

PERSPECTIVE

Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems

Michael L. Klein^{1*} and Wataru Shinoda^{2*}

Relentless increases in the size and performance of multiprocessor computers, coupled with new algorithms and methods, have led to novel applications of simulations across chemistry. This Perspective focuses on the use of classical molecular dynamics and so-called coarse-grain models to explore phenomena involving self-assembly in complex fluids and biological systems.

The proliferation of multiprocessor computers, combined with efficiently parallelized dynamics codes (1, 2) and powerful visualization software (3), has heralded an explosive growth in applications of computation across chemistry. Now multiteraflop machines with thousands of processors can be accessed for extended periods of time. This in turn has led to classical molecular dynamics (MD) simulations for systems comprising a million atoms or so (4–7), which is sufficiently large to allow the components of Nature's machinery of life to be probed. Protein folding (8) and even molecular motors (4) are all yielding to the calculation power of such machines. Perhaps the best is yet to come, with the promise of petaflop machines in the near future (9) and the prospect of being able to use cheap high-performance graphics processors for MD studies (10, 11).

These massive calculations are indeed successes for high-performance computing, notwithstanding the fact that the time scale of the MD trajectories on million-atom systems is typically less than 100 ns. However, in soft matter, the plethora of structures formed by self-assembling macromolecules typically involve micrometer-length scales embracing hundreds of millions of atoms, and phase transformations between structures occur on much longer time scales. Thus, such phenomena are currently inaccessible to brute-force classical MD treatment based on individual atoms. Also, in biological systems, phenomena such as lipid raft formation, and membrane fusion (12), system size, and especially the time scale, remain key issues.

Thus, a crucial question in carrying out large-scale MD simulations is, just how large a system does one need to access the phenomenon of

interest? Which in the present context means, can one bridge scales ranging from tens of nanometers to micrometers? Even more important, how long should the MD simulation run, or can one bridge scales from tens of nanoseconds to milliseconds? To deal with these issues in soft matter and biosystems, so-called coarse grain (CG) models have been developed that allow MD to simulate at the mesoscale with system sizes approaching the micrometer range and phenomena taking place on time scales approaching milliseconds and beyond (13). In such models, the intermolecular potentials are either fitted to all-atom (AA) MD simulations (14) and/or experimental data. In the case of atomistic models, the forces required to drive MD simulations typically are based on interatomic interaction potentials (2)

derived in part from electronic structure calculations. Today's AA potentials can handle macromolecules, proteins, and DNA with remarkable success, especially given the inherent simplicity of the force fields that are mostly based on pairwise additive interaction potentials.

To illustrate some of the current possibilities for large-scale AA-MD and CG-MD computations in chemistry, we examine two examples. The first deals with the self-assembly of surfactant molecules and illustrates how structural transformations are now accessible to MD simulations, albeit with CG models. The second illustrates how a membrane protein can induce structural changes (curvature) in a biological membrane (7). In the latter case, both classical atomistic MD simulations and CG models have been used to great effect. For work that also has pushed the field to new frontiers, see (18–20).

Complex Fluids

The polymer community long ago adopted CG models with considerable success. Seminal studies by Smit *et al.* (13) inspired subsequent work on surfactants and lipids (14, 15). The amphiphilic nature of surfactants leads to their aggregation and self-assembly into micelles, vesicles, and more complex morphologies under appropriate conditions. Understanding the role of molecular architecture in determining phase morphology and physical properties is a natural goal of MD simulations, which can then inform the design and synthesis of new molecules of interest to industry. However, even with generous access to multi-teraflop machines, the spontaneous generation of

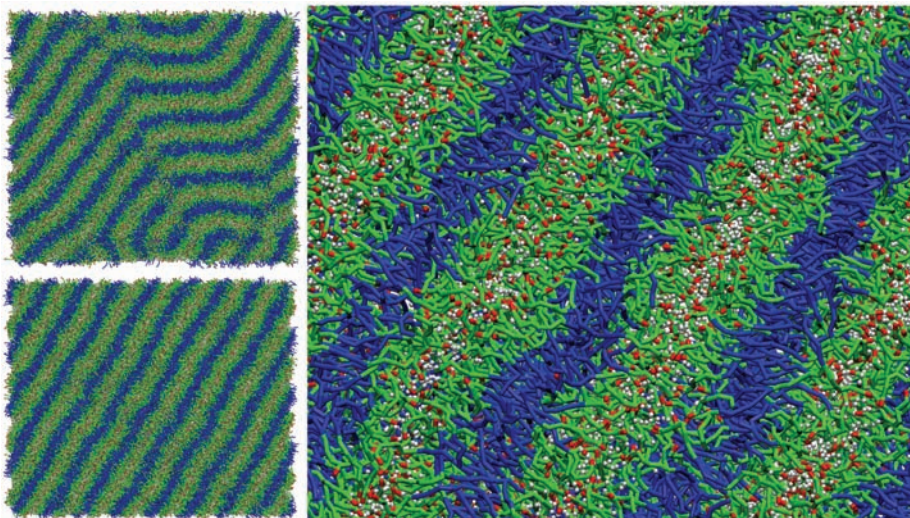


Fig. 1. Images taken from a MD simulation of an aqueous surfactant solution. The simulation system contains approximately 800,000 CG particles and represents a solution with 80% surfactant $C_{12}E_6$ and 20% water. The snapshots were taken after 100 ns (**top**) and ~ 500 ns (**bottom and right**) of MD simulation, starting from an appropriately hydrated hexagonal phase. The surfactant hydrophobic tails are shown in blue, the head groups in green and red, and the water in white. The phase transformation from initial hexagonal to final lamellar phase passes through a dislocated structure (top) before achieving the equilibrium structure (bottom). The image on the right shows expanded detail from the final structure. The CG parameters used in the MD simulations were taken from (14).

¹Center for Molecular Modeling, Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA.

²Research Institute of Computational Sciences, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba 305-8568, Japan.

*To whom correspondence should be addressed. E-mail: klein@lrsm.upenn.edu (M.L.K.); w.shinoda@aist.go.jp (W.S.)

vesicles and wormlike structures starting from randomly dispersed surfactants in solution is highly challenging because of the time and length scales involved. The study of surfactant self-assembly into, for example, a cubic phase is typically beyond the capabilities of current computational resources if one desires an AA representation. Enhanced sampling techniques and CG representations of the surfactant and polymer molecules have therefore been developed.

With the present generation of CG models (14), it is possible to readily simulate systems containing over 1 million CG particles, corresponding to atomic systems of ~10 million atoms. Moreover, with a few thousand processors, it is possible to generate in a day MD trajectories that span about 100 ns. Such simulations are thus effectively approaching the regime relevant to laboratory experiments. A typical example of such a system, from our recent work (14), is shown in Fig. 1. The CG parameters used in the MD simulation are from (14). The system included more than 800,000 CG particles, composed of ~62,000 CG $C_{12}E_6$ molecules and more than 500,000 CG water particles (14). With an AA approach, this system would require nearly 5 million atoms to represent the solvent alone. The simulation was set up by starting with an equilibrated system containing 50 weight percent (wt %) $C_{12}E_6$ in the known hexagonal phase and slowly removing the water to obtain a system composed of 80 wt % $C_{12}E_6$. Fig. 1 shows

images of the system as it evolves to the lamellar structure found in the experiment. After about 100 ns of MD simulation on the CG model, dislocation defects are clearly visible in the otherwise lamellar structure; at about 500 ns, the system has reached the lamellar phase, albeit with some minor defects still present. This CG model can also be applied to study Langmuir monolayers of surfactants at both the air/water and oil/water interfaces and the interaction of surfactants with nanoparticles and surfaces, to name just a couple of examples. Such CG computations are poised to become a predictive tool in surfactant design.

Biological Systems

A substantial beneficiary of multiteraflop machines has been chemical biology. Biological systems are intrinsically multiscale assemblies—micrometer-sized cells that contain integrated nanometer-scale functional units, such as protein oligomers. To date, the latter have been the main targets for MD. Successes have included folding studies on modest-sized proteins (8) and ion

transport through membrane-bound protein bundles. Substantial efforts have been directed at simulating light-harvesting systems, molecular motors (4), and even a whole virus (5). Of particular note is the work on elucidating the mechanism for membrane curvature being induced by

of interaction sites, the machine time required to compute the potential energy and forces for the system of interest is drastically reduced. Second, the resulting CG interaction “potentials” typically are much softer than in the AA case, thereby allowing a more ready exploration of phase space

because the time step to integrate the MD equations of motion can be increased. Of course, this gain in efficiency comes with a reduction in the level of chemical detail. Successes with applying the CG approach to biosystems include rationalizing the mechanism of molecular motors (4) and first-structure determination of lipoprotein complexes used to transport lipids *in vivo* (22).

Figure 2 illustrates results of a MD simulation, recently carried out by Arkhipov *et al.* (16, 17), of membrane sculpting by amphiphysin BAR-domain proteins. Specifically, six BAR domains (dark blue, green, yellow, and purple) were placed on top of a planar membrane patch consisting of zwitterionic dioleoylphosphatidylcholine lipids (light blue heads), with a 30% random mixture of dioleoylphosphatidylserine (DOPS) lipids (pink heads). The initial conformation of this system is shown viewed from the top and from the side. BAR domains scaffold the membrane with their concave, positively charged surface via electrostatic interactions with negatively charged DOPS lipids. A concerted action of BAR domains arranged in a lattice results in the development of a global membrane curvature on a

time scale of several microseconds, with the resulting curvature radius of ~30 nm being close to that observed experimentally (23).

Outlook

This Perspective has focused on the exciting progress achieved through the use of large-scale classical MD to probe self-assembling macromolecular systems. We indicated how the use of CG models has helped to solve the problem of simulating micrometer-sized systems at millisecond time scales. However, we have not discussed the important area of multiscale modeling, which in principle reaches from the quantum mechanical (QM) world of atoms and electrons to the macroscopic world of finite element modeling (engineering). Over the years, multiscale methods have been developed to handle the problem of bridging length and time scales in chemistry. For example, in enzyme catalysis, approaches linking the QM description of the active site and the chemical reaction to the classical molecular mechanics (MM) of the protein have enabled the

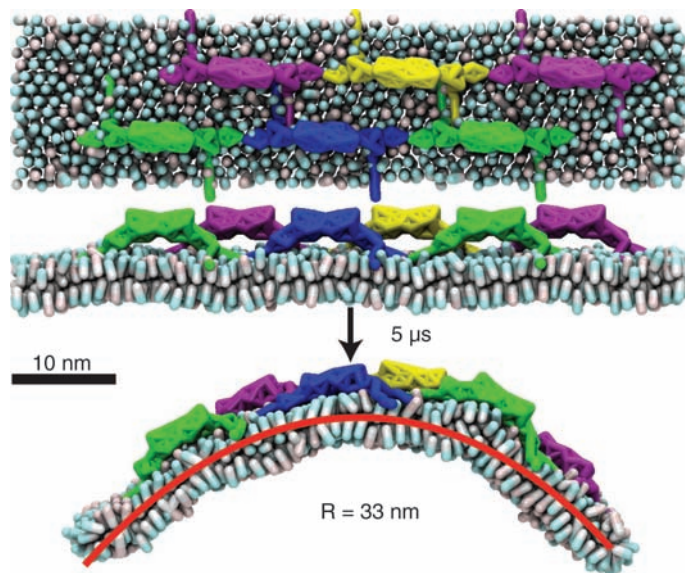


Fig. 2. Images taken from a MD simulation by Arkhipov *et al.*, (17) of biomembrane sculpting by protein-BAR domains. The simulation shown in the figure was carried out using a box with dimensions 100 by 16 by 50 nm and would correspond to a system of 10 million atoms in a fully atomistic representation. However, the computation was done using a shape-based CG model (16, 17). As a result, there were 3265 CG beads in the system. Free space is needed around the membrane to allow for the bending. A similar simulation on a system four times larger and covering ~100 μ s resulted in the spontaneous formation of a complete, stable membrane tube [private communication from the authors of (16)] with dimensions in agreement with experiment (23). R is the radius of curvature of the membrane induced by the protein-BAR domains.

so-called Bin-amphiphysin-Rvs (BAR) proteins anchored to a membrane surface (7). Understanding this process is important because this phenomenon plays a role in cell membrane restructuring associated with endocytosis, the process whereby, for example, a virus is ingested by a cell.

MD simulations of biological systems comprising about 100 million atoms, such as that needed to probe endocytosis, are on the horizon (9), but the time-scale problem will remain the biggest impediment to real progress. As with surfactants and polymers, this has led to the promulgation of CG models, which reduce the number of interacting elements by “integrating out” selected groups of atoms via various schemes. Several approaches are showing substantial progress. These include, at one extreme, whole-protein building blocks in the case of the BAR proteins (16–18) and, at the other extreme, groups of about 10 atoms typically coalesced into a single CG bead (14). In this way, amino acid residues are typically reduced to one or two beads (19–21).

Such CG models affect computational demand in two ways. First, by reducing the number

study of function via so-called QM/MM methods. Again, the issue of time scales is relevant because the QM part of the system often dominates the computation. Unless approximate QM methods are employed, it becomes very challenging to run MD trajectories long enough to achieve adequate sampling, even with multiteraflop machines (9). Thus, new methods to enhance the sampling of rare events (such as the bond breaking in catalysis) are needed to help alleviate this problem (24). Another example of a multiscale method is the linking of electronic-structure methods to finite element-engineering approaches (25) to probe phenomena occurring at defects. Such methods are also likely to find application in the world of chemical biology.

With the predicted relentless increase in available computer resources, remaining issues that arise in the investigation of complex supramolecular organization and related phenomena in biological systems may be resolved via multiscale methods. However, for the near future, many aspects of the time-scale problem are likely to persist and remain beyond the scope of brute-force AA-MD simulations, even on petaflop machines. Thus, for the near term, CG and multiscale models will proliferate and yield exciting new data (7). The application of large-scale MD to chemical problems will be dominated by chemical biology, but the key to future progress will probably reside in new methods and algorithm

development (24–26). Brute-force computation alone is going to be insufficient. Thus, the field is also poised to profit from future developments in the area of more refined methods for multiscale modeling to enable the efficient bridging between the QM, AA, CG, and finite element approaches. Finally, we have not discussed large-scale electronic structure calculations, which have been particularly successful in materials science, leading to the design of new catalysts (27). In the context of chemical biology, new functionals, and empirical corrections for van der Waals dispersion forces (28), are extending the range of validity of density functional theory to treat the interaction between large molecules such as DNA bases with high accuracy, which is likely to herald a new era for computational chemistry (29).

References and Notes

1. For example, see www.ku.uic.edu/Research/namd, <http://lammps.sandia.gov>, and www.gromacs.org.
2. For example, see www.charmm.org and <http://amber.scripps.edu>.
3. For example, see www.ku.uic.edu/Research/vmd.
4. A. Kitao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4894 (2006).
5. P. L. Freddolino, A. S. Arhipov, S. B. Larson, A. McPherson, K. Schulten, *Structure* **14**, 437 (2006).
6. K. Y. Sanbonmatsu, C.-S. Tung, *J. Struct. Biol.* **157**, 470 (2007).
7. G. S. Ayton, P. D. Blood, G. A. Voth, *Biophys. J.* **92**, 3595 (2007).
8. B. Borrell, *Nature* **451**, 240 (2008).
9. J. Markoff, *New York Times*, 9 June 2008, www.nytimes.com/2008/06/09/technology/09petaflops.html?hp.

10. J. A. Anderson, C. D. Lorenz, A. Travesset, *J. Comput. Phys.* **227**, 5342 (2008).
11. J. A. van Meel, A. Arnold, D. Frenkel, S. F. Portegies Zwart, R. G. Belleman, *Mol. Simul.* **34**, 259 (2008).
12. H.-S. Yin *et al.*, *Nature* **439**, 38 (2006).
13. B. Smit *et al.*, *Nature* **348**, 624 (1990).
14. W. Shinoda, R. DeVane, M. L. Klein, *Mol. Simul.* **33**, 27 (2007).
15. J. C. Shelley *et al.*, *J. Phys. Chem. B* **105**, 4464 (2001).
16. A. Arhipov, Y. Yin, K. Schulten, *Biophys. J.* doi:10.1529/biophysj.107.121160 (2008).
17. www.biophysj.org/cgi/rapidpdf/biophysj.108.132563v1
18. D. T. Mirijanian, G. A. Voth, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1204 (2008).
19. W. Treptow, S. J. Marrink, M. Tarek, *J. Phys. Chem. B* **112**, 3277 (2008).
20. M. Neri, M. Baaden, V. Carnevale, C. Anselmi, A. Maritan, P. Carloni, *Biophys. J.* **94**, 71 (2008).
21. L. Monticelli *et al.*, *J. Chem. Theory Comput.* **4**, 819 (2008).
22. A. Y. Shih, P. L. Freddolino, A. Arhipov, K. Schulten, *J. Struct. Biol.* **157**, 579 (2007).
23. B. J. Peter *et al.*, *Science* **303**, 495 (2004).
24. E. Vanden-Eijnden, M. Heymann, *J. Chem. Phys.* **128**, 061103 (2008).
25. E. A. Carter, *Science* **321**, 800 (2008).
26. A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
27. F. Studt *et al.*, *Science* **320**, 1320 (2008).
28. V. R. Cooper *et al.*, *J. Am. Chem. Soc.* **130**, 1304 (2008).
29. M. Head-Gordon, E. Artacho, *Phys. Today* **61**, 58 (2008).
30. We thank A. Kohlmeier, R. DeVane, and K. Anderson for discussion and K. Schulten, Y. Yin, and A. Arhipov for providing Fig. 2. This research was supported in part by NSF (grant DMR 0520020) and NIH (grant GM 40712).

10.1126/science.1157834

PERSPECTIVE

Challenges in Modeling Materials Properties Without Experimental Input

Emily A. Carter

Simulations of materials behavior are an important component of materials science research, partly because measurements are indirect, requiring theoretical interpretation, and partly because often the ideal experiment simply cannot be performed (due to technological limitations). Empirical physical models used in this context often rely on parameters drawn from experiments on simpler systems, and so introduce various inaccuracies. In contrast, a quantum mechanical model can potentially offer an independent source of data more closely attuned to the complexities of the system at hand. This Perspective reviews current quantum mechanics-based materials modeling approaches and their successes and limitations, and offers a view to the future.

Imagine a day when we can invent new materials on demand just by entering into a computer the elements we want to use, the specific property we want to optimize, and our

cost constraints. We are not there yet. Discovery and design of optimal, inexpensive materials will require both experimentalists and modelers working together to characterize the properties of new candidate materials. To this end, the theorist's role must be to develop and apply robust and accurate techniques that are first validated against well-characterized materials and then applied to new ones. This Perspective outlines

the current state of quantum mechanics (QM)-based modeling of materials behavior and the frontier challenges that remain (1).

The first question to answer is, why focus on QM-based simulations? There is clearly much to be learned from atomistic, mesoscopic, and continuum-level modeling, but all such models require assumptions to be made about the physical laws governing the behavior of the material. Those assumptions can be good, but they are usually based on measurements taken from known, often simpler, materials, with unknown uncertainty as to their accuracy when applied to more complex materials. For instance, even one level less refined than QM, the classical mechanics or "atomistic" level, requires input of interatomic interaction potentials. Such potentials constrain the physics predicted via choice of functional form and fitted data, and do not exist for all combinations of elements. By contrast, in principle QM simply requires input of the atomic numbers of the elements, with no assumption made other than that the laws of QM hold.

The second question is, what properties can QM actually predict? We're told that QM should be able to predict everything, but in practice, each QM technique has its own set of limitations, so it is important to consider what can be predicted correctly qualitatively, quantitatively,

Department of Mechanical and Aerospace Engineering and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544-5263, USA. E-mail: eac@princeton.edu