Molecular dynamics simulations of large macromolecular complexes
Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu and Klaus Schulten

Connecting dynamics to structural data from diverse experimental sources, molecular dynamics simulations permit the exploration of biological phenomena in unparalleled detail. Advances in simulations are moving the atomic resolution descriptions of biological systems into the million-to-billion atom regime, in which numerous cell functions reside. In this opinion, we review the progress, driven by large-scale molecular dynamics simulations, in the study of viruses, ribosomes, bioenergetic systems, and other diverse applications. These examples highlight the utility of molecular dynamics simulations in the critical task of relating atomic detail to the function of supramolecular complexes, a task that cannot be achieved by smaller-scale simulations or existing experimental approaches alone.

Addresses
1 Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
2 Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
3 Energy Biosciences Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
4 Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
5 Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Corresponding author: Schulten, Klaus (kschulte@ks.uiuc.edu)

Current Opinion in Structural Biology 2015, 31:64-74
This review comes from a themed issue on Theory and simulation
Edited by Claire Lesieur and Klaus Schulten

http://dx.doi.org/10.1016/j.sbi.2015.03.007
0959-440X/© 2015 Elsevier Ltd. All right reserved.

Introduction
The essential conundrum of modern biology, namely the question of how life emerges from myriad molecules whose behavior is governed by physical law alone, is embodied within a single cell — the quantum of life. As illustrated in Figure 1, the rise of scientific supercomputing has allowed for the study of the living cell in unparalleled detail, from the scale of the atom [1**,2] to a whole organism [3–5] and at all levels in between [6]. In particular, the past three decades have witnessed the evolution of molecular dynamics (MD) simulations as a "computational microscope" [7], which has provided a unique framework for the study of the phenomena of cell biology in atomic (or near-atomic) detail.

Now, in the era of petascale computing, high-performance MD software packages such as NAMD [8], GROMACS [9], and LAMMPS [10] are being optimized for scaling to an ever-increasing number of cores on cutting-edge computing hardware [2,11,10], enabling the investigation of previously unfathomable biological phenomena through the use of large-scale atomistic simulations. Moreover, the development of computational tools such as molecular dynamics flexible fitting (MDFF) [12,13] are forging an intimate connection between experiment and theory, informing the construction of atomic-level models of large-scale, supramolecular complexes through a synthesis of multiscale experimental data from cryo-EM, NMR spectroscopy, and X-ray crystallography. Complementary to all-atom MD simulations, the development of force fields for coarse-grained MD (CGMD) simulations continues to be a popular source of techniques which favor computational efficiency over atomic and chemical accuracy, permitting simulations on even larger time and length scales [14].

This opinion will focus on the ways in which large-scale MD simulations are having a profound impact in numerous diverse scientific endeavors. From the treatment of disease and development of drugs [15*,16*] to the fabrication of novel biomaterials [17*] and creation of bio-based renewable energy sources [18], large-scale MD simulations are helping to achieve a fundamental understanding of living organisms. Taken together, the work reviewed here demonstrates the maturity of the MD apparatus as a tool to progress basic science and the investigation of the molecular makeup of life.

Large-scale MD simulations of viruses
Viruses are parasitic life-forms that replicate by hijacking resources present in the cells they infect. Because of their small size compared to cells (20–1500 nm scale), observation of the viral particle during different stages of the
replication cycle is mostly limited to electron microscopy. Yet, virus particles are large in size for all-atom simulations (see Figure 2) and as a result most studies at the atomic level have been limited to isolated virus proteins or subfragments of a viral particle or capsid [19]. The satellite tobacco mosaic virus (STMV) became the first complete virus to be investigated through all-atom MD simulations [20]. Since then, MD programs have become capable of simulating systems of even larger size and complexity [1**,**11,21**], thus allowing the study of viral particles up to two orders of magnitude larger in atom count than STMV [1**].

High resolution structures of symmetrical virus capsids like poliovirus, southern bean mosaic virus and satellite tobacco necrosis virus have been available for several years, leading to routine investigations using MD simulations [22,24–26]. More recently, MDFF has been applied to elucidate the structures of yet larger and more complex virus capsids in their native environments [27,15*,1**,28]. For instance, MDFF was instrumental in the structural determination of the HIV-1 core [1**], a polymorphic capsid with no apparent symmetry. The 64-million-atom MD simulation of the mature HIV-1 capsid should enable the characterization of complex interactions between host cell factors and the assembled capsid lattice [29], thus providing an unexploited framework for the development of novel drugs targeting HIV-1. Similarly, MDFF’s application to the Rabbit hemorrhagic disease virus (RHDV), a 10-million-atom MD calculation, improved the overall fitting of the crystal structure to the cryo-EM density, leading to the development of a vaccine [15*].

Spontaneous assembly of the immature and mature forms of the capsid from viral nucleic acids and proteins is an essential step for the replication cycle of a virus. However, the rate at which assembly occurs prevents a molecular-level description of the process by experimental means, making it an attractive task for computational modeling. The virus assembly kinetics and pathways can be computationally determined using stochastic kinetics [30], elastic networks [31], simulation-based data fitting [32], and CGMD simulations [23*33,30,34]. In fact, by means of CGMD simulations, optimal configuration of the viral genome has been shown to be essential for proper assembly of the capsid [23*,33]. In the particular case of simian virus 40 (SV40), two pathways have been observed during assembly, with each of the pathways sampling different intermediate states depending on the strength of both ionic solution and protein–protein interactions [23*]. Coarse-grained simulations have also been performed to investigate the self-assembly process of the HIV-1 mature capsid [35,36,1**], allowing the identification of a trimer-of-dimers structure as a fundamental step during assembly of hexagonal lattices in a crowded environment [35,36]. However, due to the lack of structural information available at the time for informing the CG models, both studies neglected explicit interactions at the trimeric interface of the lattice, an interface now known to play an essential role in capsid curvature, assembly, and stability as well as infectivity in vivo [1**].

In order to become infectious, entire viral particles undergo a global structural rearrangement known as maturation. Such a maturation process is thought to follow two different pathways: disassembly–reassembly
Viral particles of different sizes studied using MD simulations. The viruses were arranged in the order of increasing size with the capsid diameters given in parentheses: STMV (17 nm) \([20]\), poliovirus (32 nm) \([22]\), RHDV (43 nm) \([15]\), SV40 (49 nm) \([23]\), and HIV-1 (70–100 nm) \([1*]\). For size comparison, HIV-1 protease, one of the most studied enzymes, is shown at the bottom right.

or displacive transition \([37,38]\). Mechanistic insights into the continuous displacive transition pathway were offered by investigating the swelling process of the cowpea chlorotic mottle virus capsid \([39,40]\) in which the process was identified to be symmetry-breaking. Normal mode analysis of entire viral capsids has been used to probe mechanical properties of the particles \([31,41–43]\), and should be a valuable tool in studying large conformational transitions associated with viral processes (e.g. maturation) \([19,31]\). In a remarkable study, the pH-induced maturation of bacteriophage HK97 was investigated by Brooks and coworkers using constant pH MD simulations \([44,45**]\), revealing a novel intermediate state stabilized by key residues that undergo significant pK\(_a\) shifts during maturation. Although limited presently to a smaller subsystem of bacteriophage HK97, constant pH simulations could be applied in the future to investigate the maturation process of whole capsids. To our knowledge, however, no MD simulation has described yet the disassembly–reassembly process of a virus.

In summary, large scale MD simulations have proven to be essential to the structural determination of viral particles, in particular by the unique potential to combine information from multiple experimental sources and by the ability to probe viral protein dynamics in native environments. However, processes during virus replication that are inherently slow and involve large systems, such as virus assembly and budding \([46]\), membrane fusion, virus maturation, uncoating of viral capsids, and nuclear entry, represent the current threshold for MD simulations of viruses. In particular, studies on host–pathogen interactions involving large numbers of proteins \([47]\), are still scarce, yet essential to the investigation of virus replication cycles, as well as for the development of drugs and the engineering of viruses for gene therapy.

**Large-scale MD simulation of ribosomes**

As another important pharmaceutical target, the ribosome is the most ubiquitous and complex molecular machine in living cells, responsible for decoding the genetic code into functional proteins. MD simulation has been successfully applied to ribosomal translation \([48]\). Here, we review the most recent advances in large simulations involving complete ribosome structures.
Simulations of the ribosome have advanced greatly in recent years. A key subject of such studies has been the process of translocation, which moves mRNA ahead by three bases and shifts tRNAs from A to P to E sites. Translocation is still a challenge for MD simulations because the process involves a collective motion of both ribosomal subunits as well as of tRNA substrates [52,53] (see Figure 3a), the description of which requires simulation of a complete solvated ribosome that amounts to ~2.5 million atoms. By combining cryo-EM data and crystallographic structures, Grubmüller and coworkers [49**] performed MD simulations of complete ribosome structures at 13 intermediate states of the translation process. The simulations revealed an atomic-level description of time-resolved translocation, transition rates linking translocation substrates and the underlying molecular forces, all properties essentially inaccessible to experiment. Another MD simulation of a complete ribosome performed by Sanbonmatsu and coworkers [54] captured the small subunit rotation, which is essential for moving the tRNA forward [55]. The simulation allowed the authors to identify reaction coordinates that capture the rotation and to obtain estimates of the free energy barriers associated with translocation.

The ribosome is a target of many antibiotic drugs [56]. Simulations have played a key role in elucidating antibiotic action in the ribosome [57–61]. For example, MD simulations have shown that drug-regulated base-flipping of the ribosome underlies the action of small antibiotic compounds [57,58]. A recent MD simulation involving a complete bacterial ribosome and the clinically important macrolide drug erythromycin (ERY) [62,63], shown in Figure 3c, has revealed a new perspective of how the family of macrolide antibiotic drugs act on bacterial ribosomes [16*]. It was previously believed that the presence of the nascent protein is necessary for the action of macrolide drugs [64–66]. However, the MD simulations have shown that ERY acts on ribosomes without a nascent protein present, by inducing base flipping of critical nucleotides at the ribosomal catalytic center [16*]. The change of ribosomal RNA bases predicted by the MD simulations was later supported by a cryo-EM reconstruction of a ribosome-ERY complex [67].

Computational studies of the ribosome have begun to simulate the complete ribosome over increasing timescales [49**,54,16*,68] and to further incorporate experimental data [48]. In addition, ribosome simulations have begun to capture the heterogeneous nature of translation. For example, modeling and simulations were key in determining the translocon-bound ribosome structure with a nanodisc [50] membrane (shown in Figure 3b) and allowed the characterization, in atomic detail, of how nascent proteins are inserted into a membrane [51]. The availability of the high-resolution structure for the human ribosome [69] together with simulations will have a significant impact on antibiotic drug development, as drugs

Figure 3

(a) Translocating ribosome at the pretranslocation state with an A-site tRNA (red) and a P-site tRNA (green) [49**]. A red arrow shows the direction of tRNA’s traversal motion. (b) Insertion of a nascent protein by the ribosome into a nanodisc [50] membrane working with the SecYE translocon [51]. The nascent protein and P-site tRNA are shown in green. A red arrow shows the direction of the nascent protein’s insertion motion. C. Bacterial ribosome with the antibiotic drug erythromycin (in red circle) shown at its binding site inside the ribosome [16*].
that kill bacteria, but leave human ribosomes unharmed, can be designed through computational guidance.

Large-scale MD simulations of bioenergy systems
Development of new pharmaceuticals is not the only target of large simulations, indeed, biotechnological applications related to bioenergy are also been studied by multi-million atom MD simulations. Applications range from studies of energy conversion by complex interlocking mechanisms of several proteins in photosynthetic systems, to production of biofuels out of agricultural waste.

Photosynthetic apparatus
Photosynthesis is key for life on Earth, converting light energy into various forms of chemical energy, for example, into the synthesis of ATP from ADP. The conversion takes place in an apparatus containing extensive membrane–protein systems. Even in the case of the simplest photosynthetic systems, that of so-called purple bacteria, hundreds of proteins work in concert to turn sunlight into ATP [70]. Input from several experimental techniques has been combined to construct models of entire photosynthetic membranes, which are presently being explored through quantum mechanical calculations and MD simulations. The largest photosynthetic membrane simulation published thus far is that of a 23-million-atom membrane patch from the bacterium *Rps. photometricum* [71] as depicted in Figure 4; a much larger, 100-million-atom MD simulation of a chromatophore from *Rh. sphaeroides* [72] is presently underway. Topographical data from atomic force microscopy provided the initial positions of the light harvesting complexes, which were then placed in a lipid bilayer along with quinone molecules, crucial for completing the needed series of energy conversion steps. The 150 ns MD simulation on the membrane patch gave insight into the packing of the proteins and the mobility of the quinone molecules, which must diffuse through the narrow lipid phase separating closely packed protein complexes.

MD simulations do not only support understanding of in vivo behavior, but are also capable of explaining in vitro processes and support the bioengineering of energy conversion devices. A remarkable example is the 200 ns MD simulation of the 3-million-atom system comprising the trimeric photosynthetic pigment–protein complex, photosystem I (PSI), embedded in a detergent belt of *n*-dodecyl-β-D-maltoside (DDM). The study performed by Frymier and coworkers [73] gave atomic resolution insight into the impact of the detergent on conformational dynamics of the system and its possible implications on PSI functionality.

Cellulose and second-generation biofuels
Producing energy is not only essential for biological cells, but also for human society. Computer simulations are assisting in the development of cost-effective biofuels from agricultural waste, by studying enzymes and enzymatic complexes employed [18,74,75] in cellulose degradation as well as biomass recalcitrance [76,77]. To avoid the recalcitrance problem, biomass pretreatment methods employ various chemical and thermal conditions to modify the molecular interactions and association of the different biomass components in favor of cellulose accessibility and digestibility [78]. Studying two of the main components of the biomass, namely lignin and cellulose

---

**Figure 4**

Photosynthetic membrane studied by molecular dynamics simulations. MD simulations provide insight into local and global mechanisms of light harvesting and energy conversion. The photosynthetic membrane pictured consists of three types of protein complexes, light harvesting complex 1 (light blue) and 2 (green) and photosynthetic reaction center (dark blue) as well as quinone molecules (red), which are surrounded by a phospholipid membrane (translucent gray). The full system simulated contains ~23 million atoms [71].
MD simulations of large macromolecular complexes are essential in bioenergy studies. Already reaching the 100-million-atom milestone, simulations of an entire chromatophore organelle are underway to investigate the overall mechanism of a large energy-converting system. Also taking advantage of large biomolecular complexes, biofuel production would certainly be bolstered by MD simulations of realistic plant cell wall models and efficient large multi-enzyme cellulose complexes [75], making production more cost-effective.

**Diverse applicability of large-scale simulations**

Besides the aforementioned prominent applications to pharmaceutically relevant systems and energy conversion, MD simulations have been successfully employed to study a wide variety of large and complex systems. The scope ranges from fundamental biological processes like cell motility or essential membrane processes to biotechnological applications like biomaterial development.

**Actin filament branch junction**

Cell motility is organized mainly by coordinated regulation of microtubule and actin filament dynamics [80]. Actin networks are organized by actin-related protein 2/3 (Arp2/3) complexes that form geometrically regular junctions between two actin filaments [81]. Branched junction structures were proposed based on a 26 Å resolution electron microscopy (EM) tomogram [82], addressing questions about the interaction between the Arp2/3 domain and the mother actin filament. Gathering 31 protein subunits and including ~3 million atoms for almost 200 ns, MD simulations performed by Voth and coworkers suggest that the branched structure is remarkably stable [83**]. Furthermore, the MD results were able to show that Arp2/3 complexes and the mother actin filament contain a large number of dynamic salt bridges and hydrophobic contacts that can form or break on the nanosecond timescale. MD simulations improved the fits of individual complex subunits to the EM density maps by a vast test with thousands of conformations docked by rigid body fitting into the EM map. It was possible to suggest targets for future experimental studies by examining the identified contacts between Arp2/3 complexes and actin filaments.

**Realistic cell membrane**

The cell membrane, which is the environment for manifold cellular processes, consists of many different lipid species. The minor lipid species that are present in very low concentration, yet essential in regulating membrane protein functions, can now be studied in large system simulations. Important examples of minor lipid species are polyunsaturated phospholipids that can enhance membrane fission induced by dynamin and endophilin [84*]. In a remarkable study, Marrink and coworkers performed CGMD simulations of a large plasma membrane that mimics the lipid composition of a cellular membrane [85*]. The plasma membrane model consisted of ~20 000 lipids with 63 lipid species. Owing to the size of the membrane patch, built employing the coarse-grained MARTINI lipid model [86], it was possible to include minor lipid species in a large number. After 40 μs of simulation, the plasma membrane established an asymmetrical distribution of cholesterol and formed transient domains with liquid-order character, a finding which agrees with behaviors in a plasma membrane observed experimentally [86].

**Membrane sculpting**

Large-scale simulations, complementing recent advances in structural biology techniques, provide unique opportunities to study protein-induced membrane sculpting [88]. Complementing atomistic simulations, CG methods offer a feasible approach to studying relevant large protein–membrane assembly structures due to the long timescale associated with membrane morphology transitions and the large size of the protein–membrane complex [89]. Recently, large MD simulations have been carried out to study membrane tubulation induced by lattices of membrane sculpting proteins, namely the F-BAR domains [87*]. The dynamic process of F-BAR–driven tube formation from an initially flat membrane, as shown in Figure 5, was illustrated by employing ~350 μs CG simulations based on a shape-based coarse-grained model previously developed [90,91]. The N-BAR domains are another example of membrane sculpting protein complexes. Voth and coworkers described the self-assembly mechanism of tubular networks by N-BAR domains starting from lipid vesicles [92]. Combining previous experiments [93] and CG simulations of a 26-site N-BAR domain model, the authors concluded that such tubular networks were formed by interactions between the membrane and amphiphatic helices. Simulations of large protein-membrane assemblies at the subcellular scale serve as a stepping stone towards future whole-cell modeling at the molecular level.
Nuclear pore complex
Another large protein–membrane assembly at the subcellular scale is the nuclear pore complex (NPC) [94*,95], located in the double bilayers of the nuclear envelope mediating particle exchange between nucleus and cytoplasm. Nsp1, a nuclear pore protein, is the major component of NPC. MD simulations, performed at all-atom and coarse-grained levels, showed that an Nsp1 array forms dynamic, brush-like structures of multiprotein bundles [96]. More recently, microsecond-long CGMD simulations were performed on the entire transport channel of the NPC, the description of which required a 16-million-particle CGMD calculation. Although NPC has a physical opening of 60 nm in diameter, the cross-linking between the Nsp1 creates a sieve-like structure that limits the size of freely diffusing molecules passing through the NPC to <9 nm [94*]. Larger molecules require the assistance of transport factors to be transported through the NPC.

DNA-functionalized nanoparticles
Even though widely applied to pure biological systems MD simulations are not restricted to them, and MD use in the development of biomaterials [97] and bio-functionalized materials with enzymes [98] has been widely effective. An important application concerns new systems for drug delivery. Gold nanoparticles (AuNPs) functionalized with DNA have been intensively exploited as a drug carrier even though little was known about molecular details of the complex. Performing MD simulations of systems of up to 5 million atoms, Vashishta and coworkers were able to construct two significant supercrystal conformations of AuNPs showing that DNA becomes capable in the presence of the supercrystals of smooth changes that do not arise for DNA in solution [17*]. The differences resulted in a contraction of DNA by changing its base-base stacking. The simulations permitted an estimate of system properties, such as elasticity.

Conclusion
Through the combination of state-of-the-art computing hardware and advanced software development [11,99] a true computational microscope has emerged [7]. Molecular dynamics simulation has established itself as a reliable tool to view the structure and dynamics of large protein complexes within realistic cellular environments. While all-atom MD remains the gold standard for MD simulations, CGMD approaches present a complementary technique to probe extended time and length scales [14], applicable in case one can be certain that the simplified force field captures the physical and chemical mechanism underlying the process investigated. Future physics-based multi-scale models, incorporating biochemical information and systems biology will enable the study of cellular-scale processes involving thousands of proteins, as exemplified by the pioneering work of Feig et al., related to the atomistic description of the bacterial cytoplasm [5].

Advances in computational microscopy are moving the atomic resolution descriptions of biological systems into the 1 million (10 nm) to 1 billion (100 nm) atom regime, where Ångstrom-scale interactions bring about 100-nm scale concerted dynamic processes, crucial for numerous cell functions. This regime is rich in exciting biological questions and characteristic of large biomolecular systems [100–103] beyond those described here.

In this opinion, we reviewed the progress driven by large-scale molecular dynamics simulations in applications ranging from cell biology to health sciences to biofuel production. Considering the success of these applications, we foresee large-scale simulations playing an ever-increasing role. Combined, simulation and experiment are helping to bridge the gap between atomic level properties and whole cell behavior, an endeavor which cannot be accomplished by either approach alone.
Conflict of interest
None declared.

Acknowledgements
We thank Dr Yaxin Liu for providing the figure of the RHDV virus. The authors gratefully acknowledge funding from the National Institutes of Health (NIH, 9P41GM106401, P01-GM07687, U54 GM087519), the National Science Foundation (NSF, MCB-1157615, PHY1430124), and Energy Biosciences Institute (EBI, 231 UCB BP 201400401). TR is supported by the Alexander von Humboldt Foundation.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest


The authors present an improved virus capsid model using multimillion-atom MDFF calculations that combine the structural information from cryo-EM and X-ray crystallography.


MD simulations reproducibly see the antibiotic drug erythromycin inducing conformational changes at the catalytic center of the bacterial ribosome.


This paper reveals the applicability of MD simulations to study the interaction between inorganic materials and biological molecules.


The authors present a computational model of an Influenza A virion. The model enables the study of biophysical properties of the viral particle that may be related to the fitness of the virus.


50. Combination of cryo-EM data and MD simulations reveal a full picture of ribosomal translocation at atomic-detail.


67. Tenson T, Lovmar M, Ehrenberg M: The mechanism of action of macrolides, lincosamides and streptogramin B reveals the


This study consolidates several experimental results in a MD simulation of a large multi-protein-membrane system, which served as input to study the quantum process of photosynthetic light harvesting.


MD simulations provide insight into the in vitro behavior of photosystem I (PSI) in detergent.


The landmark paper describes the structure of the cellulose-lignin complex, showing that recalcitrance of biomass results from lignin adhesion to cellulose.


MD simulations identified the dynamics of the contact area between actin filament and Arp2/3.


The authors showed that polysaturated phospholipids, even though they arise only as a small percentage of membrane lipids, enhance membrane fission induced by dynamin and endophilin.


The authors performed the first MD simulation on a membrane bilayer with realistic lipid composition.


The authors demonstrated that the F-BAR domains generate membrane sculpturing via the scaffolding mechanism; the protein lattice structure optimal for producing membrane curvature was determined.


The authors studied nuclear pore complex structure and function by means of residue-based CGMD simulations, one of the largest simulations of the kind ever carried out.


