



Published in final edited form as:

Annu Rev Cell Dev Biol. 2019 October 06; 35: 191–211. doi:10.1146/annurev-cellbio-100617-062542.

Whole-Cell Models and Simulations in Molecular Detail

Michael Feig^{1,2,*}, Yuji Sugita^{2,3}

¹Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, United States;

²Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Kobe, Japan,

³Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, 2–1 Hirosawa, Wako, Japan

Abstract

Comprehensive data about the composition and structure of cellular components has enabled the construction of quantitative whole-cell models. While kinetic network type models have already been established, it is now 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.

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 understand these processes. Therefore, a major effort in modern biology has been focused on improving the resolution of macromolecular structures at sub-nanometer 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

*Corresponding Author: Michael Feig, Department of Biochemistry & Molecular Biology, Michigan State University, 603 Wilson Road, Room BCH 218, East Lansing, MI 48824, USA, mfeiglab@gmail.com, +1-517-432-7439.

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 2017). 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 has 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 non-specific 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 only focus on narrow points in the space-time universe. For example, crystallography and cryoEM 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 living cells (Drenth 2007, Fernandez-Leiro & Scheres 2016). On the other hand, the best single-molecule fluorescence methods can track the dynamics of molecules *in vivo* on millisecond time scales 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 so far rely on empirical mathematical models that are parameterized based on experimental data and focus on a kinetic view of cellular processes (Guerrier & Holcman 2017, Karr et al 2015, Macklin et al 2014, Szigeti et al 2017, Tomita 2001). These models can access macroscopic time scales and connect directly with biological function. However, spatial resolution is often limited or absent altogether 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 such 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 are physical models that build up by assembling individual molecules based on structures at the atomistic level (Figure 1) (Im et al 2016) and then subjecting such models to computer simulations (Feig et al 2018, Feig et al 2017). 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 time 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 those that connect with molecular structure at atomistic resolution. The main emphasis will be 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 will be available.

CHALLENGES IN BUILDING PHYSICAL MODELS OF CELLS

Physical models of cellular environments that involve molecular level of detail hold much promise but face a number of challenges which 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 the following.

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 time scales and what size of systems can be investigated with 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. Often, atomistic models are explored via molecular dynamics (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 $1/10^{\text{th}}$ to $1/20^{\text{th}}$ 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 time scales that can be reached directly are generally limited to sub-millisecond scales, even on the largest computer clusters available today. The accessible time scales can be extended, though, 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 where entire molecule are represented as a single sphere. The reduced degrees of freedom typically come at a price of reduced accuracy and ability to apply physical laws. CG models are typically parameterized based on a combination of data from experiment and higher-resolution modeling such as atomistic simulations. As a result, universal use and transferability is 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 questions at larger scales are addressed that do not require atomistic resolution (Ando & Skolnick 2010).

Another strategy for reducing the computational cost is the use of continuum representations for parts of a system where macroscopic properties dominate or where experimental data does 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 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 to capture 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 the 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 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 multi-scale 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 multi-scale frameworks focus on bridging spatial scales, it is also possible to bridge time scales. The use of continuum models to avoid kinetic barriers has already been mentioned. 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). Multi-scale methods are in principle highly suitable for studying cellular environments where 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 a lack of practical implementations that perform well on high-performance computing platforms.

Interaction potentials

Interaction potentials between the sites of a given model allow 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 non-bonded 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 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 (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. Especially, available parameters

for small molecules have variable accuracy. This 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. Often, the empirical potentials are parameterized based on 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 therefore is well-suited to study 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 to study 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 been used for a long time to model the interactions of colloids (Bianchi et al 2011). These types of models can also be applied to biological macromolecules if internal structural 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. Using ML methods it is possible 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 for that matter, 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 the system composition, the concentration of each species, and structural information is 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 actually 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 environments is non-trivial (Figure 1). The initial placement of molecules needs to be consistent with the overall dimensions of a given cellular system and match lower-resolution imaging and tomography data when available. Moreover, the naïve 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 how such complex systems can be built, but two specific approaches appear to be most promising. 1) Instead of assembling a crowded system directly at full resolution, a hierarchical multi-scale scheme may be applied where biomolecules are placed initially at lower resolution to facilitate optimization and relaxation before gradually increasing resolution until the target resolution is reached (Figure 1) (Feig et al 2015). 2) A crowded system can also be built up by packing molecules into pre-defined volumes or shapes. This method is implemented in the CellPACK software, which uses sophisticated packing algorithms to assemble complex systems within short time (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). Coarse-grained models can also be simulated efficiently via Monte Carlo (MC) techniques (Kmieciak 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 study 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 up to 1 M particles, but they are generally not prepared to address the unique challenges when 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, this is only supported by some MD software packages so far, most notably GENESIS (Jung et al 2015, Jung et al 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 only possible with a few packages (Jung et al 2016, Kale et al 1999) (Hess et al 2008).

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 where many types of CG models can be run very efficiently, especially via openMM (Eastman et al 2017), which allows flexible interaction functions 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 are 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 peta-byte scales (Figure 2). This creates significant challenges for analysis and visualization (Yu et al 2018) and, the analysis itself becomes a high-performance computing challenge that requires significant computing resources and parallel scaling in itself (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 not possible anymore 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 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 which are only available for a few organisms at best. On the other hand, physical models are limited by a lack of structural information and high computational cost. As a result, successful predictive cellular models have so far only been reported for the minimal bacterium *Mycoplasma genitalium* (Karr et al 2011, Yu et al 2016) and, to a lesser extent, for *Escherichia coli* (Frembgen-Kesner & Elcock 2013, McGuffee & Elcock 2010). It appears to be certain that models of other and more complex systems will follow as preparations are already underway 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 time scales. Assuming 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 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 based on physical models and 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 non-specific 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.

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 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 destabilize native states (Ignatova et al 2007, Miklos et al 2010, Miklos et al 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) but also 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 it is likely that much is left to be discovered in that regard.

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

Diffusion of biomolecules

Most biological function involves the encounter of different biomolecules. Therefore, diffusional properties are essential to understand 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 time scales. More generally, it is also still unclear why exactly 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 non-specifically 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 to dissect the different contributions to diffusion. Especially diffusion on longer time scales, in the presence of larger cellular components, and a better understanding of anisotropic characteristics in crowded systems are good opportunities for future studies.

Phase separation of biomolecules

Biomolecules can be viewed either as 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 study such behavior. A major challenge is that it is necessary to describe the formation of macroscopic domains on macroscopic time scales based on 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 based on atomistic simulations and tuned based on available experimental data. For globular molecules, simple spherical models can be sufficient as such models have already provided much useful insight in 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 non-specific 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 penetration of other molecules in a size-dependent fashion (Hacker et al 2017, Mondal et al 2011, Yildirim & Feig 2017). 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, Schonhofs 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, on the other hand, 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 non-specific interactions of biomolecules with membranes that are not known to be membrane-bound (Aisenbrey et al 2008) and 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 *in cellulo* rational drug design so that specificity and selectivity can be considered from the very beginning when evaluating potential new drug candidates. 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 non-specifically with biomolecules. One possible effect on biomolecular structure has already been 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 non-specific 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, it is clear that 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 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 develop new, more effective therapies where side effects are accounted for from the beginning of the design process. On the practical side, although 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 time scales 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 multi-scale modeling strategies where atomistic models are used only to establish shorter-term behavior at the molecular scale and train coarse-grained models that can reach larger scales. Even just at 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 access even longer time scales. 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 just to

provide input data but to validate results from modeling and simulation. This will require advances also on the experimental side which we hope will be stimulated by progress on the modeling side.

ACKNOWLEDGEMENTS

Funding was provided by the National Institute of Health Grants R01 GM084953, GM103695, and GM126948 (to MF), from the National Science Foundation Grants MCB 1330560 and MCB 1817307 (to MF), the Fund from the FLAGSHIP 2020 project focused area 1 “Innovative drug discovery infrastructure through functional control of biomolecular systems of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) (to YS), a Grant-in-Aid for Scientific Research on Innovative Areas “Novel measurement techniques for visualizing ‘live’ protein molecules at work” (No. 26119006) (to YS), a grant from JST CREST on “Structural Life Science and Advanced Core Technologies for Innovative Life Science Research” (to YS), RIKEN QBIC/BDR, AICS/R-CCS, iTHES, and Dynamic Structural Biology (to YS). Computer time was used at NSF XSEDE facilities (TG-MCB090003) and at the High Performance Computing Infrastructure (HPCI) Strategic Program (hp120309, hp130003, hp140229, hp150233, hp150145, and hp160120)

REFERENCES

- Aisenbrey C, Bechinger B, Gröbner G. 2008 Macromolecular crowding at membrane interfaces: adsorption and alignment of membrane peptides. *J. Mol. Biol* 375: 376–85 [PubMed: 18022193]
- 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 [PubMed: 18275810]
- 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 [PubMed: 22897254]
- Ando T, Skolnick J. 2010 Crowding and Hydrodynamic Interactions Likely Dominate *in vivo* Macromolecular Motion. *Proc. Natl. Acad. Sci. U.S.A* 107: 18457–62 [PubMed: 20937902]
- 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 [PubMed: 11159421]
- Balaeff A, Mahadevan L, Schulten K. 1999 Elastic rod model of a DNA loop in the lac operon. *Phys. Rev. Lett* 83: 4900
- 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 [PubMed: 29539394]
- Banks DS, Fradin C. 2005 Anomalous Diffusion of Proteins Due to Molecular Crowding. *Biophys. J* 89: 2960–71 [PubMed: 16113107]
- Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. 2007 Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* 389: 1017–31 [PubMed: 17668192]
- 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 [PubMed: 21094684]
- 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 [PubMed: 25400522]
- 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 [PubMed: 23341755]

- Bianchi E, Blaak R, Likos CN. 2011 Patchy colloids: state of the art and perspectives. *Phys. Chem. Chem. Phys* 13: 6397–410 [PubMed: 21331432]
- Bianco S, Chiariello AM, Annunziatella C, Esposito A, Nicodemi M. 2017 Predicting chromatin architecture from models of polymer physics. *Chromosome Research* 25: 25–34 [PubMed: 28070687]
- 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 [PubMed: 30127355]
- Bowers PM, Strauss CEM, Baker D. 2000 De novo protein structure determination using sparse NMR data. *J. Biomol. NMR* 18: 311–18 [PubMed: 11200525]
- 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 [PubMed: 18173377]
- Bucciarelli S, Myung JS, Farago B, Das S, Vliegthart GA, et al. 2016 Dramatic Influence of Patchy Attractions on Short-Time Protein Diffusion under Crowded Conditions. *Sci. Adv* 2: e1601432 [PubMed: 27957539]
- Buchete NV, Straub JE, Thirumalai D. 2004 Development of novel statistical potentials for protein fold recognition. *Curr. Opin. Struct. Biol* 14: 225–32 [PubMed: 15093838]
- 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 [PubMed: 27341016]
- 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 [PubMed: 28508076]
- Ciccotti G, Ferrario M, Schuette C. 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 [PubMed: 23189168]
- 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 I. *J. Mol. Biol* 298: 937–53 [PubMed: 10801360]
- 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 [PubMed: 27977883]
- 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 [PubMed: 15362103]
- DeLano WL. 2002 The PyMOL Molecular Graphics System. pp. <http://www.pymol.org>: DeLano Scientific, San Carlos, CA
- 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.-Uk* 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 [PubMed: 22512305]
- 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 [PubMed: 22369433]
- 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 [PubMed: 18160663]
- 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. *Proc. Natl. Acad. Sci. U.S.A* 112: 7189–94 [PubMed: 26015579]
- Ellis RJ. 2001 Macromolecular Crowding: an Important but Neglected Aspect of the Intracellular Environment. *Curr. Opin. Struct. Biol* 11: 114–19 [PubMed: 11179900]
- 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. *Journal of Physics: Conference Series* 2018, 1036: 012010 IOP Publishing.
- 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 [PubMed: 22117862]
- 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 [PubMed: 28666087]
- Fernandez-Leiro R, Scheres SH. 2016 Unravelling biological macromolecules with cryo-electron microscopy. *Nature* 537: 339 [PubMed: 27629640]
- 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–L77 [PubMed: 21044566]
- Frembgen-Kesner T, Elcock AH. 2013 Computer Simulations of the Bacterial Cytoplasm. *Biophys. Rev* 5: 109–19 [PubMed: 23914257]
- 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 [PubMed: 27300587]
- Gnutt D, Ebbinghaus S. 2016 The Macromolecular Crowding Effect – From *in vitro* into the Cell. *Biol. Chem* 397: 37–44 [PubMed: 26351910]
- Goodsell DS, Autin L, Olson AJ. 2018 Lattice Models of Bacterial Nucleoids. *J. Phys. Chem. B* 122: 5441–47 [PubMed: 29338247]
- 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 [PubMed: 19967787]
- 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 [PubMed: 19965477]
- 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. *BBA 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 [PubMed: 28645155]
- 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 [PubMed: 26592890]
- 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 [PubMed: 23402619]
- 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 [PubMed: 26584231]

- Harrigan MP, Sultan MM, Hernández CX, Husic BE, Eastman P, et al. 2017 MSMBUILDER: Statistical Models for Biomolecular Dynamics. *Biophys. J* 112: 10–15 [PubMed: 28076801]
- 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 [PubMed: 26620784]
- Hospital A, Goñi JR, Orozco M, Gelpi JL. 2015 Molecular Dynamics Simulations: Advances and Applications. *Adv. Appl. Bioinform. Chem* 8: 37–47 [PubMed: 26604800]
- Huang B, Bates M, Zhuang X. 2009 Super-resolution fluorescence microscopy. *Annu. Rev. Biochem* 78: 993–1016 [PubMed: 19489737]
- 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 [PubMed: 27819658]
- 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*: e1393
- Humphrey W, Dalke A, Schulten K. 1996 VMD: Visual Molecular Dynamics. *J. Mol. Graph* 14: 33–38 [PubMed: 8744570]
- Hyman AA, Weber CA, Jülicher F. 2014 Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol* 30: 39–58 [PubMed: 25288112]
- 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 [PubMed: 17206628]
- Im W, Brooks CL III. 2004 De Novo Folding of Integral Membrane Proteins. *J. Mol. Biol* Submitted
- 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 [PubMed: 27255863]
- 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 [PubMed: 19262675]
- 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. *Physical Review 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 [PubMed: 26753008]
- Jung J, Nourse 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 [PubMed: 27631425]
- 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 [PubMed: 25723249]
- 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*: under review
- 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 [PubMed: 23997693]
- Karr JR, Sanghvi JC, Jacobs JM, Covert MW. 2011 Toward a Whole Cell Model of Mycoplasma Genitalium. *Biophys. J* 100: 32–32 [PubMed: 21190654]
- 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 [PubMed: 22817898]
- Karr JR, Takahashi K, Funahashi A. 2015 The principles of whole-cell modeling. *Curr. Opin. Microbiol* 27: 18–24 [PubMed: 26115539]

- Kirchhoff H, Haferkamp S, Allen JF, Epstein DB, Mullineaux CW. 2008 Protein diffusion and macromolecular crowding in thylakoid membranes. *Plant physiology* 146: 1571–78 [PubMed: 18287489]
- Klada 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 [PubMed: 20496934]
- Klein T, Autin L, Kozlíková B, Goodsell DS, Olson A, et al. 2018 Instant Construction and Visualization of Crowded Biological Environments. *IEEE transactions on visualization and computer graphics* 24: 862–72 [PubMed: 28866533]
- Kmieciak S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. 2016 Coarse-grained protein models and their applications. *Chem. Rev* 116: 7898–936 [PubMed: 27333362]
- 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 [PubMed: 28718930]
- Koldsø H, Sansom MSP. 2015 Organization and Dynamics of Receptor Proteins in a Plasma Membrane. *J. Am. Chem. Soc* 137: 14694–704 [PubMed: 26517394]
- Kowalczyk P, Ciach A, Gauden PA, Terzyk AP. 2011 Equilibrium Clusters in Concentrated Lysozyme Protein Solutions. *J. Colloid Interface Sci* 363: 579–84 [PubMed: 21880324]
- 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 [PubMed: 19965468]
- 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 [PubMed: 23237705]
- 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 [PubMed: 19791823]
- Li W, Zhang Y, Skolnick J. 2004 Application of Sparse NMR Restraints to Large-Scale Protein Structure Prediction. *Biophys. J* 87: 1241–48 [PubMed: 15298926]
- 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 [PubMed: 25793835]
- Macklin DN, Ruggero NA, Covert MW. 2014 The future of whole-cell modeling. *Curr. Opin. Biotech* 28: 111–15 [PubMed: 24556244]
- 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 [PubMed: 24824226]
- Manning GS. 1978 The Molecular Theory of Polyelectrolyte Solutions with Applications to the Electrostatic Properties of Polynucleotides. *Q. Rev. Biophys* 11: 179–246 [PubMed: 353876]
- Marcovitz A, Levy Y. 2013 Obstacles may facilitate and direct DNA search by proteins. *Biophys. J* 104: 2042–50 [PubMed: 23663847]
- Marsh JA, Teichmann SA. 2015 Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem* 84: 551–75 [PubMed: 25494300]
- 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 [PubMed: 16824541]
- 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 [PubMed: 22594708]
- 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*
- 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 [PubMed: 20672856]
- Miklos AC, Sarkar M, Wang Y, Pielak GJ. 2011 Protein Crowding Tunes Protein Stability. *J. Am. Chem. Soc* 133: 7116–20 [PubMed: 21506571]

- Miles EW, Rhee S, Davies DR. 1999 The molecular basis of substrate channeling. *J. Biol. Chem* 274: 12193–96 [PubMed: 10212181]
- 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 [PubMed: 7295652]
- Modi V, Xu Q, Adhikari S, Dunbrack RL Jr. 2016 Assessment of template-based modeling of protein structure in CASP11. *Proteins* 84: 200–20 [PubMed: 27081927]
- 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 [PubMed: 21641305]
- 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 [PubMed: 26621095]
- 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. *BBA 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 [PubMed: 29151345]
- 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.-Uk* 8
- Pastore A, Temussi PA. 2017 The Emperor's New Clothes: Myths and Truths of In-Cell NMR. *Arch. Biochem. Biophys*
- Patel A, Malinowska L, Saha S, Wang J, Alberti S, et al. 2017 ATP as a biological hydrotrope. *Science* 356: 753–56 [PubMed: 28522535]
- 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 [PubMed: 27063730]
- Qin S, Zhou H-X. 2017 Protein Folding, Binding, and Droplet Formation in Cell-Like Conditions. *Curr. Opin. Struct. Biol* 43: 28–37 [PubMed: 27771543]
- Renevey A, Riniker S. 2017 Improved Accuracy of Hybrid Atomistic/Coarse-Grained Simulations Using Reparametrised Interactions. *J. Chem. Phys* 146: 124131 [PubMed: 28388132]
- Rivas G, Minton AP. 2016 Macromolecular Crowding In Vitro, In Vivo, and In Between. *Trends Biochem. Sci* 41: 970–81 [PubMed: 27669651]
- Robustelli P, Piana S, Shaw DE. 2018 Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A* 115: E4758–E66 [PubMed: 29735687]
- Rohl CA, Strauss CE, Misura KM, Baker D. 2004 Protein Structure Prediction Using Rosetta. *Methods Enzymol.* 383: 66–93 [PubMed: 15063647]
- 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 [PubMed: 27434647]
- Ross CA, Poirier MA. 2004 Protein aggregation and neurodegenerative disease. *Nature medicine* 10: S10
- Roux B, Simonson T. 1999 Implicit Solvent Models. *Biophys. Chem* 78: 1–20 [PubMed: 17030302]
- Roy A, Kucukural A, Zhang Y. 2010 I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc* 5: 725–38 [PubMed: 20360767]
- 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: 102–05 [PubMed: 19262674]
- Schonhoft JD, Kosowicz JG, Stivers JT. 2013 DNA translocation by human uracil DNA glycosylase: role of DNA phosphate charge. *Biochemistry* 52: 2526–35 [PubMed: 23506309]
- 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 [PubMed: 27145874]

- Shin Y, Brangwynne CP. 2017 Liquid phase condensation in cell physiology and disease. *Science* 357: eaaf4382 [PubMed: 28935776]
- 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 [PubMed: 29570991]
- 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. *Proc. Natl. Acad. Sci. U.S.A.*: 201322611
- Stachowiak JC, Schmid EM, Ryan CJ, Ann HS, Sasaki DY, et al. 2012 Membrane bending by protein-protein crowding. *Nature cell biology* 14: 944 [PubMed: 22902598]
- 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 [PubMed: 15565151]
- Szigeti B, Roth YD, Sekar JA, Goldberg AP, Pochiraju SC, Karr JR. 2017 A blueprint for human whole-cell modeling. *Current opinion in systems biology*
- 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 [PubMed: 17181192]
- 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 [PubMed: 26575522]
- 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 [PubMed: 27309278]
- Tan D, Piana S, Dirks RM, Shaw DE. 2018 RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U.S.A* 115: E1346–E55 [PubMed: 29378935]
- 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 [PubMed: 15836408]
- 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 [PubMed: 24832260]
- 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 [PubMed: 22880817]
- 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 [PubMed: 29128709]
- Tomita M 2001 Whole-cell simulation: a grand challenge of the 21st century. *TRENDS in Biotechnology* 19: 205–10 [PubMed: 11356281]
- 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*: 1–18
- Trovato F, Tozzini V. 2014 Diffusion within the Cytoplasm: A Mesoscale Model of Interacting Macromolecules. *Biophys. J* 107: 2579–91 [PubMed: 25468337]
- Vitkup D, Melamud E, Moulton J, Sander C. 2001 Completeness in structural genomics. *Nat. Struct. Biol* 8: 559–66 [PubMed: 11373627]
- 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. *International Journal of Quantum Chemistry* 115: 1084–93
- Vorontsova Maria A, Chan Ho Y, Lubchenko V, Vekilov Peter G. 2015 Lack of Dependence of the Sizes of the Mesoscopic Protein Clusters on Electrostatics. *Biophys. J* 109: 1959–68 [PubMed: 26536272]

- Walcott S, Sun SX. 2010 A mechanical model of actin stress fiber formation and substrate elasticity sensing in adherent cells. *Proc. Natl. Acad. Sci. U.S.A*
- 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 [PubMed: 28306259]
- Wang Y, Li C, Pielak GJ. 2010 Effects of Proteins on Protein Diffusion. *J. Am. Chem. Soc* 132: 9392–97 [PubMed: 20560582]
- Wang Y, Sarkar M, Smith AE, Krois AS, Pielak GJ. 2012 Macromolecular Crowding and Protein Stability. *J. Am. Chem. Soc* 134: 16614–18 [PubMed: 22954326]
- 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 [PubMed: 28102001]
- 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–D19
- Westbrook J, Feng ZK, Chen L, Yang HW, Berman HM. 2003 The Protein Data Bank and Structural Genomics. *Nucleic Acids Res.* 31: 489–91 [PubMed: 12520059]
- 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 [PubMed: 28422503]
- Yildirim A, Feig M. 2017 High-Resolution 3D Models of *Caulobacter crescentus* Chromosome Reveal Genome Structural Variability and Organization. under review
- 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 [PubMed: 25166278]
- Yu I, Feig M, Sugita Y. *Journal of Physics: Conference Series* 2018, 1036: 012009 IOP Publishing.
- 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 [PubMed: 27801646]
- 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 [PubMed: 19965476]
- Zhang WH, Chen JH. 2014 Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. *J. Chem. Theory Comput* 10: 918–23 [PubMed: 26580171]
- Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J. 2006 On the origin and highly likely completeness of single-domain protein structures. *Proc. Natl. Acad. Sci. U.S.A* 103: 2605–10 [PubMed: 16478803]
- Zhou H-X, Pang X. 2018 Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev* 118: 1691–741 [PubMed: 29319301]
- Zhou H-X, Rivas G, Minton AP. 2008 Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annu. Rev. Biophys* 37: 375–97 [PubMed: 18573087]
- Zhuang Y, Charbonneau P. 2016 Recent Advances in the Theory and Simulation of Model Colloidal Microphase Formers. *J. Phys. Chem. B* 120: 7775–82 [PubMed: 27466702]
- Zimmerman SB, Minton AP. 1993 Macromolecular Crowding - Biochemical, Biophysical, and Physiological Consequences. *Annu. Rev. Biophys. Biomol. Struct* 22: 27–65 [PubMed: 7688609]

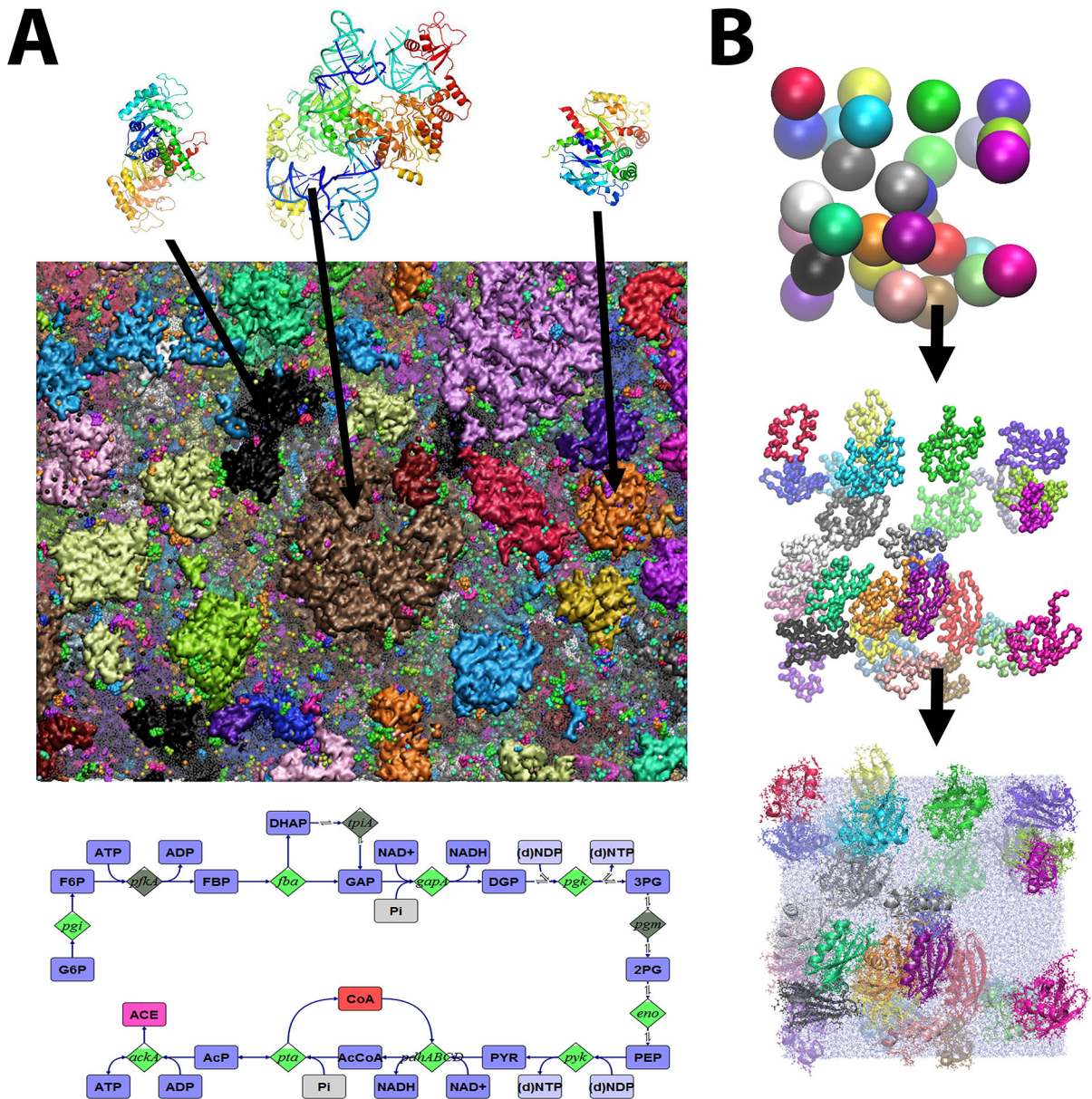


Figure 1: Construction of cellular systems from atomistic structures of individual molecules based on biochemical pathway reconstruction for the cytoplasm of *M. genitalium* (A) (Feig et al 2015). Multiscale assembly generation protocol from spherical models to fully solvated atomistic system (B) (Feig et al 2015).

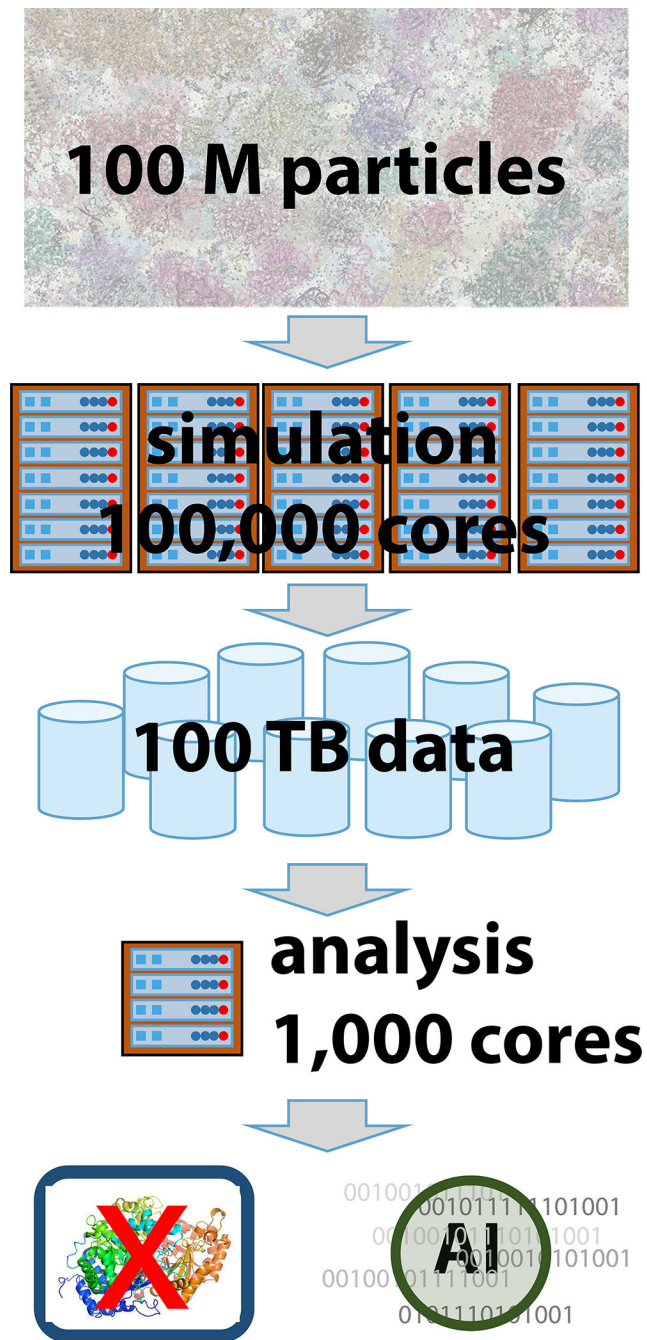


Figure 2:
Flowchart of a typical simulation of a large cellular system that involves high-performance computing, data management, and analysis challenges.

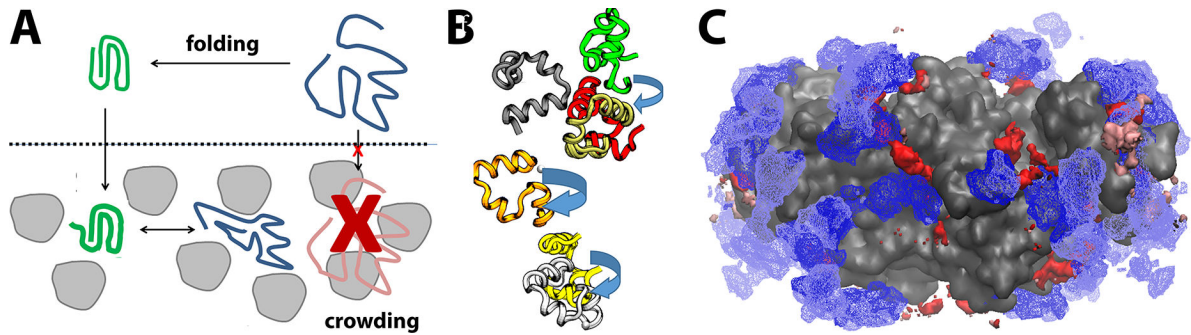


Figure 3: Effects of cellular environments on stability, dynamics, and binding. Protein stability is altered in crowded cellular environments due to volume exclusion and interactions with crowder molecules (grey) (A) (Harada et al 2013); rotational diffusion in crowded environments depends on transient cluster formation in concentrated villin solutions (B) (Nawrocki et al 2017); binding of ATP (red with crowding, blue without crowding) to acetate kinase (grey) varies in the presence of crowders (C) (Yu et al 2016).