Whole-Cell Models and Simulations in Molecular Detail

Michael Feig[1,2] and Yuji Sugita[2,3]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA; email: mfeiglab@gmail.com

[2]Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Kobe, Hyogo, 650-0047, Japan

[3]Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Wako, Saitama, 351-0198, Japan

Abstract

Comprehensive data about the composition and structure of cellular components have enabled the construction of quantitative whole-cell models. While kinetic network–type models have been established, it is also becoming possible to build physical, molecular-level models of cellular environments. This review outlines challenges in constructing and simulating such models and discusses near- and long-term opportunities for developing physical whole-cell models that can connect molecular structure with biological function.

Keywords

Crowding; systems biology; protein structure; molecular dynamics simulation; network models

INTRODUCTION

At the most fundamental level, biology arises from molecular behavior. Proteins and nucleic acids are the main actors, and their dynamics and interactions via molecular recognition and binding are key aspects of virtually any biological function. Atomistic resolution is typically necessary to fully understanding these processes. Therefore, a major effort in modern biology has been focused on improving the resolution of macromolecular structures at subnanometer scales. As the Protein Data Bank (PDB) (Westbrook et al. 2003) continues to grow rapidly, much is known today about the structures of proteins and nucleic acids. In fact, high-resolution experimental structures are available for essentially all major protein types (Vitkup et al. 2001, Zhang et al. 2006), and an increasing number of macromolecular complexes have been resolved (Marsh & Teichmann 2015). Challenges remain in determining the structures of very large and dynamic complexes, membrane proteins, and ribonucleic acid (RNA) molecules, which may be addressed soon via cryo-electron microscopy (cryo-EM) (Fernandez-Leiro & Scheres 2016). The high-resolution structure of chromosomal deoxyribonucleic acid (DNA) has also remained elusive. However, computational structure prediction is becoming increasingly powerful so that useful models can often be constructed when experimental structures are not available by template-based modeling (Rohl et al. 2004, Roy et al. 2010, Waterhouse et al. 2018) and/or by assisting in model building with lower-resolution experimental data for proteins (Alexander et al. 2008, Bowers et al. 2000, Li et al. 2004) and nucleic acids (Miao et al. 2015), including genomic DNA (Bianco et al. 2017, Di Stefano et al. 2016, Hacker et al. 2017, Tiana & Giorgetti 2018, Yildirim & Feig 2018). Both experiment and modeling continue to see rapid progress. It may therefore be expected that complete structural coverage for all macromolecules in a cell of a specific organism could be within reach.

Structure alone does not fully explain most biological function. Equally important are conformational dynamics within a single molecule and dynamic interactions between molecules. The conformational dynamics of proteins and nucleic acids have been studied extensively via experiment and molecular dynamics (MD) simulations, resulting in much insight about the mechanisms of biochemical processes (Hospital et al. 2015). However, almost all of these studies, experiment and modeling alike, invoke simplified conditions without considering the high concentrations and physicochemical complexities of cellular environments. In fact, many

questions remain about how biological macromolecules behave inside cells and navigate a spectrum of specific and nonspecific interactions in the presence of a variety of electrolytes, osmolytes, and other small molecules (Cohen & Pielak 2017, Gnutt & Ebbinghaus 2016, Rivas & Minton 2016). Molecular interactions are essential for enzyme function, signaling, and many other biological processes, but interactions can also lead to macromolecular aggregation and the development of diseases (Ross & Poirier 2004). Moreover, interactions between cellular components can stabilize or destabilize biomolecular structures (Wang et al. 2012) or lead to the formation of phase-separated states (Dumetz et al. 2008), thereby further modulating biological function. It is thus becoming increasingly clear that a complete link between molecular structure and biological function requires the integration of structure and conformational dynamics at the atomistic level with dynamics and interactions at the cellular level under realistic biological conditions.

Experimental techniques can probe many aspects of biomolecular structure and dynamics, but they typically focus only on narrow points in the space-time universe. For example, crystallography and cryo-EM can resolve structures in atomistic detail, but without any or only very limited time resolution and under artificial conditions that may have little in common with conditions in living cells (Drenth 2007, Fernandez-Leiro & Scheres 2016). In contrast, the best single-molecule fluorescence methods can track the dynamics of molecules in vivo on millisecond timescales, but only at spatial resolutions of tens of nanometers (Huang et al. 2009). NMR spectroscopy can capture dynamics at atomistic scales under in vitro and in vivo conditions (Inomata et al. 2009, Sakakibara et al. 2009), but the interpretation of such data, especially when collected in heterogeneous environments, can be difficult (Pastore & Temussi 2017). Therefore, experiment alone has not been able to provide a unified picture of biomolecular structure, dynamics, and function across the entire range of scales from the molecular to the cellular level. Modeling can fulfill this role via whole-cell models (Trovato & Fumagalli 2017) that incorporate the experimental data and aim to predict how changes at the molecular level propagate to altered function at the systems level.

WHOLE-CELL MODELS

The most successful whole-cell models have so far relied on empirical mathematical models that

are parameterized on the basis of experimental data and that focus on a kinetic view of cellular processes (Guerrier & Holcman 2017, Karr et al. 2015, Macklin et al. 2014, Szigeti et al. 2018, Tomita 2001). These models can access macroscopic timescales and connect directly with biological function. However, spatial resolution is often limited or absent, and molecular-level details are rarely, if ever, considered. As a result, there is no connection to molecular structure and interactions and the physical laws that operate at the molecular level so that predictions of how changes at the molecular level affect the cellular outcome often cannot be made. For example, it would be difficult to predict how the introduction of a new drug candidate perturbs cellular function without making specific assumptions about altered kinetic pathways.

The alternative models are physical models that build up by assembling individual molecules on the basis of structures at the atomistic level (**Figure 1**) (Im et al. 2016) and then subjecting such models to computer simulations (Feig et al. 2017, 2018). Such models connect behavior at the molecular level to cellular function. The wider range of scales—as well as the greater involvement of physical laws—provides, at least in principle, for greater predictive abilities. The main limitation of physical modeling is the computational cost for reaching cellular timescales and spatial scales with models that retain molecular detail at high resolution. However, as computational methods advance and computer hardware becomes ever more powerful, the physical modeling of cellular environments is becoming possible (Yu et al. 2016). This review focuses on the challenges and opportunities of physical models of cellular environments, especially models that connect with molecular structure at atomistic resolution. The main emphasis is on models of bacterial cells as such systems are beginning to be tractable today, but the general ideas discussed here readily transfer to more complex eukaryotic cells that will become accessible as resources increase and new experimental data are available.

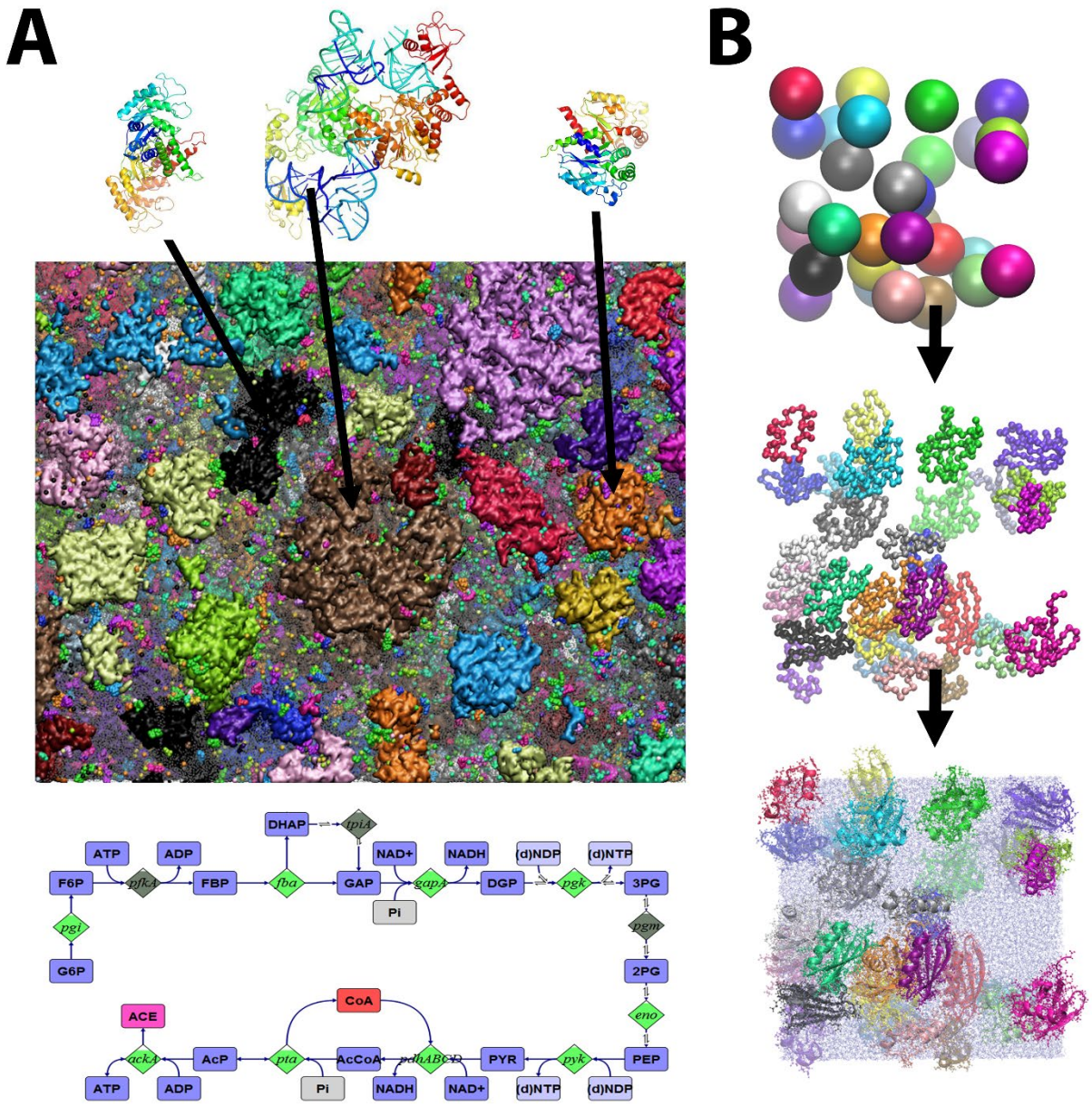

Figure 1 (*a*) Construction of cellular systems from atomistic structures of individual molecules based on biochemical pathway reconstruction for the cytoplasm of *Mycoplasma genitalium*. (*b*) Multiscale assembly generation protocol from spherical models to a fully solvated atomistic system as described in Feig et al. (2015).

CHALLENGES IN BUILDING PHYSICAL MODELS OF CELLS

Physical models of cellular environments that involve a molecular level of detail hold much promise but face a number of challenges that are reviewed in this section. The choice of model

resolution is a critical aspect, whereas the success of the modeling depends on realistic interaction potentials and the availability of experimental data to construct initial systems. Other challenges revolve around the practical issues of assembling, running, analyzing, and visualizing very large scale molecular systems. These aspects are discussed in more detail in this section.

Model Resolution

Physical models of biomolecular systems are typically based on particle-based models at resolutions ranging from atomistic detail to coarse levels, where an entire macromolecule may be represented as a single particle. Continuum models are also sometimes employed to represent certain cellular components such as bulk solvent, membranes, DNA, or cytoskeleton components when it is sufficient to capture their macroscopic or overall mechanoelastic properties. The level of resolution determines the accuracy of the models and their ability to project molecular-level effects to the cellular scale, but higher resolution demands increased computational cost and thereby limits what timescales and what size of systems can be investigated with the available computer resources.

Atomistic models of biomolecules, where atoms are represented as spherical particles, have been used for decades since they match the resolution of biomolecular structures from experiment. Atomistic models are often explored via MD simulations to provide insights into the dynamics and energetics of biomolecules that are otherwise difficult to extract from experiments (Huggins et al. 2018). Typically, atomistic modeling starts from experimental structures and involves explicit solvent molecules such as water or lipid bilayers for membrane-interacting systems. Because of the high level of detail, a moderate-size single protein surrounded by explicit solvent can easily require 100,000 particles, whereas atomistic models of a bacterial cytoplasm reaches 100 M particles for a $(100\ \text{nm})^3$ cubic volume that is still only one-tenth to one-twentieth of the smallest bacterial cell (Feig et al. 2015, Yu et al. 2016). Such large numbers of particles require substantial computer resources, and therefore, the timescales that can be directly reached are generally limited to submillisecond scales, even on the largest computer clusters available today. The accessible timescales can be extended, however, by as much as several orders of magnitude via enhanced sampling methods (Mori et al. 2016) and/or by inferring kinetics from Markov state models constructed from shorter simulations (Harrigan et al. 2017, Lane et al. 2013).

The limitations of atomistic models can be overcome at least in part with coarse-grained (CG) models (Buchete et al. 2004, Kar & Feig 2014, Takada et al. 2015). Such models combine molecular fragments into single interaction sites to reduce the number of particles and hence the computational cost. The level of resolution may vary from combining just C-H groups into unified particles to models in which an entire molecule is represented as a single sphere. The reduced degrees of freedom typically come at the price of reduced accuracy and ability to apply physical laws. CG models are typically parameterized on the basis of a combination of data from experiment and higher-resolution modeling such as atomistic simulations. As a result, universal use and transferability are more limited, and it is more difficult to derive truly new insights that do not simply recapitulate what is already known from the data or simulations that were used to parameterize the model. However, the simplicity of CG models can be a good choice when larger-scale questions that do not require atomistic resolution are addressed (Ando & Skolnick 2010).

Another strategy for reducing the computational cost is the use of continuum representations for parts of a system in which macroscopic properties dominate or in which experimental data do not provide sufficient resolution to build high-resolution particle-based models. Continuum models reduce computational costs not just by reducing the degrees of freedom but also by allowing for instantaneous relaxation of certain components that are otherwise subject to kinetic barriers due to molecular reorganization. One example is the replacement of explicit aqueous solvent with dielectric-based implicit models (Roux & Simonson 1999). Such models can be extended to lipid bilayers and, in principle, to any heterogeneously varying dielectric environment (Tanizaki & Feig 2005). While implicit solvent models are attractive for capturing solvation effects without the need for explicit solvent molecules, it is difficult to capture hydrodynamic interactions accurately within such a framework (Ando et al. 2012, Dlugosz et al. 2012, Mereghetti & Wade 2012). Continuum models can also be applied to biomolecules, e.g., in the form of elastic rod models of nucleic acids (Balaeff et al. 1999) and cytoskeletal components (Walcott & Sun 2010) to capture bending and twisting flexibility, as elastic models of lipid bilayers to allow for deformations in response to interactions with other molecular components (Brown 2008), or as models of flexible globular proteins via fluctuating finite elements that focus on overall shape fluctuations (Solernou et al. 2018).

Different resolutions can be applied simultaneously via multiscale models to limit the use of

expensive high-resolution treatments. One strategy is to apply higher-resolution models only for selected parts of a system when and where necessary instead of the entire system (Kar & Feig 2017, Renevey & Riniker 2017, Ward et al. 2017, Wassenaar et al. 2011). In another approach, low-resolution modeling may be used to broadly span conformational sampling before switching to higher resolution to obtain quantitative information of the dynamic landscape of a given system (Harada & Kitao 2012, Tempkin et al. 2014, Zhang & Chen 2014). While most multiscale frameworks focus on bridging spatial scales, it is also possible to bridge timescales. The use of continuum models to avoid kinetic barriers is discussed above. Another possibility is to switch kinetically separated states with a certain probability (Prytkova et al. 2016) instead of waiting for simulations to overcome kinetic barriers, similar to what is already in wide use in kinetic network whole-cell models (Bernstein 2005). Multiscale methods are in principle highly suitable for studying cellular environments in which different processes occur on different scales of time and space. However, practical applications have been hindered by technical challenges with coupling models at different levels of resolution and by a lack of software implementations that perform well on high-performance computing platforms.

Interaction Potentials

Interaction potentials between the sites of a given model allow for the calculation of energies and forces so that simulation can be carried out. The type of interaction potential depends on the level of resolution. At the atomistic level, the potential terms are based on physical laws or designed to approximate physical laws. Chemical bonds and bond angles are typically approximated via harmonic functions, while rotations around dihedral torsions are often described via Fourier series terms. Longer-range nonbonded interactions commonly consist of electrostatic Coulomb terms and a Lennard-Jones potential that prevents atom overlap and accounts for van der Waals dispersion attraction forces. The combination of these terms is a classical approximation of what would otherwise be largely quantum mechanical interactions, but without allowing for bond formation or breakage. Essentially all biomolecular force fields are based on these terms. After many years of development, the latest generation of force fields have arguably reached a high level of accuracy and transferability (Best et al. 2012, Galindo-Murillo et al. 2016, Harder et al. 2016, Huang et al. 2017, Klauda et al. 2010, Wang et al. 2017), including the ability to accurately simulate RNA (Tan et al. 2018), intrinsically disordered

peptides (IDPs) (Best et al. 2014, Huang et al. 2017, Robustelli et al. 2018), and interacting proteins (Best et al. 2014, Nawrocki et al. 2017). However, high-quality parameters are still not available for all types of molecules. In particular, available parameters for small molecules have variable accuracy. This variability impacts the ability to model metabolites or potential drug candidates within cellular environments.

At the CG level, empirical and statistical potentials are used more prominently, sometimes in combination with physical potentials. The empirical potentials are often parameterized on the basis of atomistic potentials or certain experimental observables, and different CG models tend to focus on specific scientific applications rather than being universally applicable. For example, the widely used MARTINI model (Monticelli et al. 2008) was initially developed to reproduce relative hydrophobicities of different amino acids and is therefore well suited to studying water-membrane partitioning but requires restraints to maintain secondary structures (Monticelli et al. 2008). The higher-resolution CG model PRIMO uses mostly physics-based interaction terms to achieve compatibility with atomistic force fields, with the drawback of higher computational costs, which limit the scale of applications that can be studied (Gopal et al. 2010, Kar et al. 2013). Structure-based models such as Go models require knowledge of target conformational states but can be useful in studying transition pathways (Clementi et al. 2000). Some of these higher-resolution CG models may also be suitable for describing protein-protein interactions (Frembgen-Kesner & Elcock 2010, Kmiecik et al. 2016). At the very coarse level, patchy particle models with angle-dependent radial interaction functions have long been used to model the interactions of colloids (Bianchi et al. 2011). These types of models can also be applied to biological macromolecules if internal structure and dynamics can be neglected (Bucciarelli et al. 2016).

A very recent trend is the use of machine-learning (ML) methods to derive interaction potentials (Chmiela et al. 2017, Li et al. 2015, von Lilienfeld et al. 2015). This can be done at all levels of resolution, given suitable training data, and therefore, this approach has the potential to blur the differences between atomistic and CG models. ML methods can be used to generate interaction potentials that match the resolution of the underlying models, especially when ML methods are trained on reproducing energetics rather than fitting parameters for the commonly used standard interaction potential forms.

Input Data

Models of cellular environments, or any biological system, could not be constructed without experimental data. To build models of cellular environments, it is a major effort to assemble the necessary data (Guell et al. 2009, Kuhner et al. 2009, Singla et al. 2018, Yus et al. 2009). At the minimum, information about both the system composition and the concentration of each species and structural information are needed. In addition, knowledge about the spatial dimensions and arrangements of cellular components such as membranes, organelles, the cytoskeleton, genomic DNA, and other compartments is also necessary. The structural organization of cells, including the distribution of large complexes, can be visualized via microscopy or electron tomography (Beck et al. 2011). The composition and concentration of cellular components, at least those present in high abundance, can be determined via mass spectrometry (Bantscheff et al. 2007) and complemented by insights from metabolic network modeling (Feig et al. 2015, Karr et al. 2012, Tomita et al. 2000). However, complete structural coverage at the molecular level is probably the most significant challenge. Although many high-resolution structures have been resolved, significant structural coverage of the genes in any single organism exists only for very few organisms. In fact, for almost all organisms, there are none or at best very few structures available in the PDB. Computational structure prediction can fill the gap and, in many cases, does so quite well (Modi et al. 2016, Waterhouse et al. 2018). However, the accuracy of the predicted structures varies, and prediction is not always possible when structural templates from related proteins or nucleic acids are not available. Further advances in experimental structural biology are expected to improve the situation, but the larger gains will likely have to come from improvements in computational structure prediction methods to provide missing structural information.

System Assembly

The assembly of a heterogeneous cellular environment is nontrivial (**Figure 1**). The initial placement of molecules needs to be consistent with the overall dimensions of a given cellular system and to match lower-resolution imaging and tomography data when available. Moreover, the naive placement of biomolecules at high concentrations leads to significant overlap between different molecules, which is difficult to resolve via subsequent minimization or simulation. There are different ways in which to build such complex systems, but two specific approaches

appear to be most promising. First, instead of assembling a crowded system directly at full resolution, a hierarchical multiscale scheme may be applied wherein biomolecules are placed initially at lower resolution to facilitate optimization and relaxation before gradually increasing in resolution until the target resolution is reached (**Figure 1**) (Feig et al. 2015). Second, a crowded system can also be built up by packing molecules into predefined volumes or shapes. This method is implemented in the cellPACK software, which uses sophisticated packing algorithms to assemble complex systems quickly (Klein et al. 2018). Additional challenges arise when systems need to incorporate densely packed genomic DNA (Goodsell et al. 2018) or cytoskeletal structures, because these elements present obstacles that limit available space for placing other molecules and cannot be approximated well by simple spherical models. Finally, the initial construction of strain-free lipid membranes surrounding cells, organelles, or vesicles is also difficult, especially when lipid compositions are heterogeneous and membrane proteins are embedded (Koldsø & Sansom 2015). More efficient and flexible modeling tools are expected to emerge in the future so that any kind of crowded cellular system can be assembled from molecular components at high levels of resolution.

Simulation Methods and Software

Subjecting molecular models to simulation or other related computational methods provides insight into dynamics and interactions with other molecules. MD and related techniques such as Brownian and Stokesian dynamics are the most common approach, especially for systems represented at the atomistic level (Brady & Bossis 1988, Ciccotti et al. 2014, Ermak & McCammon 1978). CG models can also be efficiently simulated via Monte Carlo techniques (Kmiecik et al. 2016). Other computational techniques that focus on dynamics include normal-mode or elastic network methods (Atilgan et al. 2001, Case 1994), but these methods are generally limited to studies of conformational dynamics of single molecules and are not well suited to studying interactions in a crowded environment.

Many highly efficient software packages exist for carrying out MD simulations. Most packages are optimized for the simulation of smaller systems of up to 1 M particles, but they are generally not prepared to address the unique challenges of simulating much larger systems. One important practical issue is the handling of input/output and parallel operations, which has to be fully distributed for the efficient simulation of systems with 100 M or 1 B atoms because it takes

too much time and memory for any single node to read and store the entire system (**Figure 2**). To our knowledge, parallel input/output operations are supported by only some MD software packages so far, most notably GENESIS (Jung et al. 2015, 2016; Kobayashi et al. 2017) and NAMD (Kale et al. 1999). Moreover, the high computational cost of large system sizes requires efficient parallelism up to a very large number of cores on conventional clusters or many GPUs on GPU-based clusters, which is also possible only with a few packages (Hess et al. 2008, Jung et al. 2016, Kale et al. 1999).

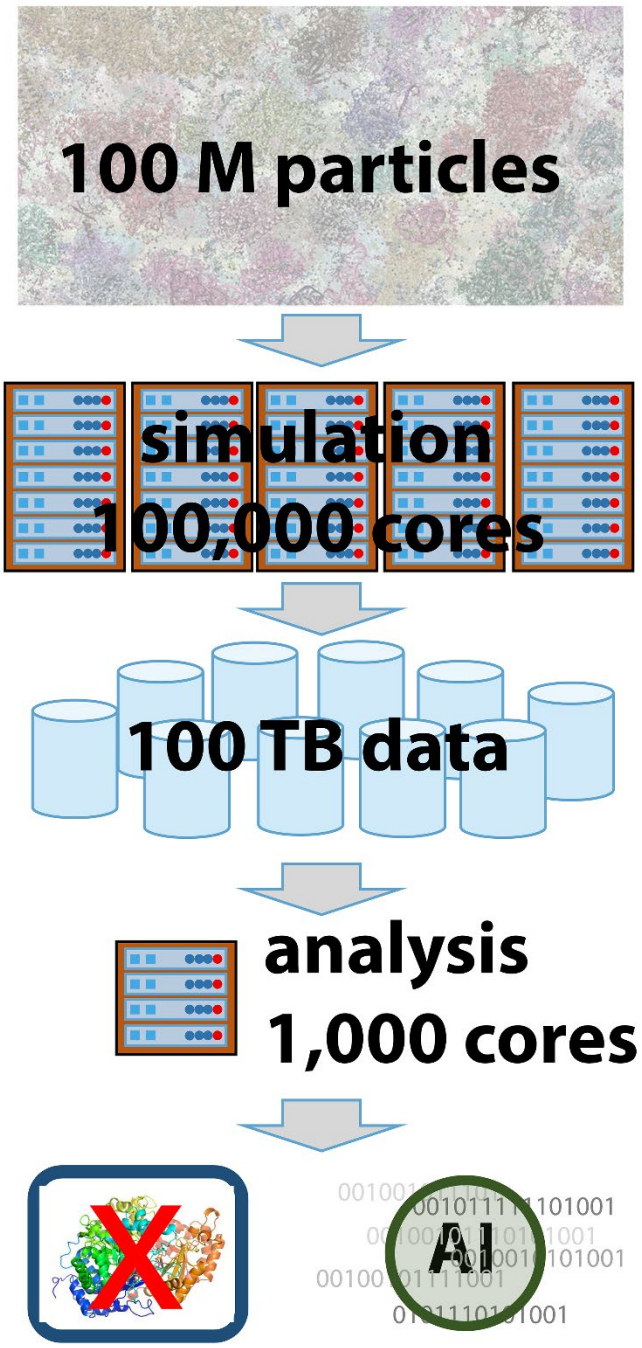

Figure 2 Flowchart of a typical simulation of a large cellular system that involves high-performance computing, data management, and analysis challenges.

Simulations of CG models or of atomistic models with continuum solvent are possible with many packages. It is especially straightforward if the CG model uses interaction potential terms

similar to the terms used in standard atomistic force fields, such as with the MARTINI CG model. However, efficient parallel scaling is a significant problem in many CG simulations, especially when the number of particles is relatively small. Many CG models do not explicitly consider solvent, which results in uneven particle distributions that create load-balancing issues. Moreover, the complex algorithms underlying continuum solvent and hydrodynamic models, if they are used, are difficult to parallelize. One solution is the use of GPU hardware in which many types of CG models can be run very efficiently, especially via OpenMM (Eastman et al. 2017), which allows flexible interaction functions to be defined at runtime. However, scaling to multiple GPUs is generally difficult with CG models, and further software developments are clearly necessary to improve the performance of large-scale CG simulations on conventional and GPU-based clusters.

Analysis and Visualization

When cellular models are subjected to simulations, the main outcome is trajectories of particle coordinates as a function of time. Even if coordinate frames are not saved often, the data generated for a cellular system with >100 M particles can easily reach petabyte scales (**Figure 2**). This creates significant challenges for analysis and visualization (Yu et al. 2018), and the analysis becomes a high-performance computing challenge that requires significant computing resources and parallel scaling (Yu et al. 2018). An even more significant problem is that the very large data sets of complex cellular systems have changed the mode of scientific discovery. While it may be straightforward to visually inspect a single-molecule trajectory via molecular graphics software such as VMD (Humphrey et al. 1996) and PyMOL (DeLano 2002), this is no longer possible for systems of 100 M atoms or more. Apart from the technical challenge of interactively animating such a large number of particles, there is simply too much information for visual analysis of detailed features by humans. Instead, data science methods are desperately needed to automatically inspect large trajectory data sets and to identify potentially interesting new scientific aspects. Only with such tools will it be possible to realize the full scientific potential of molecular-level simulations of cellular environments.

OPPORTUNITIES OF PHYSICAL WHOLE-CELL MODELING

Whole-cell modeling largely remains a promise of future scientific impact. This applies to

physical models as well as models based on kinetic networks. Network-based models require extensive kinetic data from experiment and complete functional gene annotations that are available only for a few organisms at best. However, physical models are limited by a lack of structural information and have high computational cost. As a result, successful predictive cellular models have so far been reported only for the minimal bacterium *Mycoplasma genitalium* (Karr et al. 2011, Yu et al. 2016) and, to a lesser extent, *Escherichia coli* (Frembgen-Kesner & Elcock 2013, McGuffee & Elcock 2010). Models of other and more complex systems are certain to follow, as preparations are already under way to develop whole-cell models of systems as large and complex as a human pancreatic β-cell (Singla et al. 2018).

The long-term vision of physical whole-cell modeling is that it will be possible to assemble all of the molecules in a given cell in atomistic resolution in a fully consistent fashion with experiment and then subject such an in silico cell to simulations that can reach biological timescales. If one assumes that the simulations are sufficiently accurate, the result would be a comprehensive understanding of how molecular structure and function couple at the cellular level in a way that will probably never be measurable via experiments. Moreover, such models could provide a comprehensive understanding of disease and allow for the design of new drugs in the full context of the cellular environment.

Many opportunities exist in the short term to better understand the behavior of biomolecules in cellular environments on the basis of physical models; such efforts will be motivated by recent experiments and simulations, as reviewed in detail recently (Feig et al. 2017). Fundamental questions remain about how cellular environments affect molecular stability, determine diffusive properties, give rise to nonspecific transient interactions and phase transitions, and modulate ligand binding (**Figure 3**). It is also largely unclear how the presence of genomic DNA and membrane surfaces affects biomolecular structure and dynamics.

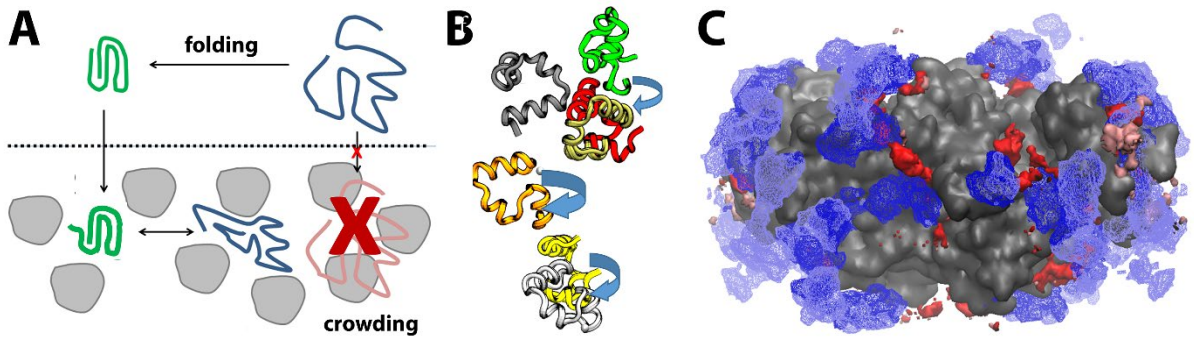

Figure 3 Effects of cellular environments on stability, dynamics, and binding. (*a*) Protein stability is altered in crowded cellular environments due to volume exclusion and interactions with crowder molecules (*gray*). Horizontal arrows indicate folding-unfolding transitions. Vertical arrows indicate transitions between dilute and crowded environments. The red 'x's indicates the state and transition presumed to be disfavored by crowding. (*b*) Rotational diffusion in crowded environments depends on transient cluster formation in concentrated villin solutions. Blue arrow indicate magnitude of rotational diffusion rates.. (*c*) Binding of ATP (*red* with crowding, *blue* without crowding) to acetate kinase (*gray*) varies in the presence of crowders.

Stability and Dynamics of Biomolecules

Most biological macromolecules carry out their function in a well-defined native state. If that state becomes destabilized, function is compromised. The question of whether crowded cellular environment affect biomolecular stability has been studied for quite some time (Minton & Wilf 1981, Zhou et al. 2008, Zimmerman & Minton 1993). Initial studies have identified a largely stabilizing role of the volume exclusion effect under crowded conditions (McPhie et al. 2006), but recent experiments suggest that crowding by proteins can have the opposite effect and can destabilize native states (Ignatova et al. 2007; Miklos et al. 2010, 2011). The role of modeling and simulation has been to provide insight into mechanistic details of the experimental observations (Candotti & Orozco 2016, Feig & Sugita 2012, Harada et al. 2013, Yu et al. 2016), as well as to generate new hypotheses for how the stability of biomolecules may be modulated under crowded conditions. One interesting observation from simulation is that metabolites at cellular concentrations appear to be able to induce more compact structures of phosphoglycerate kinase in a manner that otherwise resembles volume exclusion crowding effects (Yu et al. 2016). Other studies have begun to examine the effects of cellular environments on nucleic acid structure. One finding is a possible shift between A- and B-DNA forms as a result of cellular environments (Yildirim et al. 2014). These examples suggest that cellular environments may

impact biomolecular stability in a variety of unanticipated ways. Simulations of biomolecules in cellular environments can in principle identify all of the factors that may be at play, and much is likely left to be discovered in that regard.

The conformational dynamics of biomolecules are related to their stability for natively folded proteins but become a separate topic for highly dynamic systems and IDPs. Again, the general prediction from volume exclusion effects of crowding is that dynamic ensembles become more compact, but in particular, IDPs with polymer-like characteristics exhibit behavior that is different from that of globular proteins (Banks et al. 2018, Kang et al. 2015, Schuler et al. 2016, Soranno et al. 2014). Molecular simulations are ideally suited to studying the dynamics of biomolecules in cellular environments. IDPs under cellular or crowded conditions present rich opportunities, as very few high-resolution simulation studies have been reported (Cino et al. 2012).

Diffusion of Biomolecules

Most biological function involves the encounter of different biomolecules. Therefore, diffusional properties are essential to understanding biological function at the cellular level. It is clear that the diffusion of biomolecules is significantly retarded due to crowding (Banks & Fradin 2005, Dauty & Verkman 2004, Li et al. 2009, Szymanski et al. 2006, Wang et al. 2010), but many questions remain about how diffusional properties vary for different molecules, in different local cellular environments, and over different timescales. More generally, it is also unclear why diffusion is slowed down in cellular environments and how rotational and translational diffusion may be affected differently (Roos et al. 2016). Simulations of protein in cellular environments suggest that the propensity to transiently interact nonspecifically is a key determinant of diffusional properties (Feig & Sugita 2012, McGuffee & Elcock 2010, Nawrocki et al. 2017, Trovato & Tozzini 2014, Yu et al. 2016). It appears that translational and rotational diffusion, at least in the short-time regime, can be essentially predicted from transient cluster formation (Nawrocki et al. 2017). However, this view is at odds with other studies that ascribe hydrodynamic interactions a significant role (Ando & Skolnick 2010) or simply invoke increased solvent viscosity in crowded environments as the key factor (Ellis 2001). Again, molecular simulations of cellular environments are an ideal tool for dissecting the different contributions to diffusion. Diffusion on longer timescales and in the presence of larger cellular components and a

better understanding of anisotropic characteristics in crowded systems are particularly good opportunities for future studies.

Phase Separation of Biomolecules

Biomolecules can be viewed as either globular colloid-like particles or polymers with attractive and repulsive properties that exhibit different phase behavior, depending on their interactions and concentration (Zhou & Pang 2018). All biomolecules aggregate into gel-like, amorphous, or crystalline solid phases at concentrations above the solubility limit. At lower concentrations, finite-size clusters (Kowalczyk et al. 2011, Stradner et al. 2004, Vorontsova et al. 2015) or liquid-liquid phase-separated states may form (Elbaum-Garfinkle et al. 2015, Hyman et al. 2014). Biomolecular aggregation is often associated with disease (Ross & Poirier 2004), but recent experiments suggest that phase separations can also have important functional roles (Shin & Brangwynne 2017), e.g., to facilitate the initiation of transcription (Boehning et al. 2018). Phase separations in vivo are inherently a cellular-scale phenomenon with a physical origin at the molecular level. Therefore, physical models of cellular environments are well suited to studying such behavior. A major challenge is that it is necessary to describe the formation of macroscopic domains on macroscopic timescales on the basis of molecular interactions at the atomistic level. While this is not feasible today with fully atomistic simulations, a viable strategy is to apply CG models that have been parameterized on the basis of atomistic simulations and tuned on the basis of available experimental data. For globular molecules, simple spherical models can be sufficient, as such models have already provided much useful insight into colloid physics (Mani et al. 2014, Woldeyes et al. 2017, Zhuang & Charbonneau 2016). Phase separations involving IDPs are more challenging, as they require higher resolution to capture the characteristics of flexible polymers. As very few studies have straddled the intersection between condensed phase physics and the complexities of biological environments to understand the phase behavior of biomolecules (Nguemaha & Zhou 2018, Qin & Zhou 2017, Woldeyes et al. 2017), there is ample room for future modeling studies that can interpret the emerging experimental data.

Interactions with Cellular Components

Many studies of cellular environments focus on crowding and nonspecific interactions within the cytoplasm (Gnutt & Ebbinghaus 2016, Yu et al. 2016), neglecting the presence of genomic DNA, plasmids, membranes, organelles, and cytoskeletal elements. These large molecular

structures present obstacles to diffusing molecules and present interaction surfaces with unique properties. For example, DNA is highly charged, even after condensation of counterions (Manning 1978), whereas membrane surfaces are ionic or zwitterionic at the water interface and hydrophobic when penetrated. Chromosomal DNA also has a porous structure that allows for penetration of other molecules in a size-dependent fashion (Hacker et al. 2017, Mondal et al. 2011, Yildirim & Feig 2018). Much remains to be discovered about how biomolecules behave in the presence of such structures. For example, diffusion along DNA by DNA-binding proteins has been characterized relatively well (Marcovitz & Levy 2013, Schonhoft et al. 2013, Tan et al. 2016, Terakawa et al. 2012), but much less is known about the diffusion of proteins that do not interact specifically with DNA in the presence of chromosomal DNA (Ando & Skolnick 2014). The interaction of biomolecules with membranes, in contrast, is also well explored for proteins that are known to interact with membranes either peripherally or via insertion (Ash et al. 2004, Chavent et al. 2016, Im & Brooks 2004, Jeon et al. 2016, Kirchhoff et al. 2008), but relatively little is known about nonspecific interactions of biomolecules with membranes that are not known to be membrane bound (Aisenbrey et al. 2008) or about the effects of crowding on membranes (Guigas & Weiss 2016, Stachowiak et al. 2012). There is a clear need for modeling studies to investigate such questions in more detail, which will be possible once high-resolution structure of genomic DNA are available and realistic models of biological membranes that include membrane proteins can be constructed.

Ligand Binding in Cellular Environments

The applications of whole-cell models discussed so far focus on fundamental scientific questions. However, the most significant impact may be the ability to study ligand binding in vivo, as that would ultimately allow for in cellulo rational drug design so that specificity and selectivity can be considered from the very beginning when potential new drug candidates are evaluated. While it will be a while before practical methods for in cellulo drug design can be developed, many questions about the binding of ligands in cellular environments can be addressed in the meantime. One key question is whether cellular environments alter ligand binding affinities or pathways. Another question is to what extent ligands interact nonspecifically with biomolecules. One possible effect on biomolecular structure is mentioned above, but extensive interactions of ligands with biomolecules may also alter solvation properties and

solubility. One example is the increase of protein solubility in the presence of ATP (Patel et al. 2017). From the ligand perspective, the consequence of extensive nonspecific interactions means sequestration that results in an effectively lower concentration and lower average diffusion rates. Experiments as well as simulations have provided some evidence that this may indeed be a significant effect (Duff et al. 2012, Yu et al. 2016). Substrate channeling between enzymes, although known for a long time (Miles et al. 1999), is a related topic that is not well understood in the context of cellular environments. Therefore, further studies are needed to gain a better understanding of ligand binding in cellular environments.

CONCLUSIONS AND OUTLOOK

We are entering a time where it is possible to model and simulate cellular environments in full molecular detail. This allows for a full connection between structure and systems-level biology and in principle opens up vast predictive abilities for connecting changes at the molecular level to biological function. An important implication is the potential to better understand disease origins and to develop new, more effective therapies in which side effects are accounted for from the beginning of the design process. On the practical side, although the first models of minimal cellular environments in full atomistic detail have emerged (Feig et al. 2015, Yu et al. 2016), many challenges remain. The simulation of such systems on biologically relevant timescales is the biggest hurdle, but a lack of experimental data, especially for biomolecular structure, hinders progress as well. A viable strategy may be to apply multiscale modeling strategies in which atomistic models are used only to establish shorter-term behavior at the molecular scale and train CG models that can reach larger scales. Even just at the intermediate scales of individual biomolecules exposed to cellular environments, instead of modeling entire cells, there is actually much that remains to be learned in terms of fundamental biophysics, and this is probably where physical models of cellular environments can have the greatest impact in the near future. However, as the understanding of biomolecular behavior in cells becomes more comprehensive, physical models could be coupled with kinetic network models to add reactivity and to access even longer timescales. Ultimately, only a fully integrated approach that applies different, but connected, frameworks across different scales will likely succeed in truly capturing how molecular behavior in cellular environments leads to cellular phenotypes. The focus of this

review is on computational modeling, but the role of experiments is essential not only for providing input data but also for validating results from modeling and simulation. Experimental validation **[←**AU: Please clarify**]** will require advances also on the experimental side, which we hope will be stimulated by progress on the modeling side.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Funding was provided by the National Institutes of Health grants R01 GM084953, GM103695, and GM126948 (to M.F.); by National Science Foundation (NSF) grants MCB 1330560 and MCB 1817307 (to M.F.); by funds from FLAGSHIP 2020 project focus area 1, "Innovative Drug Discovery Infrastructure Through Functional Control of Biomolecular Systems," of the Ministry of Education, Culture, Sports, Science and Technology (to Y.S.); by a Japanese Society for the Promotion of Science Grant-in-Aid for Scientific Research on Innovative Areas "Novel Measurement Techniques for Visualizing 'Live' Protein Molecules at Work" (grant 26119006) (to YS..); by a grant from the Japan Science and Technology Agency CREST program for the area of "Structural Life Science and Advanced Core Technologies for Innovative Life Science Research" (to Y.S.); and by RIKEN Quantitative Biology Center, RIKEN Center for Biosystems Dynamics Research, RIKEN Advanced Institute for Computational Science, RIKEN Center for Computational Science, RIKEN Interdisciplinary Theoretical Science Research Group, and RIKEN Dynamic Structural Biology (to Y.S.). Computer time came at the NSF Extreme Science and Engineering Discovery Environment facilities (TG-MCB090003) and through the High Performance Computing Infrastructure Strategic Program (hp120309, hp130003, hp140229, hp150233, hp150145, and hp160120).

LITERATURE CITED

Aisenbrey C, Bechinger B, Gröbner G. 2008. Macromolecular crowding at membrane interfaces:

adsorption and alignment of membrane peptides. *J. Mol. Biol.* 375:376–85

Alexander N, Al-Mestarihi A, Bortolus M, Mchaourab H, Meiler J. 2008. De novo high-resolution protein structure determination from sparse spin-labeling EPR data. *Structure* 16:181–95

Ando T, Chow E, Saad Y, Skolnick J. 2012. Krylov subspace methods for computing hydrodynamic interactions in Brownian dynamics simulations. *J. Chem. Phys.* 137:064106

Ando T, Skolnick J. 2010. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *PNAS* 107:18457–62

Ando T, Skolnick J. 2014. Sliding of proteins non-specifically bound to DNA: Brownian dynamics studies with coarse-grained protein and DNA models. *PLOS Comp. Biol.* 10:e1003990

Ash WL, Zlomislic MR, Oloo EO, Tieleman DP. 2004. Computer simulations of membrane proteins. *Biochim. Biophys. Acta Biomembr.* 1666:158–89

Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–15

Balaeff A, Mahadevan L, Schulten K. 1999. Elastic rod model of a DNA loop in the lac operon. *Phys. Rev. Lett.* 83:4900–3

Banks A, Qin S, Weiss KL, Stanley CB, Zhou H-X. 2018. Intrinsically disordered protein exhibits both compaction and expansion under macromolecular crowding. *Biophys. J.* 114:1067–79

Banks DS, Fradin C. 2005. Anomalous diffusion of proteins due to molecular crowding. *Biophys. J.* 89:2960–71

Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. 2007. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 389:1017–31

Beck M, Topf M, Frazier Z, Tjong H, Xu M, et al. 2011. Exploring the spatial and temporal organization of a cell's proteome. *J. Struct. Biol.* 173:483–96

Bernstein D. 2005. Simulating mesoscopic reaction-diffusion systems using the Gillespie algorithm. *Phys. Rev. E* 71:041103

Best RB, Zheng W, Mittal J. 2014. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* 10:5113–24

Best RB, Zhu X, Shim J, Lopes P, Mittal J, et al. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 8:3257–73

Bianchi E, Blaak R, Likos CN. 2011. Patchy colloids: state of the art and perspectives. *Phys. Chem. Chem. Phys.* 13:6397–410

Bianco S, Chiariello AM, Annunziatella C, Esposito A, Nicodemi M. 2017. Predicting chromatin architecture from models of polymer physics. *Chromosome Res*. 25:25–34

Boehning M, Dugast-Darzacq C, Rankovic M, Hansen AS, Yu T, et al. 2018. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* 25:833–40

Bowers PM, Strauss CEM, Baker D. 2000. De novo protein structure determination using sparse NMR data. *J. Biomol. NMR* 18:311–18

Brady JF, Bossis G. 1988. Stokesian dynamics. *Annu. Rev. Fluid Mech.* 20:111–57

Brown FL. 2008. Elastic modeling of biomembranes and lipid bilayers. *Annu. Rev. Phys. Chem.* 59:685–712

Bucciarelli S, Myung JS, Farago B, Das S, Vliegenthart GA, et al. 2016. Dramatic influence of patchy attractions on short-time protein diffusion under crowded conditions. *Sci. Adv.* 2:e1601432

Buchete NV, Straub JE, Thirumalai D. 2004. Development of novel statistical potentials for protein fold recognition. *Curr. Opin. Struct. Biol.* 14:225–32

Candotti M, Orozco M. 2016. The differential response of proteins to macromolecular crowding. *PLOS Comp. Biol.* 12:e1005040

Case DA. 1994. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* 4:285–90

Chavent M, Duncan AL, Sansom MSP. 2016. Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale. *Curr. Opin. Struct. Biol.* 40:8–16

Chmiela S, Tkatchenko A, Sauceda HE, Poltavsky I, Schütt KT, Müller K-R. 2017. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* 3:e1603015

Ciccotti G, Ferrario M, Schuette C, ed. 2014. Molecular dynamics simulation. *Entropy* 16:1

Cino EA, Karttunen M, Choy W-Y. 2012. Effects of molecular crowding on the dynamics of intrinsically disordered proteins. *PLOS ONE* 7:e49876

Clementi C, Nymeyer H, Onuchic JN. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–53

Cohen RD, Pielak GJ. 2017. A cell is more than the sum of its (dilute) parts: a brief history of quinary structure. *Protein Sci*. 26:403–13

Dauty E, Verkman AS. 2004. Molecular crowding reduces to a similar extent the diffusion of small solutes and macromolecules: measurement by fluorescence correlation spectroscopy. *J. Mol. Recognit.* 17:441–47

DeLano WL. 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* 40:82-92

Di Stefano M, Paulsen J, Lien TG, Hovig E, Micheletti C. 2016. Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. *Sci. Rep.* 6:35985

Dlugosz M, Antosiewicz JM, Zielinski P, Trylska J. 2012. Contributions of far-field hydrodynamic interactions to the kinetics of electrostatically driven molecular association. *J. Phys. Chem. B* 116:5437–47

Drenth J. 2007. *Principles of Protein X-Ray Crystallography*. New York: Springer

Duff MR Jr., Grubbs J, Serpersu E, Howell EE. 2012. Weak interactions between folate and osmolytes in solution. *Biochemistry* 51:2309–18

Dumetz AC, Chockla AM, Kaler EW, Lenhoff AM. 2008. Protein phase behavior in aqueous solutions: crystallization, liquid-liquid phase separation, gels, and aggregates. *Biophys. J.* 94:570–83

Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, et al. 2017. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLOS Comp. Biol.* 13:e1005659

Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CC-H, Eckmann CR, et al. 2015. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *PNAS* 112:7189–94

Ellis RJ. 2001. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* 11:114–19

Ermak DL, McCammon JA. 1978. Brownian dynamics with hydrodynamic interactions. *J.*

Chem. Phys. 69:1352–60

Feig M, Harada R, Mori T, Yu I, Takahashi K, Sugita Y. 2015. Complete atomistic model of a bacterial cytoplasm integrates physics, biochemistry, and systems biology. *J. Mol. Graph. Modell.* 58:1–9

Feig M, Nawrocki G, Yu I, Wang P-h, Sugita Y. 2018. Challenges and opportunities in connecting simulations with experiments via molecular dynamics of cellular environments. *J. Phys. Conf. Ser.* 1036:012010

Feig M, Sugita Y. 2012. Variable interactions between protein crowders and biomolecular solutes are important in understanding cellular crowding. *J. Phys. Chem. B* 116:599–605

Feig M, Yu I, Wang P-h, Nawrocki G, Sugita Y. 2017. Crowding in cellular environments at an atomistic level from computer simulations. *J. Phys. Chem. B* 121:8009–25

Fernandez-Leiro R, Scheres SH. 2016. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* 537:339–46

Frembgen-Kesner T, Elcock AH. 2010. Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in Barnase-Barstar association. *Biophys. J.* 99: L75–77

Frembgen-Kesner T, Elcock AH. 2013. Computer simulations of the bacterial cytoplasm. *Biophys. Rev.* 5:109–19

Galindo-Murillo R, Robertson JC, Zgarbová M, Šponer J, Otyepka M, et al. 2016. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* 12:4114–27

Gnutt D, Ebbinghaus S. 2016. The macromolecular crowding effect—from in vitro into the cell. *Biol. Chem.* 397:37–44

Goodsell DS, Autin L, Olson AJ. 2018. Lattice models of bacterial nucleoids. *J. Phys. Chem. B* 122:5441–47

Gopal SM, Mukherjee S, Cheng Y-M, Feig M. 2010. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins* 78:1266–81

Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, et al. 2009. Transcriptome complexity in a genome-reduced bacterium. *Science* 326:1268–71

Guerrier C, Holcman D. 2017. Multiscale models and stochastic simulation methods for

computing rare but key binding events in cell biology. *J. Comput. Phys.* 340:617–38

Guigas G, Weiss M. 2016. Effects of protein crowding on membrane systems. *Biochim. Biophys. Acta Biomemb*. 1858:2441–50

Hacker WC, Li S, Elcock AH. 2017. Features of genomic organization in a nucleotide-resolution molecular model of the *Escherichia coli* chromosome. *Nucleic Acids Res*. 45:7541–54

Harada R, Kitao A. 2012. The fast-folding mechanism of villin headpiece subdomain studied by multiscale distributed computing. *J. Chem. Theory Comput.* 8:290–99

Harada R, Tochio N, Kigawa T, Sugita Y, Feig M. 2013. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J. Am. Chem. Soc.* 135:3696–701

Harder E, Damm W, Maple J, Wu C, Reboul M, et al. 2016. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* 12:281–96

Harrigan MP, Sultan MM, Hernández CX, Husic BE, Eastman P, et al. 2017. MSMBuilder: statistical models for biomolecular dynamics. *Biophys. J.* 112:10–15

Hess B, Kutzner C, Van Der Spoel D, Lindahl E. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–47

Hospital A, Goñi JR, Orozco M, Gelpi JL. 2015. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* 8:37–47

Huang B, Bates M, Zhuang X. 2009. Super-resolution fluorescence microscopy. *Annu. Rev. Biochem.* 78:993–1016

Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, et al. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* 14:71–73

Huggins DJ, Biggin PC, Dämgen MA, Essex JW, Harris SA, et al. 2018. Biomolecular simulations: from dynamics and mechanisms to computational assays of biological activity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* In press; https://doi.org/10.1002/wcms.1393

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38

Hyman AA, Weber CA, Jülicher F. 2014. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* 30:39–58

Ignatova Z, Krishnan B, Bombardier JP, Marcelino AMC, Hong J, Gierasch LM. 2007. From the test tube to the cell: exploring the folding and aggregation of a β-clam protein. *Biopolymers* 88:157–63

Im W, Brooks CL III. 2004. De novo folding of integral membrane proteins: an exploration of the structure and NMR properties of the fd coat protein. *J. Mol. Biol.* 337:513–19

Im W, Liang J, Olson A, Zhou H-X, Vajda S, Vakser IA. 2016. Challenges in structural approaches to cell modeling. *J. Mol. Biol.* 428:2943–64

Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, et al. 2009. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* 458:106–11

Jeon J-H, Javanainen M, Martinez-Seara H, Metzler R, Vattulainen I. 2016. Protein crowding in lipid bilayers gives rise to non-Gaussian anomalous lateral diffusion of phospholipids and proteins. *Phys. Rev. X* 6:021006

Jung J, Mori T, Kobayashi C, Matsunaga Y, Yoda T, et al. 2015. GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 5:310–23

Jung J, Naurse A, Kobayashi C, Sugita Y. 2016. Graphics processing unit acceleration and parallelization of GENESIS for large-scale molecular dynamics simulations. *J. Chem. Theory Comput.* 12:4947–58

Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, et al. 1999. NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312

Kang H, Pincus PA, Hyeon C, Thirumalai D. 2015. Effects of macromolecular crowding on the collapse of biopolymers. *Phys. Rev. Lett.* 114:068303

Kar P, Feig M. 2014. Recent advances in transferable coarse-grained modeling of proteins. *Adv. Prot. Chem. Struct. Biol.* 96:143–80

Kar P, Feig M. 2017. Hybrid all-atom/coarse-grained simulations of proteins by direct coupling of CHARMM and PRIMO force fields. *J. Chem. Theory Comput.* 13:5753–65

Kar P, Gopal SM, Cheng YM, Predeus AV, Feig M. 2013. PRIMO: a transferable coarse-grained force field for proteins. *J. Chem. Theory Comput.* 9:3769–88

Karr JR, Sanghvi JC, Jacobs JM, Covert MW. 2011. Toward a whole cell model of *Mycoplasma genitalium*. *Biophys. J.* 100:32

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, et al. 2012. A whole-cell

computational model predicts phenotype from genotype. *Cell* 150:389–401

Karr JR, Takahashi K, Funahashi A. 2015. The principles of whole-cell modeling. *Curr. Opin. Microbiol.* 27:18–24

Kirchhoff H, Haferkamp S, Allen JF, Epstein DB, Mullineaux CW. 2008. Protein diffusion and macromolecular crowding in thylakoid membranes. *Plant Physiol*. 146:1571–78

Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, et al. 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* 114:7830–43

Klein T, Autin L, Kozlíková B, Goodsell DS, Olson A, et al. 2018. Instant construction and visualization of crowded biological environments. *IEEE Trans. Vis. Comput. Graph.* 24:862–72

Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. 2016. Coarse-grained protein models and their applications. *Chem. Rev.* 116:7898–936

Kobayashi C, Jung J, Matsunaga Y, Mori T, Ando T, et al. 2017. GENESIS 1.1: a hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J. Comput. Chem.* 38:2193–206

Koldsø H, Sansom MSP. 2015. Organization and dynamics of receptor proteins in a plasma membrane. *J. Am. Chem. Soc.* 137:14694–704

Kowalczyk P, Ciach A, Gauden PA, Terzyk AP. 2011. Equilibrium clusters in concentrated lysozyme protein solutions. *J. Colloid Interface Sci.* 363:579–84

Kuhner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, et al. 2009. Proteome organization in a genome-reduced bacterium. *Science* 326:1235–40

Lane TJ, Shukla D, Beauchamp KA, Pande VS. 2013. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* 23:58–65

Li C, Wang Y, Pielak GJ. 2009. Translational and rotational diffusion of a small globular protein under crowded conditions. *J. Phys. Chem. B* 113:13390–92

Li W, Zhang Y, Skolnick J. 2004. Application of sparse NMR restraints to large-scale protein structure prediction. *Biophys. J.* 87:1241–48

Li Z, Kermode JR, De Vita A. 2015. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* 114:096405

Macklin DN, Ruggero NA, Covert MW. 2014. The future of whole-cell modeling. *Curr. Opin.*

Biotech. 28:111–15

Mani E, Lechner W, Kegel WK, Bolhuis PG. 2014. Equilibrium and non-equilibrium cluster phases in colloids with competing interactions. *Soft Matter* 10:4479–86

Manning GS. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* 11:179–246

Marcovitz A, Levy Y. 2013. Obstacles may facilitate and direct DNA search by proteins. *Biophys. J.* 104:2042–50

Marsh JA, Teichmann SA. 2015. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* 84:551–75

McGuffee SR, Elcock AH. 2010. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLOS Comp. Biol.* 6:e1000694

McPhie P, Ni YS, Minton AP. 2006. Macromolecular crowding stabilizes the molten globule form of apomyoglobin with respect to both cold and heat unfolding. *J. Mol. Biol.* 361:7–10

Mereghetti P, Wade RC. 2012. Atomic detail Brownian dynamics simulations of concentrated protein solutions with a mean field treatment of hydrodynamic interactions. *J. Phys. Chem. B* 116:8523–33

Miao Z, Adamiak RW, Blanchet M-F, Boniecki M, Bujnicki JM, et al. 2015. RNA-puzzles round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* 21:1066–84

Miklos AC, Li C, Sharaf NG, Pielak GJ. 2010. Volume exclusion and soft interaction effects on protein stability under crowded conditions. *Biochemistry* 49:6984–91

Miklos AC, Sarkar M, Wang Y, Pielak GJ. 2011. Protein crowding tunes protein stability. *J. Am. Chem. Soc.* 133:7116–20

Miles EW, Rhee S, Davies DR. 1999. The molecular basis of substrate channeling. *J. Biol. Chem.* 274:12193–96

Minton AP, Wilf J. 1981. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* 20:4821–26

Modi V, Xu Q, Adhikari S, Dunbrack RL Jr. 2016. Assessment of template-based modeling of protein structure in CASP11. *Proteins* 84:200–20

Mondal J, Bratton BP, Li Y, Yethiraj A, Weisshaar JC. 2011. Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* cells. *Biophys. J.* 100:2605–13

Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. 2008. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* 4:819–34

Mori T, Miyashita N, Im W, Feig M, Sugita Y. 2016. Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms. *Biochim. Biophys. Acta Biomemb*. 1858:1635–51

Nawrocki G, Wang P-h, Yu I, Sugita Y, Feig M. 2017. Slow-down in crowded protein solutions correlates with transient oligomer formation. *J. Phys. Chem. B* 121:11072–84

Nguemaha V, Zhou H-X. 2018. Liquid-liquid phase separation of patchy particles illuminates diverse effects of regulatory components on protein droplet formation. *Sci. Rep.* 8:6728

Pastore A, Temussi PA. 2017. The emperor's new clothes: myths and truths of in-cell NMR. *Arch. Biochem. Biophys.* 628:114–22

Patel A, Malinovska L, Saha S, Wang J, Alberti S, et al. 2017. ATP as a biological hydrotrope. *Science* 356:753–56

Prytkova V, Heyden M, Khago D, Freites JA, Butts CT, et al. 2016. Multi-conformation monte carlo: a method for introducing flexibility in efficient simulations of many-protein systems. *J. Phys. Chem. B* 120:8115–26

Qin S, Zhou H-X. 2017. Protein folding, binding, and droplet formation in cell-like conditions. *Curr. Opin. Struct. Biol.* 43:28–37

Renevey A, Riniker S. 2017. Improved accuracy of hybrid atomistic/coarse-grained simulations using reparametrised interactions. *J. Chem. Phys.* 146:124131

Rivas G, Minton AP. 2016. Macromolecular crowding in vitro, in vivo, and in between. *Trends Biochem. Sci.* 41:970–81

Robustelli P, Piana S, Shaw DE. 2018. Developing a molecular dynamics force field for both folded and disordered protein states. *PNAS* 115: E4758–66

Rohl CA, Strauss CE, Misura KM, Baker D. 2004. Protein structure prediction using Rosetta. *Methods Enzymol*. 383:66–93

Roos M, Ott M, Hofmann M, Link S, Rössler E, et al. 2016. Coupling and decoupling of rotational and translational diffusion of proteins under crowding conditions. *J. Am. Chem. Soc.* 138:10365–72

Ross CA, Poirier MA. 2004. Protein aggregation and neurodegenerative disease. *Nat. Med.*

10:S10–17

Roux B, Simonson T. 1999. Implicit solvent models. *Biophys. Chem.* 78:1–20

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5:725–38

Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, et al. 2009. Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* 458:102–5

Schonhoft JD, Kosowicz JG, Stivers JT. 2013. DNA translocation by human uracil DNA glycosylase: role of DNA phosphate charge. *Biochemistry* 52:2526–35

Schuler B, Soranno A, Hofmann H, Nettels D. 2016. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu. Rev. Biophys.* 45:207–31

Shin Y, Brangwynne CP. 2017. Liquid phase condensation in cell physiology and disease. *Science* 357:eaaf4382

Singla J, McClary KM, White KL, Alber F, Sali A, Stevens RC. 2018. Opportunities and challenges in building a spatiotemporal multi-scale model of the human pancreatic β cell. *Cell* 173:11–19

Solernou A, Hanson BS, Richardson RA, Welch R, Read DJ, et al. 2018. Fluctuating finite element analysis (FFEA): a continuum mechanics software tool for mesoscale simulation of biomolecules. *PLOS Comp. Biol.* 14:e1005897

Soranno A, Koenig I, Borgia MB, Hofmann H, Zosel F, et al. 2014. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *PNAS* 111:4874–79

Stachowiak JC, Schmid EM, Ryan CJ, Ann HS, Sasaki DY, et al. 2012. Membrane bending by protein–protein crowding. *Nat. Cell Biol.* 14:944–49

Stradner A, Sedgwick H, Cardinaux F, Poon WCK, Egelhaaf SU, Schurtenberger P. 2004. Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature* 432:492–95

Szigeti B, Roth YD, Sekar JA, Goldberg AP, Pochiraju SC, Karr JR. 2018. A blueprint for human whole-cell modeling. *Curr. Opin. Syst. Biol.* 7:8–15

Szymanski J, Patkowski A, Wilk A, Garstecki P, Holyst R. 2006. Diffusion and viscosity in a crowded environment: from nano- to macroscale. *J. Phys. Chem. B* 110:25593–97

Takada S, Kanada R, Tan C, Terakawa T, Li W, Kenzaki H. 2015. Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. *Acc. Chem. Res.* 48:3026–35

Tan C, Terakawa T, Takada S. 2016. Dynamic coupling among protein binding, sliding, and DNA bending revealed by molecular dynamics. *J. Am. Chem. Soc.* 138:8512–22

Tan D, Piana S, Dirks RM, Shaw DE. 2018. RNA force field with accuracy comparable to state-of-the-art protein force fields. *PNAS* 115: E1346–55

Tanizaki S, Feig M. 2005. A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. *J. Chem. Phys.* 122:124706

Tempkin JOB, Qi B, Saunders MG, Roux B, Dinner AR, Weare J. 2014. Using multiscale preconditioning to accelerate the convergence of iterative molecular calculations. *J. Chem. Phys.* 140:184114

Terakawa T, Kenzaki H, Takada S. 2012. p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J. Am. Chem. Soc.* 134:14555–62

Tiana G, Giorgetti L. 2018. Integrating experiment, theory and simulation to determine the structure and dynamics of mammalian chromosomes. *Curr. Opin. Struct. Biol.* 49:11–17

Tomita M. 2001. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol*. 19:205–10

Tomita M, Hashimoto K, Takahashi K, Matsuzaki Y, Matsushima R, et al. 2000. The E-CELL project: towards integrative simulation of cellular processes. *New Generat. Comput*. 18:1–12

Trovato F, Fumagalli G. 2017. Molecular simulations of cellular processes. *Biophys. Rev.* 9:941–58

Trovato F, Tozzini V. 2014. Diffusion within the cytoplasm: a mesoscale model of interacting macromolecules. *Biophys. J.* 107:2579–91

Vitkup D, Melamud E, Moult J, Sander C. 2001. Completeness in structural genomics. *Nat. Struct. Biol.* 8:559–66

von Lilienfeld OA, Ramakrishnan R, Rupp M, Knoll A. 2015. Fourier series of atomic radial distribution functions: a molecular fingerprint for machine learning models of quantum chemical properties. *Int. J. Quantum Chem.* 115:1084–93

Vorontsova MA, Chan HY, Lubchenko V, Vekilov PG. 2015. Lack of dependence of the sizes of

the mesoscopic protein clusters on electrostatics. *Biophys. J.* 109:1959–68

Walcott S, Sun SX. 2010. A mechanical model of actin stress fiber formation and substrate elasticity sensing in adherent cells. *PNAS* 107:7757–62

Wang L-P, McKiernan KA, Gomes J, Beauchamp KA, Head-Gordon T, et al. 2017. Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. *J. Phys. Chem. B* 121:4023–39

Wang Y, Li C, Pielak GJ. 2010. Effects of proteins on protein diffusion. *J. Am. Chem. Soc.* 132:9392–97

Wang Y, Sarkar M, Smith AE, Krois AS, Pielak GJ. 2012. Macromolecular crowding and protein stability. *J. Am. Chem. Soc.* 134:16614–18

Ward MD, Nangia S, May ER. 2017. Evaluation of the hybrid resolution PACE model for the study of folding, insertion, and pore formation of membrane associated peptides. *J. Comput. Chem.* 38:1462–71

Wassenaar TA, Ingolfsson HI, Priess M, Marrink SJ. 2011. Mixing MARTINI: electrostatic coupling in hybrid atomistic-coarse-grained biomolecular simulations. *J. Phys. Chem. B* 117:3516–30

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 45:D313–19

Westbrook J, Feng ZK, Chen L, Yang HW, Berman HM. 2003. The Protein Data Bank and structural genomics. *Nucleic Acids Res*. 31:489–91

Woldeyes MA, Calero-Rubio C, Furst EM, Roberts CJ. 2017. Predicting protein interactions of concentrated globular protein solutions using colloidal models. *J. Phys. Chem. B* 121:4756–67

Yildirim A, Feig M. 2018. High-resolution 3D models of *Caulobacter crescentus* chromosome reveal genome structural variability and organization. *Nucleic Acids Res*. 46:3937–52

Yildirim A, Sharma M, Varner BM, Fang L, Feig M. 2014. Conformational preferences of DNA in reduced dielectric environments. *J. Phys. Chem. B* 118:10874–81

Yu I, Feig M, Sugita Y. 2018. High-performance data analysis on the big trajectory data of cellular scale all-atom molecular dynamics simulations. *J. Phys. Conf. Ser.* 1036:012009

Yu I, Mori T, Ando T, Harada R, Jung J, et al. 2016. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife*

5:e19274

Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, et al. 2009. Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326:1263–68

Zhang WH, Chen JH. 2014. Accelerate sampling in atomistic energy landscapes using topology-based coarse-grained models. *J. Chem. Theory Comput.* 10:918–23

Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J. 2006. On the origin and highly likely completeness of single-domain protein structures. *PNAS* 103:2605–10

Zhou H-X, Pang X. 2018. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* 118:1691–741

Zhou H-X, Rivas G, Minton AP. 2008. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37:375–97

Zhuang Y, Charbonneau P. 2016. Recent advances in the theory and simulation of model colloidal microphase formers. *J. Phys. Chem. B* 120:7775–82

Zimmerman SB, Minton AP. 1993. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* 22:27–65