Cell Modeling Across Disciplines: A Review of Representative Models and Approaches

12212859 Sijie Li

Abstract

Numerous cell models have been developed to simulate cellular behavior, but researchers from different fields often refer to "cell models" with varying emphases. There is a lack of a unified discussion that bridges these disciplinary gaps. This review summarizes several widely used types of cell models along with their most commonly used construction methods, aiming to facilitate interdisciplinary communication and guide more effective model selection for specific tasks.

Introduction

Cells are the fundamental units of life, and understanding their mechanisms and predicting their responses are essential for advancing biology, medicine, and bioengineering. To this end, scientists have been devoted to developing cell models. By quantitatively constructing these models, we can formalize biological processes, thereby uncovering potential general laws that govern cellular behavior—analogous to Newton's laws in physics. Moreover, such formalization allows biological dynamics to be encoded into binary silicon systems, enabling virtual perturbations that accelerate experimental cycles and scientific discovery.

Multiple types of models have been proposed, each focusing on different cellular processes such as gene regulation, metabolism, and mechanical dynamics. However, these models are often developed and discussed in isolation, leading to inconsistencies in how the term "cell model" is interpreted across disciplines. To bridge this gap, this review aims to summarize four highly researched and relatively mature modeling approaches, offering a holistic and integrative perspective on cell modeling.

Gene Regulatory Network (GRN) Model

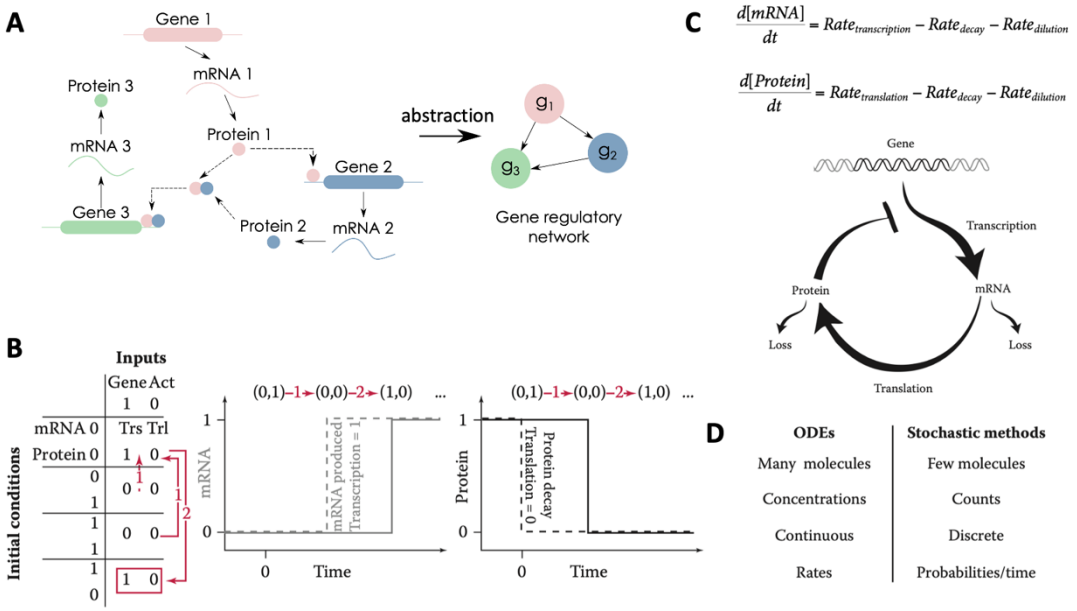


Figure 1. Concepts and approaches in modeling gene regulatory networks (GRNs)

Gene regulatory networks (GRNs) are an abstraction of the relationships and interactions among genes (Figure 1A), and GRN models aim to quantitatively formalize these. Initial models primarily focus on simple, classical regulatory motifs such as the autoregulation of *lac* operon (Covert, 2014). However, with advancements in artificial intelligence (AI) and high throughput omics technologies, GRN models have scaled up, enabling the description of complex, previously unknown networks. In parallel, the development of single-cell technologies provides high-resolution, cell-specific, and spatially resolved data for GRN modeling. Representative methods such as LINGER (Yuan & Duren, 2024), scTransNet (Kommu et al., 2024), and STGRNS (Xu et al., 2023) exemplify the integration of deep learning with single-cell data for large-scale GRN inference.

Typically, input data for GRN models include gene expression profiles, transcription factor binding information, and chromatin accessibility data, which are used to either fit the parameters of mechanistic models or to infer network structures in data-driven approaches. The detailed requirements of data vary depending on the modeling method employed. The final output includes the regulatory architecture or the predicted network dynamics, which can offer new biological insight or support engineering and biomedical applications. For instance, in synthetic biology,

rational gene circuit design relies heavily on accurate network predictions. Also, in biomedical research, GRN models are widely used for disease mechanism analysis, biomarker identification, and drug target discovery.

Boolean Method

Each gene is considered to be either in an active (ON, 1) or inactive (OFF, 0) state. The state of each gene "after some time" is determined by a set of logical functions (e.g., AND, OR, NOT) based on the known regulation (Covert, 2014) (Figure 1B). This model assumes gene expression is binary and evolves in discrete time steps. It is conceptually simple and does not require high-resolution or large-scale data, making it suitable for systems where only qualitative dynamics is needed.

ODE Method

The rate of gene expression is modeled using ordinary differential equations (ODEs), generally as a function of regulator concentration and mRNA degradation rates (Covert, 2014) (Figure1 C). Gene expression in this framework is treated as deterministic, and nonlinear regulatory influences are often captured using Hill functions. The model parameters are estimated or fitted from experimental data. However, when scaling up to larger networks, solving the ODEs can become computationally expensive and often requires simplifying assumptions, such as separating fast and slow timescales.

Stochastic Method

Gene expression is treated as a stochastic process, where each biochemical reaction occurs with a certain probability (Figure1 D). These models typically assume the system is memoryless and follows Markovian dynamics. Stochastic models are particularly valuable in capturing intrinsic noise in gene expression, especially in systems where molecule numbers are low and random fluctuations play a significant role. However, due to their inherently probabilistic nature, the results are often less intuitive to interpret. Additionally, the simulation of individual reaction events step by step can be computationally intensive.

Statistical Inference Method

This approach primarily relies on data-driven methods (Huynh-Thu & Sanguinetti, 2019) but can incorporate prior biological knowledge to enhance accuracy. Specifically, it aims to infer regulatory relationships directly from observational data—such as gene expression matrices—without explicitly modeling the underlying biochemical mechanisms. However, a major limitation of this method lies in its reliance on a strong—and potentially misleading—assumption: that a statistical correlation between two genes implies a causal regulatory relationship. Moreover, these models often operate as black boxes, providing limited mechanistic interpretability.

Metabolic Network Model

Metabolic flux represents the rate at which metabolic reactions occur in vivo. Metabolic network models aim to quantitatively describe these fluxes, elucidating how metabolites flow through metabolic pathways under specific conditions. Current models primarily focus on well-characterized pathways, utilizing extensive prior knowledge of known reactions and constraints. However, generalizable methods for de novo discovery of novel metabolic pathways remain limited (Li, 2022). Advances in spectrometry-based detection, as well as deeper understanding of biochemical thermodynamics and enzyme kinetics, have improved model precision. Recently, novel approaches have also begun integrating biomolecular simulations (Noor, 2025) and multi-dimensional omics data, offering more mechanistic insights.

In general, the reconstruction of a metabolic network model requires a list of all known metabolic reactions with corresponding stoichiometric coefficients, the genes encoding the enzymes that catalyze these reactions, flux bounds derived from biological knowledge or experimental measurements, and an objective function defined according to biological principles or engineering goals. By incorporating additional layers of information—such as gene–protein–reaction associations, quantitative metabolomics data from mass spectrometry (MS), and physicochemical constraints—the Flux Balance Analysis (FBA) (Orth et al., 2010) framework can be extended. Isotope tracing data are used in recent metabolic flux analysis (MFA) methods to quantitatively infer flux distributions by fitting simulated isotope labeling patterns to experimental data (Figure 2A). The model's output is a predicted flux distribution across all reactions under defined environmental and physicochemical constraints. By modifying the input—such as

adjusting nutrient availability or redefining the objective function—researchers can investigate how metabolic fluxes adapt to various scenarios. This modeling flexibility has led to broad applications in biotechnology and metabolic engineering, particularly in identifying rate-limiting steps and optimizing reaction conditions to maximize the yield of target products.

Flux Balance Analysis (FBA)

Flux Balance Analysis (FBA) (Orth et al., 2010) assumes that the metabolic system is under steady state, meaning that the net accumulation of internal metabolites is zero. Mathematically, this is expressed as $S \cdot v = 0$, where S is the stoichiometric matrix derived from prior biological knowledge, and v is the flux vector representing the rate of metabolic reactions. Linear programming, supported by a range of computational packages, is employed to optimize a predefined objective function , subject to physiological flux bounds. The solution yields an optimal flux distribution under the specified constraints. This steady-state assumption is reasonable because, in many biological systems, metabolite concentrations are tightly regulated and remain relatively constant over time. However, to describe dynamic behaviors, more complex dynamic modeling approaches are required.

Cell Mechanics Model

Mechanics models are primarily based on Newtonian mechanics and aim to predict the cellular deformation, morphological changes, and mechanical responses under internal or external forces. Early models relied heavily on simplifying assumptions and were relatively coarse. For instance, cells were often treated as homogeneous elastic droplets. Recently, the integration of AI for data structure extraction and insights from statistical physics has enabled the construction of models that account for finer-scale subcellular mechanical behaviors (Zhu et al., 2024) and capture emergent phenomena such as self-organization and spatial patterning. Meanwhile, the development of high-resolution experimental techniques—including optical tweezers (Vos et al., 2024), atomic force microscopy (AFM), and microfluidic platforms—has enabled precise measurement of cellular mechanical properties and forces (Molnar & Manneville, 2025).

The modeling data contains geometric and mechanical properties of cells, subcellular structures, and applied forces. The integration of omics and imaging data further enhances these models by linking mechanical responses with biochemical activities. Depending on the modeling objective, appropriate methods are selected, with model parameters determined by input data. The resulting outputs can predict how cells deform or respond to specific mechanical stimuli, which is particularly valuable in applications such as disease diagnosis, tissue engineering, and drug screening, where cellular mechanics serve as important biophysical indicators

Formula-based Method

Formula-based method use analytical equations derived from classical mechanical laws to describe cellular mechanics. These include common linear elastic models and viscoelastic models (Alert & Trepat, 2019) (Figure 2B). The formulation of these equations depends on our prior understanding of the system and involves simplifying assumptions to make the problem tractable, such as material homogeneity, linearity, and isotropy. Due to these heavy approximations, such models are only suited for small-scale systems or scenarios where approximate, rapid estimations are sufficient.

Numerical-based Method

This method applies material laws to small discrete elements or agents and uses computational techniques (solver) to reconstruct the continuous fields that describe stress, strain, and other mechanical quantities with high spatial resolution. The most commonly used approaches include the finite element method (FEM), agent-based models, cytoskeletal network models, and coarse-grained molecular simulations. However, such models typically demand significant computational resources due to the complexity of the solvers. CompuCell3D and PhysiCell are two of the most widely adopted platforms for implementing numerical-based cell mechanics models.

Whole-cell Model

Whole-cell model integrates multiple biological processes into a unified computational framework to simulate the complete physiology of a single cell. A breakthrough is the 2012 Mycoplasma

genitalium model developed by Markus Covert's group, which comprises 28 distinct process-specific submodules representing DNA replication, transcription, translation, metabolism, and other cellular functions (Karr et al., 2012). This model marked the first attempt to comprehensively incorporate multi-dimensional information, producing simulations that were both relatively accurate and biologically informative. More recently, AI-driven approaches have been proposed to advance toward a more generalized and scalable virtual cell framework (Bunne et al., 2024) (Figure 2C). These methods leverage deep neural network architectures to infer complex biological interactions and simulate cellular behavior in a data-driven manner, aiming to overcome the limitations of human-curated mechanistic models.

To achieve more comprehensive whole-cell models, larger and more diverse datasets are required. In addition to data quantity, the type and quality of data are equally important—particularly multi-modal, multi-scale, and multi-context biological data collected under various intervention combinations, which align with the model's goal of capturing cellular complexity. However, it is important to note that such datasets may contain redundant or irrelevant information. Therefore, identifying which data are truly informative becomes an integral part of model construction.

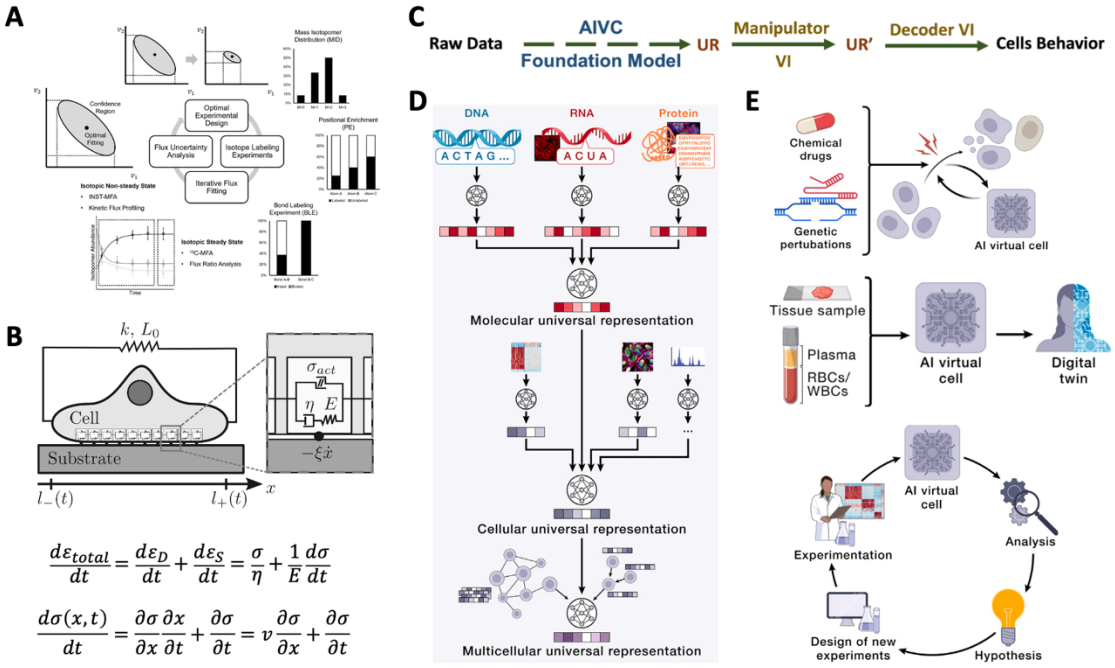
Despite significant advancements, achieving fully functional and accurate whole-cell models remains a distant goal, demanding sustained interdisciplinary collaboration and technological innovation. Nevertheless, their potential applications in predictive modeling, new biology laws discovery, digital twin development, and experimental hypotheses generation remain profoundly promising (Figure 2E).

Process-based Modular Method

This is an explicit modeling approach that decomposes cellular behavior into modularized sub-models, each representing a specific biological process. A central coordinator integrates these sub-models and ensures their compatibility. For each biological process, the corresponding modeling techniques introduced earlier are applied to build the individual submodules. While this modular design improves interpretability and flexibility, the straightforward integration of sub-models often fails to fully capture the complex interdependencies, synergistic effects, and system-wide heterogeneity present in living cells, which are essential for reproducing emergent dynamics.

AI-based Method

The AI-based model seeks to establish a foundational architecture capable of taking in diverse biological data and transforming them into a self-consistent Universal Representation (UR)—an embedding in a latent space for subsequent transformations—to enable integrated modeling (Figure 2D). This architecture is trained to learn the mapping between pre- and post-intervention cellular states using training data, thereby constructing predictive models based on learned patterns. However, a major limitation of this approach lies in its black-box nature, which significantly hampers interpretability and constrains our ability to understand the underlying biological mechanisms inferred by the model.



$$\frac{d\varepsilon_{total}}{dt} = \frac{d\varepsilon_D}{dt} + \frac{d\varepsilon_S}{dt} = \frac{\sigma}{\eta} + \frac{1}{E}\frac{d\sigma}{dt}$$

$$\frac{d\sigma(x,t)}{dt} = \frac{\partial\sigma}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial\sigma}{\partial t} = v\frac{\partial\sigma}{\partial x} + \frac{\partial\sigma}{\partial t}$$

Figure 2. Examples of modeling frameworks and future prospectives

Discussion

When referring to “cell models,” researchers may be speaking of any of the aforementioned modeling frameworks. Although these models differ in their objectives, underlying assumptions,

and methodologies, they are fundamentally similar in structure: each functions as a system that processes input conditions and generates specific outputs that characterize cellular behavior. In essence, they abstract biological complexity into interpretable computational or theoretical forms.

Despite this shared abstraction, each model type and associated methods possess distinct strengths and limitations, making them more or less suited for different applications. Therefore, understanding the characteristics of various models (Figure 3) is essential for aligning model selection with the nature of the biological question being addressed. For instance, in drug screening for a specific disease, a metabolic network model may be particularly effective in identifying flux alterations as indicators of disease-related metabolic dysfunction, whereas mechanical response models may offer less informative insights in this context.

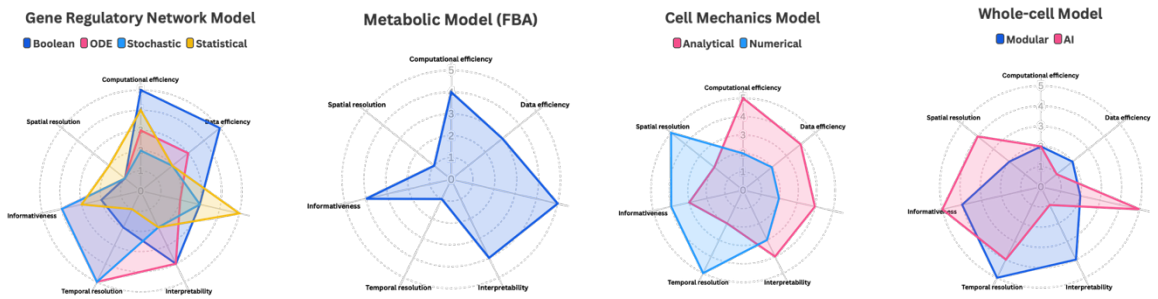


Figure 3. Comparative evaluation of four major modeling frameworks and their common methodological approaches

Future

Looking ahead, cell modeling is moving toward developing more accurate, comprehensive, and interpretable models that leverage emerging data types and computational technologies. This includes expanding model coverage across both biological scales and processes. However, current mature models remain largely task-specific, and constructing general-purpose cell models that can coherently integrate multi-modal and multi-scale biological data remains a significant challenge.

Progress in this area will require addressing limitations in data availability and quality, improving computational frameworks, and deepening our understanding of biological and physical

principles. Achieving these goals demands sustained interdisciplinary collaboration across biology, physics, computer science, and engineering. Ultimately, advancing cell modeling holds great promise for accelerating scientific discovery and improving human health.

References

Alert, R., & Trepat, X. (2019). Physical models of collective cell migration. *Annual Review of Condensed Matter Physics*, *11*(1), 77–101. https://doi.org/10.1146/annurev-conmatphys-031218-013516

Bunne, C., Roohani, Y., Rosen, Y., Gupta, A., Zhang, X., Roed, M., Alexandrov, T., AlQuraishi, M., Brennan, P., Burkhardt, D. B., Califano, A., Cool, J., Dernburg, A. F., Ewing, K., Fox, E. B., Haury, M., Herr, A. E., Horvitz, E., Hsu, P. D., … Quake, S. R. (2024). How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell*, *187*(25), 7045–7063. https://doi.org/10.1016/j.cell.2024.11.015

Covert, M. W. (2014). *Fundamentals of systems biology: From synthetic circuits to whole-cell models*. https://www.amazon.com/Fundamentals-Systems-Biology-Synthetic-Whole-cell/dp/1420084100

Dai, Z., & Locasale, J. W. (2017). Understanding metabolism with flux analysis: From theory to application. *Metabolic Engineering*, *43*, 94–102. https://doi.org/10.1016/j.ymben.2016.09.005

Huynh-Thu, V. A., & Sanguinetti, G. (2019). Gene regulatory network inference: An introductory survey. In G. Sanguinetti & V. A. Huynh-Thu (Eds.), *Gene Regulatory Networks* (Vol. 1883, pp. 1–23). Springer New York. https://doi.org/10.1007/978-1-4939-8882-2_1

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I., & Covert, M. W. (2012). A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, *150*(2), 389–401. https://doi.org/10.1016/j.cell.2012.05.044

Kommu, S., Wang, Y., Wang, Y., & Wang, X. (2024). *Gene regulatory network inference from pre-trained single-cell transcriptomics transformer with joint graph learning*. arXiv. https://arxiv.org/abs/2407.18181

Li, F. (2022). Filling gaps in metabolism using hypothetical reactions. *Proceedings of the National Academy of Sciences*, *119*(49), e2217400119. https://doi.org/10.1073/pnas.2217400119

Molnar, K., & Manneville, J.-B. (2025). Emerging mechanobiology techniques to probe intracellular mechanics. *Npj Biological Physics and Mechanics*, *2*(1), 12. https://doi.org/10.1038/s44341-025-00016-4

Noor, M. S. (2025). Next-generation metabolic models informed by biomolecular simulations. *Current Opinion in Biotechnology*.

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614

Vos, B. E., Muenker, T. M., & Betz, T. (2024). Characterizing intracellular mechanics via optical tweezers-based microrheology. *Current Opinion in Cell Biology*, *88*, 102374. https://doi.org/10.1016/j.ceb.2024.102374

Xu, J., Zhang, A., Liu, F., & Zhang, X. (2023). STGRNS: An interpretable transformer-based method for inferring gene regulatory networks from single-cell transcriptomic data. *Bioinformatics*, *39*(4), btad165. https://doi.org/10.1093/bioinformatics/btad165

Yuan, Q., & Duren, Z. (2024). Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nature Biotechnology*. https://doi.org/10.1038/s41587-024-02182-7

Zhu, H., Miao, R., Wang, J., & Lin, M. (2024). Advances in modeling cellular mechanical perceptions and responses via the membrane-cytoskeleton-nucleus machinery. *Mechanobiology in Medicine*, *2*(1), 100040. https://doi.org/10.1016/j.mbm.2024.100040