Published in final edited form as:

Curr Opin Struct Biol. 2019 April ; 55: 59–65. doi:10.1016/j.sbi.2019.03.012.

Computational approaches to macromolecular interactions in the cell

Ilya A. Vakser, Eric J. Deeds

Computational Biology Program and Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas 66045, USA

Abstract

Structural modeling of a cell is an evolving strategic direction in computational structural biology. It takes advantage of new powerful modeling techniques, deeper understanding of fundamental principles of molecular structure and assembly, and rapid growth of the amount of structural data generated by experimental techniques. Key modeling approaches to principal types of macromolecular assemblies in a cell already exist. The main challenge, along with the further development of these modeling approaches, is putting them together in a consistent, unified whole cell model. This opinion piece addresses the fundamental aspects of modeling macromolecular assemblies in a cell, and the state-of-the-art in modeling of the principal types of such assemblies.

Keywords

molecular modeling; structure prediction; molecular recognition

Introduction

A rapidly emerging strategic direction in computational structural biology is modeling of a cell through development of powerful computational approaches, based on fundamental principles of molecular structure and interaction [1]. There have been important advances in modeling of individual biomolecules and their interactions [2,3]. This modeling faces numerous challenges that need to be addressed, from the intricate details of ligand-receptor interaction to the large-scale conformational dynamics of macromolecular systems. Still, taking a broad (and inherently approximate), big-picture view of the field, one cannot escape realization that when it comes to the structural modeling of individual macromolecules and their assemblies, we already can do quite a lot. To a significant extent due to the rapid growth of the amount of structural data generated by experimental techniques, and because of the major progress in often knowledge-driven modeling methods, we can now predict at

*Corresponding author: Ilya Vakser, vakser@ku.edu, and Eric Deeds, deeds@ku.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure

The authors declare no conflict of interest.

meaningful structural resolution/approximation the equilibrium states of most proteins and their complexes. From this standpoint, the focus is inevitably shifting toward larger systems. Surprisingly or not, a really large system, which is still amenable to the rapidly evolving structural modeling methods, is a cell. Whole cell modeling is important for many reasons, such as prediction of phenotype from genotype, filling the gaps in our knowledge of cellular processes, and development of our ability to modulate them. However, the main, overarching reason is the intellectual challenge of modeling life *in silico*.

When it comes to such an ambitious goal, one of course should think of starting with a simplest case, such as prokaryotic *Mycoplasma genitalium* (525 genes). Interestingly, key modeling approaches to the principal types of macromolecular assemblies in such a cell (Figure 1) already exist. Emerging experimental techniques provide data and validation for the modeling [4]. Thus, the main challenge, along with further development of the modeling approaches, in our view, is putting these approaches together, in a consistent, unified whole cell simulation. In this opinion piece, we briefly address the fundamental aspects of modeling macromolecular assemblies in a cell, and the state-of-the-art in modeling of the principal types of such assemblies.

Need for structural modeling

Most non-structural, cell-scale models work by breaking cellular processes into putatively modular components (e.g. “metabolism” and “cell signaling”) and modeling them separately (e.g. by flux balance analysis for metabolism or systems of ordinary differential equations for cell signaling). These models are then integrated into a model of the entire cell by creating links between these separate sub-models [5–7]. While such whole-cell modeling had some success at predicting results of experimental perturbations to the cell (e.g. mutations) [5], there are limitations to this approach. For one, crosstalk between various pathways is rampant in the cell, and this modular approach may miss crucial phenomena that arise due to crosstalk [8]. There could even be crosstalk between putatively separate biological processes (e.g. metabolism and cell signaling) mediated by transient molecular interactions that arise from the fact that these processes are occurring within the same crowded cell compartments. Such interactions may be difficult to detect experimentally and would never be discovered if these processes are described as modular *a priori*. Secondly, the formation of macromolecular complexes is crucial to most cellular functions, from transcription and translation to metabolism and cell signaling. The complexes that form can be structurally and compositionally heterogeneous, and this heterogeneity may be key to the underlying molecular mechanisms [9,10]. Thus, while modular whole-cell models can be informative, it is clear that complementary, structure-based approaches to modeling cellular dynamics are needed to link molecular mechanisms and cellular function.

Principles of large-scale structural modeling

Low-resolution (or coarse-grained) modeling is essential for large, heterogeneous, and dynamic biomolecular systems like cells [11]. Conceptually it is based on the fundamental paradigm of studying physical systems - approximation. Any structural biomolecular modeling that does not explicitly use the formalism of quantum physics (this formalism

cannot be applied in practice to most systems in biomolecular modeling due to their large size and complexity) is an approximation. Thus, the only question is: how far this approximation may and should go. The validity of the low-resolution approach to macromolecular recognition is well-established. It is based on the fundamental characteristics of the intermolecular energy landscape, such as the binding funnel in protein-protein [12] and protein-nucleic acid [13] interactions, and the distribution of energy basins correlated with experimentally-observed binding modes [14]. The large-scale features of the intermolecular energy landscape are phenomenologically reflected in the macrostructural recognition characteristics of macromolecules - structural complementarity of proteins deprived of atom-size features and the anisotropy of the protein shape [11]. The practical importance of the large recognition factors in modeling of biomolecular complexes is that they allow one to ignore local conformational changes and structural inaccuracy (e.g. in modeled macromolecules). The conformational change in the protein backbone upon binding is often insignificant. However, the ability of coarse-grained, multiscale approaches to model large structural flexibility is important for understanding the dynamics of biomolecular systems [15]. The coarse-graining allows exploration of large macromolecular systems behavior at long time scales, which is key to the modeling of a cell.

Most known proteins do not have experimentally-determined structures. Even less structurally characterized are protein interactions; proteins often participate in multiple interactions, and the structure of a protein complex is generally more difficult to determine experimentally than the structure of an individual protein. Thus, structural modeling of protein interactions primarily has to rely on models of the proteins, which typically are less accurate than the experimentally determined structures. In comprehensive validation of docking on a large set of protein models, a significant part of the predictions was of good quality [16]. The results of this and other studies [17,18], showed that existing docking methodologies can be successfully applied to protein models with a broad range of structural accuracy.

The amount of experimentally-determined information on macromolecules and their interactions [19] is rapidly growing, especially due to the new advances in techniques such as cryo-EM (Figure 2), boosting development of data-driven modeling of macromolecular assemblies [20]. Emerging knowledge-based techniques range from template-based modeling of protein-protein (PPI) [21–25] and protein-RNA interactions [26], to predictions with statistical potentials [27]. Experimental data on macromolecular structures and interactions is used as constraints in the “integrative” modeling [28].

A structure-based simulation of the cell has to originate from some starting point where the subjects of the simulation are packed in a biologically adequate configuration. Given that simulations may require multiple starting points, automated packing procedures are essential. The cellPACK suite of programs allows fast generation of such packing, making it possible to construct ensembles of structural configurations [1]. An alternative cytoplasmic assembly algorithm of Feig and co-workers [29] approximates macromolecules by spheres and gradually replaces the spheres with greater structural details, followed by insertion of solvent, metabolites and ions.

PPI in cytoplasm

Comparative modeling can predict the structure of 50–70% of proteins in a typical proteome [30]. Such modeling has long been known to be more reliable and efficient than *ab initio* predictions [3]. The pool of templates for comparative modeling is constantly increasing, following rapid growth of the PDB. In addition, there have been groundbreaking developments in the *ab initio* structure prediction due to major increase in accuracy of residue contact predictions [3]. Protein-protein docking methodology, until recently, had been based primarily on energy-based considerations, although knowledge-based approaches are gaining popularity [31,32]. An essential part of the protein docking approaches is the search procedure. To make sampling of the docking degrees of freedom feasible, docking algorithms use such techniques as geometric hashing, Monte Carlo sampling and, most frequently, correlation by fast Fourier transform (FFT) [32].

An essential aspect of cell modeling is accounting for a crowded environment in the cytoplasm (Figure 3). Macromolecules occupy ~1/3 of the cytoplasmic volume. A significant part of the environment is occupied by the DNA in prokaryotic cells, and by the nucleus and other organelles in eukaryotes. Because of the complexity of the solvent environment [33]. Biomolecular interactions in the crowded environment modulate macromolecular structure and dynamics [34]. Crowding affects protein folding [35], and transient protein interactions slow down protein diffusion [36]. Hydrodynamic interactions were shown to be important for the subcellular dynamics of macromolecules [37].

PPI in membrane

Membrane proteins participate in transmembrane signaling and transport and perform enzymatic activities. They constitute about one third of all proteins in a typical organism. Similar to soluble proteins, the function of most membrane proteins is based on interaction with other proteins. However, the big difference with the soluble proteome is that due to the limitations of experimental techniques, structural characterization of membrane proteins and their complexes is far less extensive than that of the soluble ones. High-resolution structures of the membrane proteins have been determined by X-ray crystallography and, more recently, by single particle Cryo-EM. Despite rapid growth of the PDB, only 2 – 3% of the entries are transmembrane. The relative scarcity of the experimentally-determined structures of the membrane proteins leads to a lack of structural templates for their modeling, putting more emphasis on *de novo* structure prediction [38,39]. The use of metagenomic data [40] and improved contact prediction further extend the scope of structural modeling [41]. Docking of the membrane proteins has been handled primarily by modifications of the generic docking protocols to take into account the spatial constraints imposed by the membrane [42,43].

Protein-RNA interactions

Experimental techniques have been developed to characterize protein-RNA interactions [44]. However, only a fraction of identified protein-RNA interactions has been structurally characterized, further increasing the importance of the structural modeling [45]. Similar to

the protein interactome, most RNA structures in the RNA interactome have to be models. Thus, advancements in RNA modeling [46,47] are essential. The principles of protein-RNA interaction are based on structural and physicochemical complementarity [48,49], and are similar to those of protein-protein interactions [32]. This includes the similarity of the fundamental paradigms of structure prediction: free and comparative modeling of the complexes. Protein-RNA docking employs a variety of scoring functions and FFT-based search algorithms [50,51], as well as combined computational/experimental approaches [52]. A systematic study established structure alignment-based strategy for comparative modeling of protein-RNA complexes [26].

Protein-DNA interactions

Modeling of DNA packing in prokaryotic cells, and the more complicated chromatin organization in eukaryotes, is a challenge due to the great length of the polymer chain [53]. Experimental data provides spatial constraints for the modeling of DNA in prokaryotic nucleoids [54] and eukaryotic chromosomes [55–57]. Despite the rapidly growing number of experimentally determined protein-DNA structures, computational methods are essential to fill the gap between a set of static structures from the PDB and the enormous multitude of protein-DNA recognition and binding phenomena. The increasing amount of experimentally-determined data allowed investigation of structural and physicochemical factors in protein-DNA interaction through dedicated data resources [58], and conformational changes in protein-DNA interaction upon binding [59]. Protein-DNA recognition factors were also elucidated by comparing results from all-atom molecular dynamics and simulations based on simplified protein models [60]. Characteristics of the energy landscape, including the binding funnel, provide guidance for the development of computational approaches [13]. Coarse-graining is important for large-scale structural modeling of DNA and its interactions [61].

Kinetics of macromolecular assembly

Structural models of cellular processes must account for the dynamics of the formation of macromolecular complexes. Structural models of assembly largely focus on viral capsids, and found non-trivial relationships between the concentration of capsid monomers and the types of structures formed [62,63]. Non-structural models of complex formation also suggest that both kinetic trapping and compositional heterogeneity may play a role in shaping the evolution and function of macromolecular complexes [9,64]. It is clear that highly detailed molecular simulation approaches, such as molecular dynamics, cannot reach the time scales relevant for studying complex formation and dynamics in a crowded cellular environment [34]. While coarse-grained approaches like Brownian Dynamics were applied to study the thermodynamics of dimerization in the presence of generic crowders [65], a computational framework that describes the structural details of assembly at a cellular scale remains to be developed.

Experiment

Experimental techniques, ranging from atomic-resolution X-ray crystallography and NMR spectroscopy, to electron and light microscopy, can determine structural data from atomic resolution of biomacromolecules to organelles and larger cellular components. NMR can now be applied within living cells, which allowed probing biomolecular structure *in vivo*, including the structure of individual proteins as well as some molecular details of interactions [66,67]. Spatial distributions and dynamics of macromolecules can be determined by various fluorescence microscopy methods, including measurements of molecular diffusion at the single-molecule level. Application of such techniques revealed intriguing relationships between cellular metabolic state and diffusion [68]. Additional whole-cell data on protein expression and concentration levels can be obtained via techniques ranging from RNA-seq methods to quantitative mass spectrometry and other proteomic analyses. While these experimental developments are promising, new approaches aimed at the biological mesoscale will need to be developed in order to compliment modeling in exploring the dynamics of cellular processes in structural detail.

Perspective

The continued progress in experimental techniques will likely keep generating large amounts of structural data, to feed into the macromolecular modeling approaches and their applications to the whole cell simulation, as well as to provide benchmarks for such simulations. This progress, accompanied by advancement of computational methods, along with the rapid development of computing platforms, firmly puts the structure-based whole-cell modeling on the agenda of the research community. Progress in the structural cell modeling should inevitably involve a coordinated effort of a broad spectrum of computational and experimental research programs, at various levels, by bringing together researchers with complementary expertise, through scientific community-wide initiatives, and international collaborations, supported by large-scale funding opportunities.

Acknowledgments

This study was supported by grants R01GM074255 from the NIH and DBI1565107 from the NSF to IAV, and MCB1412262 from the NSF to EJD. The authors thank Petras Kundrotas, Nathan Jenkins, and Devlina Chakravarty for help in preparation of the manuscript.

References and recommended reading

1. Im W, Liang J, Olson A, Zhou HX, Vajda S, Vakser IA: Challenges in structural approaches to cell modeling. *J. Mol. Biol* 2016, 428:2943–2964. [PubMed: 27255863] •• The review presents a cross section of approaches illustrating the developing shift from the structural modeling of individual molecules to that of a cell. Studies in several related areas are covered: biological networks; automated packing of cellular components; modeling of protein complexes; prediction of transient protein interactions; thermodynamic and kinetic effects of crowding; membrane modeling; and modeling of chromosomes.
2. Lensink MF, Velankar S, Wodak SJ: Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* 2017, 85:359–377. [PubMed: 27865038] • Report on evaluation of methods for predicting the atomic resolution structures of protein complexes offered as targets to the community-wide initiative on the Critical Assessment of Predicted Interactions (CAPRI). The

evaluation suggests that the progress stems from better integration of different modeling tools with docking procedures, as well as the use of more sophisticated evolutionary information for scoring.

3. Moulton J, Fidelis K, Kryshchuk A, Schwede T, Tramontano A: Critical assessment of methods of protein structure prediction (CASP)—Round XII. *Proteins* 2018, 86:7–15. [PubMed: 29082672] • This article reports the outcome of the 12th round of Critical Assessment of Structure Prediction (CASP12) - a community experiment to determine the state of the art in modeling protein structure. This article describes the experiment and summarizes the results.
4. Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A: Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* 2012, 10:e1001244. [PubMed: 22272186]
5. 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. [PubMed: 22817898]
6. Szigeti B, Roth YD, Sekar JAP, Goldberg AP, Pochiraju SC, Karr JR: A blueprint for human whole-cell modeling. *Curr. Opin. Syst. Biol* 2018, 7:8–15. [PubMed: 29806041]
7. Blinov ML, Schaff JC, Vasilescu D, Moraru II, Bloom JE, Loew LM: Compartmental and spatial rule-based modeling with Virtual Cell. *Biophys. J* 2017, 113:1365–1372. [PubMed: 28978431]
8. Rowland MA, Greenbaum JM, Deeds EJ: Crosstalk and the evolvability of intracellular communication. *Nat. Commun* 2017, 8:16009. [PubMed: 28691706]
9. Suderman R, Deeds EJ: Machines vs. ensembles: Effective MAPK signaling through heterogeneous sets of protein complexes. *PLoS Comp. Biol* 2013, 9:e1003278.
10. Sato PM, Yoganathan K, Jung JH, Peisajovich SG: The robustness of a signaling complex to domain rearrangements facilitates network evolution. *PLoS Biol* 2014, 12:e1002012. [PubMed: 25490747]
11. Vakser IA: Low-resolution structural modeling of protein interactome. *Curr. Opin. Struct. Biol* 2013, 23:198–205. [PubMed: 23294579]
12. Hunjan J, Tovchigrechko A, Gao Y, Vakser IA: The size of the intermolecular energy funnel in protein-protein interactions. *Proteins* 2008, 72:344–352. [PubMed: 18214966]
13. Cencini M, Pigolotti S: Energetic funnel facilitates facilitated diffusion. *Nucl. Acids Res* 2018, 46:558–567. [PubMed: 29216364]
14. O'Toole N, Vakser IA: Large-scale characteristics of the energy landscape in protein-protein interactions. *Proteins* 2008, 71:144–152. [PubMed: 17932937]
15. Voth GA: A multiscale description of biomolecular active matter: The chemistry underlying many life processes. *Acc. Chem. Res* 2017, 50:594–598. [PubMed: 28945406]
16. Anishchenko I, Kundrotas PJ, Vakser IA: Modeling complexes of modeled proteins. *Proteins* 2017, 85:470–478. [PubMed: 27701777] •• The paper presents a comprehensive validation of template-based and free docking on a large set of protein complexes, where each protein model has six levels of structural accuracy. Many predictions fell into acceptable quality category, according to the CAPRI criteria, even for highly inaccurate proteins. The results show that the docking methodologies can be successfully applied to protein models with a broad range of structural accuracy.
17. Bohnuud T, Luo L, Wodak SJ, Bonvin AMJJ, Weng Z, Vajda S, Schueler-Furman O, Kozakov D: A benchmark testing ground for integrating homology modeling and protein docking. *Proteins* 2017, 85:10–16. [PubMed: 27172383]
18. Yu J, Guerois R: PPI4DOCK: Large scale assessment of the use of homology models in free docking over more than 1000 realistic targets. *Bioinformatics* 2016, 32:3760–3767. [PubMed: 27551106]
19. Kundrotas PJ, Anishchenko I, Dauzhenka T, Kotthoff I, Mnevets D, Copeland MM, Vakser IA: DOCKGROUND: A comprehensive data resource for modeling of protein complexes. *Protein Sci* 2018, 27:172–181. [PubMed: 28891124]
20. Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ: The challenge of modeling protein assemblies: The CASP12-CAPRI experiment. *Proteins* 2018, 86:257–273. [PubMed: 29127686]

21. Kundrotas PJ, Anishchenko I, Dauzhenka T, Vakser IA: Modeling CAPRI targets 110–120 by template-based and free docking using contact potential and combined scoring function. *Proteins* 2018, 86:302–310. [PubMed: 28905425]
22. Xue LC, Rodrigues JPGLM, Dobbs D, Honavar V, Bonvin AMJJ: Template-based protein-protein docking exploiting pairwise interfacial residue restraints. *Brief. Bioinform* 2017, 18:458–466. [PubMed: 27013645]
23. Estrin M, Wolfson HJ: SnapDock-template-based docking by Geometric Hashing. *Bioinformatics* 2017, 33:i30–i36. [PubMed: 28881968]
24. Kuzu G, Keskin O, Nussinov R, Gursoy A: PRISM-EM: Template interface-based modelling of multi-protein complexes guided by cryo-electron microscopy density maps. *Acta Cryst* 2016, D72:1137–1148.
25. Hwang H, Petrey D, Honig B: A hybrid method for protein–protein interface prediction. *Protein Sci* 2016, 25:159–165. [PubMed: 26178156]
26. Zheng J, Kundrotas PJ, Vakser IA, Liu S: Template-based modeling of protein-RNA interactions. *PLoS Comp. Biol* 2016, 12:e1005120.
27. Anishchenko I, Kundrotas PJ, Vakser IA: Contact potential for structure prediction of proteins and protein complexes from Potts model. *Biophys. J* 2018, 115:809–821. [PubMed: 30122295]
28. Joseph AP, Polles G, Alber F, Topf M: Integrative modelling of cellular assemblies. *Curr. Opin. Struct. Biol* 2017, 46:102–109. [PubMed: 28735107]
29. Feig M, Harada R, Mori T, Yue I, Takahashi K, Sugita Y: Complete atomistic model of a bacterial cytoplasm for integrating physics, biochemistry, and systems biology. *J. Mol. Graph. Mod* 2015, 58:1–9.
30. Perdigo N, Heinrich J, Stolte C, Sabir KS, Buckley MJ, Tabor B, Signal B, Gloss BS, Hammang CJ, Rost B, et al.: Unexpected features of the dark proteome. *Proc. Natl. Acad. Sci. USA* 2015, 112:15898–15903. [PubMed: 26578815]
31. Porter KA, Desta I, Kozakov D, Vajda S: What method to use for protein-protein docking? *Curr. Opin. Struct. Biol* 2019, this issue.
32. Vakser IA: Protein-protein docking: From interaction to interactome. *Biophys. J* 2014, 107:1785–1793. [PubMed: 25418159]
33. Feig M, Yu I, Wang P, Nawrocki G, Sugita Y: Crowding in cellular environments at an atomistic level from computer simulations. *J. Phys. Chem* 2017, 121:8009–8025. • The paper reviews recent simulations that address the effects of crowding in biological environments on biomolecular structure, dynamics, and function. The simulations are discussed in the context of relevant experiments.
34. 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. [PubMed: 27801646] •• The paper describes molecular dynamics simulations of a comprehensive atomistic model cytoplasm. The results show that protein-protein interactions may destabilize native protein structures, whereas metabolite interactions may induce more compact states due to electrostatic screening. Protein-protein interactions also resulted in significant variations in reduced macromolecular diffusion under crowded conditions.
35. Qin S, Zhou HX: Protein folding, binding, and droplet formation in cell-like conditions. *Curr. Opin. Struct. Biol* 2017, 43:28–37. [PubMed: 27771543]
36. Nawrocki G, Wang P, Yu I, Sugita Y, Feig M: Slow-down in diffusion in crowded protein solutions correlates with transient cluster formation. *J. Phys. Chem. B* 2017, 121:11072–11084. [PubMed: 29151345]
37. Skolnick J: Perspective: On the importance of hydrodynamic interactions in the subcellular dynamics of macromolecules. *J. Chem. Phys* 2016, 145:100901. [PubMed: 27634243]
38. Stansfeld PJ: Computational studies of membrane proteins: From sequence to structure to simulation. *Curr. Opin. Struct. Biol* 2017, 45:133–141. [PubMed: 28511148]
39. Ovchinnikov S, Kinch L, Park H, Liao Y, Pei J, Kim DE, Kamisetty H, Grishin NV, Baker D: Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife* 2015, 4:e09248. [PubMed: 26335199]

40. Ovchinnikov S, Park H, Varghese N, Huang PS, Pavlopoulos GA, Kim DE, Kamisetty H, Kyripides NC, Baker D: Protein structure determination using metagenome sequence data. *Science* 2017, 355:294–298. [PubMed: 28104891] •• The paper shows that Rosetta structure prediction guided by residue-residue contacts inferred from evolutionary information can accurately model proteins that belong to large families and that metagenome sequence data more than triple the number of protein families with sufficient sequences for accurate modeling.
41. Michel M, Hurtado DM, Uziela K, Elofsson A: Large-scale structure prediction by improved contact predictions and model quality assessment. *Bioinformatics* 2017, 33:i23–i29. [PubMed: 28881974]
42. Koehler-Leman J, Mueller BK, Gray JJ: Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics* 2017, 33:754–756. [PubMed: 28011777]
43. Hurwitz N, Schneidman-Duhovny D, Wolfson HJ: Memdock: An alpha-helical membrane protein docking algorithm. *Bioinformatics* 2016, 32:2444–2450. [PubMed: 27153621]
44. Ramanathan M, Majzoub K, Rao DS, Neela PH, Zarnegar BJ, Mondal S, Roth JG, Gai H, Kovalski JR, Siprashvili Z, et al.: RNA-protein interaction detection in living cells. *Nat. Methods* 2018, 15:207–212. [PubMed: 29400715]
45. Nithin C, Ghosh P, Bujnicki JM: Bioinformatics tools and benchmarks for computational docking and 3D structure prediction of RNA-protein complexes. *Genes* 2018, 9:432.
46. Miao Z, Westhof E: RNA structure: Advances and assessment of 3D structure prediction. *Annu. Rev. Biophys* 2017, 46:483–503. [PubMed: 28375730]
47. Tan D, Piana S, Dirks RM, Shaw DE: RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. USA* 2018, 115:E1346–E1355. [PubMed: 29378935]
48. Iwakiri J, Tateishi H, Chakraborty A, Patil P, Kenmochi N: Dissecting the protein-RNA interface: The role of protein surface shapes and RNA secondary structures in protein-RNA recognition. *Nucl. Acids Res* 2012, 40:3299–3306. [PubMed: 22199255]
49. Ghaemi Z, Guzman I, Gnutt D, Luthey-Schulten Z, Gruebele M: Role of electrostatics in protein–RNA Binding: The global vs the local energy landscape. *J. Phys. Chem* 2017, 121:8437–8446.
50. Arnaudova YA, Abagyan R, Totrov M: Protein-RNA docking using ICM. *J. Chem. Theory Comput* 2018, 14:4971–4984. [PubMed: 30016588]
51. Huang Y, Li H, Xiao Y: 3dRPC: A web server for 3D RNA-protein structure prediction. *Bioinformatics* 2018, 34:1238–1240. [PubMed: 29186336]
52. Dorn G, Leitner A, Boudet J, Campagne S, von Schroetter C, Moursy A, Aebersold R, Allain FHT: Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat. Methods* 2017, 14:487–490. [PubMed: 28346450]
53. Tiana G, Giorgetti L: Integrating experiment, theory and simulation to determine the structure and dynamics of mammalian chromosomes. *Curr. Opin. Struct. Biol* 2018, 49:11–17. [PubMed: 29128709]
54. Goodsell DS, Autin L, Olson AJ: Lattice models of bacterial nucleoids. *J. Phys. Chem. B* 2018, 122:5441–5447. [PubMed: 29338247]
55. Di Pierro M, Cheng RR, Aiden EL, Wolynes PG, Onuchic JN: De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc. Natl. Acad. Sci. USA* 2017, 114:12126–12131. [PubMed: 29087948]
56. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, et al.: 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 2017, 544:59–64. [PubMed: 28289288]
57. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA: Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. USA* 2018, 115:E6697–E6706. [PubMed: 29967174]
58. Sagendorf JM, Berman HM, Rohs R: DNAproDB: An interactive tool for structural analysis of DNA-protein complexes. *Nucl. Acids Res* 2017, 45:W89–W97. [PubMed: 28431131]
59. Poddar S, Chakravarty D, Chakrabarti P: Structural changes in DNA-binding proteins on complexation. *Nucl. Acids Res* 2018, 46:3298–3308. [PubMed: 29534202]
60. Etheve L, Martin J, Lavery R: Decomposing protein-DNA binding and recognition using simplified protein models. *Nucl. Acids Res* 2017, 45:10270–10283. [PubMed: 28973439]

61. Shimizu M, Takada S: Reconstruction of atomistic structures from coarse-grained models for protein–DNA complexes. *J. Chem. Theory Comput* 2018, 14:1682–1694. [PubMed: 29397721]
62. Hagan MF, Elrad OM: Understanding the concentration dependence of viral capsid assembly kinetics - the origin of the lag time and identifying the critical nucleus size. *Biophys. J* 2011, 98:1065–1074.
63. Nguyen HD, Reddy VS, Brooks CL: Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Lett* 2007, 7:338–344. [PubMed: 17297998]
64. Deeds EJ, Bachman JA, Fontana W: Optimizing ring assembly reveals the strength of weak interactions. *Proc. Natl. Acad. Sci. USA* 2012, 109:2348–2353. [PubMed: 22308356]
65. Ando T, Yu I, Feig M, Sugita Y: Thermodynamics of macromolecular association in heterogeneous crowding environments: Theoretical and simulation studies with a simplified model. *J. Phys. Chem. B* 2016, 120:11856–11865. [PubMed: 27797534]
66. Luchinat E, Banci L: In-cell NMR: A topical review. *IUCrJ* 2017, 4:108–118.
67. Breindel L, Burz DS, Shekhtman A: Interaction proteomics by using in-cell NMR spectroscopy. *J. Proteomics* 2018, doi:10.1016/j.jprot.2018.02.006.
68. Parry BR, Surovtsev IV, Cabeen MT, O'Hern CS, Dufresne ER, Jacobs-Wagner C: The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell* 2014, 156:183–194. [PubMed: 24361104]

Highlights

- Structural modeling of a cell is an evolving strategic direction
- Experimental techniques provide data and validation for the modeling
- Key modeling approaches to macromolecular assemblies in a cell already exist
- The main challenge is putting these approaches together in a unified cell modeling

Figure 1.
Major classes of targets for structural cell modeling.

Figure 2. Structural data on PPI.

Protein and protein complexes are non-redundant at 95% sequence identity, and interfaces - at TM-score 0.5. The inset shows the number of interfaces split according to different methods for their determination - X-ray crystallography, electron microscopy, and other (NMR, electron crystallography, and such).

Figure 3.
PPI in cytosol.