Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Integrating cellular and molecular structures and dynamics into whole-cell models

Zaida Luthey-Schulten[1,2,3], Zane R. Thornburg[1] and Benjamin R. Gilbert[1]

Abstract

A complete description of the state of the cell requires knowledge of its size, shape, components, intracellular reactions, and interactions with its environment—all of these as a function of time and cell growth. Adding to this list is the need for theoretical models and simulations that integrate and help to interpret this daunting amount of experimental data. It seems like an overwhelming list of requirements, but progress is being made on many fronts. In this review, we discuss the current challenges and problems in obtaining sufficient information about each aspect of a dynamical whole-cell model (DWCM) for simple and well-studied bacterial systems.

Addresses

[1] Department of Chemistry, University of Illinois at Urbana-Champaign, USA

[2] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, USA

[3] Center for the Physics of the Living Cell, University of Illinois at Urbana-Champaign, USA

Corresponding author: Luthey-Schulten, Zaida (zan@illinois.edu)

URL: http://faculty.scs.illinois.edu/schulten/

Current Opinion in Structural Biology 2022, **75**:102392

This review comes from a themed issue on **Biophysical Methods**

Edited by **David Fushman** and **Dagmar Ringe**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2022.102392

0959-440X/© 2022 Elsevier Ltd. All rights reserved.

Introduction

The goal of whole-cell modeling is to generate predictions about cell behavior from cell states evolving forward in time. From a basic science perspective, such an endeavor involves the integration of a daunting amount of experimental data along with theoretical and algorithmic studies into whole-cell models (WCM) with the promise of revealing simultaneously results and correlations that will deepen our understanding of the principles of life. To describe the state of a cell, requires information about the system's initial states, the evolution rules, and a method to apply the time-evolution. In the context of WCM we refer to these as the composition, interactions, and propagation, respectively. Composition includes knowledge of the cellular architecture and the concentrations of the biomolecules and metabolites in three-dimensional space. Interactions include physical and chemical interactions among the chemical species and with the extracellular environment. Propagation refers to the necessary computational methodologies to time-evolve a given cell state according to the interactions over biologically relevant length and timescales. The existing WCMs vary in the details of composition (entire cell or just a selected subvolume of the cell), physical interactions (atomistic force-fields to crowding by excluded-volume, each resulting in a different representation of chemical species), chemical reactions (ranging from thousands of genome-wide chemical transformations to none), and propagation (molecular dynamics, stochastic diffusion and chemical kinetics, and PDEs/ODEs). With advances in biophysical experimental and computational methods, progress is being made on all fronts. In this article, we discuss the challenges and problems in obtaining sufficient information from the standpoint of our own experience in creating 3D and well-stirred dynamical whole-cell models (DWCMs) for a living minimal cell [1].

Cellular shapes and architecture

The physical and chemical methods that allow us to simulate conformational changes in biomolecules, determine functional forms of large complexes in solution, and locate labeled protein complexes and RNAs in dense media, respectively have been recognized by Chemistry Nobel Prizes for molecular dynamics (2013), cryo-electron microscopy (CEM; 2017), and super-resolved fluorescence microscopy (SRFM; 2014). These Nobel Prizes rest upon previous prizes awarded for methods of obtaining structures of biomolecules from X-ray crystallography and in solution from NMR. Recent advances in cryo-electron tomography (CET) and super-resolution imaging (SRI) have led to an increasing amount of information about the structures of single cells and the dynamics of the molecular actors within them, respectively.

☆ Fully documented templates are available in the elsarticle package on CTAN.

CET of entire small bacterial cells like *Spiroplasma melliferum*, *Mycoplasma pneumoniae*, and the synthetic genetically minimal cell, JCVI-syn3A, can reveal the three-dimensional cell shape and spatial distributions of ribosomes within the cell [2–4,1]. In the case of *M. pneumoniae* [3], the in-cell architecture of a transcribing and translating expressome was resolved at subnanometer resolution through crosslinking mass spectrometry, subtomogram averaging, and integrative modeling. The expressome comprises RNA polymerase (RNAP), the ribosome, and the transcription elongation factors NusG and NusA. With NusA being pinpointed at the interface between a NusG-bound elongating RNAP and the ribosome, it was proposed that the expressome can mediate transcription-translation coupling. The frequency of such coupling is still uncertain and seems to vary among bacterial species [5]. Typically, bacterial nucleoids are not directly observed in CET, but their spatial localization can be inferred from the ribosome distributions. Generating ensembles of DNA configurations consistent with the ribosome distributions and bacterial genome organization principles inferred from sequencing techniques such as chromosome conformation capture (3C) [4,6,7] and CHIP-seq [8,9] methods is necessary for whole-cell simulations [1]. Nucleotide-resolution molecular models of the condensed 4.6 mbp *Escherichia coli* chromosome [8] are presently the highest-resolution cell-scale structural models of condensed bacterial nucleoids. Dynamically altering the local DNA structure during a WCM simulation at any level approaching nucleotide representation has yet to be achieved. When numerous CET are obtained from a population of cells, then correlations between cell size and ribosome distributions can be observed, within which chromosomes at different points in replication can be placed [4].

For CET of larger bacterial and eukaryal cells, focused ion beam (FIB) milling produces lamellae a few hundred nanometers in width from a population [10–12], and in the case of HeLa cells sufficient sampling has produced invaluable information on the organelles neighboring the nucleus and the nuclear pore complexes [13]. The goal to establish visual proteomics from CET has been recently reviewed in the study by Bäuerlein et al. [14] along with the remaining technical challenges to achieve subnanometer resolution. These goals will mostly be achieved with advances in technology, but there is an additional challenge to have structures or accurate predicted structures of all the biomolecules to support integrative modeling. The interpretations of the CET and SRI processing will certainly be aided by machine learning approaches such as AlphaFold2 [15], which can utilize the large and growing data sets of known protein structures. Since its impressive performance in CASP14, it is already being leveraged to predict or complete protein structures used in molecular dynamics simulations of Cas9 proteins [16].

Concentrations of proteins, mRNAs, lipids, and metabolites that provide the initial states of a cell are typically estimated from proteomics, transcriptomics, lipidomics, and metabolomics [17] data. In our own experience, counts of membrane proteins from proteomics and the complexity of the lipid composition from lipidomics can be integrated to model an initial membrane composition [1]. Complexity of the lipidome includes the degree of saturation of the fatty acids, the number of different lipids, and dynamical information about membrane remodeling during cell growth. The composition of cell membranes not only affects the overall architecture of cells, it also affects structure, organization, and function for both RNAs [18] and proteins, where some perform their functions in the membrane periphery (ex. Gram-positive degradosomes [19]).

The remaining proteomics, transcriptomics, and metabolomics data can then provide the initial concentrations of molecules present in the cytoplasm. Unfortunately, while there are numerous -omics studies on bacterial organisms, these studies are typically limited to a few growth conditions and report quantities at the population level rather than the single-cell level. As a further complication, many of these studies do not provide an exhaustive survey of complete proteomes, transcriptomes, lipidomes, or metabolomes and report only subsets of chemical species or incomplete information. Consequently, initial states of the membrane and cytoplasm need to be prepared using datasets consolidated from multiple organisms, which frequently require scaling laws for concentrations based on cell volume and genetic lineage; akin to how homology modeling has allowed us to generalize our knowledge of protein and nucleic acid structures. Presently, organization of cytoplasmic molecules beyond compartmentalization in organelles has been treated using uniform distributions in the absence of 3D localization in the -omics data. As discussed above, 3D localization in proteomics is a future goal of CET. The resolutions of techniques for 3D single-cell transcriptomics in eukaryal cells [20] are being extended to smaller and denser biological systems [21].

Kinetic models of subcellular dynamics

Combinations of crystallographic structures and single-molecule (sm) FRET can provide mechanistic insights and assist in the development and parameterization of kinetic models for fundamental processes such as bacterial DNA replication initiation [1]. Crystal structures of DnaA bound states to double- and single-stranded DNA revealed sequence-dependent binding near the origin [22,23] and smFRET provided on/off rates of DnaA filament formation on single-stranded DNA [24]. With the limited photons emitted from fluorophore labels inside living cells, capturing subcellular dynamics is restricted by a trade-off between spatial/temporal

resolution and phototoxicity. Developments in the super resolution microscopy techniques of PALM, STORM, and STED for imaging of processes within the nucleus to obtain minimal phototoxicity was reviewed [25]. A new fluorescence microscopy technique, MINFLUX (MINimal photon FLUXes) [26,27], promises to provide a particle tracking technology with a resolution 10 times better than other single-molecule techniques such as STORM and PALM. MINFLUX uses a donut-shaped excitation beam to surround a single fluorescing molecule with light. The position of the molecule is then triangulated by moving the beam to targeted locations and minimizing the fluorescence intensity while shrinking search path of the beam. As a result of needing very few photons (800-2500) to localize the fluorophore, its 3D imaging can locate molecules to 2 nm with a tracking resolution of 100 μs while retaining the multicolor and sample preparation advantages of fluorescence microscopy. In the study by Schmidt et al. [27], the spatial-temporal resolution in single fluorophore tracking was reported for the diffusion of single labeled lipids in lipid-bilayer model membranes, and MINFLUX images of nuclear pores in living U-2 OS cells expressing Nup96−mMaple were obtained in the study by Gwosch et al.[26]. With the similar spatial and temporal resolutions of MINFLUX and whole-cell simulations [1], hopefully more examples from live cells will soon be available to make the integration of experimental kinetic data into WCMs more straightforward.

All-atom models have not yet achieved cell-scale simulations. However, models of cytoplasmic spaces, such as the 100 nm cube model of DNA-free *Mycoplasma genitalium* cytoplasm [28,29] with biologically accurate protein, RNA, metabolite, and ion concentrations, are substantial efforts that provide valuable dynamic data including diffusion coefficients, conformational changes, and protein-protein contact dynamics over tens to hundreds of nanoseconds [30].

While the majority of molecular dynamics simulations are carried out over hundreds of nanoseconds, there have been some notable achievements in extensions into the microsecond regime to examine the diffusive interactions of transcription factors along short pieces of DNA using all-atom and coarse-grained simulations [31,32]. Coarse-grained models in which approximately four heavy atoms in the proteins and lipids are represented by a single bead reduces the number of equations of motion and hold the promise of pushing simulations into more biologically relevant time scales. The Martini 2 and 3 force fields [33] have been used to model lipids of an entire mitochondrion [34] and lipids with hundreds of membrane proteins (14 types) in protocells [35]. Even though they are not all-atom, the coarse-grained simulations can still be used as an additional validation of WCMs.

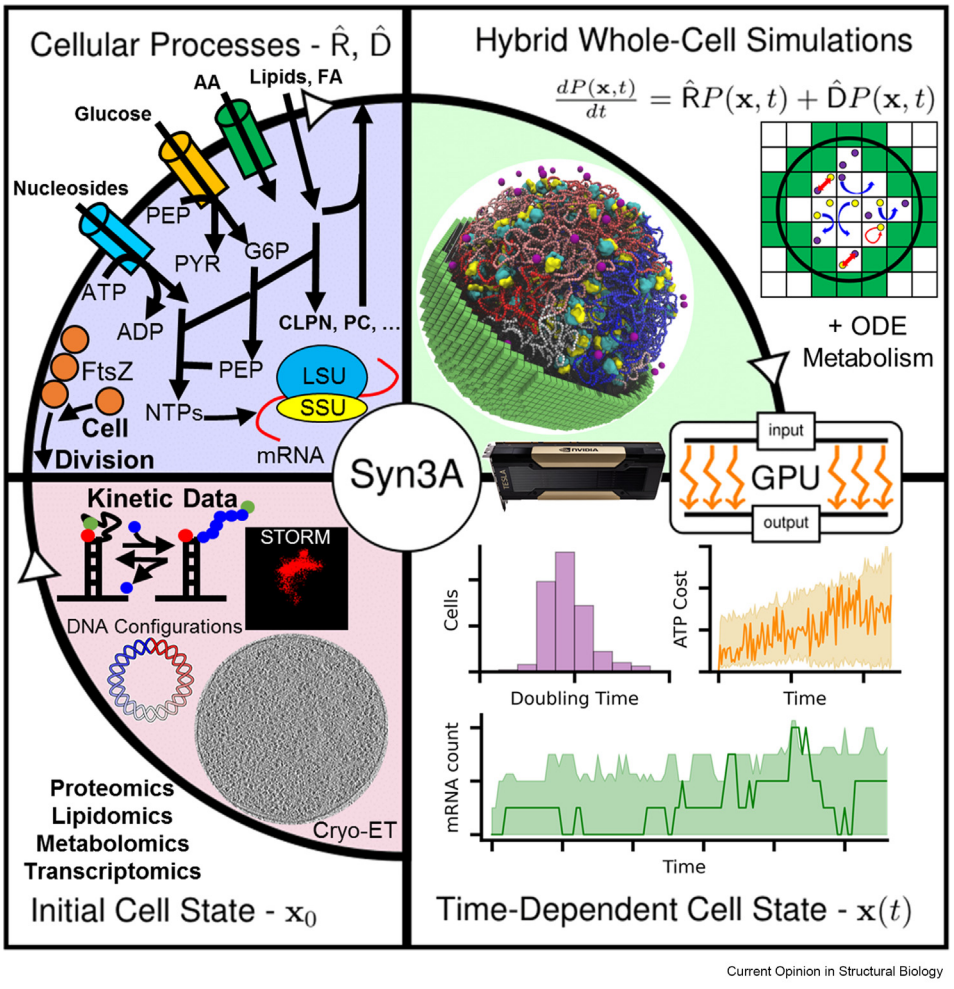
Whole-cell structure-based models that coarse-grain entire macromolecules have been created for cytoplasmic and membrane macromolecules in a *M. genitalium* cell [36] using well-stirred states generated from the WCM in the study by Karr et al. [37]. The placement of the macromolecules accounted for steric interactions and was used to test the viability of predicted cell states. The structural model of *M. genitalium* faced the further challenge of assembling a set of individual structures for the proteome. Structural models of proteins with known function were assembled using a combination of AlphaFold2 [15], the Protein Data Bank, Electron Microscopy Data Bank, UniProt, and homology modeling.

Integrating structure and reaction networks

CET provides a static cell-wide picture of large cellular structures, SRI provides single-molecule dynamics of individual molecular systems, and WCMs serve as a mean to bridge the gap between these techniques until future methods to observe more complete dynamical cell states are developed. Importantly chemical reactions associated with metabolism and information processing networks need to be included in order to capture the subcellular dynamics. The development of WCMs and how they have progressed from genome-scale metabolic models (GSMMs) [38] and the calculation of their steady-state fluxes has been recently reviewed [39,40]. Due to large numbers of genes with unknown function and the complexity in model systems such as *E. coli*, along with the broad range of length-, concentration-, and time-scales that need to be considered, simulating a complete description of the state of a cell has been challenging and requires hybrid computational methods.

Comprehensive WCMs have been developed for *M. genitalium* [37] and *E. coli* [41] where the subsystems were treated in terms of flux balance analysis, ordinary differential equations, and stochastic simulations. We recently simulated a synthetic minimal cell consisting of nearly exclusively the genes and reactions necessary for the cell to proliferate in a laboratory environment and we believe this is an ideal system for a WCM [1,42]. JCVI-syn3A is a living genetically minimal bacterial cell with a chemically synthesized genome consisting of only 493 genes on a 543 kbp chromosome, 452 of these genes are protein-coding [43]. Syn3A's genome and physical size are approximately one-tenth those of the model bacterial organism *E. coli*. The simulations of this organism in the study by Thornburg et al. [1] are based on fully dynamical kinetic models where subsystem networks and chemical species are interconnected continuously over time on a single-cell basis (Figure 1). The 3D spatial models are reconstructed at the single-cell level using ribosomes coordinates from CET and self- and ribosome-avoiding DNA configurations. Common challenges among these whole-cell modeling efforts are

Figure 1



Diverse experimental techniques and data characterizing the cellular state and processes of Syn3A are integrated into a DWCM for simulations of time-dependent cell states. Ribosomes were placed in the 3D spatial cell model using PDB 5NJT and the instancing visualization method in the study by Sener et al. [54] (Adapted from the study by Thornburg et al. [1] and includes unpublished data.).

characterizing cell architecture, establishing the reaction networks, and the limited availability of kinetic parameters and -omics data to use as initial conditions. No one bacterium has a complete set of structures, parameters, -omics data, and annotated genes with known functions. The development of any WCM relies upon synthesizing information from other organisms.

The *M. genitalium* [37] and *E. coli* [41] models tackled the challenge of curating large datasets consisting of thousands of kinetic parameters and demonstrated methods of evaluating parameters from other species and the variations that have been measured in even a single species. In our own WCM of Syn3A, initial estimates for every kinetic parameter and metabolite concentration in the metabolic reactions were determined [1] using a Bayesian estimator [44] with prior information consisting of a survey of concentrations from *E. coli* [17] and *M. pneumoniae* [45] and parameters measured in related organisms through decades of biochemical, smFRET, and spectroscopic studies reported in the literature and kinetic databases like [46], BRENDA [47], and equilibrium constants reported in NIST's TECRdb [48] and Equilibrator [49]. Each of the WCMs discussed has taken a different approach to utilizing the vast amounts of available kinetic, biochemical, and structural data. While there is not yet a clear answer to the best method of curation, the systematic organization and reporting of these parameters and structures in federated and connected databases will continue to be critical for assembling WCMs that provide unified descriptions of cell behaviors.

Simulation methodologies

Simulation methodologies for cellular processes are reviewed in the study by Smith et al. [50] so we provide only a brief summary here. In the case of our DWCM of a living minimal cell, we employ hybrid multiscale methods to propagate the cell state [1]. At the greatest level of complexity in time- and length-scales, intracellular kinetics in a three-dimensional and spatially heterogeneous environment are modeled by the reaction-diffusion master equation (RDME) approach, where particles stochastically diffuse between and react within subvolumes. For the homogeneous case in which the particles are considered to be well-stirred due to fast diffusion, the chemical master equation (CME) is used to describe stochastic reactions. In both cases, the particles are assumed to be in rather low concentrations. And in the limits of high concentrations and fast diffusion typically associated with metabolites in metabolic reactions, the reactions are described with ordinary differential equations (ODE). Given the current state of the art in whole-cell modeling, we feel our approach to WCMs remains faithful to the known interactions within the genome-wide biochemical reaction networks, while resolving the spatial heterogeneity within the cell (positions of ribosomes, circular DNA, and membrane) and enabling simulations over portions of the entire cell cycle. However, this choice in methodology assumes simple physical interactions between the particles and crowding, and an important future step includes the backtransformation for the state of a cell in such a particle picture to one that includes either all atom or coarse-grained structures of the cytoplasmic and membrane species, see Figure 1.

Concluding remarks

The DWCM in the study by Thornburg et al. [1] presented a three-dimensional model of a living minimal cell Syn3A, and used (RDME/CME/ODE) simulations to study cell behavior over 20 min before any substantial cell growth or DNA replication had occurred. From these relatively short runs, probabilistic factors were estimated that allowed predictions about responses to the environment, complexifying pathways and time correlations between the cellular processes of DNA replication and cell growth over the full cell cycle using well-stirred approximations. Determining the balance of energy generation and costs for each process was an important first step to measure certain metabolites and cellular intermediates as a function of time. As more experiments and theoretical models on DNA replication and segregation become available, more complex growth behavior in three-dimensions can be addressed over the entire cell cycle. Because of the large variation in time-scales and concentrations, the three-dimensional DWCM for Syn3A required combined hybrid stochastic-deterministic simulations using the GPU-based simulation software Lattice Microbes [51,1]. GPU computer clusters represent yet another necessary revolution to bring structural biology and its wealth of information to understanding how even a minimal cell functions.

We identify two complementary directions to further advance whole-cell modeling: aspects of unknown compositions and interactions in cells must be addressed by experimental methods and computational modelers must develop novel techniques to accommodate more complex compositions and interactions. For example, MEDYAN was recently extended to include mechanochemical interactions between the cytoskeleton and membranes [52]. No organism has a gene set where every function is known. While the number of genomes being sequenced has increased, the substantial fraction of genes with products of unknown functions (for example, roughly 40% in *E. coli* [43]) leaves gaps in models and limits their predictive power by leaving out possible chemical and physical interactions. Among the genes in Syn3A, there are seven genes known to be necessary to maintain regular spherical morphology upon cell division, but their exact functions are unknown [53]. Determining structures and performing functional kinetic studies on such genes is going to be critical for the construction of complete WCMs. In general, with the increasing abundance of dynamic and structural data of large-scale reaction networks, a more systematic approach is needed to determine or estimate the organism-specific kinetic parameters.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgments

The authors are all partially supported by NSF MCB 1818344, NSF MCB 1840320, and the Center for the Physics of Living Cells (NSF PHY 1430124).

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Thornburg Zane R, Bianchi DM, Brier TA, Gilbert BR, Earnest TM, Melo MC, Safronova N, Sáenz JP, Cook AT, Wise KS, Hutchison CA, Smith HO, Glass JI, Luthey-Schulten Z: **Fundamental behaviors emerge from simulations of a living minimal cell**. *Cell* 2022, **185**:345–360.e28, https://doi.org/10.1016/j.cell.2021.12.025.
Structural and dynamical data of many experimental techniques including CET and smFRET are integrated into a dynamic whole-cell model. These measurements were essential in constructing cell architecture and coupled metabolic-gene expression reaction networks to simulate both the spatially heterogeneous cell environment and well-stirred kinetics.

2. Ortiz JO, Förster F, Kürner J, Linaroudis AA, Baumeister W: **Mapping 70s ribosomes in intact cells by cryoelectron tomography and pattern recognition**. *J Struct Biol* 2006, **156**: 334–341, https://doi.org/10.1016/j.jsb.2006.04.014.

3. • O'Reilly Francis J, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, Singh N, Hagen WJH, Cramer P, Stülke J, Mahamid J, Rappsilber J: **In-cell architecture of an actively transcribing-translating expressome**. *Science* 2020, **369**: 554–557, https://doi.org/10.1126/science.abb3758.

Subnanometer resolution of the expressome architecture (RNA polymerase, ribosome, NusG, and NusA) is determined by combining cryo-electron tomography, whole-cell cross-linked mass spectrometry, and integrative modeling.

4. • Gilbert Benjamin R, Thornburg ZR, Lam V, Rashid F-ZM, Glass JI, Villa E, Dame RT, Luthey-Schulten Z: **Generating chromosome geometries in a minimal cell from cryo-electron tomograms and chromosome conformation capture maps**. *Front Mol Biosci* 2021, **8**:644133, https://doi.org/10.3389/fmolb.2021.644133.

Simultaneously integrated cryo-electron tomography and chromosome conformation capture data to construct cell-scale structures of the circular chromosome consistent with the ribosome distribution within a genetically minimal bacterial cell, JCVI-syn3A.

5. • Johnson Grace E, Lalanne J-B, Peters ML, Li G-W: **Functionally uncoupled transcription–translation in bacillus subtilis**. *Nature* 2020, **585**:124–128, https://doi.org/10.1038/s41586-020-2638-5.

Challenged the notion of transcription and translation coupling in bacteria which is frequently considered a defining feature of bacterial gene expression. This claim is backed up with kinetic data and time-dependent mRNA/protein levels in *B. subtilis*.

6. • Lioy Virginia S, Junier I, Boccard F: **Multiscale dynamic structuring of bacterial chromosomes**. *Annu Rev Microbiol* 2021, **75**:541–561, https://doi.org/10.1146/annurev-micro-033021-113232.

An outstanding review on bacterial chromosome organization and a discussion of recent advances that allow researchers to investigate the mechanisms governing the organization.

7. Lioy VS, Cournac A, Marbouty M, Duigou S, Mozziconacci J, Espéli O, Boccard F, Koszul R: **Multiscale structuring of the *e. coli* chromosome by nucleoid-associated and condensin proteins**. *Cell* 2018, **172**:771–783.e18, https://doi.org/10.1016/j.cell.2017.12.027.

8. Hacker WC, Li S, Elcock AH: **Features of genomic organization in a nucleotide-resolution molecular model of the escherichia coli chromosome**. *Nucleic Acids Res* 2017, **45**:7541–7554, https://doi.org/10.1093/nar/gkx541. URL https://doi.org/10.1093/nar/gkx541.

9. Guo MS, Kawamura R, Littlehale ML, Marko JF, Laub MT: **High-resolution, genome-wide mapping of positive supercoiling in chromosomes**. *Elife* 2021, **10**, e67236, https://doi.org/10.7554/elife.67236.

10. • Turk Martin, Baumeister W: **The promise and the challenges of cryo-electron tomography**. *FEBS (Fed Eur Biochem Soc) Lett* 2020, **594**:3243–3261, https://doi.org/10.1002/1873-3468.13948.

This presents one of the most comprehensive and up-to-date descriptions of cryo-electron tomography and its associated techniques.

11. Klumpe S, Fung HK, Goetz SK, Zagoriy I, Hampoelz B, Zhang X, Erdmann PS, Baumbach J, Müller CW, Beck M, Plitzko JM, Mahamid J: **A modular platform for automated cryo-FIB workflows**. *Elife* 2021, **10**, e70506, https://doi.org/10.7554/elife.70506.

12. • Khanna Kanika, Lopez-Garrido J, Sugie J, Pogliano K, Villa E: **Asymmetric localization of the cell division machinery during bacillus subtilis sporulation**. *Elife* 2021, **10**:62204, https://doi.org/10.7554/elife.62204.

Utilized cryo-electron tomography to characterize the division architectures and localization of division protein structures in both vegetative growth and sporulation in *B. subtilis*.

13. Mahamid J, Pfeffer S, Schaffer M, Villa E, Danev R, Cuellar LK, Förster F, Hyman AA, Plitzko JM, Baumeister W: **Visualizing the molecular sociology at the HeLa cell nuclear periphery**. *Science* 2016, **351**:969–972, https://doi.org/10.1126/science.aad8857.

14. •• Bäuerlein Felix JB, Baumeister W: **Towards visual proteomics at high resolution**. *J Mol Biol* 2021, **433**:167187, https://doi.org/10.1016/j.jmb.2021.167187.

Describes the current limitations and state of the art of cryo-electron tomography as well as future measures that will help to overcome the limitations to enable greater resolution and interpretability of holistic structures and cell architectures.

15. •• John Jumper, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589, https://doi.org/10.1038/s41586-021-03819-2.

The clear winner of the CASP14 structure prediction contest. AlphaFold could be used to determine structures for proteins of unknown function present in any genome and even provide insight into possible functional classifications.

16. Belato HB, D'Ordine AM, Nierzwicki L, Arantes PR, Jogl G, Palermo G, Lisi GP: **Structural and dynamic insights into the HNH nuclease of divergent cas9 species**. *J Struct Biol* 2022, **214**:107814, https://doi.org/10.1016/j.jsb.2021.107814.

17. Park JO, Rubin SA, Xu Y-F, Amador-Noguez D, Fan J, Shlomi T, Rabinowitz JD: **Metabolite concentrations, fluxes and free energies imply efficient enzyme usage**. *Nat Chem Biol* 2016, **12**:482–489, https://doi.org/10.1038/nchembio.2077.

18. Czerniak T, Saenz JP: **Lipid membranes modulate the activity of RNA through sequence-dependent interactions**. *Proc Natl Acad Sci Unit States Am* 2022, **119**:2119235119, https://doi.org/10.1073/pnas.2119235119.

19. Cho KH: **The structure and function of the gram-positive bacterial RNA degradosome**. *Front Microbiol* 2017, **8**, 00154, https://doi.org/10.3389/fmicb.2017.00154.

20. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X: **Spatially resolved, highly multiplexed RNA profiling in single cells**. *Science* 2015, **348**, aaa6090, https://doi.org/10.1126/science.aaa6090.

21. Wang G, Moffitt JR, Zhuang X: **Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy**. *Sci Rep* 2018, **8**:4847, https://doi.org/10.1038/s41598-018-22297-7.

22. Erzberger JP, Mott ML, Berger JM: **Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling**. *Nat Struct Mol Biol* 2006, **13**:676–683, https://doi.org/10.1038/nsmb1115.

23. Duderstadt KE, Chuang K, Berger JM: **DNA stretching by bacterial initiators promotes replication origin opening**. *Nature* 2011, **478**:209–213, https://doi.org/10.1038/nature10455.

24. Cheng H-M, Gröger P, Hartmann A, Schlierf M: **Bacterial initiators form dynamic filaments on single-stranded DNA monomer by monomer**. *Nucleic Acids Res* 2014, **43**:396–405, https://doi.org/10.1093/nar/gku1284.

25. Chu L-A, Chang S-W, Tang W-C, Tseng Y-T, Chen P, Chen B-C: **5d superresolution imaging for a live cell nucleus**. *Curr Opin Genet Dev* 2021, **67**:77–83, https://doi.org/10.1016/j.gde.2020.11.005.

26. • Gwosch Klaus C, Pape JK, Balzarotti F, Hoess P, Ellenberg J, Ries J, Hell SW: **MINFLUX nanoscopy delivers 3d multicolor nanometer resolution in cells**. *Nat Methods* 2020, **17**:217–224, https://doi.org/10.1038/s41592-019-0688-0.

Showed that MINFLUX is capable of localizing more than one kind of fluorophore through two color channels. This extension is applied to visualize two-color 3D images of labeled nuclear pore complexes in fixed and live U-2 OS cells at 1–3 nm resolution.

27. • Schmidt Roman, Weihs T, Wurm CA, Jansen I, Rehman J, Sahl SJ, Hell SW: **MINFLUX nanometer-scale 3d imaging and microsecond-range tracking on a common fluorescence microscope**. *Nat Commun* 2021, **12**:1478, https://doi.org/10.1038/s41467-021-21652-z.

Demonstrated that MINFLUX can achieve 1–3 nm resolution in all 3 spatial dimensions with 800–2500 photons. For dynamical data, it was deomstrated that precision was <20 nm with time localizations of ~ 100s.

28. Yu I, Mori T, Ando T, Harada R, Jung J, Sugita Y, Feig M: **Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm**. *Elife* 2016, **5**, e19274, https://doi.org/10.7554/elife.19274.

29. Heo L, Sugita Y, Feig M: **Protein assembly and crowding simulations**. *Curr Opin Struct Biol* 2022, **73**:102340, https://doi.org/10.1016/j.sbi.2022.102340.

30. Rickard MM, Zhang Y, Gruebele M, Pogorelov TV: **In-cell protein–protein contacts: transient interactions in the crowd**. *J Phys Chem Lett* 2019, **10**:5667–5673, https://doi.org/10.1021/acs.jpclett.9b01556.

31. Dai L, Yu J: **Inchworm stepping of myc-max heterodimer protein diffusion along DNA**. *Biochem Biophys Res Commun* 2020, **533**:97–103, https://doi.org/10.1016/j.bbrc.2020.08.004.

32. Dai L, Xu Y, Du Z, dong Su X, Yu J: **Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA**. *Proc Natl Acad Sci Unit States Am* 2021, **118**:2102621118, https://doi.org/10.1073/pnas.2102621118.

33. Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, Patmanidis I, Abdizadeh H, Bruininks BMH, Wassenaar TA, Kroon PC, Melcr J, Nieto V, Corradi V, Khan HM, Domański J, Javanainen M, Martinez-Seara H, Reuter N, Best RB, Vattulainen I, Monticelli L, Periole X, Tieleman DP, de Vries AH, Marrink SJ: **Martini 3: a general purpose force field for coarse-grained molecular dynamics**. *Nat Methods* 2021, **18**:382–388, https://doi.org/10.1038/s41592-021-01098-3.

34. • Pezeshkian Weria, König M, Wassenaar TA, Marrink SJ: **Back-mapping triangulated surfaces to coarse-grained membrane models**. *Nat Commun* 2020, **11**:2296, https://doi.org/10.1038/s41467-020-16094-y.
Introduced a method and freely available software to make complete large-scale lipid membranes in a coarse-grained atomistic representation (MARTINI force field) from triangulated surfaces.

35. Vermaas JV, Mayne CG, Shinn E, Tajkhorshid E: **Assembly and analysis of cell-scale membrane envelopes**. *J Chem Inf Model* 2021, **62**:602–617, https://doi.org/10.1021/acs.jcim.1c01050.

36. Maritan M, Autin L, Karr J, Covert MW, Olson AJ, Goodsell DS: **Building structural models of a whole mycoplasma cell**. *J Mol Biol* 2022, **434**:167351, https://doi.org/10.1016/j.jmb.2021.167351.

37. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW: **A whole-cell computational model predicts phenotype from genotype**. *Cell* 2012, **150**:389–401, https://doi.org/10.1016/j.cell.2012.05.044.

38. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *escherichia coli* W3110**. *Appl Environ Microbiol* 1994, **60**:3724–3731, https://doi.org/10.1128/aem.60.10.3724-3731.1994.

39. Goldberg AP, Szigeti B, Chew YH, Sekar JA, Roth YD, Karr JR: **Emerging whole-cell modeling principles and methods**. *Curr Opin Biotechnol* 2018, **51**:97–102, https://doi.org/10.1016/j.copbio.2017.12.013.

40. Marucci L, Barberis M, Karr J, Ray O, Race PR, de Souza Andrade M, Grierson C, Hoffmann SA, Landon S, Rech E, *et al.*: **Computer-aided whole-cell design: taking a holistic approach by integrating synthetic with systems biology**. *Front Bioeng Biotechnol* 2020, **8**:942, https://doi.org/10.3389/fbioe.2020.00942.

41. • Macklin Derek N, Ahn-Horst TA, Choi H, Ruggero NA, Carrera J, Mason JC, Sun G, Agmon E, DeFelice MM, Maayan I, Lane K, Spangler RK, Gillies TE, Paull ML, Akhter S, Bray SR, Weaver DS, Keseler IM, Karp PD, Morrison JH, Covert MW: **Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation**. *Science* 2020, **369**:eaav3751, https://doi.org/10.1126/science.aav3751.
Examines the heterogeneity of biological data and highlights the difficulties of assembling a compatable set of kinetic parameters.

42. Luthey-Schulten Z: **Integrating experiments, theory and simulations into whole-cell models**. *Nat Methods* 2021, **18**:446–447, https://doi.org/10.1038/s41592-021-01150-2.

43. Breuer M, Earnest TM, Merryman C, Wise KS, Sun L, Lynott MR, Hutchison CA, Smith HO, Lapek JD, Gonzalez DJ, de Crécy-Lagard V, Haas D, Hanson AD, Labhsetwar P, Glass JI, Luthey-Schulten Z: **Essential metabolism for a minimal cell**. *Elife* 2019, **8**, e36842, https://doi.org/10.7554/elife.36842.

44. Lubitz T, Liebermeister W: **Parameter balancing: consistent parameter sets for kinetic metabolic models**. *Bioinformatics* 2019, **35**:3857–3858, https://doi.org/10.1093/bioinformatics/btz129.

45. Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen W-H, Wodke JAH, Güell M, Martínez S, Bourgeois R, Kühner S, Raineri E, Letunic I, Kalinina OV, Rode M, Herrmann R, Gutiérrez-Gallego R, Russell RB, Gavin A-C, Bork P, Serrano L: **Impact of genome reduction on bacterial metabolism and its regulation**. *Science* 2009, **326**:1263–1268, https://doi.org/10.1126/science.1177263.

46. Bremer H, Dennis PP: **Modulation of chemical composition and other parameters of the cell at different exponential growth rates**. *EcoSal Plus* 2008, **3**:1–48, https://doi.org/10.1128/ecosal.5.2.3.

47. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D: **BRENDA, the ELIXIR core data resource in 2021: new developments and updates**. *Nucleic Acids Res* 2020, **49**:D498–D508, https://doi.org/10.1093/nar/gkaa1025.

48. Goldberg RN, Tewari YB, Bhat TN: **Thermodynamics of enzyme-catalyzed reactions–a database for quantitative biochemistry**. *Bioinformatics* 2004, **20**:2874–2877, https://doi.org/10.1093/bioinformatics/bth314.

49. Flamholz A, Noor E, Bar-Even A, Milo R: **eQuilibrator–the biochemical thermodynamics calculator**. *Nucleic Acids Res* 2011, **40**:D770–D775, https://doi.org/10.1093/nar/gkr874.

50. Smith S, Grima R: **Spatial stochastic intracellular kinetics: a review of modelling approaches**. *Bull Math Biol* 2018, **81**:2960–3009, https://doi.org/10.1007/s11538-018-0443-1.

51. Hallock MJ, Stone JE, Roberts E, Fry C, Luthey-Schulten Z: **Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-GPU workstations**. *Parallel Comput* 2014, **40**:86–99.

52. Ni H, Papoian GA: **Membrane-MEDYAN: simulating deformable vesicles containing complex cytoskeletal networks**. *J Phys Chem B* 2021, **125**:10710–10719, https://doi.org/10.1021/acs.jpcb.1c02336.

53. Pelletier JF, Sun L, Wise KS, Assad-Garcia N, Karas BJ, Deerinck TJ, Ellisman MH, Mershin A, Gershenfeld N, Chuang R-Y, Glass JI, Strychalski EA: **Genetic requirements for cell division in a genomically minimal cell**. *Cell* 2021, **184**:2430–2440.e16, https://doi.org/10.1016/j.cell.2021.03.008.

54. Sener M, Levy S, Stone JE, Christensen A, Isralewitz B, Patterson R, Borkiewicz K, Carpenter J, Hunter CN, Luthey-Schulten Z, Cox D: **Multiscale modeling and cinematic visualization of photosynthetic energy conversion processes from electronic to cell scales**. *Parallel Comput* 2021, **102**:102698, https://doi.org/10.1016/j.parco.2020.102698.