Perspective: 2013 AIChE Institute Lecturer Award

Recent Advances in Molecular Simulation: A Chemical Engineering Perspective

Jeremy C. Palmer
Dept. of Chemical and Biomolecular Engineering, University of Houston, Houston, TX 77204

Pablo G. Debenedetti
Dept. of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544

DOI 10.1002/aic.14706
Published online January 5, 2015 in Wiley Online Library (wileyonlinelibrary.com)

Keywords: chemical engineering, free energy, molecular dynamics, molecular simulation, Monte Carlo

Introduction

Molecular simulation has become increasingly prominent in scientific research over the last several decades, gaining widespread utilization across physics, chemistry, materials science, and engineering disciplines.[1,2] Its rise in popularity has been driven by advances in the design of computational hardware and simulation algorithms, which have resulted in an astronomical increase in the computational complexity of problems that can be addressed with such methods, with estimates suggesting that the effective improvement has been more than nine orders of magnitude since the late 1970s[3,4] [Karplus and coworkers[3] performed the first molecular dynamics (MD) calculations of a model protein in 1977, simulating an approximately 500-atom system comprised of bovine pancreatic trypsin inhibitor *in vacuo* for about 100 ps. In 2010, Shaw et al.[4] performed millisecond long MD simulations of bovine pancreatic trypsin inhibitor in explicit solvent using a system containing more than 10^4 atoms. A conservative estimate with the assumption that the computational complexity (i.e., number of mathematical operations) scales linearly with system size and the simulated time, therefore, yields an effective improvement of 10^9.] This remarkable trend has enabled molecular simulation to complement experiments in the development of rational strategies for engineering molecular interactions to control the nanoscopic and macroscopic (thermophysical) properties of matter.

Because this bottom-up design paradigm holds great promise for systematically improving processes ranging from petroleum refining to pharmaceutical formulations, chemical engineers have embraced molecular simulation and made it an integral part of the discipline. Chemical engineers, working in parallel and often collaboratively with physicists and chemists, have significantly advanced the field of molecular simulation. They have done so by improving algorithm design and developing new ways of using available computational hardware to solve problems of increasing physical complexity. As a result, the role of molecular simulation in chemical engineering has evolved rapidly. Less than 3 decades ago, the state of the art was embodied by the first direct simulations of phase equilibria for simple fluids such as argon.[1,5] Today, chemical engineers routinely apply molecular simulation to study challenging problems related to drug design[6] and formulation;[7–9] biomolecular crowding;[10] protein folding[11] and aggregation;[12,13] wetting phenomena[14,15] and hydration thermodynamics;[16–20] nucleation and growth processes;[21–24] the thermophysical properties of complex fluids[25–29] such as ionic liquids[30,31] and liquid crystals;[32] the phase behavior of polymeric, colloidal, and self-assembled systems;[33–43] and the synthesis, design and characterization of advanced materials[44–53] for applications ranging from adsorption-separation processes[54–59] to energy conversion and storage.[60–62]

In this Perspective, we highlight recent computational studies by chemical engineers that in our view, capture the state of the art in the field of molecular simulation. A similar Perspective by de Pablo and Escobedo in 2002[1] serves as a historical point of reference. A comparison with this Perspective illustrates the fact that although major strides have been made in the intervening period, many of the challenges remain qualitatively, albeit not quantitatively, the same. One of the most significant challenges is overcoming the computationally demanding nature of simulating physical systems and processes governed by mesoscopic length and timescales.[1,2] Because the computational complexity of problems addressed with molecular simulation continues to increase and keep pace with progress in developing new strategies to address length and timescale constraints, major challenges constantly arise at the frontiers of molecular simulation and will continue to do so for the foreseeable future.

We begin by briefly discussing the overarching goals of molecular simulation and the main classes of computational algorithms that are used in the field, which fall into two

Correspondence concerning this article should be addressed to J. C. Palmer at jcpalmer@uh.edu or P. G. Debenedetti at pdebene@princeton.edu

categories: MD and Monte Carlo (MC). The length and timescale problem, which has historically been a significant impediment to the application of molecular simulation to many systems, is also discussed in the context of such algorithms. We then highlight contributions from chemical engineers that use advanced strategies designed to circumvent the length and timescale limitations of conventional approaches. One prominent strategy is coarse-graining,[2] which aims to overcome the length and timescale problem through the development of physically accurate models of minimal computational complexity. We also discuss the application of algorithms designed to compute free energy to identify equilibrium states[63–65] and rare event methods[66] used to study processes that occur infrequently on the timescales accessible with molecular simulation. Some of the advanced algorithms used in these studies were previously discussed by de Pablo and Escobedo[1] at a point when they were still in their early stages of development. As these authors correctly predicted,[1] such methods have seen significant improvements that have played a key role in the evolution of the field and the design of new, more powerful algorithms. Because the technical details of the methods are beyond the scope of this Perspective, we focus our attention on the physical insights that can be gained with such approaches rather than on the techniques themselves. Although our discussion is limited to systems that can be modeled with classical mechanics, we note that chemical engineers have also made significant advances in the application and development of quantum chemistry techniques to study chemical reactions and the electronic properties of materials (e.g., refs. 67–70).

Fundamentals of Molecular Simulation

The aim of molecular simulation is to explore the role of molecular interactions in driving dynamical processes and shaping the structural and thermodynamic properties of matter. Molecular models are constructed from empirical functions, or force fields, that are used to describe interactions arising between molecules (e.g., van der Waal's and Coulomb forces) and those responsible for stabilizing their internal structure, such as chemical bonds.[2] The properties of model systems, which range from a few dozen to trillions of molecules in size, are investigated by using such force fields in conjunction with computer algorithms designed to numerically solve the equations of classical statistical mechanics (MC) or Newtonian mechanics (MD).

At equilibrium, statistical mechanics dictates that the probability density associated with the observation of a particular molecular arrangement in a system containing N molecules at a constant temperature and volume is given by the Boltzmann distribution:

$$\wp\left(\boldsymbol{x}^N\right)=Z^{-1}\exp\left[-\beta U\left(\boldsymbol{x}^N\right)\right] \tag{1}$$

where U is the potential, or configurational energy, of the system; $\boldsymbol{x}^N$ is a vector containing the Cartesian coordinates of all the atoms in the system; $\beta=(k_BT)^{-1}$ is the inverse temperature; k_B is Boltzmann's constant; and $Z=\int\exp\left[-\beta U(\boldsymbol{x}^N)\right]d\boldsymbol{x}^N$ is the partition function. Molecular simulation techniques use numerical algorithms to generate a large number, or an ensemble, of configurations distributed according to Eq. 1, evaluating U and the forces acting on molecules from the atomic coordinates using an empirical force field. Properties such as energy, pressure, and various structural quantities are computed by the direct averaging of their instantaneous value over a series of statistically uncorrelated configurations generated during the course of sampling.[2,71] Other thermophysical properties, such as the heat capacity and isothermal compressibility, are evaluated from the variance associated with instantaneous properties.[71] There are also important kinetic contributions to many equilibrium properties that are related to the average momentum of the molecules in the system. Such contributions, however, can be evaluated without the use of simulation because they are independent of atomic configurations and can, therefore, be found by the exact integration of the equilibrium momentum distribution (i.e., Maxwell's distribution).[2,71]

Molecular simulation methods use two general types of algorithms to solve the equations of statistical mechanics. MD generates configurations by the numerical solution of Newton's equations of motion to propagate the system in time, following a deterministic trajectory that is fully specified by the momenta and the positions of the molecules at the beginning of the simulation.[2,71,72] Because Newton's equations of motion naturally conserve the total energy of the system (kinetic plus potential), they are often modified with a thermostat that adds and dissipates energy during the simulation to maintain a constant temperature in accord with Eq. 1.[2,71,72] MC methods, on the other hand, produce stochastic trajectories by the movement of molecules in the system at random to generate new configurations that are either accepted or rejected on the basis of criteria designed to ensure that the correct distribution (i.e., Eq. 1) is sampled at the desired temperature.[2,64,71] Other well-known distributions that describe, for instance, systems at constant temperature and pressure,[2,71] can also be sampled with MD or MC. Both methods are rigorous in the sense that they will asymptotically sample from the exact target distribution.[2,71] The errors associated with the estimates of equilibrium properties obtained from MD and MC are, therefore, statistical in nature and will, in principle, decrease as the duration of the simulation is extended.[71]

Even though MD and MC can sample from the same equilibrium distributions, it is often advantageous to choose one method over the other depending on the application. Because MD follows the natural time evolution of the system, it can be used to investigate nonequilibrium processes and to compute properties related to the system's relaxation dynamics at equilibrium (e.g., diffusion coefficients, viscosity, etc.).[2] In contrast, there is no natural measure of time in MC simulations because molecules move in an artificial, stochastic manner that is not constrained by physical transport processes. As a result, MC methods are in general only used for examining thermophysical and structural properties at equilibrium.[2] The fact that MC simulations are not restricted by physical timescales, however, can be a significant advantage. Through the design of clever ways to move molecules in MC simulations, sluggish physical processes can be artificially accelerated or even bypassed.[1,64] This allows equilibrium to be reached much faster than in conventional MD simulations, potentially

reducing the computational effort required to perform sampling by orders of magnitude. Improvements in the sampling efficiency have also been realized by advanced methods that combine aspects of MC and MD.[71] An in-depth discussion of MD and MC algorithms, the relative merits of the two methods, and more advanced simulation techniques can be found in several reviews and perspectives that have been written by chemical engineers for a general audience.[1,2,64,72,73]

Simulations are typically initiated from configurations that are not representative of the system's most probable state because it is generally not known *a priori*. Consequently, they must evolve toward sampling regions of configuration space that are more statistically favorable according to Eq. 1. This process can be tracked by periodic monitoring of the simulation with a set of order parameters:

$$\boldsymbol{\phi}(\boldsymbol{x}^N)=\{\phi_1(\boldsymbol{x}^N),\phi_2(\boldsymbol{x}^N),...,\phi_n(\boldsymbol{x}^N)\}$$

These order parameters are calculated from the atomic coordinates of the system,[63,74] and they are usually chosen on the basis of physical intuition to provide information regarding the system's state that is useful for understanding the problem at hand. Examples include the system's density, which is useful for tracking the progress of phase transitions in constant temperature and pressure simulations, or the number of native residue contacts in a protein, which allow the folding/unfolding process to be monitored. The free-energy surface (FES) parameterized by $\boldsymbol{\phi}$ can also be computed from the probability density associated with the observation of specific values of the order parameters at equilibrium:[74]

$$\beta F(\boldsymbol{\phi})=-\ln\left[\wp(\boldsymbol{\phi})\right]+\text{constant} \tag{2}$$

where

$$\wp(\boldsymbol{\phi})=Z^{-1}\int \exp\left[-\beta U\left(\boldsymbol{x}^N\right)\right]\delta\left(\boldsymbol{\phi}-\boldsymbol{\phi}\left(\boldsymbol{x}^N\right)\right)d\boldsymbol{x}^N \tag{3}$$

and δ is a multidimensional version of Kronecker's δ that is used to sum over configurations satisfying $\boldsymbol{\phi}=\boldsymbol{\phi}(\boldsymbol{x}^N)$. Because the additive constant in Eq. 2 is independent of $\boldsymbol{\phi}$, it can be disregarded during the computation of relative free energies, which are the central quantities of interest for understanding a system's physical behavior. For reasons that will be discussed later, the FES is almost always calculated with advanced sampling methods. In principle, however, it can be computed directly from a histogram of the $\boldsymbol{\phi}$ values attained by the system over the course of a long equilibrium MD or MC simulation.

Order-parameter-based free-energy analysis has become increasingly common in molecular simulation because the FES not only provides insight into the relative stability of different states of the system at equilibrium, but it also serves as a useful roadmap for understanding processes that drive the system between states. Consider, for example, the model system with the FES shown in Figure 1(a), which is described by two order parameters, $\boldsymbol{\phi}=\{\phi_1,\phi_2\}$. The contours denote regions on the surface that differ in free energy by k_BT, the system's thermal energy scale. If a simulation is initiated near state A, which is relatively high in free energy, the system will eventually relax until it reaches state B, which lies at the global minimum in the FES. Although this process may happen slowly if the system exhibits sluggish dynamics, it will likely occur in a continuous fashion, following a downhill path on the FES connecting states A and B. The reverse process can also occur, but it is much less likely because state A is higher in free energy.

Surfaces with a single minimum, such as the one shown in Figure 1(a), occur under conditions where only one state is stable. An example of a system exhibiting such behavior is liquid water at room temperature and pressures higher than the stability limit (spinodal) of the vapor phase. Figure 1(b) illustrates an FES with two minima corresponding to states A and B, respectively. As in Figure 1(a), state A has a higher free energy than state B, which resides at the global minimum of the FES. There is, however, a free-energy barrier that the system must climb before it can transition between A and B. State A is, therefore, metastable with respect to B, and a simulation initialized in this region will

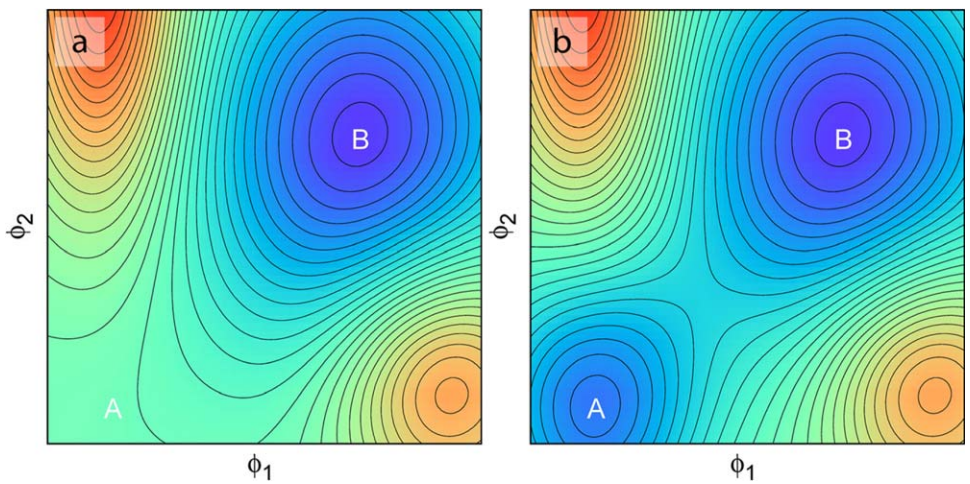

Figure 1. (a) Model FES that exhbits a single minimum corresponding the to the thermodynamically stable state of the system. (b) Model FES illustating the metastability of state A with respect to state B. States A and B are separated by a free-energy barrier that must be overcome for the system to transiton between the two regions. Countours denote free-energy isosurfaces separated by k_BT, which is the thermal energy of the system.

reside there until the system experiences a fluctuation large enough to overcome the free-energy barrier, at which point it will proceed toward state B. Such behavior occurs, for example, when liquid water is cooled slightly below its freezing point. Although ice is thermodynamically favored, the water will not freeze until the system overcomes the free-energy barrier associated with the formation of an ice nucleus that is large enough to grow. When water is at its freezing point, however, liquid and ice coexist. The corresponding FES will exhibit two minima of equal depth at such conditions because the two states (phases) have an equal probability of being observed at equilibrium.[74] As a result, a simulation performed at coexistence will eventually transition between the minima and spend an equal amount of time on average visiting each state. Much more exotic behavior is possible and also quite common. Proteins, for instance, typically exhibit many minima in their FES, each corresponding to a different conformational state of the molecule.[11] This is also true for mixtures or pure substances[29] that exhibit complex phase behaviors.

Length and Timescale Limitations

In principle, an MD or MC simulation that is run for an infinitely long time will explore all possible regions of its underlying equilibrium FES. This limit, however, cannot be approached in practice because of the finite nature of available computational resources. The computational effort required to model a system depends on several factors, including its physical size (i.e., number of atoms or molecules required to represent the system), the mathematical complexity of the force field used to describe its molecular-level interactions, and the kinetics of the processes that allow it to explore its FES.[2] As a result, it is still challenging to apply molecular simulation techniques to many systems.

Phenomena that are governed by mesoscopic or macroscopic length scales that exceed characteristic molecular dimensions by orders of magnitude are challenging to model because of the computational cost associated with the simulation of systems that are large enough to describe such behavior. The formation of mesoscopic domains in self-assembled systems is challenging to simulate, for example, because of the sheer number of molecules required to model each domain.[75] Consequently, even if this process is unimpeded by a free-energy barrier [qv, Figure 1(a)], a significant computational effort will be required to run a simulation long enough to observe self-assembly.

In other instances, the timescale over which a process occurs, not the size of the system that must be modeled, is the fundamental challenge. Processes such as protein folding and nucleation, for instance, are governed by an FES, such as the one shown in Figure 1(b), in which the system must overcome a free-energy barrier to transition from one state to another. If the barrier is many times larger than the thermal energy, fluctuations that allow the system to transition between states will be rare events. Depending on the system, such events will occur on timescales that can range from microseconds (e.g., protein folding[66]) to times longer than the age of the known universe (e.g., the diamond-to-graphite transition at ambient conditions[76]). MD simulations, on the other hand, use a fundamental time step on the order of 1 fs (10^{-15} s) to ensure numerical stability while integrating the equations of motion.[2,72] As a result, the study of such processes with standard MD simulations may be extremely computationally expensive or even impossible because of the disparity between these timescales.

Because the computational cost increases with the model complexity and simulated length and timescale, these factors cannot be considered independently in the design of a computational study. Standard MD simulations, for example, have been performed on massively parallel supercomputers with more than 140,000 processing cores to model systems containing several trillion molecules described with the Lennard-Jones force field[77], which is relatively inexpensive to compute and accurate for simple fluids such as noble gases.[5] Such very large systems, however, have only been simulated for a few picoseconds of physical time;[77] this is generally insufficient to reach equilibrium and perform sampling. In contrast, processes such as protein folding, which occur on timescales that are orders of magnitude longer and involve relatively complex molecular interactions, can be studied with specialized computational hardware to perform millisecond long MD simulations of relatively small systems containing only a single protein and a few thousand water molecules.[4] These examples represent the current state of the art in the application of standard MD simulations to investigate extremely large length and long timescales with world-class computational resources. Because most researchers have limited access to such resources, the brute-force approach of the use of more powerful computational hardware to circumvent length and timescale limitations is usually not a viable option. As a result, clever modeling strategies and advanced computational techniques are often required to address such challenges.

Coarse Graining

Molecular simulation studies typically use force fields that treat individual atoms as the fundamental building block of matter. This involves the use of simple empirical functions to provide a coarse-grained description of interatomic and intramolecular forces that ultimately originate as a consequence of the quantum behavior of subatomic particles.[2] Although atomistic force fields are generally less accurate in describing such interactions than quantum mechanical models, which explicitly describe the electronic degrees of freedom of the system, they are orders of magnitude less computationally expensive; this facilitates the study of phenomena occurring on length and timescales that are challenging, if not currently impossible, to study with first-principles methods.[2] A similar coarse-graining strategy can also be used to help overcome length and timescale limitations in molecular simulation, in which the computational complexity of the model is reduced by the replacement of two or more atoms on the same molecule with a single bead or pseudoatom.[2] As illustrated in Figure 2, for example, methane (CH_4) is often modeled as a single, coarse-grained bead; this reduces the computational cost by about an order of magnitude in comparison to the simulation of each atom explicitly. Because methane is nonpolar and roughly spherically symmetric, this approximation works well, as evidenced by the fact that such models are able to accurately predict many of

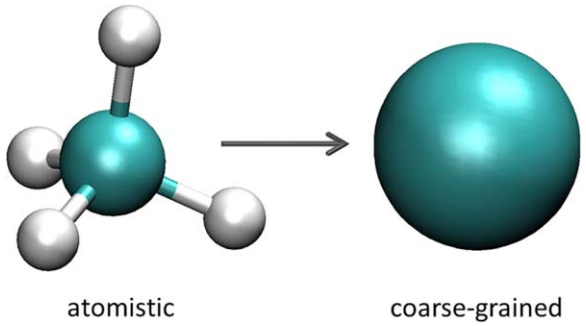

Figure 2. Molecular structure of methane (left) and a CGM (right) comprised of a single bead.

this substance's thermophysical properties in the liquid and vapor state.[78] Coarse graining, however, becomes significantly more challenging for large, complex molecules, such as proteins that contain a variety of functional groups and structural motifs. In general, the accuracy of a coarse-grained model (CGM) depends on how the structure of the molecule is represented, the method used to parameterize the force field, and the specific properties that are being investigated. Understanding how to simultaneously optimize the computational efficiency and accuracy of a CGM, is therefore, a significant challenge and a very active area of research.

Despite the aforementioned challenges, CGMs have been successfully developed for a variety of complex systems. Significant efforts have been made, for example, to develop CGMs for water because of its ubiquitous nature. Water's complex hydrogen-bonded network, however, is notoriously difficult to model even at the atomistic level, let alone with a coarse-grained representation. Despite this challenge, Molinero and Moore[79] have successfully developed a CGM (the mW model of water), which treats water as a single bead and implicitly describes hydrogen bonding through a three-body potential that constrains the angles formed by triplet sets of neighboring molecules. Not only is this CGM more accurate than many atomistic models of water, but it is more than 100 times more computationally efficient.[79]

The superior computational efficiency of the mW model of water allowed Peters, Molinero and collaborators[21] to study the homogeneous nucleation of natural gas hydrates at 273 K and 900 atm with large-scale MD simulations containing about 70,000 molecules. Such simulations were used to examine the growth/dissolution of hydrate nuclei ranging from 200 to 700 clathrate cages in size. As illustrated in Figure 3, nucleus growth was found to resemble a diffusive process,[21] with the mean squared difference of the nucleus size eventually exhibiting linear behavior as a function of time. From the simulation data, the authors estimated that the critical nucleus contained between 300 and 400 clathrate cages (5,000-6,000 molecules) and that the solution-hydrate surface tension was about 31 mJ/m^2,[21] which is in excellent agreement with the experimentally measured value of 32 ± 3 mJ/m^2.[80] The large value of the surface tension implies that the homogeneous nucleation rate is extremely small ($\sim 10^{-111}$ nuclei cm^{-3} s^{-1}) at the temperatures and pressures where methane hydrates are observed in nature, pipelines, and laboratory experiment; this indicates that another mechanism must be responsible for their formation under such conditions.[21] To put the scope of this study in perspective, computational investigations of hydrates with atomistic representations of water have examined systems containing a few thousand molecules;[24,81] this is an order of magnitude smaller and only slightly larger than the critical nucleus size estimated in ref. 21.

The three force field parameters adjusted in the development of the mW water model were determined by simulations with different values for the parameters and by the selection of the ones that best reproduced targeted experimental properties for water, including its melting temperature, vaporization enthalpy, and liquid density.[79] Although such an approach can be used to develop accurate CGMs when there are only a few adjustable parameters in the force field, more systematic methods are generally desirable because the parameter space that must be simulated grows rapidly as model systems increase in complexity. To this end, Müller and Jackson[82–85] developed a systematic coarse-graining method in which force field parameters are determined directly from an equation of state (EOS) that has been fitted to match experimental data. The key aspect of their method is the use of an EOS that is derived from a molecular perspective based on statistical associating fluid theory (SAFT).[86] The EOS is, therefore, formulated with coarse-grained molecular representations and interaction parameters than can be directly converted to a molecular simulation force field.[82] CGMs can be developed very rapidly with this approach because the fitting of the EOS parameters to experimental data is significantly less computationally expensive than the performance of numerous MD or MC simulations to optimize a force field by trial and error.[82] Such models are also accurate over a wide range of state conditions because the EOS is typically fitted over large portions of the experimentally determined phase diagram.[82]

Figure 4 shows an example of Müller and Jackson's method applied to a highly asymmetric mixture comprised of CO_2 and *n*-eicosane (*n*-$C_{20}H_{42}$) at 323.15 K. Coarse-grained

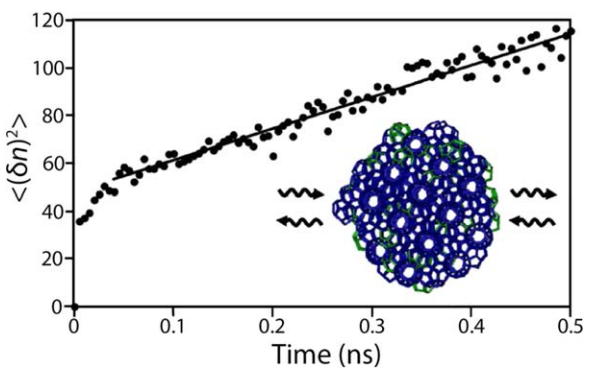

Figure 3. Mean square difference of the size of a methane hydrate nucleus as a function of the time computed from MD simulations of the coarse-grained mW water model at 273 K and 900 atm.

The MSD is computed using $\langle(\delta n^2)\rangle = \langle(n(t) - \langle n(t)\rangle)^2\rangle$, where $\langle n(t)\rangle$ is the time-dependent mean drift in nucleus size averaged over an ensemble of trajectories. The linear behavior at longer times indicates that the growth processes can be modeled as a random walk or diffusive process. Adapted with permission from ref. 21. Copyright 2012 American Chemical Society.

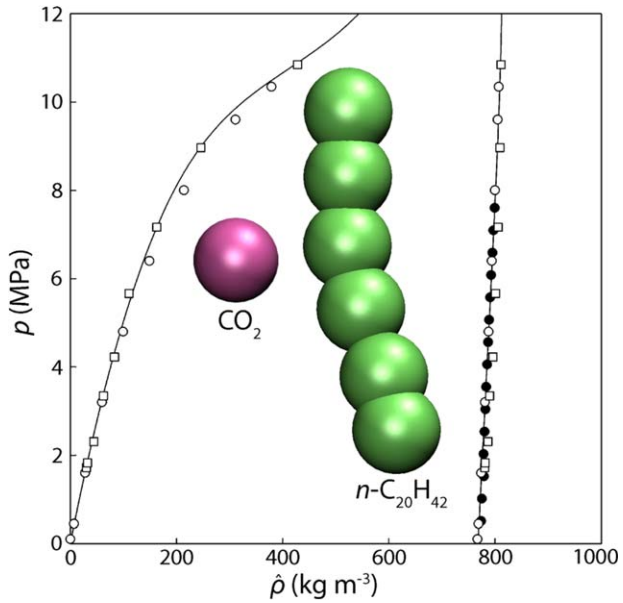

Figure 4. Pressure-mass density phase diagram of CO_2 and *n*-eicosane (n-$C_{20}H_{42}$) at 323.15 K.

The solid line shows the SAFT-based EOS, which was fitted to the experimental data from refs. 83 and 88 (open and filled circles, respectively). The results from MD simulations (squares) of the SAFT-derived CGMs[83] are in excellent agreement with the experiment. The inset shows the coarse-grained representations of CO_2 and *n*-eicosane used in the EOS calculations and MD simulations. Adapted with permission from ref. 83. Copyright 2014 American Chemical Society.

force fields for these species were derived by the application of the SAFT-based approach to pure component data with the modeling of CO_2 as a single, coarse-grained bead[84] and *n*-eicosane as a chainlike structure composed of six identical beads.[84] An adjustable binary interaction parameter was then determined by the fitting of the EOS to the experimental pressure-mass density phase diagram[83] for the mixture shown in Figure 4. As Figure 4 illustrates, both the SAFT-based EOS calculations and subsequent MD simulations that use the SAFT-derived CGMs were found to be in excellent agreement with the experimental target data over the entire range of pressures examined for the mixture. Similar agreement between simulation and experiment has also been observed for CGMs developed with this technique for smaller alkanes, sulfur hexafluoride, carbon tetrafluoride, polyethylene surfactants, and carbon dioxide-water mixtures.[82] In addition to capturing phase equilibrium behavior, SAFT-derived CGMs have also been found to yield excellent predictions of the transport properties for substances such as carbon dioxide and methane.[87] Such results demonstrate the potential of this approach for bridging the microscopic and macroscopic length scales by aiding in the development of accurate, experimentally informed coarse-grained force fields that can be used in molecular simulation studies.

Several methods have also been developed to systematically derive coarse-grained force fields directly from simulations of atomistic models.[89–91] Such approaches require no experimental data, and they are advantageous in cases where well-established atomistic models have already been developed. Because the final model is less detailed than the original, however, information about the system is irreversibly lost as a consequence of coarse graining.[89] If this information is relevant to the physical properties of the system under investigation, the accuracy of the CGM may be adversely affected. On the basis of this idea, Shell[89] developed a variational method for coarse graining that aims to systematically minimize information loss. The amount of loss is quantified with an entropy-like function that measures the statistical overlap between the coarse-grained and atomistic equilibrium ensembles.[89] Standard numerical techniques are then applied to minimize this function with respect to the adjustable force field parameters in the CGM.[89,92] For a given

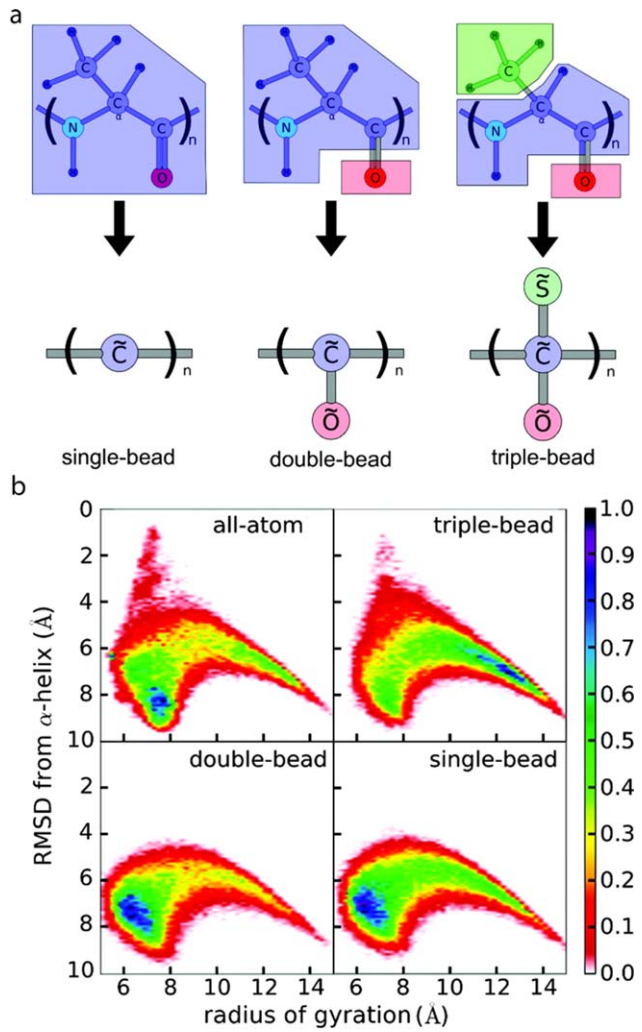

Figure 5. (a) Mapping of an atomistic representation of an alanine residue onto CGMs comprised of one, two, and three beads, respectively. (b) FESs for a 15-residue alanine peptide computed with the atomistic model and three CGMs. The FES was parameterized by the RMSD of the peptide from an ideal α-helical structure and R_g. The conformation characteristics of the α-helical and β-hairpin structures are located at 1.0 RMSD (7.2 R_g) and 8.5 RMSD (7.8 R_g) on the FES, respectively. Adapted with permission from ref. 92 Copyright 2012 American Chemical Society.

coarse-grained representation of the system, Shell's approach,[89] therefore, seeks to find force field parameters that best reproduce the equilibrium ensemble generated by fully atomistic simulations.

Charmichael and Shell[92] recently applied this novel method to develop CGMs to study the self-assembly of peptides. Their target system was a 15-residue alanine peptide that was simulated with an atomistic force field in conjunction with an implicit solvent model for water.[92] This system was chosen because alanine-rich peptides are model systems for amyloids, which form aggregates that are thought to play a central role in the onset of various neurodegenerative diseases.[93] Consequently, the properties of alanine peptides have been extensively investigated with computer simulation and experiment.[92] As illustrated in Figure 5(a), Charmichael and Shell[92] considered three different coarse-grained peptide models, in which each alanine residue was represented with one, two, or three beads. Optimal force field parameters for each CGM were derived with the variational approach described previously to ensure that the equilibrium ensemble of the target system was matched as closely as possible. The accuracy of each CGM was subsequently examined by a comparison of the predicted peptide structure against the results for the atomistic model. Figure 5(b) shows the FES for each of the alanine peptide models as function of two order parameters, the root mean square deviation (RMSD) of the peptide from an ideal α-helical structure and the radius of gyration (R_g). All three coarse-grained representations successfully captured the shape of the FES predicted by the atomistic model. Only the coarse-grained representation with three beads per residues, however, successfully captured both the characteristic α-helical and β-hairpin conformations observed in the atomistic model (1.0 RMSD and 7.2 R_g and 8.5 RMSD and 7.8 R_g on the FES, respectively).[92] Despite the deficiency of the two-bead model in capturing the α-helical structure, it was found to aggregate and successfully form β-like parallel and antiparallel chain alignments, in accordance with more detailed CGMs for the alanine peptides.[94] Thus, although information loss in coarse graining to the one- and two-bead models significantly impacts their ability to describe the conformations of the alanine peptide, such models may be adequate for examining other properties related to peptide self-assembly.

Rare Event and Free-Energy Methods

The development of robust coarse-graining techniques has helped to reduce the computational cost of modeling complex systems and has, thereby, significantly increased the length and timescales that can be investigated with molecular simulation. The timescales accessible with CGMs, however, are still typically too short to investigate systems governed by an FES like the one shown in Figure 1(b), in which the long-lived states are separated by a large free-energy barrier. As discussed previously, in such instances, it is unlikely that the system will explore its underlying FES and visit all

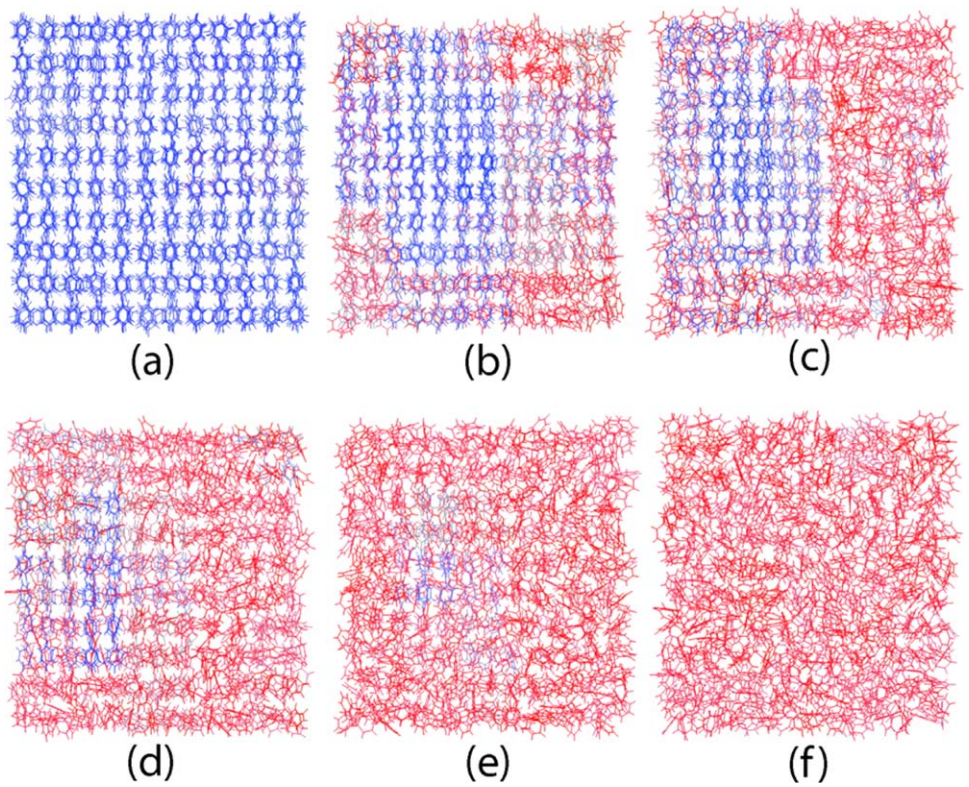

Figure 6. (a-f) Series of snapshots illustrating intermediate growth stages along trajectories that connect the crystal and melt phases of benzene.

The molecules with crystalline and liquidlike local environments are colored blue and red, respectively. Snapshots at stages b-e highlight the crystal nucleus and liquid-melt interface. Reprinted with permission from ref. 23. Copyright 2011 American Chemical Society.

relevant states in a single simulation. Moreover, even if transitions between states occur occasionally during long MD simulations, it is usually too computationally expensive to simulate a sufficient number of these rare events to characterize their kinetics and underlying mechanisms. Because most CGMs are only designed to reproduce static equilibrium properties of the target system, there is also no guarantee that they will accurately describe rare, activated processes. As a result, a number of advanced techniques have been developed to overcome timescale limitations by other means to investigate rare events and examine the long-time, equilibrium behavior of systems. Although many of these advanced techniques were originally conceived and developed by physicists and chemists, the examples below illustrate that chemical engineers have been highly successful in improving upon these methods and applying them to complex physical systems of interest in modern engineering practice.

Shah et al.[23] recently used a rare event simulation technique known as *aimless shooting*[95] to study the homogeneous nucleation of benzene crystals from the melt with atomistic MD simulations. In this case, the crystallization of benzene is a rare event because it is impeded by the free-energy barrier associated with the formation of an interface between the crystal nucleus and the melt phase. The aimless shooting method of Peters and Trout,[95] like other transition path sampling techniques,[66] overcomes timescale limitations by using an MC procedure to focus the computational effort on the generation of MD trajectories that exhibit a transition between the long-lived states of interest. This MC procedure is similar to the one described in the Introduction, but instead of generating atomic configurations according to Eq. 1, it samples from a transition path ensemble, which contains all possible trajectories of a given duration that begin in one state and end in the other.[66]

Using this method, the authors of ref. 23 generated 1650 candidate MD trajectories, of which 874 successfully transitioned between the melt and crystal. Thus, although a single nucleation event would be extremely unlikely to be observed in 1650 normal MD simulations of comparable duration, the aimless shooting method enhanced the sampling of this process and allowed more than half of the total computational effort to be spent gathering meaningful statistics regarding the transition mechanism. A series of snapshots from one of the successful trajectories is shown in Figure 6. Benzene molecules with crystalline and liquidlike local environments are colored blue and red, respectively, to highlight the crystal nucleus at intermediate stages of growth. The local environment of the molecules and the global state of the system were characterized with order parameters for molecular crystals developed by Santiso and Trout.[96] Such order parameters are necessary to identify the initial and final states of the system during the implementation of rare event methods such as the aimless shooting algorithm.[95] An important feature of many rare event methods, however, is that the order parameter choice does not bias the resulting MD trajectories.[66] Shah et al.[23] were, therefore, able to subsequently analyze saved MD trajectories in greater detail to examine the nucleation mechanism.

Velez-Vega et al.[97] used similar rare-event techniques to investigate the unfolding pathways for the Trp-cage, a 20-residue miniprotein. Trp-cage has been extensively investigated as model systems for understanding protein folding/unfolding because these processes occur on microsecond timescales that can be accessed with both experimental and computational techniques. Trp-cage's unfolding pathway is known to involve at least one intermediate state, in which the native protein (N) first transitions into a loop structure (L) before it reaches the unfolded state (U).[98] Previous computational studies have also suggested that that the N-L transition is the rate-limiting step in the unfolding process.[98] To scrutinize this prediction, Velez-Vega et al.[97] studied the N-L transition with an atomistic model of a Trp-cage using an improved version of the forward flux sampling (FFS) technique of Allen et al.[99] that had been previously developed by Borrero and Escobedo.[100] Like the aimless shooting technique mentioned previously, the objective of FFS is to harvest a large number of trajectories that transition from one long-lived state to another. It has the advantage over other similar techniques of being designed specifically to compute transition rates.[66] Thus, the authors of ref. 97 used their modified FFS technique to estimate the N-L transition rate, which they predicted to be $(8\ \mu s)^{-1}$. This estimate is in reasonable agreement with the experimentally measured value of $(12.7\ \mu s)^{-1}$ for the N-U transition[101] and supports

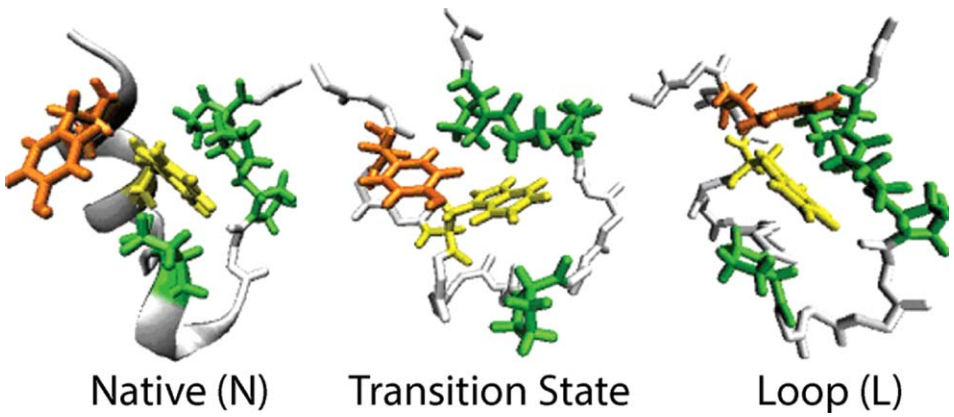

Figure 7. Trp-cage configurations representative of the N and L states are shown in the left and right panels, respectively, whereas the center panel shows the transition state.

Reprinted with permission from ref. 97. Copyright 2010, AIP Publishing LLC.

the assertion[98] that the N-L transition is the rate-limiting step in the unfolding process.

In examining the N-L transition, Velez-Vega et al.[97] used the RMSD of residues 2–8 on a Trp-cage from an ideal helical structure as an order parameter to distinguish between N and L states. Although this order parameter was found to be suitable for this purpose, they observed that it did not provide significant insight into the mechanism underlying the N-L transition.[97] This finding illustrates the important distinction between order parameters that allow the initial and final states to be distinguished and reaction coordinates, which are useful for describing the entire transition process. An ideal reaction coordinate should not only describe the mechanism underlying the transition process and discriminate between the initial and final states, but it should also be useful for predicting which long-lived state the system will first visit when initiated from any arbitrary intermediate point along the transition path. Using these ideas, Velez-Vega et al.[97] examined seven additional order parameters that were monitored during the FFS simulations to search for a suitable reaction coordinate. They found that a nonlinear combination of two of the order parameters provided the best overall description of the transition processes.[97] Figure 7 shows snapshots of the Trp-cage protein in the N and L states and a configuration of the protein at the transition state along the N-L path that was identified with this new order parameter.

In addition to transient phenomena, long-time equilibrium behavior is also essential to understanding molecular processes. The central quantity relevant to characterizing such behavior is free energy. As suggested by Eq. 2, however, the computation of free energy requires that all relevant states of the system be visited over the course of a simulation. As previously discussed, this is practically impossible, especially for systems where the states of interest are disjoint and separated by large free-energy barriers, as illustrated in Figure 1(b). Advanced computational methods have, therefore, been developed to facilitate the calculation of free energy.[1,63] The basic idea behind many of these approaches is to apply a bias potential to the simulations to compensate for, or flatten, the underlying FES.[63] In methods such as umbrella sampling,[102] an attractive, time-independent bias potential is used to restrain the simulation to enhance sampling of a specific targeted region of the order parameter space. The data collected from many such simulations targeting different regions can then be combined[103] to obtain an accurate estimate of the overall FES. In contrast, other methods, such as metadynamics,[63,104] use a repulsive bias that is updated periodically to drive the system away from regions that it has already visited to encourage near-uniform exploration of the FES over the course of a single simulation. A crucial aspect of all of these techniques is that the bias, in principle, does not impact the computed free energy. The free energy computed with such methods should, therefore, be an estimate of the one that would be obtained by the performance of very long unbiased simulations to evaluate Eq. 2.

One recent application of advanced free-energy techniques has been the study of the potential of ionic liquids as solvents for the dissolution and separation of lignocellulose biomass into its principal components.[105] Although biomass dissolution is not well understood, it has been posited that solvent-induced conformational changes in the structure of cellulose might play an important role in determining the thermodynamics of this process.[106] To investigate this possibility, Jarin and Pfaendtner[105] used an advanced free-energy method to examine the

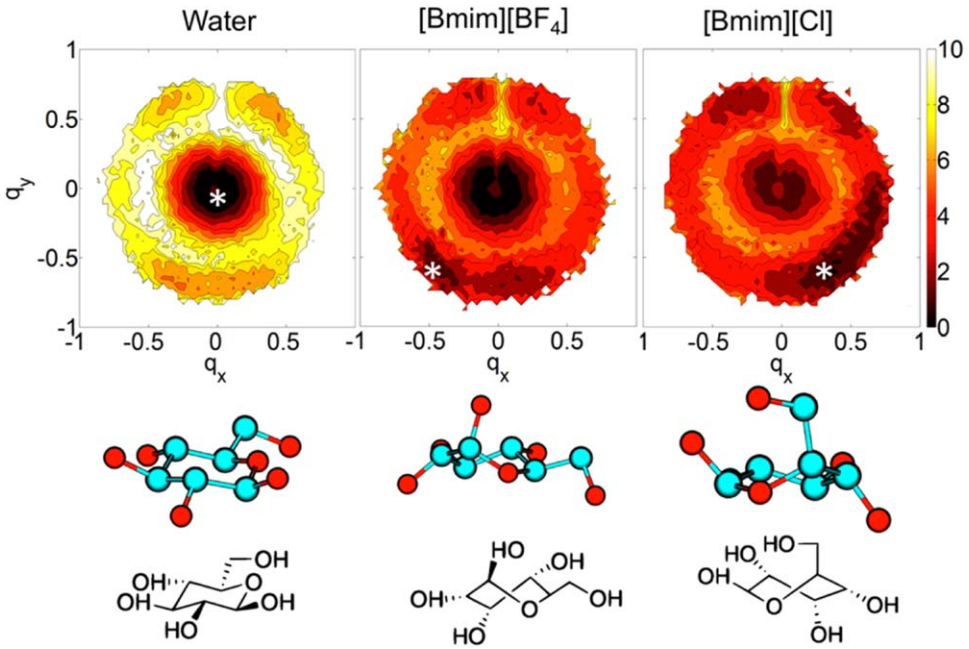

Figure 8. FESs (kcal/mol) parameterized by ring-puckering coordinates (q_x, q_y) for glucose solvated with water, [Bmim][BF_4], and [Bmim][Cl], respectively.

The asterisk on each FES marks the region in the puckering coordinate space corresponding to the stable glucose conformer. An illustration of the stable conformer is shown directly below the FES for each solvent. Reprinted with permission from ref. 105. Copyright 2014 American Chemical Society.

relative stability of different conformations of saccharide ring structures when they were solvated with water and the ionic liquids 1-butyl-3-methylimidazolium chloride ([Bmim][Cl]) and butyl-3-methylimidazoulim boron tetrafluoride ([Bmim][BF_4]). Glucose was used as the model saccharide in their simulations because of its simple structure and the fact that it is the fundamental building block of cellulose.[105] The free energy was computed with a variant of the metadynamics technique of Parrinello and coworkers[63,104] that was augmented with an advanced parallel tempering scheme.[107] As described previously, the metadynamics method uses a history-dependent repulsive bias potential to drive the system away from regions of the order parameter space that it has already visited.[63,104] The bias, therefore, encourages the system to overcome free-energy barriers and explore regions that have been undersampled. Because the free energy is proportional to the logarithm of the probability of observing the system in a given state (Eq. 2), it can be computed directly from the history-dependent bias potential,[63,104] which in effect, keeps track of the states visited during the simulation.

To study glucose's ring conformations, Jarin and Pfaendtner[105] performed metadynamics simulations using order parameters known as *puckering coordinates,* which are specifically designed to describe the puckering of ringlike structures as they deviate from a planar geometry.[108] Figure 8 shows the FES computed as a function of these order parameters for glucose in three different solvents, with the asterisk on each plot indicating the global free-energy minimum. The stable conformers of glucose corresponding to these minima are also shown directly below each surface in Figure 8. The results demonstrate that the stable chair conformer of glucose in water (C_1) became metastable with respect to the skewed conformers S_3 and S_0 when the solvent was changed to [Bmim][BF_4] and [Bmim][Cl], respectively. Subsequent analysis revealed that the skewed conformers were stabilized in the ionic liquid solvents by energetic and not entropic driving forces. Jarin and Pfaendtner[105] also posited that the anions play an important role in the dissolution of biomass in ionic liquids by disrupting the hydrogen bonding between the chains of the cellulose microfibril. They suggest that this widely accepted mechanism is supported by the fact that different stable conformers were observed in their simulations when [Bmim][BF_4] and [Bmim][Cl] were used as solvents; this indicated that interactions between anions and the hydrogen atoms on glucose were likely responsible.[105]

Free-energy methods have also been recently applied to examine the interactions between colloidal nanoparticles embedded in nematic liquid crystals.[32] Such particles can exhibit unusual effective interactions that are mediated by the structure of liquid-crystal molecules near their surface.[32] This structure is, in turn, controlled by anchoring effects between the nanoparticles and liquid-crystal molecules that arise because of their mutual interactions. The ability to judiciously tune such interactions to develop new colloidal liquid-crystal composites could result in improved materials for designing technologies ranging from semiconductors to sensors.[109] To this end, de Pablo and coworkers[32] performed MD simulations to study the behavior of nanoparticles in nematic liquid crystals. In their CGM, the nanoparticles and liquid-crystal molecules were described as simple spheres and oblate ellipsoidal particles, respectively. Two types of anchoring potentials were considered. The first was used to encourage planar anchoring, where the liquid-crystal molecules align parallel to the surface of the nanoparticles. Homeotropic anchoring was also considered, in which the perpendicular alignment of liquid-crystal molecules is favorable. The effective interactions between the colloidal nanoparticles were systematically examined in these models with metadynamics to compute the free energy as a function of their center-to-center separation distance.

de Pablo and coworkers[32] observed a rich variety of physical behaviors in their simulations that were found to be sensitive to the density of the liquid-crystal phase and the relative strength of the planar and homeotropic anchoring interactions. Figure 9, for example, shows the free-energy profiles as a function of the particle separation distances that were computed for systems with weakly planar and strongly homeotropic anchoring interactions. The left and right images in the bottom panel of Figure 9 also show configurations for these two cases, respectively. In the case of weak planar anchoring, there was a single minimum in the free energy corresponding to configurations where the nanoparticles were in close proximity, sharing layers of liquid-crystal molecules that wet their surface. For strongly homeotropic anchoring interactions, however, the free energy exhibited two minima caused by interactions between the halolike structures of bound liquid-crystal molecules that formed around each nanoparticle and the surrounding interstitial fluid.[32] The metastable minimum at approximately 10σ, where σ is the minor axis of a liquid-crystal molecule, corresponded to the halos coming into contact, as illustrated

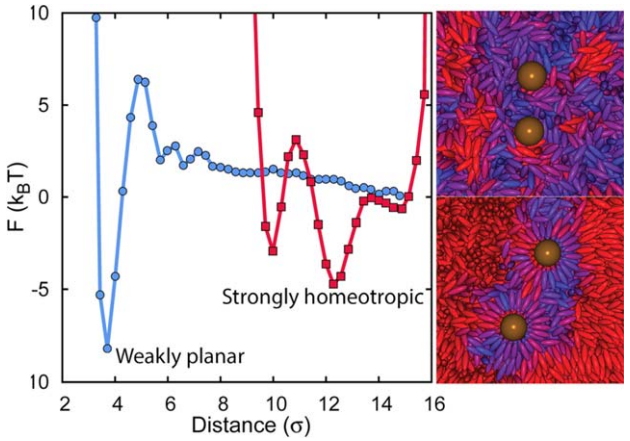

Figure 9. Free energy as a function of separation distance for nanoparticles embedded in liquid crytals with weakly planar and strongly homeotropic anchoring interactions, respectively.

The x-axis of the left figure is reported in units of the diameter of the minor axis of a liquid crystal molecule (σ). The figure on the top right shows the nanoparticles (gold spheres) near their favored separation distances in the case of weakly planar anchoring interactions, whereas the figure on the bottom right illustrates the halos of liquid crystal molecules that form around the nanoparticles in the case of strongly homeotropic anchoring interactions. Liquid crystal particles with strong and weak nematic ordering are colored blue and red, respectively. Adapted with permission from ref. 32. Copyright 2013, AIP Publishing LLC.

by the image on the right in the bottom panel of Figure 9. Because the free-energy profiles exhibited barriers in excess of about $6k_BT$, it was extremely unlikely that even a long, unbiased simulation would visit all of the relevant states of interest for these systems. In the case of weakly planar anchoring interactions, for instance, unbiased simulations initiated with the two nanoparticles separated by more than 5σ would be unlikely to overcome the free-energy barrier necessary to explore regions near the global free-energy minimum where the particles are in close proximity. The fact that such behavior could be missed completely demonstrates that advanced free-energy methods, such as the one used by de Pablo and coworkers, are necessary to explore the full range of relevant behaviors, even in a highly coarse-grained system.

Conclusions

The range of physical systems that can be studied with molecular simulations has increased dramatically over the last several decades. This progress has been driven not only by major advances in the speed of computational hardware but also by the development of new computational algorithms that have allowed increasingly complex problems to be examined. Despite these great strides, however, many challenges still remain in the adaptation of molecular simulation techniques to study the broad range of systems that fall within the scope of modern chemical engineering. One of the most significant outstanding challenges is the development of strategies to tackle systems that are governed by mesoscopic length and timescales. As this Perspective demonstrates, chemical engineers have been at the forefront in the development and application of state-of-the-art computational techniques to overcome the length and timescale limitations of traditional molecular simulation methods. These sustained efforts have been extremely fruitful and have led to an improved molecular-level perspective of complex processes ranging from nucleation to protein folding.

Although our Perspective has focused on advanced computational strategies for overcoming length and timescale limitations, it should be stressed that chemical engineers have also worked in parallel and in collaboration with physicists and chemists to make significant contributions to other important areas of molecular simulation. One related area is the development of efficient code and algorithms to use new types of computational hardware. Graphical processing units, for instance, allow simulations to be readily parallelized; this accelerates the calculations and helps to overcome potential length and timescale limitations. Because of their relatively high performance-to-cost ratio, graphical processing units have enabled many researchers to gain access to superior computation resources. Chemical engineers have been among those leading serious efforts to develop publically available codes and improved algorithms to take full advantage of these new resources.[110–113] Another particularly noteworthy area is the application of molecular simulation to the solution of problems that are broad in scope or combinatorial in nature. Such problems pose challenges that are distinct from those faced when one deals with length and timescale limitations. Chemical engineers have developed, for instance, new approaches to screen massive libraries of materials to identify promising candidates for applications such as carbon capture and natural gas storage.[46,58] Finally, chemical engineers have also made great advances in the field of quantum chemistry, a closely related discipline, by developing and applying such methods to improve catalyst design and enhance the performance of materials in a variety of environmental and energy-related applications.[67–70]

As history demonstrates,[1,2] computational studies that represent the current state of the art will become routine over the next decade as improvements in hardware and algorithm design continue to propel molecular simulation forward and rapidly expand the range of physical length and timescales that can be investigated with such methods. In looking forward, we, therefore, anticipate that molecular simulation will play an increasingly important role in chemical engineering, chemistry, physics, biology, and other scientific disciplines, complementing the next generation of experimental techniques designed to probe the molecular-level interactions and driving forces that govern the behavior of matter. The insights gained from combined computational and experimental investigations will help to achieve an improved understanding of nature's design and develop new strategies and technologies aimed at addressing many of the grand challenges faced by humanity in the next century, for example, in the areas of global climate change, sustainability, and medicine. Chemical engineers are uniquely positioned to play a central role in these exciting developments thanks to sustained efforts of the last several decades that have helped to make molecular simulation an integral part of the discipline.

Acknowledgments

J.C.P. gratefully acknowledges seed funds provided by the University of Houston and P.G.D. acknowledges the support of the National Science Foundation (contract grant numbers CBET-1263565 and CHE-1213343).

Literature Cited

1. de Pablo JJ, Escobedo FA. Molecular simulations in chemical engineering: present and future. *AIChE J*. 2002;48(12):2716–2721.
2. Gubbins KE, Moore JD. Molecular modeling of matter: impact and prospects in engineering. *Ind Eng Chem Res*. 2010;49(7):3026–3046.
3. Mc Cammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature*. 1977;267(5612):585–590.
4. Shaw DE, Maragakis P, Lindorff-Larsen K, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*. 2010;330(6002):341–346.
5. Panagiotopoulos AZ. Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol Phys*. 1987;61(4):813–826.
6. Wang K, Chodera JD, Yang YZ, Shirts MR. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J Comput Aid Mol Des*. 2013;27(12):989–1007.
7. Telesco SE, Shih AJ, Jia F, Radhakrishnan R. A multiscale modeling approach to investigate molecular mechanisms of pseudokinase activation and drug resistance in the HER3/ErbB3 receptor tyrosine kinase signaling network. *Mol Biosyst*. 2011;7(6):2066–2080.

8. Hatch HW, Debenedetti PG. Molecular modeling of mechanical stresses on proteins in glassy matrices: formalism. *J Chem Phys*. 2012;137(3):035103.
9. Santiso EE, Musolino N, Trout BL. Design of linear ligands for selective separation using a genetic algorithm applied to molecular architecture. *J Chem Inf Model*. 2013;53(7):1638–1660.
10. Kim YC, Mittal J. Crowding induced entropy-enthalpy compensation in protein association equilibria. *Phys Rev Lett*. 2013;110(20):208102.
11. Hatch HW, Stillinger FH, Debenedetti PG. Computational study of the stability of the miniprotein Trp-cage, the GB1 β-hairpin, and the AK16 peptide, under negative pressure. *J Phys Chem B*. 2014;118(28): 7761–7769.
12. Blanco MA, Sahin E, Robinson AS, Roberts CJ. Coarse-grained model for colloidal protein interactions, B-22, and protein cluster formation. *J Phys Chem B*. 2013; 117(50):16013–16028.
13. Wagoner VA, Cheon M, Chang I, Hall CK. Computer simulation study of amyloid fibril formation by palindromic sequences in prion peptides. *Proteins*. 2011; 79(7):2132–2145.
14. Santiso EE, Herdes C, Muller EA. On the calculation of solid-fluid contact angles from molecular dynamics. *Entropy*. 2013;15(9):3734–3745.
15. Kumar V, Errington JR. Impact of small-scale geometric roughness on wetting behavior. *Langmuir*. 2013;29(38): 11815–11820.
16. Sharma S, Debenedetti PG. Evaporation rate of water in hydrophobic confinement. *Proc Natl Acad Sci U S A*. 2012;109(12):4365–4370.
17. Chaudhari MI, Holleran SA, Ashbaugh HS, Pratt LR. Molecular-scale hydrophobic interactions between hard-sphere reference solutes are attractive and endothermic. *Proc Natl Acad Sci U S A*. 2013;110(51):20557–20562.
18. Vembanur S, Patel AJ, Sarupria S, Garde S. On the thermodynamics and kinetics of hydrophobic interactions at interfaces. *J Phys Chem B*. 2013;117(35):10261–10270.
19. Patel AJ, Garde S. Efficient method to characterize the context-dependent hydrophobicity of proteins. *J Phys Chem B*. 2014;118(6):1564–1573.
20. Palmer JC, Debenedetti PG. Computer simulation of water sorption on flexible protein crystals. *J Phys Chem Lett*. 2012;3(18):2713–2718.
21. Knott BC, Molinero V, Doherty MF, Peters B. Homogeneous nucleation of methane hydrates: unrealistic under realistic conditions. *J Am Chem Soc*. 2012;134(48): 19544–19547.
22. Uline MJ, Torabi K, Corti DS. Homogeneous nucleation and growth in simple fluids. I. Fundamental issues and free energy surfaces of bubble and droplet formation. *J Chem Phys*. 2010;133(17):174511.
23. Shah M, Santiso EE, Trout BL. Computer simulations of homogeneous nucleation of benzene from the melt. *J Phys Chem B*. 2011;115(35):10400–10412.
24. Sarupria S, Debenedetti PG. Homogeneous nucleation of methane hydrate in microsecond molecular dynamics simulations. *J Phys Chem Lett*. 2012;3(20):2942–2947.
25. Eggimann BL, Sunnarborg AJ, Stern HD, Bliss AP, Siepmann JI. An online parameter and property database for the TraPPE force field. *Mol Simulat*. 2014;40(1-3):101–105.
26. Ketko MH, Kamath G, Potoff JJ. Development of an optimized intermolecular potential for sulfur dioxide. *J Phys Chem B*. 2011;115(17):4949–4954.
27. Garcia-Cuellar AJ, Chapman WG. Theory and simulation for associating cyclic molecules. *Mol Phys*. 2011; 109(14):1813–1820.
28. Kim HM, Schultz AJ, Kofke DA. Second through fifth virial coefficients for model methane-ethane mixtures. *Fluid Phase Equilib*. 2013;351:69–73.
29. Palmer JC, Martelli F, Liu Y, Car R, Panagiotopoulos AZ, Debenedetti PG. Metastable liquid-liquid transition in a molecular model of water. *Nature*. 2014;510(7505):385–388.
30. Wu H, Maginn EJ. Water solubility and dynamics of CO_2 capture ionic liquids having aprotic heterocyclic anions. *Fluid Phase Equilib*. 2014;368:72–79.
31. Liu HN, Zhang ZT, Bara JE, Turner CH. Electrostatic potential within the free volume space of imidazole-based solvents: insights into gas absorption selectivity. *J Phys Chem B*. 2014;118(1):255–264.
32. Whitmer JK, Joshi AA, Roberts TF, de Pablo JJ. Liquid-crystal mediated nanoparticle interactions and gel formation. *J Chem Phys*. 2013;138(19):194903.
33. Abbott LJ, Hart KE, Colina CM. Polymatic: a generalized simulated polymerization algorithm for amorphous polymers. *Theor Chem Acc*. 2013;132(3):1334.
34. Harmandaris V, Doxastakis M. Molecular dynamics of polyisoprene/polystyrene oligomer blends: the role of self-concentration and fluctuations on blend dynamics. *J Chem Phys*. 2013;139(3):034904.
35. Bayramoglu B, Faller R. Modeling of polystyrene under confinement: exploring the limits of iterative Boltzmann inversion. *Macromolecules*. 2013;46(19):7957–7976.
36. Vogiatzis GG, Theodorou DN. Local segmental dynamics and stresses in polystyrene-C-60 mixtures. *Macromolecules*. 2014;47(1):387–404.
37. Kalathi JT, Yamamoto U, Schweizer KS, Grest GS, Kumar SK. Nanoparticle diffusion in polymer nanocomposites. *Phys Rev Lett*. 2014;112(10):108301.
38. Lafitte T, Kumar SK, Panagiotopoulos AZ. Self-assembly of polymer-grafted nanoparticles in thin films. *Soft Matter*. 2014;10(5):786–794.
39. Jayaraman A. Polymer grafted nanoparticles: effect of chemical and physical heterogeneity in polymer grafts on particle assembly and dispersion. *J Polym Sci Part A Polym Phys*. 2013;51(7):524–534.
40. Avendano C, Watson CML, Escobedo FA. Directed self-assembly of spherical caps via confinement. *Soft Matter*. 2013;9(38):9153–9166.
41. Long AW, Ferguson AL. Nonlinear machine learning of patchy colloid self-assembly pathways and mechanisms. *J Phys Chem B*. 2014;118(15):4228–4244.
42. Muller M, de Pablo JJ. Computational approaches for the dynamics of structure formation in self-assembling polymeric materials. *Annu Rev Mater Res*. 2013;43:1–34.
43. Guo S, Moore TC, Iacovella CR, Strickland LA, McCabe C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid

headgroup chemistry. *J Chem Theory Comput*. 2013; 9(11):5116–5126.
44. Ferreiro-Rangel CA, Gelb LD. Investigation of the bulk modulus of silica aerogel using molecular dynamics simulations of a coarse-grained model. *J Phys Chem B*. 2013;117(23):7095–7105.
45. Fichthorn KA. Atomic-scale theory and simulations for colloidal metal nanocrystal growth. *J Chem Eng Data*. 2014;59:3113–3119.
46. Wilmer CE, Leaf M, Lee CY, et al. Large-scale screening of hypothetical metal-organic frameworks. *Nat Chem*. 2012;4(2):83–89.
47. Marson RL, Phillips CL, Anderson JA, Glotzer SC. Phase behavior and complex crystal structures of self-assembled tethered nanoparticle telechelics. *Nano Lett*. 2014;14(4):2071–2078.
48. Landers J, Gor GY, Neimark AV. Density functional theory methods for characterization of porous materials. *Colloid Surf A*. 2013;437:3–32.
49. Jin L, Auerbach SM, Monson PA. Simulating the formation of surfactant-templated mesoporous silica materials: a model with both surfactant self-assembly and silica polymerization. *Langmuir*. 2013;29(2):766–780.
50. Palmer JC, Moore JD, Brennan JK, Gubbins KE. Simulating local adsorption isotherms in structurally complex porous materials: a direct assessment of the slit pore model. *J Phys Chem Lett*. 2011;2(3):165–169.
51. Long Y, Palmer JC, Coasne B, Sliwinska-Bartkowiak M, Gubbins KE. Pressure enhancement in carbon nanopores: a major confinement effect. *Phys Chem Chem Phys*. 2011;13(38):17163–17170.
52. Shen VK, Siderius DW. Elucidating the effects of adsorbent flexibility on fluid adsorption using simple models and flat-histogram sampling methods. *J Chem Phys*. 2014;140(24):244106.
53. Jain A, Bollinger JA, Truskett TM. Inverse methods for material design. *AIChE J*. 2014;60(8):2732–2740.
54. Amrouche H, Creton B, Siperstein F, Nieto-Draghi C. Prediction of thermodynamic properties of adsorbed gases in zeolitic imidazolate frameworks. *RSC Adv*. 2012;2(14):6028–6035.
55. Builes S, Vega LF. Effect of immobilized amines on the sorption properties of solid materials: impregnation versus grafting. *Langmuir*. 2013;29(1):199–206.
56. Xiong RC, Sandler SI, Vlachos DG. Molecular screening of alcohol and polyol adsorption onto MFI-type zeolites. *Langmuir*. 2012;28(9):4491–4499.
57. Bucior BJ, Chen DL, Liu JC, Johnson JK. Porous carbon nanotube membranes for separation of H_2/CH_4 and CO_2/CH_4 mixtures. *J Phys Chem C*. 2012;116(49):25904–25910.
58. Lin LC, Berger AH, Martin RL, et al. In silico screening of carbon-capture materials. *Nat Mater*. 2012;11(7):633–641.
59. Fang HJ, Kamakoti P, Ravikovitch PI, Aronson M, Paur C, Sholl DS. First principles derived, transferable force fields for CO_2 adsorption in Na-exchanged cationic zeolites. *Phys Chem Chem Phys*. 2013;15(31):12882–12894.
60. Rajput NN, Monk J, Hung FR. Ionic liquids confined in a realistic activated carbon model: a molecular simulation study. *J Phys Chem C*. 2014;118(3):1540–1553.
61. Clancy P. Application of molecular simulation techniques to the study of factors affecting the thin-film morphology of small-molecule organic semiconductors. *Chem Mater*. 2011;23(3):522–543.
62. Li S, Van Aken KL, McDonough JK, Feng G, Gogotsi Y, Cummings PT. The electrical double layer of dicationic ionic liquids at onion-like carbon surface. *J Phys Chem C*. 2014;118(8):3901–3909.
63. Singh S, Chopra M, de Pablo JJ. Density of states-based molecular simulations. *Annu Rev Chem Biomol*. 2012;3: 369–394.
64. Theodorou DN. Progress and outlook in Monte Carlo simulations. *Ind Eng Chem Res*. 2010;49(7):3047–3058.
65. Kofke DA. Free energy methods in molecular simulation. *Fluid Phase Equilib*. 2005;228:41–48.
66. Escobedo FA, Borrero EE, Araque JC. Transition path sampling and forward flux sampling. Applications to biological systems. *J Phys-Condens Mater*. 2009;21(33):333101.
67. Ghorbanpour A, Rimer JD, Grabow LC. Periodic, vdW-corrected density functional theory investigation of the effect of Al siting in H-ZSM-5 on chemisorption properties and site-specific acidity. *Catal Commun*. 2014;52:98–102.
68. Saavedra J, Doan HA, Pursell CJ, Grabow LC, Chandler BD. The critical role of water at the gold-titania interface in catalytic CO oxidation. *Science*. 2014;345:1599–1602.
69. Keith JA, Carter EA. Theoretical insights into pyridinium-based photoelectrocatalytic reduction of CO_2. *J Am Chem Soc*. 2012;134(18):7580–7583.
70. Zope BN, Hibbitts DD, Neurock M, Davis RJ. Reactivity of the gold/water interface during selective oxidation catalysis. *Science*. 2010;330(6000):74–78.
71. Frenkel D, Smit B. *Understanding Molecular Simulation: From Algorithms to Applications*. 2nd ed. San Diego, CA: Academic Press; 2002.
72. Maginn EJ, Elliott JR. Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Ind Eng Chem Res*. 2010;49(7):3059–3078.
73. Gubbins KE, Liu YC, Moore JD, Palmer JC. The role of molecular modeling in confined systems: impact and prospects. *Phys Chem Chem Phys*. 2011;13(1):58–85.
74. Radhakrishnan R, Trout B. Order parameter approach to understanding and quantifying the physico-chemical behavior of complex systems. In: Yip S, ed. *Handbook of Materials Modeling*. Dordrecht, the Netherlands: Springer; 2005:1613–1626.
75. Saunders MG, Voth GA. Coarse-graining of multiprotein assemblies. *Curr Opin Struct Biol*. 2012;22(2):144–150.
76. Moller P, Oddershede L. Small-scale nonequilibrium systems. In: Sattler KD, ed. *Handbook of Nanophysics*. Boca Raton, FL: CRC Press; 2010:1–16.
77. Eckhardt W, Heinecke A, Bader R, et al. 591 TFLOPS multi-trillion particles simulation on SuperMUC. In: Kunkel J, Ludwig T, Meuer H, eds. *Supercomputing*. Vol 7905. Berlin, Germany: Springer; 2013:1–12.
78. Martin MG, Siepmann JI. Transferable potentials for phase equilibria. 1. United-atom description of *n*-alkanes. *J Phys Chem B*. 1998;102(14):2569–2577.
79. Molinero V, Moore EB. Water modeled as an intermediate element between carbon and silicon. *J Phys Chem B*. 2009;113(13):4008–4016.

80. Anderson R, Llamedo M, Tohidi B, Burgass RW. Experimental measurement of methane and carbon dioxide clathrate hydrate equilibria in mesoporous silica. *J Phys Chem B*. 2003;107(15):3507–3514.
81. Vatamanu J, Kusalik PG. Observation of two-step nucleation in methane hydrates. *Phys Chem Chem Phys*. 2010; 12(45):15065–15072.
82. Muller EA, Jackson G. Force-field parameters from the SAFT-γ equation of state for use in coarse-grained molecular simulations. *Annu Rev Chem Biomol Eng*. 2014;5: 405–427.
83. Mejía A, Cartes M, Segura H, Müller EA. Use of equations of state and coarse grained simulations to complement experiments: describing the interfacial properties of carbon dioxide + decane and carbon dioxide + eicosane mixtures. *J Chem Eng Data*. 2014;59:2928–2941.
84. Avendano C, Lafitte T, Galindo A, Adjiman CS, Jackson G, Muller EA. SAFT-γ force field for the simulation of molecular fluids. 1. A single-site coarse grained model of carbon dioxide. *J Phys Chem B*. 2011;115(38):11154–11169.
85. Avendano C, Lafitte T, Adjiman CS, Galindo A, Muller EA, Jackson G. SAFT-γ force field for the simulation of molecular fluids: 2. Coarse-grained models of greenhouse gases, refrigerants, and long alkanes. *J Phys Chem B*. 2013;117(9):2717–2733.
86. Chapman WG, Gubbins KE, Jackson G, Radosz M. SAFT: equation-of-state solution model for associating fluids. *Fluid Phase Equilib*. 1989;52:31–38.
87. Aimoli CG, Maginn EJ, Abreu CRA. Transport properties of carbon dioxide and methane from molecular dynamics simulations. *J Chem Phys*. 2014;141(13):134101.
88. Huie NC, Luks KD, Kohn JP. Phase-equilibriums behavior of systems carbon dioxide-*n*-eicosane and carbon dioxide-*n*-decane-*n*-eicosane. *J Chem Eng Data*. 1973;18(3):311–313.
89. Shell MS. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J Chem Phys*. 2008;129(14):144108.
90. Izvekov S, Voth GA. A multiscale coarse-graining method for biomolecular systems. *J Phys Chem B*. 2005; 109(7):2469–2473.
91. Reith D, Putz M, Muller-Plathe F. Deriving effective mesoscale potentials from atomistic simulations. *J Comput Chem*. 2003;24(13):1624–1636.
92. Carmichael SP, Shell MS. A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly. *J Phys Chem B*. 2012;116(29):8383–8393.
93. Koo EH, Lansbury PT, Kelly JW. Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc Natl Acad Sci U S A*. 1999;96(18):9989–9990.
94. Nguyen HD, Hall CK. Phase diagrams describing fibrillization by polyalanine peptides. *Biophys J*. 2004;87(6): 4122–4134.
95. Peters B, Trout BL. Obtaining reaction coordinates by likelihood maximization. *J Chem Phys*. 2006;125(5):054108.
96. Santiso EE, Trout BL. A general set of order parameters for molecular crystals. *J Chem Phys*. 2011;134(6):064109.
97. Velez-Vega C, Borrero EE, Escobedo FA. Kinetics and mechanism of the unfolding native-to-loop transition of Trp-cage in explicit solvent via optimized forward flux sampling simulations. *J Chem Phys*. 2010;133(10): 105103.
98. Juraszek J, Bolhuis PG. Rate constant and reaction coordinate of Trp-Cage folding in explicit water. *Biophys J*. 2008;95(9):4246–4257.
99. Allen RJ, Warren PB, ten Wolde PR. Sampling rare switching events in biochemical networks. *Phys Rev Lett*. 2005;94(1):018104.
100. Borrero EE, Escobedo FA. Reaction coordinates and transition pathways of rare events via forward flux sampling. *J Chem Phys*. 2007;127(16):164101.
101. Qiu LL, Pabit SA, Roitberg AE, Hagen SJ. Smaller and faster: the 20-residue Trp-cage protein folds in 4 μs. *J Am Chem Soc*. 2002;124(44):12952–12953.
102. Torrie GM, Valleau JP. Non-physical sampling distributions in Monte-Carlo free-energy estimation: umbrella sampling. *J Comput Phys*. 1977;23(2):187–199.
103. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys*. 2008;129(12):124105.
104. Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett*. 2008;100(2):020603.
105. Jarin Z, Pfaendtner J. Ionic liquids can selectively change the conformational free-energy landscape of sugar rings. *J Chem Theory Comput*. 2014;10(2):507–510.
106. Tadesse H, Luque R. Advances on biomass pretreatment using ionic liquids: an overview. *Energ Environ Sci*. 2011;4(10):3913–3929.
107. Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble. *Phys Rev Lett*. 2010;104(19): 190601.
108. Cremer D, Pople JA. General definition of ring puckering coordinates. *J Am Chem Soc*. 1975;97(6):1354–1358.
109. Kato T, Hirai Y, Nakaso S, Moriyama M. Liquid-crystalline physical gels. *Chem Soc Rev*. 2007;36(12):1857–1867.
110. Anderson JA, Lorenz CD, Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J Comput Phys*. 2008; 227(10):5342–5359.
111. Nguyen TD, Phillips CL, Anderson JA, Glotzer SC. Rigid body constraints realized in massively-parallel molecular dynamics on graphics processing units. *Comput Phys Commun*. 2011;182(11):2307–2313.
112. Mick J, Hailat E, Russo V, Rushaidat K, Schwiehert L, Potoff J. GPU-accelerated Gibbs ensemble Monte Carlo simulations of Lennard-Jonesium. *Comput Phys Commun*. 2013;184(12):2662–2669.
113. Daly KB, Benziger JB, Debenedetti PG, Panagiotopoulos AZ. Massively parallel chemical potential calculation on graphics processing units. *Comput Phys Commun*. 2012; 183(10):2054–2062.