Electronic Supplementary Material (ESI) for Soft Matter.

SAXS-guided unbiased coarse-grained Monte Carlo simulation for identification of self-assembly nanostructure and dimension, supplementary information

Silabrata Pahari,[†,‡] Shuhao Liu,[†] Mustafa Akbulut,[†] and Joseph Sang-Il Kwon[*,†,‡]

†*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, USA*

‡*Texas A&M Energy Institute, 1617 Research Pkwy, College Station, TX 77843, USA*

E-mail: kwonx075@tamu.edu

Appendix A1: SAXS profile calculation

Ab-initio SAXS profile calculations have been immensely successful for obtaining high-resolution target structures of proteins and colloids. Therefore, the same has also been implemented in this work.[1–3] Specifically, the *ab-initio* SAXS profiles were evaluated from the coordinates of the amphiphiles by utilizing the Debye's equation as follows:

$$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} \phi_i(q)\phi_j(q) \frac{sin(qr_{ij})}{qr_{ij}} \tag{1}$$

where N is the number of atoms, and ϕ_i and ϕ_j are the shape factors for the corresponding atoms. The shape factors are derived from the quantum mechanical calculations,[4,5] $q = \frac{sin\theta}{\lambda}$ is the scattering vector, and r_{ij} is the pair-wise distance between the atoms. The values of

ϕ are obtained as a function of scattering vector, and q is in the range of 0 to 3.50 Å^{-1}. The value of ϕ_i is obtained via the following equations:[4]

$$\phi_i(q) = f_{v,i}(q) - c_1 f_{s,i}(q) + c_2 S_{w,i} f_{w,i}(q) \tag{2}$$

where $f_{v,i}(q)$ is the shape factor of an atom in vacuo, $f_{s,i}(q)$ is the shape factor of the hypothetical atom that represents the displaced water (excluded volume), $S_{w,i}$ is the solvent accessible surface area of the atoms present, and $f_{w,i}$ is the shape factor of water. Here, c_1 and c_2 are the model parameters. Specifically, c_1 is used to adjust the electron density contrast, and c_2 is used to adjust the hydration shell thickness. The overall computation of the *ab-initio* SAXS profiles are done using the ATSAS CRYSOL package.[1]

Appendix A2: MC simulation details

The overall goal of MC simulations is to incorporate MC moves that can generate a Markov chain of states consistent with the system's statistical mechanics.[6–11] Specifically, for the OAPB and DTA molecules, rotation, dihedral angle change, translation, bond angle variation, and configurational biased segment-wise regrowth of molecules are the set of MC moves implemented to perform the simulations. The probability of accepting a particular MC move is obtained from the standard Metropolis and Hasting's algorithm (the explicit law is derived in this section and is highlighted in Eq. 2)[12] by following the approach highlighted in the work done by Shah et al[8]. Specifically, multiple MC moves present in the literature were studied to propose the combination of simple MC moves that can drive the DBC system close to equilibrium in a computationally tractable fashion.

First, the rotation movement involves the rigid body rotation of the entire molecule about one of the axis, i.e., the x, y, or z-axis, by an angle chosen from a uniform distribution of angles in the range of 0-180°. Second, the dihedral angular variation movement include selecting one of the dihedral angles in the flexible chains and then rotating it by a value sampled in

the interval between 0-30°. Third, the bond angle variations involve randomly selecting a bond angle of the long flexible chains and changing the angle to a definite value sampled from a uniform distribution of 0-180°. The translation move involves the movement of the center-of-mass of the entire flexible molecule by a value of δ. The value of δ is highlighted in Table 1. The aforementioned MC moves a schematic highlighting them is presented in Fig. 1.

Table 1: Values of the crucial parameters implemented in the MC simulations

Parameters	Parameter values
δ	0.15 nm
ϵ	$8.8 \cdot 10^{-12}$ F/m
r_{cut}	10 nm

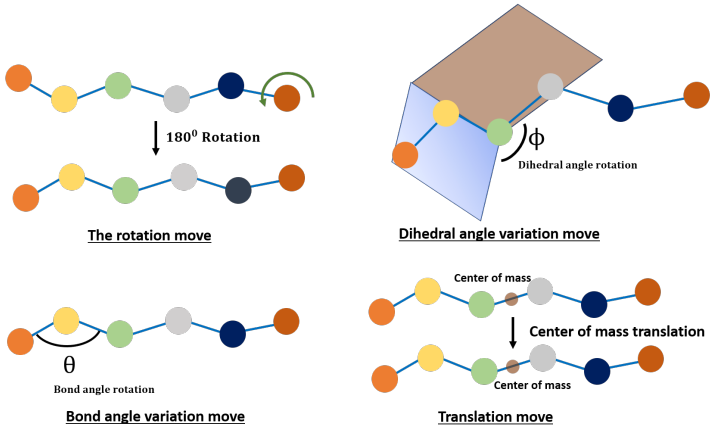

Figure 1: The schematic of the different MC moves implemented.

Additionally, a segment-wise CBMC move is performed to sample the intramolecular configuration.[8] Specifically, CBMC moves involve a step-wise regrowth of the flexible molecules by combining segments of molecules. In the step-wise regrowth of the OAPB and DTA molecules during the CBMC simulations, segments are sampled from a repository that is created beforehand. This repository is known as the library of segments. Generating this library involves performing the atom-wise regrowth of each segment. Therefore, configurations with a wide range of bond angles are present in this library.

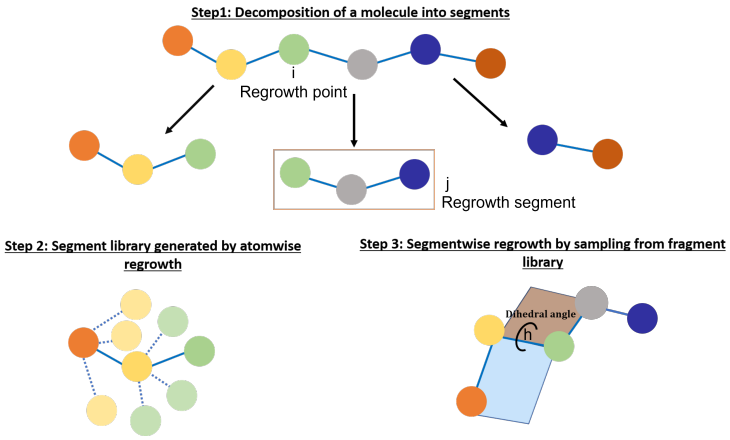

Figure 2: The schematic of the CBMC moves considered for the segment-wise regrowth of molecules.

Steps to implement the segment-wise CBMC move is mentioned as follows:

1. The probability of removing segments from either side is proportional to the number of segments present on that side which is given as follows:

$$P_{removal,k} = \frac{\sum_k^N 1}{\sum_1^k 1 + \sum_k^N 1} \tag{3}$$

where $P_{removal,k}$ is the probability to remove all the segments after the k^{th} segment in the polymer. Specifically, this is simply the fraction of the segments that are present on either side of the kth segment. In subsequent steps, the deleted segments are regrown.

2. The specific end from which the molecules regrow is referred to as i. At a given regrowth point i, two decisions are made. The first decision is the choice of the regrowth segment from the repository, and the second is the choice of the segment dihedral angle during the regrowth process. The probability of choosing the j^{th} segment while regrowing the molecule from point i is P_{ij}. Additionally, the likelihood of selecting the h^{th} dihedral angle at the regrowth point i is given as P_{ih}.

3. The probability for selecting the j^{th} segment, P_{ij}, is mentioned as follows:

$$P_{ij} = \frac{exp(-\frac{U_j}{k_bT})}{\sum_{N_{frag}} exp(-\frac{U_j}{k_bT})} \tag{4}$$

where U_j is the potential energy associated with one particular segment, and N_{frag} is the total number of segments present in the segment library.

4. The probability of choosing the h^{th} dihedral angle, P_{ih}, is given by the following equation:

$$P_{ih} = \frac{exp(-\delta U_h)}{\sum_k exp(-\delta U_h)} \tag{5}$$

where δU_h is the change in potential energy on choosing one particular regrowth dihedral angle, and k is the total number of dihedral angles that the growth segment can orient towards.

5. When all the removed segments are regrown by following the aforementioned strategy, two probabilities are calculated. The first is the probability of moving from the initial state m to the final state n. The second is the probability for the reverse event, i.e., the probability of moving from n to m. The acceptance probability is calculated by following the detailed balance criterion of the Markov chains and is obtained as follows:

$$P_{m-n} = min\Bigg(1, \frac{(\prod_{i=1}^{N_{del}} P_{ij}^{frag} P_{ih}^{dih})_{nm}}{(\prod_{i=1}^{N_{del}} P_{ij}^{frag} P_{ih}^{dih})_{mn}} exp\Bigg(- \beta\Big(U_n - U_m\Big)\Bigg)\Bigg) \tag{6}$$

where U_n and U_m are the potential energies of the state before and after the CBMC moves, respectively, and $\beta = \frac{1}{k_bT}$. It is to be noted that in this work, all the simulations are done with an NVT ensemble if not mentioned otherwise. An illustration of the implemented CBMC moves is highlighted in Fig. 2.

Appendix A3: Coarse-grained force-field details

The second crucial feature of the proposed MC simulations is to consider coarse-grained force-field based potential energy. The primary interaction potentials considered in the force-fields are the bonded interaction and non-bonded long-range (i.e., Coulombic and the van der Waals (VDW)) interaction. The total bonded potential considered in this work is comprised of the harmonic potential due to bond-length stretching, potential due to bond angle variation, and potential due to dihedral angle change. The 12-6 Lennard-Jones (LJ) potential is used to model the VDW interaction. A cut-off distance of r_{cut} (value highlighted in Table 1) is fixed to restrict the calculation of the exact pair-wise potential. The LJ interactions beyond the cut-off distance is approximated by the tail potentials and for the Coulombic interaction a Ewald summation is used. In the calculation of Coulombic interaction, the value of the permitivity of vacuum ϵ, used in this work, is highlighted in Table 1.

The coarse-grained model is built based on the atomistic model and is done via a trial-and-error strategy. Specifically, an NPT ensemble of the all atom and coarse-grained molecules with periodic boundary conditions are subjected to the MC simulation with a CHARMM all-atom force-field and Martini-3 coarse-grained force-field, respectively.[13,14] Then, the results of bond length distribution and density obtained from the two simulations are compared. Based on the mismatch obtained in the results, judicious modifications are made to the coarse-grained model. Here, the coarse-grained model is built by utilizing beads of three different sizes: tiny, small, and regular. These beads represent the all-atom to coarse-grained maps in the ratio of 2:1, 3:1, or 4:1, respectively, with sizes of 0.34 nm, 0.41 nm, and 0.47 nm. Furthermore, beads of three different types, the apolar (C); intermediate/non-polar (N); and monovalent ionic (Q), are considered to capture the varying polarity for beads of each size. It is to be noted that the varying LJ interaction parameters among different coarse-grained beads helps capture the varying polarity of different atom groups. Subsequently, the explicit solvent (i.e., water) molecules are modeled by the separate regular size water beads (W).

References

(1) Svergun, D.; Barberato, C.; Koch, M. H. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **1995**, *28*, 768–773.

(2) Konarev, P. V.; Petoukhov, M. V.; Volkov, V. V.; Svergun, D. I. ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **2006**, *39*, 277–286.

(3) Li, M.; Wang, W.; Yin, P. A general approach to access morphologies of polyoxometalates in solution by using SAXS: An ab initio modeling protocol. *Chemistry* **2018**, *24*, 6639–6644.

(4) Putnam, D. K.; Lowe Jr, E. W.; Meiler, J. Reconstruction of SAXS profiles from protein structures. *Computational and Structural Biotechnology Journal* **2013**, *8*, e201308006.

(5) Cromer, D. T.; Mann, J. B. X-ray scattering factors computed from numerical Hartree–Fock wave functions. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **1968**, *24*, 321–324.

(6) Pangali, C.; Rao, M.; Berne, B. On a novel Monte Carlo scheme for simulating water and aqueous solutions. *Chemical Physics Letters* **1978**, *55*, 413–417.

(7) Pant, P. K.; Theodorou, D. N. Variable connectivity method for the atomistic Monte Carlo simulation of polydisperse polymer melts. *Macromolecules* **1995**, *28*, 7224–7234.

(8) Shah, J. K.; Marin-Rimoldi, E.; Mullen, R. G.; Keene, B. P.; Khan, S.; Paluch, A. S.; Rai, N.; Romanielo, L. L.; Rosch, T. W.; Yoo, B., et al. Cassandra: An open source Monte Carlo package for molecular simulation. 2017.

(9) Bereau, T.; Rudzinski, J. F. Accurate structure-based coarse graining leads to consistent barrier-crossing dynamics. *Physical Review Letters* **2018**, *121*, 256002.

(10) Hoffmann, D.; Knapp, E.-W. Polypeptide folding with off-lattice Monte Carlo dynamics: the method. *European Biophysics Journal* **1996**, *24*, 387–403.

(11) Mavrantzas, V. G.; Theodorou, D. N. Atomistic simulation of polymer melt elasticity: Calculation of the free energy of an oriented polymer melt. *Macromolecules* **1998**, *31*, 6310–6332.

(12) Oelschlaeger, C.; Suwita, P.; Willenbacher, N. Effect of counterion binding efficiency on structure and dynamics of wormlike micelles. *Langmuir* **2010**, *26*, 7045–7053.

(13) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al. CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614.

(14) Souza, P. C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M.; Wassenaar, T. A., et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods* **2021**, *18*, 382–388.