-atom force field. We calculated the integral in eq 3 by conducting all-atom molecular dynamics simulations of the smallest molecular systems adequately constraining U^α.

Our use of effective potentials $U^\alpha(\{R_i^\alpha\})$ depending on more than one variable is not typical among coarse-grained models. Typically, coarse-grained models built from isotropic sites use single-variable effective potentials analogous to those found in all-atom force fields: bonded interactions summed to depend separately on bond lengths, bond angles, and dihedral angles, and spherically symmetric nonbonded interactions depending only on radial separations. Since all-atom distribution functions typically do not factorize with respect to these variables, direct Boltzmann inversion would not be expected to produce accurate effective potentials. Iterative fitting procedures like iterative Boltzmann inversion,[66] the multiscale coarse-graining method,[28−31] and the relative entropy method[33−36] are typically used instead. These methods use a series of coarse-grained simulations to converge to an optimized coarse-grained model. Their success relies on the ability of single-variable effective potentials to collectively yield accurate many-body distribution functions, despite the fact that these distribution functions do not factorize into a form commensurate with the single-variable terms of the potential energy function.

Figure 3 illustrates our approach for the bonded interaction between backbone and side chain sites, contrasting our approach with a typical approach for an isotropic coarse-grained model. Neglecting dihedral terms for simplicity, a typical coarse-grained model would write the interaction as a sum of terms each depending on only one geometric parameter (see Figure 3a),

$$U^{sc}(\{|r|, \theta_1, \theta_2\}) = U^r(|r|) + U^1(\theta_1) + U^2(\theta_2) \quad (2)$$

Figure 3a shows the minimal molecule needed to converge the potentials of eq 2: a dimer with three backbone nitrogens

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

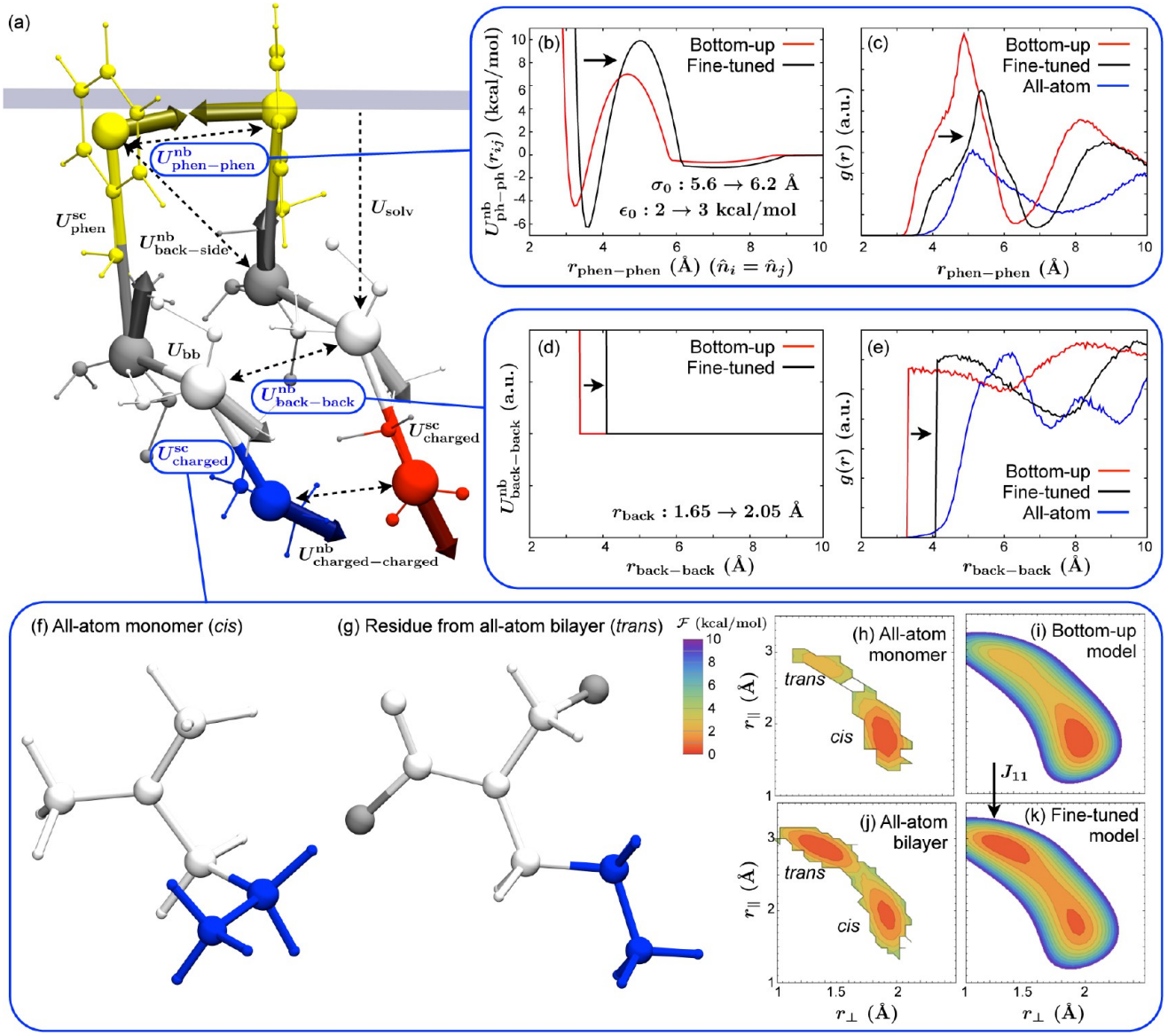


Figure 4. (a) Schematic illustrating the effective interactions of the model, labeled by the effective interaction functions defined in eqs 4−11. (b−k) Effect of fine-tuning four of the 115 parameters defining these functions. (b-c) Changing the length and energy scale of the phenethyl−phenethyl interaction from 5.6 Å and 2 kcal/mol to 6.2 Å and 3 kcal/mol resulted in (b) a shift in the interaction potential (shown here for parallel sites), greater agreement between simulated and experimental X-ray scattering spectra (see Supporting Information Methods), and (c) greater agreement of pair distribution functions for coarse-grained and all-atom simulations of peptoid bilayers (shown here for phenethyl side chain sites in the same bilayer leaf but different peptoid chains). (d-e) Changing the hard-core radius of the backbone sites (as shown in d) resulted in greater agreement between simulated and experimental X-ray scattering spectra (see Supporting Information Methods) and (e) greater agreement of pair distribution functions for coarse-grained and all-atom simulations (shown here for backbone sites in the same bilayer leaf but different peptoid chains). (f−k) Adjusting the parameter J_{11} in the side chain bonded interaction for aminoethyl side chains shifted the most favored configuration from the (d) *cis* configuration favored by isolated all-atom aminoethyl monomers to the (e) *trans* configuration favored by aminoethyl monomers in all-atom bilayers.

(three backbone sites) and one bulky side chain (one side chain site) branched from the central nitrogen. As shown in Figure 3b, the two-dimensional potential of mean force depending on θ_1 and $|r|$ does *not* factorize into θ_1 and $|r|$ terms: the basin of attraction at large $|r|$ (corresponding to an extended, *trans* side chain) is more narrowly distributed in θ than the basin attraction at small $|r|$ (corresponding to a contracted, *cis* side chain). Thus, one would not expect direct Boltzmann inversion of single-variable effective interactions to yield an accurate model. Instead, one would rely on an iterative procedure to converge targeted features of the model with those of the all-atom system. These procedures would rely on collective effects from other interactions, such as a long-range nonbonded interaction between the side chain site and the terminal backbone sites, to converge the model.

In our model, we write the bonded side chain interaction as a multiple-variable function that depends on the position *and orientation* of the bonded side chain and backbone sites but does *not* depend on positions of neighboring backbone sites. Considering only the side chain site's position for simplicity (ignoring terms introduced to control the side chain site's orientation), the interaction has the form

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

$$U^{sc}(r_\perp, r_\parallel) = k_1(((r_\perp - r_{\perp 0})^2 + (r_\parallel - r_{\parallel 0})^2)^{1/2} - r_0)^2 + \sum_{j=0}^{4} J_{1j}(\arctan(r_\parallel/r_\perp))^j \quad (3)$$

where $r_\parallel$ and $r_\perp$ are the components of the side chain separation parallel and perpendicular to the backbone director (Figure 3a) and the other parameters are constants. The first term in eq 3 describes an arc traced out by the side chain site as the middle of three side chain hydrocarbon bonds twists, and the second term in eq 3 describes the bimodal distribution of dihedral angles with minima corresponding to *cis* and *trans* configurations (Figure 3c). Dihedral rotation of the third hydrocarbon bond does not affect the side chain site's position (center of mass of the aromatic ring), and we model the first hydrocarbon bond as freely rotatable, motivated by the relatively weak dihedral interaction in our reference all-atom force field.[67] Thus, a simple pairwise interaction between oriented backbone and side chain sites captures the dominant atomic-level energetics of the underlying all-atom system. (Modeling the weak dihedral angle dependence for the first hydrocarbon bond could be achieved by using fully oriented sites[61] rather than sites with only one symmetry axis.)

Since the bonded interaction (eq 3) depends only on the position and orientation of the backbone and side chain sites in a single monomer, we parametrized the interaction by matching potentials of mean force (e.g., Figure 3e) to all-atom potentials of mean force (e.g., Figure 3f) for an isolated monomer dissolved in water (Figure 3d), using a CHARMM-based peptoid force field developed in our lab.[67,68] Comparing Figure 3f to Figure 3c shows that the all-atom potential of mean force for the monomer agrees well with the same potential of mean force for the trimer, justifying our use of the monomer to constrain the side chain bonded interaction.

We parametrized the remaining effective interactions using a similar Boltzmann inversion procedure. Figure 4a illustrates the eight types of effective interactions, and their functional forms are written below in eqs 4−11. Equation 4 is the backbone bonded interaction U_{bb} among three adjacent backbone sites, which uses the same coefficients regardless of monomer type. Equation 5 is the side chain bonded interaction between backbone and phenethyl side chain sites and eq 6 is the side chain bonded interaction between backbone and charged side chain sites, which uses different coefficients for aminoethyl and carboxyethyl side chains. Equations 5 and 6 have different functional forms because the symmetry axis of the side chain sites are perpendicular and parallel to the side chain's hydrocarbon chain for phenethyl and charged side chains, respectively. Equation 7 is the nonbonded interaction between phenethyl side chain sites; eq 8 is the nonbonded interaction between charged side chain sites (with charges and geometric parameters depending on the identities of the two involved sites); eq 9 is the nonbonded interaction between backbone and any other site (with hard-core radii depending on the identities of the sites); and eq 10 is the interaction between nonbonded phenethyl and side chain sites (also with hard-core radii depending on the identities of the sites). Equation 11 is the one-body, height-dependent solvation interaction, with coefficients depending on the monomer type.

$$U_{bb} = \sum_{i=1}^{N-1}\sum_{j=0}^{4} K_{1j}|\vec{r}^{\,bb}_{i,i+1}|^j + \sum_{i=2}^{N-1}\sum_{j=0}^{4} K_{2j}(\hat{r}^{bb}_{i-1,i}\cdot\hat{r}^{bb}_{i,i+1} - (\hat{r}^{bb}_{i-1,i}\cdot\hat{n}_i^{bb})(\hat{r}^{bb}_{i,i+1}\cdot\hat{n}_i^{bb}))^j + \sum_{i=2}^{N} k_l\left(\hat{r}^{bb}_{i-1,i}\cdot\hat{n}_i^{bb} - \frac{|\vec{r}^{\,bb}_{i,i-1}| - r_{0l}}{s_l}\right)^2 + \sum_{i=1}^{N-1} k_r\left(\hat{r}^{bb}_{i,i+1}\cdot\hat{n}_i^{bb} - \frac{|\vec{r}^{\,bb}_{i,i+1}| - r_{0r}}{s_r}\right)^2 \quad (4)$$

$$U^{sc}_{phen} = k_1(((r_\perp - r_{\perp 0})^2 + (r_\parallel - r_{\parallel 0})^2)^{1/2} - r_0)^2 + \sum_{j=0}^{4} J_{1j}(\arctan(r_\parallel\{/\}r_\perp))^j + \sum_{j=0}^{6} J_{20j}{r_\parallel}^j + \sum_{j=0}^{2} J_{22j}{n_\parallel}^2{r_\parallel}^j + J_{240}{n_\parallel}^4 + \sum_{j=0,2,4} J_{30j}{n_\parallel}^j + \sum_{j=1,3} J_{31j}(\vec{r}\cdot\hat{n}){n_\parallel}^j + \sum_{j=0,2} J_{32j}(\vec{r}\cdot\hat{n})^2{n_\parallel}^j + J_{331}(\vec{r}\cdot\hat{n})^3 n_\parallel + J_{340}(\vec{r}\cdot\hat{n})^4 \quad (5)$$

$$U^{sc}_{charged} = k_1(((r_\perp - r_{\perp 0})^2 + (r_\parallel - r_{\parallel 0})^2)^{1/2} - r_0)^2 + \sum_{j=0}^{4} J_{1j}(\arctan(r_\parallel/r_\perp))^j + k_2(n_\parallel - \sum_{j=0}^{2} r_{2j}{r_\parallel}^j))^2 + k_3(\vec{r}\cdot\hat{n} - \sum_{j=0}^{2} n_{3j}{n_\parallel}^j))^2 \quad (6)$$

$$U^{nb}_{ph-ph}(X_i, X_j) = U_{GB}(X_i, X_j) + \frac{1}{2}(S(r_{iz}) + S(r_{jz}))U_{sm}(X_i, X_j) + \left(1 - \frac{1}{2}(S(r_{iz}) + S(r_{jz}))\right)U_q(X_i, X_j) \quad (7)$$

$$U^{nb}_{charged-charged}(X_i, X_j) = U_{ex}(|\vec{r}_{ij}| - (R_i + R_j)) + U_{el}(|(\vec{r}_j + \delta_j\hat{n}_j) - (\vec{r}_i + \delta_i\hat{n}_i)|) \quad (8)$$

$$U^{nb}_{back-any}(r_{ij}) = \begin{cases}\infty, & r_{ij} \le r_0\\ 0, & r_{ij} > r_0\end{cases} \quad (9)$$

$$U^{nb}_{phen-charged}(r_{ij}) = \begin{cases}\infty, & r_{ij} \le r_0\\ 0, & r_{ij} > r_0\end{cases} \quad (10)$$

$$U_{solvation}(r_z) = U_{bs} - k_B T_r \ln\left(1 + \frac{\exp(U_{bs}/k_B T_r) - 1}{1 + \exp(-4(r_z - z_{int})/w)}\right) + U_{int}\exp\left(-\frac{1}{2}\left(\frac{r_z - z_{int}}{\sigma_{int}}\right)^2\right) \quad (11)$$

In eq 7 the expressions entering the nonbonded phenethyl−phenethyl interaction are

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

$$U_{GB}(X_i, X_j) = 4\varepsilon_0\varepsilon(X_i, X_j; \kappa')\left(\left(\frac{\xi\sigma_0}{|r_{ij}| - \sigma_0\sigma(X_i, X_j) + \xi\sigma_0}\right)^{12} - \left(\frac{\xi\sigma_0}{|r_{ij}| - \sigma_0\sigma(X_i, X_j) + \xi\sigma_0}\right)^{6}\right) \tag{12}$$

$$\varepsilon(X_i, X_j; \kappa') = (1 - \chi(\kappa)^2(\vec{n}_i\cdot\vec{n}_j)^2) \times \left(1 - \frac{\chi'(\kappa')}{2}\left(\frac{\hat{r}_{ij}\cdot\vec{n}_i + \hat{r}_{ij}\cdot\vec{n}_j}{1 + \chi'(\kappa')\vec{n}_i\cdot\vec{n}_j} + \frac{\hat{r}_{ij}\cdot\vec{n}_i - \hat{r}_{ij}\cdot\vec{n}_j}{1 - \chi'(\kappa')\vec{n}_i\cdot\vec{n}_j}\right)\right) \tag{13}$$

$$\sigma(X_i, X_j) = \left(1 - \frac{\chi(\kappa)}{2}\left(\frac{(\hat{r}_{ij}\cdot\vec{n}_i + \hat{r}_{ij}\cdot\vec{n}_j)^2}{1 + \chi(\kappa)\vec{n}_i\cdot\vec{n}_j} + \frac{(\hat{r}_{ij}\cdot\vec{n}_i - \hat{r}_{ij}\cdot\vec{n}_j)^2}{1 - \chi(\kappa)\vec{n}_i\cdot\vec{n}_j}\right)\right)^{-1/2} \tag{14}$$

$$\chi(\kappa) = \frac{\kappa^2 - 1}{\kappa^2 + 1} \tag{15}$$

$$\chi'(\kappa') = \frac{1 - \kappa'^{1/\mu}}{1 + \kappa'^{1/\mu}} \tag{16}$$

$$S(z) = \frac{1}{1 + \exp(4(z - z_{int})/\sigma_{int})} \tag{17}$$

$$U_{sm}(X_i, X_j) = \begin{cases} \varepsilon_r\varepsilon(X_i, X_j; \kappa_r)\left(1 - \frac{4}{w^2}\left(|\vec{r}_{ij}| - \sigma_0\sigma(X_i, X_j) - \frac{w}{2}\right)^2\right), & |\vec{r}_{ij}| - \sigma_0\sigma(X_i, X_j) < w \\ -\varepsilon_{ss}\varepsilon(X_i, X_j; \kappa_{ss})\left(1 - \frac{4}{w^2}\left(|\vec{r}_{ij}| - \sigma_0\sigma(X_i, X_j) - \frac{3w}{2}\right)^2\right), & w < |\vec{r}_{ij}| - \sigma_0\sigma(X_i, X_j) < 2w \\ 0, & |\vec{r}_{ij}| - \sigma_0\sigma(X_i, X_j) > 2w \end{cases} \tag{18}$$

In eq 8, the expressions entering the nonbonded charged−charged interaction are

$$U_{ex}(\Delta r) = \begin{cases} \infty, & \Delta r < 0 \\ f_{sh}(U_{bsi} + U_{bsj})\left(1 - 4\left(\frac{\Delta r - r_{sh}/2}{r_{sh}}\right)^2\right), & 0 < \Delta r < r_{sh} \\ 0, & \Delta r > r_{sh} \end{cases} \tag{19}$$

$$U_{el}(r) = \frac{Cq_iq_j\exp(-r/\lambda_D)}{\varepsilon(r - (R_i - \delta_i + R_j - \delta_j))r} \tag{20}$$

$$\varepsilon(r) = \begin{cases} \varepsilon_0 + (\varepsilon_\infty - \varepsilon_0)r/\lambda_\varepsilon, & r < \lambda_\varepsilon \\ \varepsilon_\infty, & r > \lambda_\varepsilon \end{cases} \tag{21}$$

Sections 1−3 of the Supporting Information Methods discuss in detail the Boltzmann inversion procedures we used to parametrize each term. We parametrized bonded side chain interactions (eqs 5 and 6) by inverting potentials of mean force from all-atom simulations of isolated monomers dissolved in water using the MF-TOID all-atom peptoid force field,[67] as discussed above for eq 3 but including the dependence on the side chain sites' orientations. We used MF-TOID simulations of a peptoid dimer with three backbone nitrogens to invert the backbone bonded interaction (eq 4), which includes both two- and three-body terms among neighboring backbone sites.

We parametrized the nonbonded interactions by matching orientation-dependent potentials of mean force to published all-atom potentials of mean force for small-molecule analogs of the coarse-grained sites.[69,70] We used a Gay−Berne potential to model the nonbonded interaction between phenethyl side chain sites,[71−74] targeting potentials of mean force for toluene,[69] and we used an anisotropic electrostatic potential (hard spheres with off-center point charges) with an ad-hoc solvent-mediated term to model the interaction between oppositely charged aminoethyl and carboxyethyl side chain sites, targeting potentials of mean force for methylammonium and acetate.[70] We used the same electrostatic parameters to model the interactions between equally charged side chain sites. (As far as we know, no one has conducted all-atom simulations for equally charged molecules like methylammonium and acetate that are analogous to our charged side chains.) For simplicity, we let the remaining, weaker interactions (backbone−side chain, phenethyl−aminoethyl, and phenethyl−carboxyethyl) be purely repulsive, using hard-core repulsions with conservatively small radii to prevent erroneous exclusions of viable all-atom configurations. We let the hard-core radii of the charged sites be equal to the hard-core radii used in the interaction between charged sites, we let the hard-core radius of the phenethyl side chain site be equal to the shorter of the principle radii for the Gay−Berne potential, and we initially set the hard-core radii of the backbone site equal to a conservatively small value, the van der Waals radius of the central nitrogen atom.

Finally, we parametrized a height-dependent solvation free energy for each type of site, targeting experimental bulk solvation free energies[75−77] and height-dependent all-atom potentials of mean force[78−80] for the molecular analogs of each site. Although the use of molecular analog solvation free energies *by themselves* to predict macromolecule free energies has been criticized because it neglects solvent exclusion and "self-solvation",[81] we explicitly accounted for those effects by including both solvation interactions and solvent-mediated nonbonded interactions.

In total, our model has 115 parameters, 114 of which were parametrized to match experimental free energies or all-atom potentials of mean force for small-molecule analogs, and one of which (the backbone site hard-core radius) was initially assigned a conservatively small value.

In a second stage of parametrization, we fine-tuned four of the 115 parameters, including the one that was not para-

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

metrized from the bottom up, to improve the agreement with experiments and all-atom simulations[82] of peptoid bilayers. We fine-tuned these parameters to simultaneously improve the agreement with (1) experimental X-ray scattering spectra, which showed a somewhat sharper dominant peak at a somewhat smaller wavenumber, (2) all-atom pair distribution functions, some of which showed peaks at larger distances, and (3) distributions for the position and orientation of the aminoethyl side chains, which were shifted more toward the *trans* configuration.

Figure 4 highlights the effect of fine-tuning the four parameters, and a detailed discussion appears in Section 7 of the Supporting Information Methods. As shown in Figure 4b-c, increasing the size (by 11%) and interaction strength (by 50%) of the phenethyl side chain sites (the dominant nonbonded interaction in the system) improved the agreement of phenethyl−phenethyl pair distribution functions between all-atom and course-grained simulations. As shown in Figure 4d-e, increasing the hard-core radius of the backbone sites from the conservatively small nitrogen van der Waals radius improved the agreement of backbone−backbone pair distribution functions. We chose particular values of the phenethyl size and the backbone length scale from a two-dimensional grid to optimize the agreement with the experimental X-ray scattering peak height and location, and we found that our choice was also among the best choices for improving the agreement with all-atom distribution functions. We did not adjust the parameters controlling the shape of the phenethyl-phenethyl side chain interaction, because we found that orientational pair distribution functions (see Supporting Information Methods) agreed well with the all-atom simulations throughout the range of phenethyl sizes that we investigated.

Finally, as shown in Figure 4f−k, we found that it was necessary to adjust the side chain bonded interaction for the phenethyl side chains so that the side chains would predominantly adopt a *trans* configuration, as in all-atom simulations of peptoid bilayers (Figure 4g), rather than a *cis* configuration, as in all-atom simulations of isolated monomers (Figure 4f). Adjusting the linear coefficient J_{11} in the fourth-order polynomial describing the bimodal *cis*−*trans* transition (see eq 6) resulted in free energies (Figure 4k) that agreed with the all-atom bilayer (Figure 4j) more closely than the isolated all-atom monomer (Figure 4h). The bonded interactions for the carboxyethyl and phenethyl side chains did not need such an adjustment, because their all-atom free energies are very similar for bilayers and isolated monomers.

We call our model the Molecular Foundry Coarse-grained Model for Peptoids (MF-CG-TOID). Source code in C for initializing, simulating, and analyzing the model are available as Supporting Information.

We expect that combining our current model with careful treatment of fluctuations of the air−water interface will allow us to investigate the large-scale dynamic processes of adsorption and collapse. To gain confidence that these studies will connect directly to experiment, we first establish in the current paper that our model captures the structural properties of the three metastable states relevant for the nanosheet production cycle (Figure 1): solvated polymers, adsorbed monolayers, and free-floating bilayers. To calculate these properties, we performed Monte Carlo simulations with periodic boundary conditions in the appropriate ensembles, as detailed in Sections 4 and 5 of the Supporting Information Methods. As discussed in Section 6 of the Supporting Information Methods, we compared

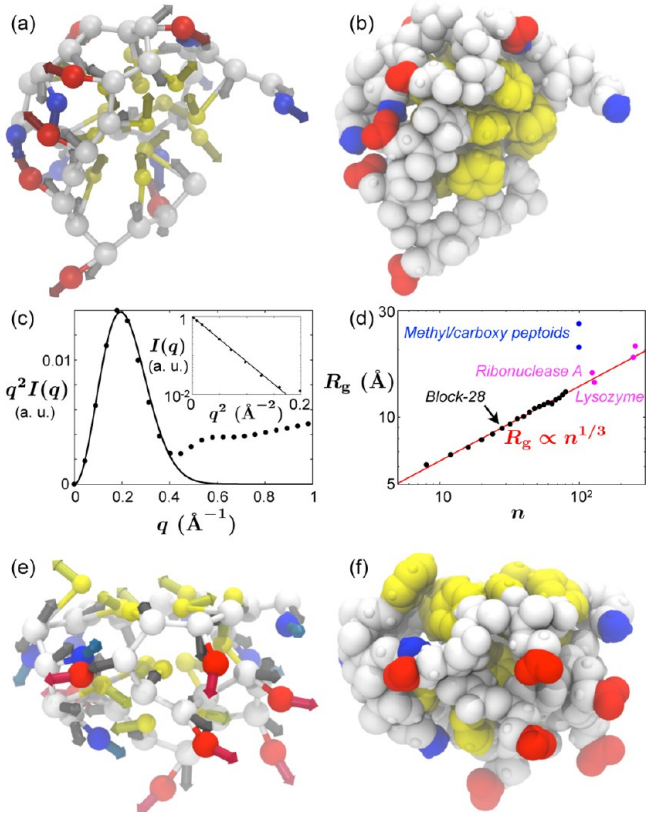


Figure 5. (a-b) Snapshots of a single solvated block-28 peptoid in (a) coarse-grained and (b) backmapped all-atom representations. (c) Kratky (main) and Guinier (inset) plots of the radially averaged X-ray scattering spectrum $I(q)$ for the solvated block-28 peptoid. The solid curves are the fit to Guinier's law, $I(q) = I_0 \exp(-(qR_g)^2/3)$. We arbitrarily rescale $I(q)$ by I_0. (d) Radius of gyration R_g (from the fit to Guinier's law) vs number of monomers for block-n peptoids. The red line is a fit to the scaling expected for spherical globules, restricting the fit to $n \geq 16$. For comparison, the blue points are experimental values for 100-monomer peptoids with either alternating (top point) or "protein-like" (bottom point) patterns of 80% methyl and 20% carboxyethyl side chains,[83] and the magenta points are experimental values for globular proteins from ref 84. (e-f) Snapshots of the block-28 peptoid exposed to a horizontal air−water interface. The atoms in b and f are shown with twice their van der Waals radii to show the exposed aromatic surfaces (yellow).

simulated and experimental structures by calculating scattering spectra on all-atom configurations generated from our coarse-grained simulations. Although all-atom scattering spectra have been generated from other coarse-grained models,[65,85,86] previous approaches have included energy minimization and annealing steps to remove unphysical local configurations. We believe that our work, using accurate all-atom configurations generated by anisotropic coarse-grained sites, is the first example of a coarse-grained model able to generate accurate all-atom scattering spectra directly.

COMPARISON WITH EXPERIMENT

We start our exploration of the nanosheet production cycle with a study of the equilibrium structure of solvated peptoids. This structure presumably strongly influences the ability of peptoids to adsorb to the air−water interface. Experimental X-ray[83] and neutron[87,88] scattering has indicated that solvated peptoids tend to collapse into single-chain globules, except for highly charged peptoids that form extended conformations in

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

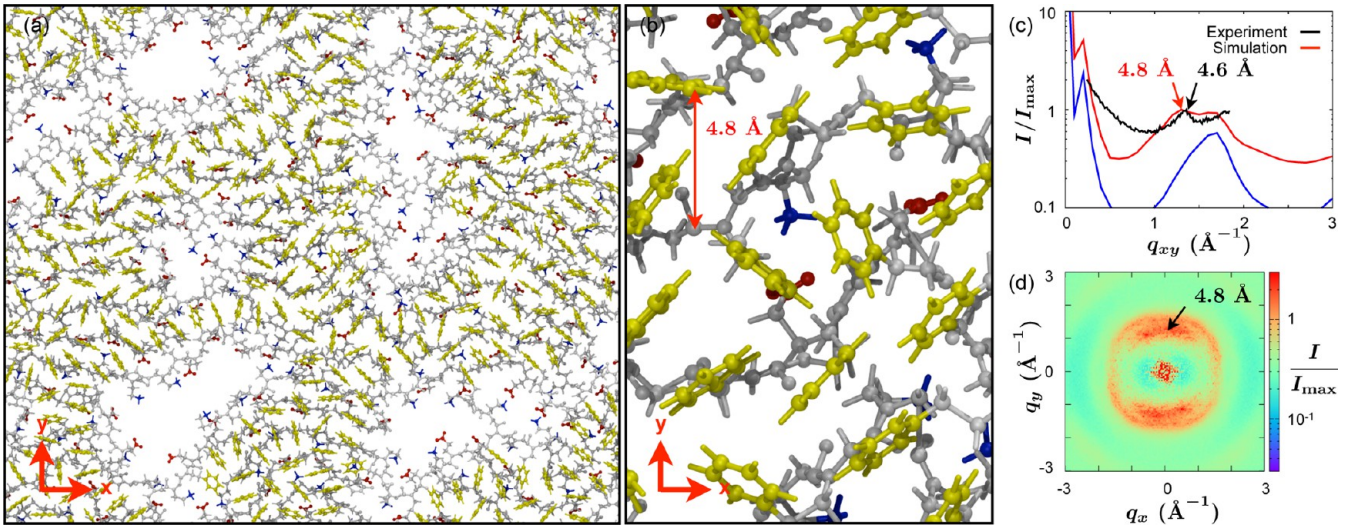


Figure 6. (a) Wide-angle and (b) close-up snapshots of a periodic monolayer of 48 block-28 peptoids simulated at the experimental equilibrium surface pressures of 31 mN/m. The peptoids are stable at the air−water interface, filling most of the interface and exhibiting small voids near some peptoid termini. Peptoids remain predominantly parallel to their neighbors, separated by a characteristic distance of 4.8 Å (red arrow in b). (c) Comparison of the experimental grazing-incidence X-ray scattering[13] (black) and simulated (red) radially averaged in-plane X-ray spectra show a similar peak location and amplitude at 4.6 (experiment) and 4.8 (simulation) Å. The spectra are plotted on a log scale and normalized by the amplitude I_{max} of the dominant peak, in order to allow quantitative comparison without knowing the incident X-ray intensity in experiments (varying which would only shift the log plot up or down). (d) Two-dimensional in-plane spectrum of the radially anisotropic simulated monolayer confirms that the dominant peak comes from correlations in the y direction.

low ionic strength solutions.[88] Our model block peptoid collapses into a single-chain globule when equilibrated in water because of the strong attraction between aromatic rings in phenylethyl side chains, as seen in experiments and simulations of other synthetic polymers with suitably designed amphiphilic patterns.[89−94] Snapshots of the block-28 peptoid at the end of the simulation (Figure 5a-b) show its roughly spherical shape, with most of the yellow aromatic side chains sequestered in the core. The Gaussian peak in the main panel of Figure 5c indicates a roughly spherical shape characterized by a radius of gyration R_g. The dependence on n for n-monomer block peptoids (Figure 5d) indicates that, for large peptoids, the globule size grows with molecular weight as expected (red line) for spherical globules; peptoids smaller than block-16 do not pack as spherically. Comparing our simulated results with experimental results for more weakly hydrophobic peptoids consisting of 80% methyl and 20% carboxyethyl side chains[83] (blue points in Figure 5b) suggests that the negatively charged methyl/carboxy peptoids pack less tightly than the block peptoids, likely due to repulsions between negative charges. Instead, the agreement of the red scaling with the magenta data points for globular proteins[84] suggests that our 50% phenylethyl peptoids pack as tightly as globular proteins.

When our model peptoid is exposed to an air−water interface, as in the snapshots of Figure 5e-f, part of the hydrophobic core flattens out along the interface as the peptoid strikes a balance between maintaining aromatic interactions within its core, and exposing nonpolar groups to the air. The persistence of the aromatic core may help explain why experimental monolayer formation proceeds slower than would be expected for a diffusion-limited process.[13]

The next stage of nanosheet production is the formation at the air−water interface of structured peptoid monolayers, a crucial intermediate between unassociated peptoids and nanosheets.[12,13] Figure 6 demonstrates that our coarse-grained model captures the essential structural features of the monolayer. We modeled the bulk monolayer by equilibrating a periodic monolayer at fixed values of the surface pressure,[13] including the experimental equilibrium surface pressure p_S = 31 mN/m. We plan to use our model to explore the equilibrium established between a peptoid solution and an air−water interface; in the current work, we used the surface pressure as a control parameter.

As shown in Figure 6a, our simulation box equilibrates as a dense monolayer containing small voids near the terminii of some of the peptoids. Nonpolar side chains (yellow) tend to segregate away from these voids. Inspection of close-up images such as Figure 6b reveals that peptoids tend to align parallel with their neighbors, separated by a characteristic distance of 4.8 Å. The distance 4.8 Å appears as a peak in the radially averaged X-ray spectrum (Figure 6c). Both the location and amplitude of the peak agree well with the experimental peak at 4.6 Å obtained from grazing incidence X-ray scattering in ref 13. In the simulation, peptoids remain predominantly oriented in the x direction, wrapping around the simulation box, so the simulation sample is radially anisotropic. This anisotropy allows us to separate the in-plane spectrum into x and y components, helping to confirm that the dominant peak comes from separations (predominantly in the y direction) between parallel peptoids. The more isotropic halo arises primarily from correlations between nonpolar side chains, which tend to organize isotropically except for the constraints imposed by their bonds to the backbones. Calculating the radial average only for atoms within the nonpolar side chains (blue curve in Figure 6c) illustrates that this contribution leads to the second peak observed in the simulation, and may explain the peak seen at the right-hand side of the experimental curve (black).

Focusing next on the final part of the nanosheet production cycle, we show in Figure 7 that our model reproduces the key structural features of the bilayer nanosheets themselves. We modeled the interior of a large, free-floating nanosheet by simulating a periodic bilayer at zero tension (Figure 7a). As

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

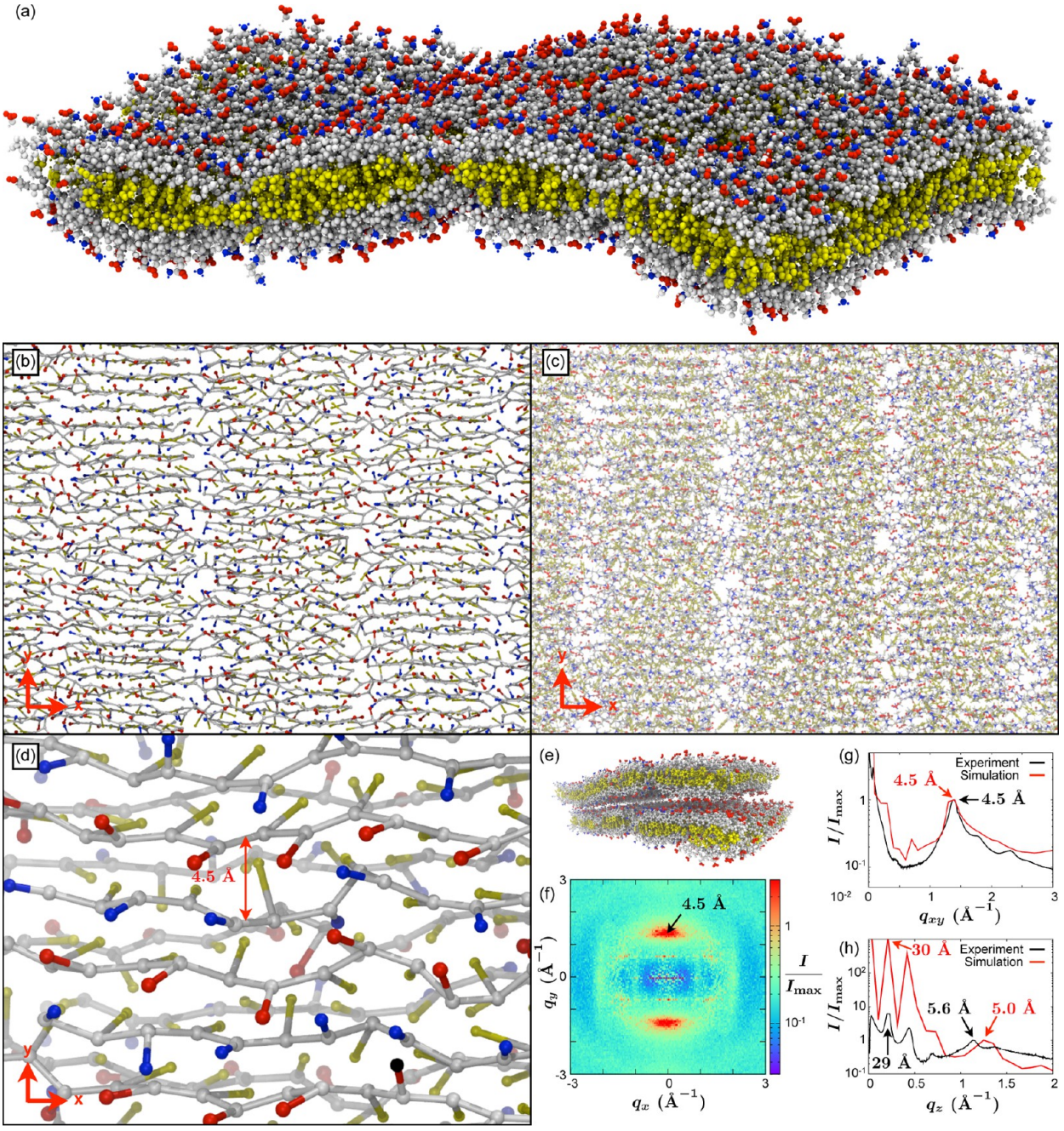


Figure 7. (a) All-atom representation of one periodic cell of a bilayer nanosheet at zero tension (96 block-28 peptoids). (b-c) Wide-angle snapshots looking down on the nanosheet, shown in (a) coarse-grained and (b) all-atom representations. The bilayer is stable in a rectilinear configuration, with peptoids running in the x direction and seams of terminii running in the y direction. (d) Close-up snapshots in the coarse-grained representation illustrate the typical 4.5 Å spacing between parallel peptoid backbones in each leaf of the bilayer. (e) We used a periodic cell of two stacked bilayers to calculate the in-plane and out-of-plane X-ray spectra, in order to compare to those obtained experimentally[13] from stacked bilayers. (f) Two-dimensional simulated in-plane X-ray spectrum reveals a dominant peak in the y direction, corresponding to a typical spacing of 4.5 Å between parallel peptoids. (g) The radially averaged in-plane spectrum agrees well with the experimental spectrum. (h) Experimental and simulated transverse X-ray spectra exhibit dominant peaks at 29 and 30 Å, respectively, corresponding to the lamellar spacing between stacked nanosheets, as well as peaks at 5.6 and 5.0 Å, respectively, whose origin is more subtle.[13] The stacking peaks are larger in simulation due the perfect stacking in the periodic z direction.

shown in Figure 7b-c, the bilayer stabilizes in a rectilinear configuration, with parallel peptoids wrapping around the x direction and seams of terminii running in the y direction. To calculate X-ray scattering profiles that could be compared with experimental ones, we placed two vertically separated bilayers in a periodic box and allowed them to stack together by allowing the attractive forces between the bilayers to drive out intervening solvent (see Section 6 of the Supporting Information Methods). As shown in Figure 7e, the bilayers equilibrated at a lamellar spacing of 30 Å. Then, we mimicked the experiments of ref 7 by performing X-ray scattering in and out of the bilayer plane. (Isolated nanosheets showed very little difference in their in-plane spectra.)

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

As shown in Figure 7f, the in-plane spectrum reveals a dominant peak in the y direction corresponding to a separation of 4.5 Å between neighboring peptoids. The distance is somewhat shorter than in the monolayer, probably due to the additional attractive interactions of the bilayer's second leaf, and the peak is more confined to the y direction due to the bilayer's ability to stabilize a rectilinear configuration. The radially averaged in-plane spectrum agrees very well with experiment (Figure 7g). The agreement is better than for the monolayer because we fine-tuned four model parameters to match this target.

The transverse spectrum in Figure 7h also agrees well with experiment, though the amplitudes of the lamellar peak at 30 Å and its two higher harmonics are larger in simulation due to the perfect stacking in the z direction. As discussed in ref 13, the dominant short-range peak at 5.0 Å (compared to 5.6 Å in the experiments) arises from a subtle combination of many contributions, so the agreement in its shape and location is a good indicator of the quality of the coarse-grained model.

INTERFACING WITH ALL-ATOM SIMULATIONS

The results described above demonstrate the accuracy afforded by our coarse-grained modeling scheme, which can be achieved in a computationally efficient manner. Our coarse-grained model is at least 10^4-fold faster than an all-atom calculation: the model possesses 10 times fewer interaction sites per unit volume (e.g., 57 coarse-grained sites versus 543 atoms for a block-28 peptoid, not counting the water molecules necessary for the all-atom simulation); it therefore requires 100 times fewer calculations per unit time per unit volume; and it can be propagated using a time step roughly 100 times larger than is used for all-atom calculations (10^{-13} seconds versus 10^{-15} seconds; see Section 4 of the Supporting Information Methods).

Furthermore, we are able to pass configurations between coarse-grained and all-atom simulations so as to access the efficiency of the coarse-grained representation and the high resolution of the atomic-scale model. Such multiscale simulation has been done by embedding small all-atom regions of interest within coarse-grained simulations,[43,95−99] by performing replica-exchange simulations across scales,[100−103] or by initializing all-atom simulations with coordinates generated from coarse-grained simulations.[104−106]

To establish the feasibility of this approach for the present model, we determined the compatibility of backmapped all-atom configurations of our coarse-grained model with our previously published CHARMM-based[68] all-atom peptoid force field, MFTOID.[67] Following the approach used for the PRIMO protein coarse-grained model,[42] we calculated how far atoms move when mapped "roundtrip" from an all-atom configuration (AA) to our coarse-grained model (CG) and back to an all-atom configuration (AA). Since the AA → CG mapping reduces the number of degrees of freedom, and the "backmapping" CG → AA is deterministic, atoms must move during the AA → CG → AA roundtrip. Because multiscale simulation schemes are based on either the AA → CG mapping, the CG → AA backmapping, or both, these schemes work best when the roundtrip distance moved is as small as possible.

Using an all-atom bilayer nanosheet as a test case,[82] we found that the "roundtrip" root-mean-square displacement (RMSD) per heavy atom is 0.291, 0.300, and 0.469 Å for Npe, Nae, and Nce monomers, respectively. Table 1 compares these values to those acquired from reconstructing protein test sets using two coarse-grained protein models: SICHO/CA, which has two isotropic sites per monomer and can generate all-atom configurations using the Molecular Modeling Tools for Structural Biology toolset,[42,107] and PRIMO, an intermediate-scale model designed specifically to interface directly with all-atom simulations.[42] Note that the excellent RMSD values for the PRIMO model, on the order of 0.1 Å, are made possible by using nearly as many coarse-grained sites as heavy atoms (e.g., 6 vs 9 for glutamic acid and 6 vs 11 for phenylalanine). Our RMSD values lie intermediate between the two models, demonstrating that considerable information that can be stored in the orientations of our coarse-grained sites.

Table 1. Root Mean Square Displacement Per Heavy Atom between Initial and Regenerated All-Atom Configurations for Coarse-Grained Peptoid and Protein Models

Model	Monomer	N_{Sites}	$N_{Heavy\ atoms}$	RMSD (Å)
SICHO/CA	Phe	2	11	1.149[42]
SICHO/CA	Glu	2	9	0.984[42]
PRIMO	Phe	6	11	0.059[42]
PRIMO	Glu	6	9	0.098[42]
MF-CG-TOID	Npe	2	12	0.291
MF-CG-TOID	Nae	2	7	0.300
MF-CG-TOID	Nce	2	9	0.469

Ongoing work on a related protein model suggests that storing the full orientation of each coarse-grained site (rather than just the principal symmetry axis) may increase the resolution of a coarse-grained model beyond the intermediate-resolution model: mapping backbone heavy atoms from the Protein Data Bank[108] to the protein model and back again yields an RMSD of 0.051 Å.[61] Although using a full orientation would only require increasing the number of degrees of freedom per site from five to six, we chose not to do so for our peptoid model because of the added complexity that would introduce to the effective interaction parametrization.

The accuracy of our coarse-grained model, reflected both in its ability to capture atomic-scale interactions and its ability to generate all-atom configurations, relies on the use of anisotropic coarse-grained sites. Although such anisotropy makes the model roughly six times more costly to simulate than an equivalent model comprising only simple isotropic interactions of a similar range (see Section 4 of the Supporting Information), it allows us to estimate all-atom configurations with sufficient accuracy to perform accurate scattering calculations and interface directly with all-atom simulations. As discussed above, producing such configurations with all-atom simulations would result in an approximately 10^4-fold slowdown. Producing them with a high-resolution coarse-grained model such as PRIMO, with three times as many sites as our coarse-grained model, would result in at least a 9-fold slowdown (due to a 3-fold increase in sites and 3-fold increase in force calculations, plus a shorter time step due to stiffer interactions).

CONCLUSION

We have shown that using a coarse-grained model with anisotropic coarse-grained sites permits efficient and accurate simulation of sequence-defined polymers. Using a minimal number of coarse-grained sites but including an independently fluctuating symmetry axis for each, we are able to efficiently sample ensembles of coarse-grained configurations that map to

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

detailed all-atom configurations with reasonable accuracy. Although we fine-tuned four parameters to improve agreement with the experimental in-plane X-ray spectrum for the bilayer, other emergent features of the model (the transverse bilayer spectrum and the in-plane monolayer spectrum) matched experiment without fine-tuning. We suggest that future efforts to optimize the accuracy of coarse-graining schemes should consider the symmetry of coarse-grained sites as an important variable.

We plan to use MF-CG-TOID to investigate the large-scale dynamic processes involved in the peptoid nanosheet production cycle, using our coarse-grained model both as a stand-alone tool and as a key component of a multiscale simulation protocol. We expect that MF-CG-TOID (augmented by parametrization of the necessary side chains) may also be useful in creating a new class of protein-mimetic materials based on the precision assembly of sequence-defined peptoid polymers, building on applications already developed in therapeutics,[109] diagnostics,[110−112] transfection,[113,114] and antibiotics.[115]

■ ASSOCIATED CONTENT

Supporting Information

(1) Supporting Information Methods describing the model parametrization, Monte Carlo simulations, initialization, generation of all-atom configurations and scattering spectra, and model fine-tuning. (2) Source code in C for initializing, simulating, and analyzing the model. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: tkhaxton@lbl.gov.

*E-mail: swhitelam@lbl.gov.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This project was funded by the Defense Threat Reduction Agency under Contract No. IACRO-B1144571. Work at the Molecular Foundry and the National Energy Research Scientific Computing Center was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

■ REFERENCES

(1) Collier, G.; Ortiz, V. *Arch. Biochem. Biophys.* **2013**, *538*, 6.
(2) Cui, Q.; Karplus, M. *Protein Sci.* **2008**, *17*, 1295.
(3) McGeagh, J. D.; Ranaghan, K. E.; Mulholland, A. J. *Biochim. Biophys. Acta* **2011**, *1814*, 1077.
(4) Tuffery, P.; Derreumaux, P. *J. R. Soc. Interface* **2012**, *9*, 20.
(5) England, J.; Lucent, D.; Pande, V. *Curr. Opin. Struct. Biol.* **2008**, *18*, 163.
(6) Wang, H.; Cui, J.; Hong, W.; Paterson, I. C.; Laughton, C. A. *J. Mol. Model* **2013**, *19*, 4997.
(7) Bowman, G. R.; Voelz, V. A.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4.
(8) Olivier, G. K.; Cho, A.; Sanii, B.; Connolly, M. D.; Tran, H.; Zuckermann, R. *ACS Nano* **2013**, *7*, 9276.
(9) Sun, J.; Zuckermann, R. N. *ACS Nano* **2013**, *7*, 4715−4732.
(10) Nam, K. T.; Shelby, S. A.; Choi, P. H.; Marciel, A. B.; Chen, R.; Tan, L.; Chu, T. K.; Mesch, T. A.; Lee, B.-C.; Connolly, M. D.; Kisielowski, C.; Zuckermann, R. N. *Nat. Mater.* **2010**, *9*, 454.
(11) Kudirka, R.; Tran, H.; Sanii, B.; Nam, K. T.; Choi, P. H.; Venkateswaran, N.; Chen, R.; Whitelam, S.; Zuckermann, R. N. *Peptide Sci.* **2011**, *96*, 586.
(12) Sanii, B.; Kudirka, R.; Cho, A.; Venkateswaran, N.; Olivier, G. K.; Olson, A. M.; Tran, H.; Harada, R. M.; Tan, L.; Zuckermann, R. N. *J. Am. Chem. Soc.* **2011**, *133*, 20808.
(13) Sanii, B.; Haxton, T. K.; Olivier, G. K.; Barton, B.; Proulx, C.; Whitelam, S.; Zuckermann, R. N. *ACS Nano* **2014**, *8*, 11674.
(14) Nielson, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J. Phys.: Condens. Matter* **2004**, *16*, R481.
(15) Clementi, C. *Curr. Opin. Struct. Biol.* **2008**, *18*, 10.
(16) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869.
(17) Tozzini, V. *Q. Rev. Biophys.* **2010**, *43*, 3.
(18) Trylska, J. *J. Phys.: Condens. Matter* **2010**, *22*, 453101.
(19) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41.
(20) Hyeon, C.; Thirumalai, D. *Nat. Commun.* **2011**, *2*, 487.
(21) Takada, S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130.
(22) Shinoda, W.; DeVane, R.; Klein, M. L. *Curr. Opin. Struct. Biol.* **2012**, *22*, 175.
(23) Saunders, M. G.; Voth, G. A. *Annu. Rev. Biophys.* **2013**, *42*, 73.
(24) Noid, W. G. *J. Chem. Phys.* **2013**, *139*, 090901.
(25) Louis, A. A. *J. Phys.: Condens. Matter* **2002**, *14*, 9187.
(26) Stillinger, F. H.; Sakai, H.; Torquato, S. *J. Chem. Phys.* **2002**, *117*, 288.
(27) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. *J. Chem. Phys.* **2007**, *126*, 144509.
(28) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.
(29) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izkekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
(30) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
(31) Izvekov, S.; Chung, P. W.; Rice, B. M. *J. Chem. Phys.* **2010**, *133*, 064109.
(32) Müller, M. *J. Stat. Phys.* **2011**, *145*, 967.
(33) Shell, M. S. *J. Chem. Phys.* **2008**, *129*, 144108.
(34) Chaimovich, A.; Shell, M. S. *Phys. Rev. E* **2010**, *81*, 060104.
(35) Chaimovich, A.; Shell, M. S. *J. Chem. Phys.* **2011**, *134*, 094112.
(36) Carmichael, S. P.; Shell, M. S. *J. Phys. Chem. B* **2012**, *116*, 8383.
(37) Kowalczyk, P.; Gauden, P. A.; Ciach, A. *J. Phys. Chem. B* **2009**, *113*, 12988.
(38) Kowalczyk, P.; Gauden, P. A.; Ciach, A. *J. Phys. Chem. B* **2011**, *115*, 6985.
(39) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Comput. Chem.* **2008**, *4*, 819.
(40) Hills, R. D.; Lu, L.; Voth, G. A. *PLoS Comput. Biol.* **2010**, *6*, e1000827.
(41) Kar, P.; Gopal, S. M.; Cheng, Y.-M.; Predeus, A.; Feig, M. *J. Chem. Theory Comput.* **2013**, *9*, 3769.
(42) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. *Proteins* **2010**, *78*, 1266.
(43) Predeus, A. V.; Gul, S.; Gopal, S. M.; Feig, M. *J. Phys. Chem. B* **2012**, *116*, 8610.
(44) Liwo, A.; Oldziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849.
(45) Liwo, A.; Oldziej, S.; Czaplewski, C.; Kozlowska, U.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 9421.
(46) Yap, E.-H.; Fawzi, N. L.; Head-Gordon, T. *Proteins* **2008**, *70*, 626.
(47) Májek, P.; Elber, R. *Proteins* **2009**, *76*, 822.
(48) Alemani, D.; Collu, F.; Cascella, M.; Dal Peraro, M. *J. Chem. Theory Comput.* **2010**, *6*, 315.
(49) Enciso, M.; Rey, A. *J. Chem. Phys.* **2010**, *132*, 235102.
(50) Enciso, M.; Rey, A. *J. Chem. Phys.* **2012**, *136*, 215103.
(51) Spiga, E.; Alemani, D.; Degiacomi, M. T.; Cascella, M.; Dal Peraro, M. *J. Chem. Theory Comput.* **2013**, *9*, 3515.

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315

(52) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. *Phys. Rev. Lett.* **2010**, *104*, 178101.
(53) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. *J. Chem. Phys.* **2010**, *132*, 035105.
(54) Linak, M. C.; Tourdot, R.; Dorfman, K. D. *J. Chem. Phys.* **2011**, *135*, 205102.
(55) Sulc, P.; Romano, F.; Ouldridge, T. E.; Rovigatti, L.; Doye, J. P. K.; Louis, A. A. *J. Chem. Phys.* **2012**, *137*, 135101.
(56) Orsi, M.; Essex, J. W. *PLoS One* **2011**, *6*, e28637.
(57) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192.
(58) Praprotnik, M.; Delle Site, L.; Kremer, K. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545.
(59) Peter, C.; Kremer, K. *Soft Matter* **2009**, *5*, 4347.
(60) Meier, K.; Choutko, A.; Dolenc, J.; Eichenberger, A. P.; Riniker, S.; van Gunsteren, W. F. *Angew. Chem., Int. Ed.* **2013**, *52*, 2820.
(61) Haxton, T. K. High-resolution coarse-grained modeling using oriented coarse-grained sites. Submitted; available online at arxiv.org/abs/1409.8658.
(62) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.
(63) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812.
(64) Nielson, S. O.; Srinivas, G.; Lopez, C. F.; Klein, M. L. *Phys. Rev. Lett.* **2005**, *94*, 228301.
(65) Tschöp, W.; Kremer, K.; Hahn, O.; Batoulis, J.; Bürger, T. *Acta Polym.* **1998**, *49*, 75.
(66) Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.
(67) Mirijanian, D. T.; Mannige, R. V.; Zuckermann, R. N.; Whitelam, S. *J. Comput. Chem.* **2014**, *35*, 360.
(68) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400.
(69) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217.
(70) Zhu, S.; Elcock, A. H. *J. Chem. Theory Comput.* **2010**, *6*, 1293.
(71) Gay, J. G.; Berne, B. J. *J. Chem. Phys.* **1981**, *74*, 3316.
(72) Gupta, S.; Sediawan, W. B.; McLaughlin, E. *Mol. Phys.* **1988**, *65*, 961.
(73) Walsh, T. R. *Mol. Phys.* **2002**, *100*, 2867.
(74) Cacelli, I.; Cinacchi, G.; Prampolini, G.; Tani, A. *J. Chem. Phys.* **2004**, *120*, 3648.
(75) Wolfenden, R. *Biochemistry* **1978**, *17*, 201.
(76) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664.
(77) Kang, Y. K.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4118.
(78) Shaytan, A. K.; Ivanov, V. A.; Shaitan, K. V.; Khokhlov, A. R. *J. Comput. Chem.* **2010**, *31*, 204.
(79) Minofar, B.; Vácha, R.; Wahab, A.; Mahiuddin, S.; Kunz, W.; Jungwirth, P. *J. Phys. Chem. B* **2006**, *110*, 15939.
(80) Jungwirth, P. private communication.
(81) K onig, G.; Bruckner, S.; Boresch, S. *Biophys. J.* **2013**, *104*, 453.
(82) Mannige, R. V.; Haxton, T. K.; Proulx, C.; Butterfoss, G. L.; Zuckermann, R. N.; Whitelam, S. submitted.
(83) Murnen, H. K.; Khokhlov, A. R.; Khalatur, P. G.; Segalman, R. A.; Zuckermann, R. N. *Macromolecules* **2012**, *45*, 5229.
(84) Mylonas, E.; Svergun, D. I. *J. Appl. Crystallogr.* **2007**, *40*, s245.
(85) Shih, A. Y.; Freddolino, P. L.; Sligar, S. G.; Schulten, K. *Nano Lett.* **2007**, *7*, 1692.
(86) Perlmutter, J. D.; Sachs, J. N. *Biochim. Biophys. Acta* **2009**, *1788*, 2284.
(87) Rosales, A. R.; Murnen, H. K.; Kline, S. R.; Zuckermann, R. N.; Segalman, R. A. *Soft Matter* **2012**, *8*, 3673.
(88) Murnen, H. K.; Rosales, A. M.; Dobrynin, A. V.; Zuckermann, R. N.; Segalman, R. A. *Soft Matter* **2013**, *9*, 90.
(89) Lozinsky, V. I. *Adv. Polym. Sci.* **2006**, *196*, 87.
(90) Altintas, O.; Barner-Kowollik, C. *Macromol. Rapid Commun.* **2012**, *33*, 958.
(91) Akagi, T.; Piyapakorn, P.; Akashi, M. *Langmuir* **2012**, *28*, 5249.
(92) Moreno, A. J.; Lo Verso, F.; Sanchez-Sanchez, A.; Arbe, A.; Colmenero, J.; Pomposo, J. A. *Macromolecules* **2013**, *46*, 9748.
(93) Lo Verso, F.; Pomposo, J. A.; Colmenero, J.; Moreno, A. J. *Soft Matter* **2014**, *10*, 4813.
(94) Terashima, T.; Sugita, T.; Fukae, K.; Sawamotot, M. *Macromolecules* **2014**, *47*, 589.
(95) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. *Phys. Rev. Lett.* **2005**, *95*, 218102.
(96) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045.
(97) Machado, M. R.; Dans, P. D.; Pantano, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18134.
(98) Mamonov, A. B.; Lettieri, S.; Ding, Y.; Sarver, J. L.; Palli, R.; Cunningham, T. F.; Saxena, S.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2012**, *8*, 2921.
(99) di Pasquale, N.; Marchisio, D.; Carbone, P. *J. Chem. Phys.* **2012**, *137*, 164111.
(100) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. *Phys. Rev. Lett.* **2006**, *96*, 028105.
(101) Lyman, E.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2006**, *2*, 656.
(102) Christen, M.; van Gunsteren, W. F. *J. Chem. Phys.* **2006**, *124*, 154106.
(103) Moritsugu, K.; Terada, T.; Kidera, A. *J. Chem. Phys.* **2010**, *133*, 244105.
(104) Thogersen, L.; Schiott, B.; Vosegaard, T.; Nielsen, N. C.; Tajkhorshid, E. *Biophys. J.* **2008**, *95*, 4337.
(105) Stansfeld, P. J.; Sansom, M. S. P. *J. Chem. Theory Comput.* **2011**, *7*, 1157.
(106) Permutter, J. D.; Drasler, W. J.; Xie, W.; Gao, L.; Popot, J.-L.; Sachs, J. N. *Langmuir* **2011**, *27*, 10523.
(107) Feig, M.; Karanicolas, J.; Brooks, C. L. *J. Mol. Graph. Model.* **2004**, *22*, 377.
(108) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(109) Zuckermann, R. N.; Kodadek, T. *Curr. Opin. Mol. Ther.* **2009**, *11*, 299.
(110) Gao, C. M.; Yam, A. Y.; Wang, X.; Magdangal, E.; Salisbury, C.; Peretz, D.; Zuckermann, R. N.; Connolly, M. D.; Hansson, O.; Minthon, L. *PLoS One* **2010**, e15725.
(111) Yam, A. Y.; Wang, X.; Gao, C. M.; Connolly, M. D.; Zuckermann, R. N.; Bleu, T.; Hall, J.; Fedynyshyn, J. P.; Allauzen, S.; Peretz, D. *Biochemistry* **2011**, *50*, 4322.
(112) Reddy, M. M.; Wilson, R.; Wilson, J.; Connell, S.; Gocke, A.; Hynan, L.; German, D.; Kodadek, T. *Cell* **2011**, *144*, 132.
(113) Utku, Y.; Dehan, E.; Oeurfelli, O.; Piano, F.; Zuckermann, R. N.; Pagano, M.; Kirshenbaum, K. A. *Mol. BioSyst.* **2006**, *2*, 312.
(114) Lobo, B. A.; Vetro, J. A.; Suich, D. M.; Zuckermann, R. N.; Middaugh, C. R. *J. Pharm. Sci.* **2003**, *92*, 1905.
(115) Chongsiriwatana, N. P.; Patch, J. A.; Czyzewksi, A. M.; Dohm, M. T.; Ivankin, A.; Gidalevitz, D.; Zuckermann, R. N.; Barron, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2794.

DOI: 10.1021/ct5010559
J. Chem. Theory Comput. 2015, 11, 303−315