Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Emerging whole-cell modeling principles and methods

Arthur P Goldberg[1,2,3], Balázs Szigeti[1,2,3], Yin Hoon Chew[1,2,3], John AP Sekar[1,2,3], Yosef D Roth[1,2] and Jonathan R Karr[1,2]

Whole-cell computational models aim to predict cellular phenotypes from genotype by representing the entire genome, the structure and concentration of each molecular species, each molecular interaction, and the extracellular environment. Whole-cell models have great potential to transform bioscience, bioengineering, and medicine. However, numerous challenges remain to achieve whole-cell models. Nevertheless, researchers are beginning to leverage recent progress in measurement technology, bioinformatics, data sharing, rule-based modeling, and multi-algorithmic simulation to build the first whole-cell models. We anticipate that ongoing efforts to develop scalable whole-cell modeling tools will enable dramatically more comprehensive and more accurate models, including models of human cells.

Addresses
[1] Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Corresponding author: Karr, Jonathan R (karr@mssm.edu)
[3] These authors contributed equally to the work.

Current Opinion in Biotechnology 2018, **51**:97–102

This review comes from a themed issue on **Systems biology**

Edited by **Nathan Price** and **Eran Segal**

https://doi.org/10.1016/j.copbio.2017.12.013

Introduction

Whole-cell (WC) computational models aim to predict cellular phenotypes from genotype and the environment by representing the function of each gene, gene product, and metabolite [1••]. WC models could unify our understanding of cell biology and enable researchers to perform *in silico* experiments with complete control, scope, and resolution [2•,3••]. WC models could also help bioengineers rationally design microorganisms that can produce useful chemicals and act as biosensors, and help physicians design personalized therapies tailored to each patient's genome.

Despite their potential, there is little consensus on how WC models should represent cells, what phenotypes WC models should predict, or how to achieve WC models. Nevertheless, we and others are beginning to leverage advances in measurement technology, bioinformatics, rule-based modeling, and multi-algorithmic simulation to develop WC models [4•,5•,6,7••,8•,9•]. However, substantial work remains to achieve WC models [10••,11••].

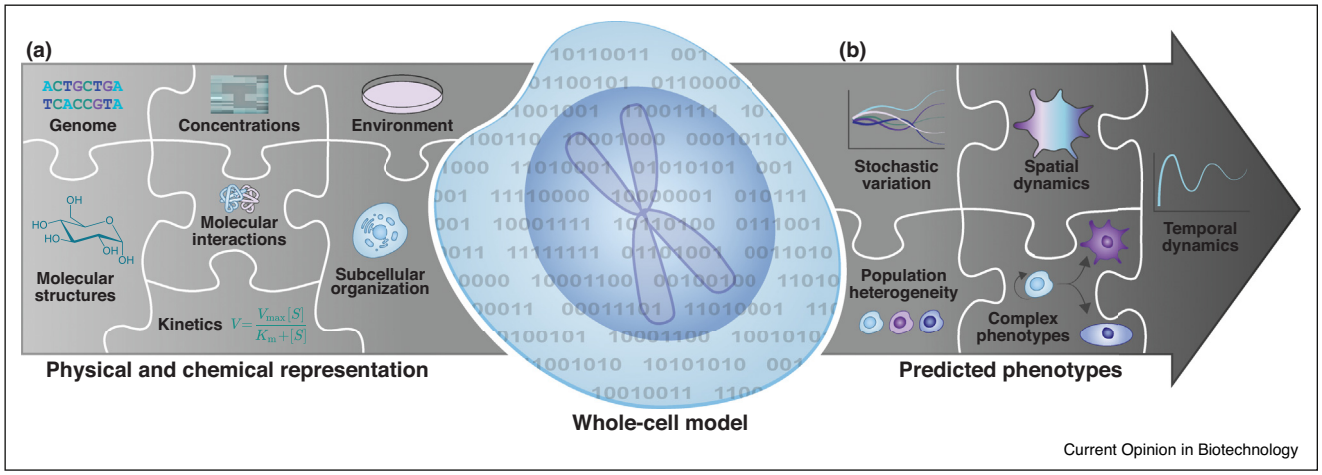
To build consensus on WC modeling, we propose a set of key physical and chemical mechanisms that WC models should aim to represent, and a set of key phenotypes that WC models should aim to predict. We also summarize the experimental and computational progress that is making WC modeling feasible, and outline several technological advances that would help accelerate WC modeling.

Note, our proposals focus on defining WC models that are needed for research studies and applications such as bioengineering and personalized medicine which depend on understanding the molecular details of the majority of intracellular processes. However, research that depends on fewer intracellular processes could be served by smaller, focused models.

Physics and chemistry that WC models should aim to represent

We propose that WC models aim to represent all of the chemical reactions in a cell and all of the physical processes that influence their rates (Figure 1a). This requires representing (a) the **sequence** of each chromosome, RNA, and protein; the location of each chromosomal feature, including each gene, operon, promoter, and terminator; and the location of each site on each RNA and protein; (b) the **structure** of each molecule, including atom-level information about small molecules, the domains and sites of macromolecules, and the subunit composition of complexes; (c) the **subcellular organization** of cells into organelles and microdomains; (d) the participants and effect of each **molecular interaction**, including the molecules that are consumed, produced, and transported, the molecular sites that are modified, and the bonds that are broken and formed, (e) the **kinetic parameters** of each interaction; (f) the **concentration** of each species in each organelle and microdomain; and (g) the concentration of each species in the **extracellular environment**. In addition, to enable WC models to be rigorously tested, each WC model should represent a single, well-defined experimental system. To minimize the complexity of WC models, we recommend modeling small, fast-growing, non-adherent, autonomous, self-renewing cells growing on defined, rich,

Figure 1



The physical and chemical mechanisms that WC models should aim to represent **(a)** and the phenotypes that WC models should aim to predict **(b)**.

homogeneous media. Together, this would enable WC models to describe how cellular behavior emerges from the combined function of each gene and genetic variant, and capture how cells respond to changes in their internal and external environments.

Phenotypes that WC models should aim to predict

We also propose that WC models aim to predict the behavioral trajectories of single cells over their life cycles, with each simulation representing a different cell within a heterogeneous clonal population (Figure 1b). This should include behaviors within individual cells such as the **stochastic dynamics** of each molecular interaction; the **temporal dynamics** of the concentration of each species; the **spatial dynamics** of the concentration of each species in each organelle and microdomain; and **complex phenotypes** such as cell shape, growth rate, motility, and fate, as well as the **variation** in the behavior of single cells within clonal populations. Together, this would enable WC models to capture how stochastic and single-cell variation can generate phenotypic diversity; how a cell responds to external cues such as nutrients, growth factors and drugs; and how a cell coordinates critical events such as the G1/S transition. This would also enable WC models to generate predictions that could be embedded into higher-order multiscale models. For example, WC models could predict the timing and speed of chemotaxis, which could help multiscale models predict tumor metastasis.

Available resources

Achieving WC models will require extensive data to constrain every parameter. Fortunately, measurement technology is rapidly advancing. Here, we review the latest methods for generating data for WC models, and highlight repositories and other resources that contain useful data for WC modeling.

Measurement methods

Advances in single-cell and genomic measurement are rapidly generating data that could be used for WC modeling [12–14] (Table S1). For example, Meth-Seq can assess epigenetic modifications [15], Hi-C can determine chromosome structures [16], ChIP-seq can determine protein-DNA interactions [17], fluorescence microscopy can determine protein localizations, mass-spectrometry can quantitate metabolite and protein concentrations, FISH [18] and scRNA-seq [19] can quantitate the dynamics and single-cell variation of RNA abundances, and fluorescence microscopy and mass cytometry [20] can quantitate the dynamics and single-cell variation of protein abundances. In particular, WC models can be constrained by combining high-dimensional measurement methods with multiple genetic and environmental perturbations, frequent temporal observations, and cutting-edge distributed parameter estimation methods. However, substantial work remains to develop methods that can measure non-model organisms including small, slow-growing, and unculturable cells.

Data repositories

Researchers are also rapidly aggregating much of the data needed for WC modeling into public repositories (Table S2). For example, UniProt contains a multitude of information about proteins [21]; BioCyc contain extensive information about interactions [22]; ECMDB [23],