Driving biological insight from genome-scale and multiscale to whole-cell modeling

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Driving biological insight from genome-scale and multiscale to whole-cell modeling

Yuhang Li[1,2,3], Jens Nielsen[4,5], Feiran Li[6] and Yu Chen[1,2,7]

Mathematical modeling is a powerful tool for gaining insights into diverse cellular processes and for guiding rational design of biological systems. A growing trend from genome-scale and multiscale to whole-cell models has emerged and these large-scale models serve as valuable knowledge platforms for the reconstruction of cellular systems across diverse biological scales, ultimately enabling comprehensive understanding of cellular systems. Here, we briefly summarize the primary principles of these models and highlight how they drive biological insights. Additionally, we discuss the respective advantages of them and outline potential opportunities represented by data and algorithms for advancing this paradigm.

Addresses
[1] State Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[2] Faculty of Synthetic Biology, Shenzhen University of Advanced Technology, Shenzhen, China
[3] State Key Laboratory of Pharmaceutical Biotechnology, Institute of Functional Biomolecules, School of Life Sciences, Nanjing University, Nanjing, China
[4] BioInnovation Institute, Copenhagen N, Denmark
[5] Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden
[6] Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China
[7] Shenzhen Key Laboratory of Genome Manipulation and Biosynthesis, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Corresponding authors: Nielsen, Jens (nielsenj@chalmers.se); Chen, Yu (y.chen3@siat.ac.cn); Li, Feiran (feiranli@sz.tsinghua.edu.cn)

Current Opinion in Systems Biology 2026, **45**:100596

This review comes from a themed issue on **Whole Cell Modeling 2026**

Edited by **Alexander Hoffman**, **Chiara Damiani** and **Markus Covert**

For a complete overview see the Issue and the Editorial

Available online 25 May 2026

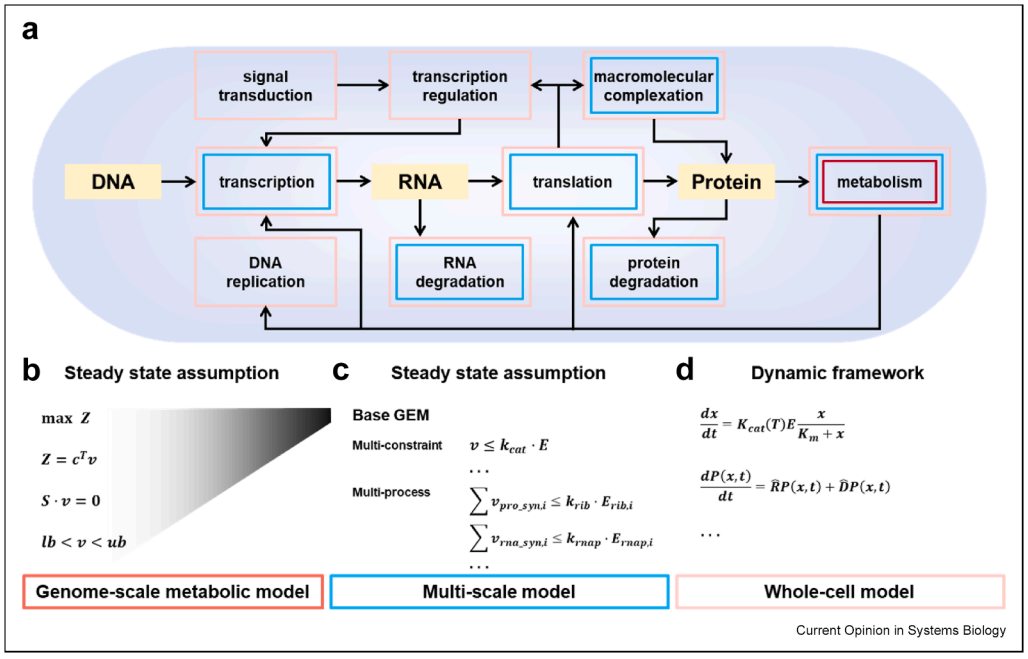
https://doi.org/10.1016/j.coisb.2026.100596

Introduction

Mathematical models employing diverse algorithms and covering different aspects of cell biology are pivotal for elucidating cellular complexity (Figure 1a). For example, these include network analysis for gene regulatory network [1], stochastic simulation for gene expression [2], differential equations for signal transduction [3], and flux balance analysis (FBA) for metabolism [4]. Each of these models consists of numerous molecular interactions and associated parameters. While the specifics differ, systematically incorporating, evaluating, and analyzing heterogeneous knowledge from public databases and the literature remains unstandardized and manual, thus leading to a labor-intensive and time-consuming task.

Over the past two decades, a growing trend from genome-scale and multiscale to whole-cell models (WCMs; hereafter collectively termed as large-scale models) has highlighted the role of these models as integrative knowledge platforms for systematically reconstructing, evaluating, and simulating organism-specific cellular systems. The genome-scale metabolic model (GEM) of *Haemophilus influenzae* was the earliest effort which described the metabolic genotype containing 488 metabolic reactions operating on 343 metabolites [5]. GEMs use FBA to predict cellular phenotype (metabolic fluxes of each reaction) from the genotype under specific genetic and environmental perturbations (Figure 1b) [6]. Later, considering the sparse solution space of GEMs, multi-constraint models, which combine with thermodynamic and enzyme constraints, have been developed [7,8]. Meanwhile, another direction for extending GEMs is to incorporate other cellular processes and thus lead to multiprocess models, e.g., metabolism and gene expression models (ME-models; Figure 1c) [9]. Together, multi-constraint models and multiprocess models are now often referred to as multiscale models [10]. Ultimately, accumulated knowledge and the ambition to predict all cellular states drive the development of WCMs with a highly hybrid mathematical framework, including

Figure 1



Overview of modeling processes and principles from genome-scale and multiscale to whole-cell modeling. (**a**). Major biochemical processes and components in a cell, modified from Reference [13]. Modeling processes are multiple colored by modeling types. For instance, metabolism rectangles are colored by red, blue, and pink, indicating that this process is contained in all three types of large-scale models. (**b**), (**c**), and (**d**) describe the mathematical principles of genome-scale metabolic model, multiscale model, and whole-cell model, respectively, with corresponding colored rectangles.

deterministic equations, stochastic simulations, and FBA (Figure 1d) [11–15].

The key value of large-scale models lies in their ability to provide a systematic perspective to explore underlying mechanisms across diverse cellular processes. These increasingly developing models can be applied independently or in conjunction with other experimental data, supporting hypothesis generation and systems-level insights into multiple disciplines. Previous studies have thoroughly reviewed the principles of the aforementioned models and their applications in aiding rational design in metabolic engineering and biomedicine [16–18]. In this review, we introduce the basic principles of large-scale models from genome-scale and multiscale to WCMs and highlight their roles in generating novel insights into diverse biological systems. Note that there are other valuable analysis methods for these models, such as kinetic metabolic modeling [19] and network-based approaches [20], which are not discussed in this review.

Genome-scale metabolic modeling

Reconstruction of a GEM starts from collecting a large set of enzyme-catalytic biochemical reactions, which has been greatly facilitated by public databases and automated tools [21]. Following the law of mass conservation, a linear programming problem, so called FBA, could be formed using stoichiometric information of modeling reactions. The mathematical formalism is defined as follows:

$$max : Z \tag{1}$$

$$Z = c^T v \tag{2}$$

$$S \cdot v = 0 \tag{3}$$

$$lb < v < ub \tag{4}$$

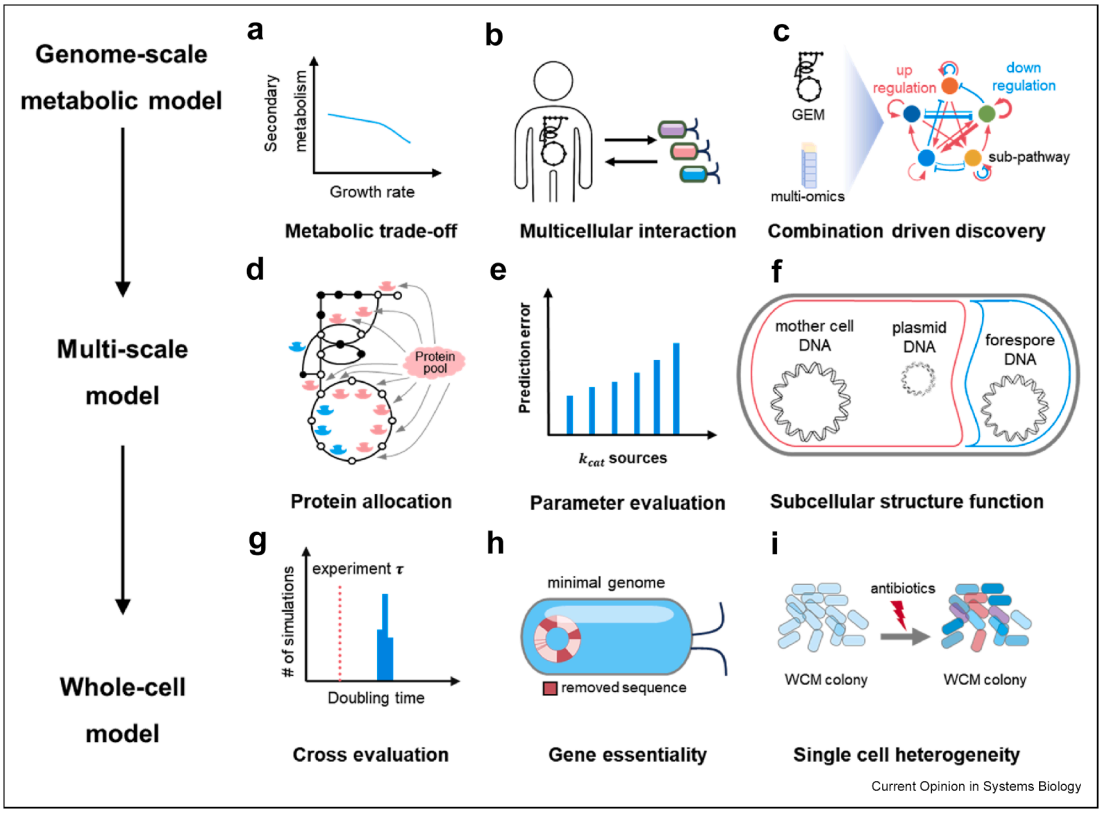
where Z is the objective function, v is the reaction flux vector, $c^T v$ is the combination of target reaction fluxes, and $S \cdot v = 0$ enforces mass conservation by coupling the stoichiometric matrix S with the reaction flux vector v, lb, and ub are lower bound and upper bound of v, respectively. Under this steady state framework, GEM gives a flux distribution in the network that is associated with the objective function, typically maximum flux toward biomass, through multiple algorithm variants, such as parsimonious FBA, and flux variability analysis [6]. However, the flux distribution is not unique as several different flux distributions may be associated with the objective function and flux variability analysis can be used to identify the variability of fluxes in the network.

GEMs can be applied to modeling individual cells, populations, or integrated with multi-omics data, thereby supporting systems-level insights into multiple disciplines (Figure 2a–c). The primary use of GEMs is to understand trade-offs within metabolism, such as secondary metabolism and growth [22], and redox trade-offs under methanotrophic and autotrophic conditions [23]. An increasing number of available GEMs and development of automated reconstruction tools enable investigation of intricate metabolic interactions in multicellular systems [21], such as human–microbe interactions [24], animal–microbe interactions [25], and plant–microbe interactions [26]. Combining with multi-omics datasets, including transcriptomics [27], proteomics [28], and single-cell transcriptomics [29], GEMs support further in-depth understanding. In particular, when combining with perturbed transcriptome datasets, GEM iCEL1314 has been used to calculate condition-specific fluxes, revealing systems-level design principles of metabolic rewiring in *Caenorhabditis elegans*, called the compensation-repression model [27].

Multiscale modeling

Generally, reconstructing a multiscale model begins with a base GEM and integrates additional constraints or processes. For multi-constraint models, e.g., enzyme-constrained models (ecModels), the fundamental principle is to account for the limited catalytic capacity and abundance of enzymes. This is achieved by imposing constraints on reaction fluxes based on the measured or estimated turnover numbers (k_{cat}) of enzymes and their abundance [30,31]. The mathematical formalism can be represented as an extension of the base GEM (Eqs. (1)–(4)),

Figure 2



Applications in generating biological insights from genome-scale and multiscale to whole-cell modeling. Illustrative examples are described as follows: (**a**) Potato-GEM has been used to explore the mechanisms of growth-defense trade-offs [22]; (**b**) APOLLO, 247,092 microbiome models for different body site, age groups, and host health statuses, supporting future investigation of host–microbe interactions [24]; (**c**) The compensation-repression model is revealed through GEM and transcriptome dataset. Each nodes indicates subsystem function and each edge indicates the upregulation or downregulation of target node genes in response to perturbations of source node genes [27]; (**d**) Enzyme-constrained model was used to explore proteome allocation strategy heterogeneity across pathways account for more than half of the *E. coli* proteome during exponential growth [33]; (**e**) PRESTO was proposed to evaluate turn over numbers by simultaneous consideration of condition-specific proteomics and physiological datasets [37]; (**f**) Multiscale models revealing division of metabolic labor of mother and son cells during sporulation and trade-off between bacterial DNA and plasmid [39,40]; (**g**) WCM enables cross-evaluation of heterogenous datasets, such as RNA polymerase efficiency [13]; (**h**) Machine learning surrogate training on WCM simulations enables evaluation of gene essentiality and design of a minimal genome [45]; (**i**) Random simulations of WCMs revealing single-cell heterogeneity [43]. GEM, genome-scale metabolic model; WCMs, whole-cell models.

$$v \leq k_{cat} \cdot E \tag{5}$$

where k_{cat} is the turnover number of the enzyme, E is the abundance of the corresponding catalytic enzyme. For multiprocess models, the stoichiometric matrix $S\cdot\ v$ is expanded to include reactions and components in various cellular processes. These models explicitly couple metabolic fluxes to the energetic and metabolic demands of other cellular processes while constraining metabolic fluxes with the availability of the corresponding enzymes synthesized by other processes, requiring a self-consistent allocation of resources among multiple cellular processes [9]. As an illustrative example, the extended mathematical formalism is presented as follows,

$$\sum v_{pro\_syn,i} \leq k_{rib} \cdot E_{rib} \tag{6}$$

$$\sum v_{rna\_syn,i} \leq k_{rnap} \cdot E_{rnap} \tag{7}$$

where $v_{pro\_syn,i}$ is the protein synthesis rate, $v_{rna\_syn,i}$ is the RNA synthesis rate, and $\sum$ is the sum of synthesis rates. Similar to equation (5), k_{rib} and k_{rnap} are the translation and transcription efficiency, respectively. E_{rib} and E_{rnap} are the ribosome and RNA polymerase abundance, respectively. With expanded knowledge and parameters, multiscale models obtain not only a reduced solution space of the metabolic fluxes [8,30], i.e. the flux variability is significantly reduced, but additional allocation strategy of proteome and ribosome [9], hence are sometimes referred to as resource allocation models [32].

Multiscale models have been developed to understand cellular functions beyond metabolism either independently or in combination with other constraints and processes (Figure 2d–f). With sufficient coverage of mechanistic information (e.g., enzyme kinetics, macromolecular complexation stoichiometry), ecModels and ME-models are capable of calculating condition-specific resource allocation strategies in model organisms, such as *Escherichia coli* [33], *Saccharomyces cerevisiae* [34], and *Arabidopsis thaliana* [35]. Additionally, integrating other mechanistic constraints, such as predictive thermodynamics [36], corrected enzyme kinetics [37], and multiple stresses (including temperature, pH, and oxygen) [38], extends the capacity of multiscale modeling frameworks and enables evaluation of their critical influence on cellular physiology. Incorporating additional cellular processes has further enhanced our understanding of subcellular structure functions, such as spore differentiation [39] and plasmid system [40]. A representative example is that *E. coli* ME-models were constructed for wild-type and mutant strains to simulate differences in extracellular electron transfer pathways, which helped illustrate a potentially widespread new type of anaerobic energy metabolism [41].

Whole-cell model

WCMs often incorporate a wealth of mechanistic knowledge from heterogeneous sources. Differences in the complexity and our knowledge of modeled organisms have led to distinct mathematical frameworks, yet these approaches converge on a shared objective of capturing the temporal dynamics of concentrations of molecules at the single-cell level, and in some cases, their spatial organization [11–15]. Details of their formulation and modeling frameworks have been summarized previously [17], hence we only present the simplified equation of RNA, protein, and metabolite from Reference [13] as follows:

$$\frac{dr}{dt} = v_{syn,r} - \left(\frac{ln\ 2}{t_{r,1/2}} + \frac{ln\ 2}{\tau}\right) r \tag{8}$$

$$\frac{dp}{dt} = v_{syn,p} - \left(\frac{ln\ 2}{t_{p,1/2}} + \frac{ln\ 2}{\tau}\right) p \tag{9}$$

$$\frac{dm}{dt} = \sum Sv = \sum S \cdot k_{cat} \cdot E \cdot \frac{m}{K_m + m} \tag{10}$$

where $\frac{dr}{dt}$, $\frac{dp}{dt}$, and $\frac{dm}{dt}$ are the net rates of concentration change of RNA, protein, and metabolite, respectively. For RNA (Eq. (8)), $v_{syn,r}$ is the simplified synthesis rate, $t_{r,1/2}$ is the half life of each RNA, τ is the doubling time, r is the concentration of each RNA. For protein (Eq. (9)), the formulism is similar to RNA, expect for the synthesis rate of protein $v_{syn,p}$ and the half life of protein $t_{p,1/2}$. For metabolite (Eq. (10)), S is the stoichiometry matrix in the FBA framework, k_{cat} is enzyme turnover number, E is the enzyme copy number, m is the concentration of metabolite, and K_m is Michaelis constant. In this ordinary differential equation (ODE)-based framework, WCM comprises thousands of coupled differential equations and simulates a full cell cycle with substantial numerical integration. As a result, it captures the dynamic evolution of all molecular concentrations over time and thus provides the most fine-grained information.

So far, WCMs have been mainly serving as cross-evaluation platforms to understand the function of all molecular entities (Figure 2g–i). In particular, *E. coli* WCM has been used to vet large datasets from disparate sources, revealing that the measured efficiencies of RNA polymerases and ribosomes were insufficient to support observed cell-doubling times [13]. A subsequent update of the *E. coli* WCM, incorporating 788 polycistronic operons and 1231 transcription units, enabled the discovery of new principles of operon function [42]. Furthermore, single-cell level and community-level simulations of *E. coli* WCMs enabled the identification of key genes contributing to antibiotic resistance underscoring the potential of WCMs in understanding cellular heterogeneity [43]. Interestingly,

WCMs have been leveraged for identifying gene essentiality. The earlier effort, *Mycoplasma genitalium* WCM, predicted a bacterium with the smallest genome *in silico* [44]. More recently, combining *E. coli* WCM with a machine learning surrogate allowed prediction of a reduced genome with only 737 genes while preserving division capacity [45].

Challenges and perspectives

From genome-scale and multiscale to WCM, this paradigm provides systematic understanding of cellular systems at diverse scales. Below, we summarize key directions regarding how these models enable biological insights in recent years.

Application breadth is inversely proportional to model complexity, largely due to knowledge gaps and the degree of automation in model construction. As aforementioned, GEMs have been applied to multicellular systems across multiple disciplines; multiscale models account for diverse cellular processes in many organisms; whereas WCMs remain available only for a few organisms, with few downstream applications. The reason for the limited use of WCMs compared with the wide application of GEMs is that the latter can be reconstructed using gene-protein-reaction association [46], and a variety of automated pipelines have made community-level automated reconstruction possible by virtue of this association [21]. In contrast, multiscale models are constrained by quantitative knowledge gaps, most notably the lack of enzyme kinetic parameters. Consequently, parameter prediction tools [36,47] have become essential components of automated reconstruction pipelines. Though differences exist *in vivo* and *in vitro* [48], alternative or effective parameters are adopted to prioritize model availability [31,49]. Notably, this pragmatic strategy has recently enabled the first community-level automated reconstruction, curation, and simulation of ME-models [50]. More recently, large language models have been leveraged for automated knowledge integration and curation, which enables the generation of enzyme-constrained whole-body human metabolic models and the simulation of inter-organ metabolite exchanges under varying nutritional states using a dynamic framework [51]. Despite higher mechanistic resolution, WCMs are reconstructed in the presence of substantial qualitative and quantitative knowledge gaps. For instance, the well-annotated gene fraction was low for *E. coli* WCM [13] and completely organism-specific parameterization was difficult [11,12,14,15]. Moreover, WCMs are further constrained by the heavy computational burden associated with large-scale numerical integration, which severely limits scalability and downstream application [17,45].

Given that GEMs and multiscale models containing vast stoichiometric and kinetic information are valuable for establishing WCMs [11–15], we emphasize potential opportunities for further advancing WCMs to overcome two major limitations: knowledge gaps and computational constraints. From a mechanistic perspective, the most essential ingredients for a WCM are biochemical reaction equations and corresponding parameters, such as Michaelis–Menten equations and k_{cat}, which describe complex molecular interactions qualitatively and quantitatively. Except retrieving from public databases directly, combining liquid chromatography-mass spectrometry with high-throughput experimental technologies now enables rapid identification of unknown molecular interactions *in vivo* and *in vitro* [52,53]. In parallel, through artificial intelligence (AI)-assisted methods, multiple computational tools have been developed for macromolecular interaction prediction [54], enzyme specificity prediction [55], and biochemical reaction kinetic parameter prediction [56]. Beyond biochemical knowledge gaps, physical factors regarding how spatial location and mechanical force influence cellular behaviors have been recently studied or reviewed [57,58]. Considering the heavy computational burden of large-scale simulations, neural operators (e.g., physics-informed neural network) are widely used to learn solution operators of high-dimensional differential equations, providing fast and generalizable surrogates for traditional numerical solvers [59]. From an integrative perspective, hybrid approaches that integrate AI with mechanistic models, often referred to as gray-box models [60], constitute a rising field that helps fill knowledge gaps and reduce computational burden of large-scale modeling [61]. Moreover, standard datasets and competition, e.g., Virtual Cell Challenge [62], may seem to be a great driving force for advancing WCM. Overall, advances in both dry-lab and wet-lab technologies for filling knowledge gaps, together with automated reconstruction pipelines and pragmatic simulation strategies, pave the way toward improving model performance from genome-scale and multiscale to WCM, thereby accelerating life sciences in the foreseeable future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work was financially supported by the National Key R&D Program of China (2023YFA0913900), the Shenzhen Medical Research Fund (A2303026), the National

Natural Science Foundation of China (22478223, 32571654 and 62532017), and the Shenzhen Science and Technology Program (ZDSYS20220303153551001).

Data availability

No data was used for the research described in the article.

References

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

* of special interest
** of outstanding interest

1. Ali SIM, Alrashid SZ: **A review of methods for gene regulatory networks reconstruction and analysis**. *Artif Intell Rev* 2025, **58**:256, https://doi.org/10.1007/s10462-025-11257-z.
2. Miles CE: **Mechanistic inference of stochastic gene expression from structured single-cell data**. *Curr Opin Syst Biol* 2025, **42**, 100555, https://doi.org/10.1016/j.coisb.2025.100555.
3. Erdem C, Mutsuddy A, Bensman EM, Dodd WB, Saint-Antoine MM, Bouhaddou M, *et al*.: **A scalable, open-source implementation of a large-scale mechanistic model for single cell proliferation and death signaling**. *Nat Commun* 2022, **13**: 3555, https://doi.org/10.1038/s41467-022-31138-1.
4. Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245–248, https://doi.org/10.1038/nbt.1614.
5. Edwards JS, Palsson BO: **Systems properties of the haemophilus influenzaeRd metabolic genotype**. *J Biol Chem* 1999, **274**:17410–17416, https://doi.org/10.1074/jbc.274.25.17410.
6. Carter EL, Constantinidou C, Alam MT: **Applications of genome-scale metabolic models to investigate microbial metabolic adaptations in response to genetic or environmental perturbations**. *Briefings Bioinf* 2023, **25**:bbad439, https://doi.org/10.1093/bib/bbad439.
7. Li F, Chen Y, Qi Q, Wang Y, Yuan L, Huang M, *et al*.: **Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints**. *Nat Commun* 2022, **13**:2969, https://doi.org/10.1038/s41467-022-30689-7.
8. Henry CS, Broadbelt LJ, Hatzimanikatis V: **Thermodynamics-based metabolic flux analysis**. *Biophys J* 2007, **92**:1792–1805, https://doi.org/10.1529/biophysj.106.093138.
9. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ: **Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction**. *Mol Syst Biol* 2013, **9**:693, https://doi.org/10.1038/msb.2013.52.
10. Lu H, Kerkhoven EJ, Nielsen J: **Multiscale models quantifying yeast physiology: towards a whole-cell model**. *Trends Biotechnol* 2022, **40**:291–305, https://doi.org/10.1016/j.tibtech.2021.06.010.
11. Thornburg ZR, Bianchi DM, Brier TA, Gilbert BR, Earnest EE, Melo MCR, *et al*.: **Fundamental behaviors emerge from simulations of a living minimal cell**. *Cell* 2022, **185**:345–360.e328, https://doi.org/10.1016/j.cell.2021.12.025.
12. Ye C, Xu N, Gao C, Liu G, Xu J, Zhang W, *et al*.: **Comprehensive understanding of *Saccharomyces cerevisiae* phenotypes with whole-cell model WM_S288C**. *Biotechnol Bioeng* 2020, **117**:1562–1574, https://doi.org/10.1002/bit.27298.
13. Macklin DN, Ahn-Horst TA, Choi H, Ruggero NA, Carrera J, Mason JC, *et al*.: **Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation**. *Science* 2020, **369**:eaav3751, https://doi.org/10.1126/science.aav3751.
14. Karr Jonathan R, Sanghvi Jayodita C, Macklin Derek N, Gutschow Miriam V, Jacobs Jared M, Bolival B, *et al*.: **A whole-cell computational model predicts phenotype from genotype**. *Cell* 2012, **150**:389–401, https://doi.org/10.1016/j.cell.2012.05.044.
15. ** Thornburg ZR, Maytin A, Kwon J, Brier TA, Gilbert BR, Fu E, *et al*.: **Bringing the genetically minimal cell to life on a computer in 4D**. *Cell* 2026, **189**:1–16, https://doi.org/10.1016/j.cell.2026.02.009.
This work integrates reaction diffusion master equation, global chemical mater equation, Brownian dynamics, and ordinary differential equation to construct a 4D (space and time) whole cell model of the entire cell cycle of the minimal bacterium JCVI-syn3A. The model captures chromosome dynamics, metabolism, growth, and division, and leverages 4D visualization to reveal how spatial heterogeneity of the intracellular environment influences biochemical reactions and phenotypic heterogeneity.
16. Schroeder WL, Suthers PF, Willis TC, Mooney EJ, Maranas CD: **Current state, challenges, and opportunities in genome-scale resource allocation models: a mathematical perspective**. *Metabolites* 2024, **14**:365, https://doi.org/10.3390/metabo14070365.
17. Georgouli K, Yeom J-S, Blake RC, Navid A: **Multi-scale models of whole cells: progress and challenges**. *Front Cell Dev Biol* 2023, **11**, 1260507, https://doi.org/10.3389/fcell.2023.1260507.
18. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY: **Current status and applications of genome-scale metabolic models**. *Genome Biol* 2019, **20**:121, https://doi.org/10.1186/s13059-019-1730-3.
19. Toumpe I, Choudhury S, Hatzimanikatis V, Miskovic L: **The dawn of high-throughput and genome-scale kinetic modeling: recent advances and future directions**. *ACS Synth Biol* 2025, **14**:2906–2919, https://doi.org/10.1021/acssynbio.4c00868.
20. Shinar G, Feinberg M: **Structural sources of robustness in biochemical reaction networks**. *Science* 2010, **327**: 1389–1391, https://doi.org/10.1126/science.1183372.
21. Srinak N, Krüger F, Kaleta C, Taubenheim J: **Metabolic modeling of host-microbe interactions**. *Comput Struct Biotechnol J* 2025, **27**:4304–4319, https://doi.org/10.1016/j.csbj.2025.10.006.
22. Zrimec J, Correa S, Zagorščak M, Petek M, Bleker C, Stare K, *et al*.: **Evaluating plant growth–defense trade-offs by modeling the interaction between primary and secondary metabolism**. *Proc Natl Acad Sci USA* 2025, **122**, e2502160122, https://doi.org/10.1073/pnas.2502160122.
23. Saldivar A, Ruiz-Ruiz P, Revah S, Zuñiga C: **Genome-scale flux balance analysis reveals redox trade-offs in the metabolism of the thermoacidophile Methylacidiphilum fumariolicum under auto-, hetero-and methanotrophic conditions**. *Frontiers in Systems Biology* 2024, **4**, 1291612, https://doi.org/10.3389/fsysb.2024.1291612.
24. Heinken A, Hulshof TO, Nap B, Martinelli F, Basile A, O'Brolchain A, *et al*.: **A genome-scale metabolic reconstruction resource of 247,092 diverse human microbes spanning multiple continents, age groups, and body sites**. *Cell Syst* 2025, **16**, 101196, https://doi.org/10.1016/j.cels.2025.101196.
25. Best L, Dost T, Esser D, Flor S, Gamarra AM, Haase M, *et al*.: **Metabolic modelling reveals the aging-associated decline of host–microbiome metabolic interactions in mice**. *Nat Microbiol* 2025, **10**:973–991, https://doi.org/10.1038/s41564-025-01959-z.
26. Holland BL, Matthews ML, Bota P, Sweetlove LJ, Long SP, diCenzo GC: **A genome-scale metabolic reconstruction of soybean and *Bradyrhizobium diazoefficiens* reveals the cost–benefit of nitrogen fixation**. *New Phytol* 2023, **240**: 744–756, https://doi.org/10.1111/nph.19203.
27. ** Li X, Zhang H, Hodder T, Wang W, Myers CL, Yilmaz LS, *et al*.: **Systems-level design principles of metabolic rewiring in an animal**. *Nature* 2025, **640**:203–211, https://doi.org/10.1038/s41586-025-08636-5.
Combing GEM and perturbation transcriptome, this work proposes a system level design principle of metabolic rewiring in *Caenorhabditis elegans*, and may also human.
28. Li X, Walhout AJM, Yilmaz LS: **Enhanced flux potential analysis links changes in enzyme expression to metabolic flux**.

Mol Syst Biol 2025, **21**:413–445, https://doi.org/10.1038/s44320-025-00090-9.

29. Huang Y, Mohanty V, Dede M, Tsai K, Daher M, Li L, *et al.*: **Characterizing cancer metabolism from bulk and single-cell RNA-seq data using METAFlux**. *Nat Commun* 2023, **14**:4883, https://doi.org/10.1038/s41467-023-40457-w.

30. Sánchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven EJ, Nielsen J: **Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints**. *Mol Syst Biol* 2017, **13**:935, https://doi.org/10.15252/msb.20167411.

31. Chen Y, Gustafsson J, Rangel AT, Anton M, Domenzain I, Kittikunapong C, *et al.*: **Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO toolbox 3.0**. *Nat Protoc* 2024, **19**:629–667, https://doi.org/10.1038/s41596-023-00931-7.

32. Teusink B, Grigaitis P, Remeijer M, Bruggeman F, Steuer R: **Resource allocation models: theory and applications in microbial biotechnology**. *Curr Opin Biotechnol* 2026, **97**, 103391, https://doi.org/10.1016/j.copbio.2025.103391.

33. Hu X-P, Schroeder S, Lercher MJ: **Proteome efficiency of metabolic pathways in *Escherichia coli* increases along the nutrient flow**. *mSystems* 2023, **8**, e00760, https://doi.org/10.1128/msystems.00760-23. –00723.

34. Dinh HV, Maranas CD: **Evaluating proteome allocation of Saccharomyces cerevisiae phenotypes with resource balance analysis**. *Metab Eng* 2023, **77**:242–255, https://doi.org/10.1016/j.ymben.2023.04.009.

35. Goelzer A, Rajjou L, Chardon F, Loudet O, Fromion V: **Resource allocation modeling for autonomous prediction of plant cell phenotypes**. *Metab Eng* 2024, **83**:86–101, https://doi.org/10.1016/j.ymben.2024.03.009.

36. Fan W, Hao Y, Hou X, Ding C, Huang D, Zheng W, *et al.*: **Unraveling principles of thermodynamics for genome-scale metabolic networks using graph neural networks**. *Cell Syst* 2025, **16**, 101393, https://doi.org/10.1016/j.cels.2025.101393.

37. Wendering P, Arend M, Razaghi-Moghadam Z, Nikoloski Z: **Data integration across conditions improves turnover number estimates and metabolic predictions**. *Nat Commun* 2023, **14**:1485, https://doi.org/10.1038/s41467-023-37151-2.

38. Zhao J, Chen K, Palsson BO, Yang L: **StressME: unified computing framework of Escherichia coli metabolism, gene expression, and stress responses**. *PLoS Comput Biol* 2024, **20**, e1011865, https://doi.org/10.1371/journal.pcbi.1011865.

39. Tibocha-Bonilla JD, Lyda J, Riley E, Pogliano K, Zengler K: **Deciphering metabolic differentiation during *Bacillus subtilis* sporulation**. *Nat Commun* 2025, **16**:129, https://doi.org/10.1038/s41467-024-55586-z.

40. Oftadeh O, Hatzimanikatis V: **Genome-scale models of metabolism and expression predict the metabolic burden of recombinant protein expression**. *Metab Eng* 2024, **84**:109–116, https://doi.org/10.1016/j.ymben.2024.06.005.

41. ** Kundu BB, Krishnan J, Szubin R, Patel A, Palsson BO, Zielinski DC, *et al.*: **Extracellular respiration is a latent energy metabolism in *Escherichia coli***. *Cell* 2025, **188**:p2907–p2924, https://doi.org/10.1016/j.cell.2025.03.016.
E. coli ME models have been constructed for wild-type and mutant strains to simulate differences in extracellular electron transfer pathways, which helped illustrate a potentially widespread new type of anaerobic energy metabolism.

42. Sun G, DeFelice MM, Gillies TE, Ahn-Horst TA, Andrews CJ, Krummenacker M, *et al.*: **Cross-evaluation of E. coli's operon structures via a whole-cell model suggests alternative cellular benefits for low- versus high-expressing operons**. *Cell Syst* 2024, **15**:227–245.e227, https://doi.org/10.1016/j.cels.2024.02.002.

43. Skalnik CJ, Cheah SY, Yang MY, Wolff MB, Spangler RK, Talman L, *et al.*: **Whole-cell modeling of E. coli colonies enables quantification of single-cell heterogeneity in antibiotic responses**. *PLoS Comput Biol* 2023, **19**, e1011232, https://doi.org/10.1371/journal.pcbi.1011232.

44. Rees-Garbutt J, Chalkley O, Landon S, Purcell O, Marucci L, Grierson C: **Designing minimal genomes using whole-cell models**. *Nat Commun* 2020, **11**:836, https://doi.org/10.1038/s41467-020-14545-0.

45. * Gherman IM, Sharma K, Rees-Garbutt J, Pang W, Abdallah ZS, Gorochowski TE, *et al.*: **Accelerated design of Escherichia coli reduced genomes using a whole-cell model and machine learning**. *Cell Syst* 2025, **16**, 101392, https://doi.org/10.1016/j.cels.2025.101392.
This work uses random forest surrogate model to reduce 95% computational time requirement compared with the original WCM in gene knock off *in silico* simulation and obtains a final genome retains ~60% of WCM genes.

46. Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction**. *Nat Protoc* 2010, **5**:93–121, https://doi.org/10.1038/nprot.2009.203.

47. Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, *et al.*: **Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction**. *Nat Catal* 2022, **5**:662–672, https://doi.org/10.1038/s41929-022-00798-z.

48. Van Eunen K, Kiewiet JAL, Westerhoff HV, Bakker BM: **Testing biochemistry revisited: how in vivo metabolism can be understood from in vitro enzyme kinetics**. *PLoS Comput Biol* 2012, **8**, e1002483, https://doi.org/10.1371/journal.pcbi.1002483.

49. Lloyd CJ, Ebrahim A, Yang L, King ZA, Catoiu E, O'Brien EJ, *et al.*: **COBRAme: a computational framework for genome-scale models of metabolism and gene expression**. *PLoS Comput Biol* 2018, **14**, e1006302, https://doi.org/10.1371/journal.pcbi.1006302.

50. Tibocha-Bonilla JD, Santibáñez-Palominos R, Weng Y, Kumar M, Zengler K: **Metabolism and gene expression models for the microbiome reveal how diet and metabolic dysbiosis impact disease**. *Cell Syst* 2025, **17**, 101451, https://doi.org/10.1016/j.cels.2025.101451.

51. * Luo J, Wang H, Moyer D, Guo Z, Robinson JL, Gustafsson J, *et al.*: **Reconstruction of human metabolic models with large language models**. *Proc Natl Acad Sci USA* 2026, **123**, e2516511123, https://doi.org/10.1073/pnas.2516511123.
Human2 establishes a multi-scale human metabolic modeling framework that combines LLM-assisted curation with enzyme-constrained, dynamic whole-body simulations to enable tissue-, sex-, age- and nutrition-specific analysis of human metabolism and inter-organ metabolic coordination.

52. Peng H, Kotelnikov S, Egbert ME, Ofaim S, Stevens GC, Phanse S, *et al.*: **Ligand interaction landscape of transcription factors and essential enzymes in *E. coli***. *Cell* 2025, **188**:1441–1455, https://doi.org/10.1016/j.cell.2025.01.003.

53. Gruber CH, Noor E, Buffing MF, Sauer U: **Systematic identification of allosteric effectors in metabolism**. *Proc Natl Acad Sci USA* 2025, **122**, e2423767122, https://doi.org/10.1073/pnas.2423767122.

54. Zhang J, Humphreys IR, Pei J, Kim J, Choi C, Yuan R, *et al.*: **Predicting protein-protein interactions in the human proteome**. *Science* 2025, **390**:eadt1630, https://doi.org/10.1126/science.adt1630.

55. Cui HY, Su YF, Dean TJ, Yu TH, Zhang ZY, Peng J, *et al.*: **Enzyme specificity prediction using cross-attention graph neural networks**. *Nature* 2025, **647**, https://doi.org/10.1038/s41586-025-09697-2.

56. * Choudhury S, Narayanan B, Moret M, Hatzimanikatis V, Miskovic L: **Generative machine learning produces kinetic models that accurately characterize intracellular metabolic states**. *Nat Catal* 2024, **7**, https://doi.org/10.1038/s41929-024-01220-6.
This work proposes the RENAISSANCE framework which uses machine learning and natural evolution strategies to generate dynamic metabolic models without training data, greatly improving the construction efficiency.

57. Cheng B, Li M, Lin M, Guo H, Xu F: **Mechanobiology across timescales**. *Nat Rev Phys* 2025, **7**:621–644, https://doi.org/10.1038/s42254-025-00874-w.

58. Klumpp S, Pang TY, Lercher MJ: **Optimal density of bacterial cells**. *PLoS Comput Biol* 2023, **19**, https://doi.org/10.1371/journal.pcbi.1011177.

59. Azizzadenesheli K, Kovachki N, Li Z, Liu-Schiaffini M, Kossaifi J, Anandkumar A: **Neural operators for accelerating scientific simulations and design**. *Nat Rev Phys* 2024, **6**:320–328, https://doi.org/10.1038/s42254-024-00712-5.

60. * Ahmadi Daryakenari N, De Florio M, Shukla K, Karniadakis GE: **AI-Aristotle: a physics-informed framework for systems biology gray-box identification**. *PLoS Comput Biol* 2024, **20**, e1011916, https://doi.org/10.1371/journal.pcbi.1011916.
This work formulizes gray-box identification in systems biology by embedding machine learning within physics-informed constraints, directly addressing scalability and identifiability challenges in large scale modeling.

61. Noordijk B, Garcia Gomez ML, Ten Tusscher KHWJ, De Ridder D, Van Dijk ADJ, Smith RW: **The rise of scientific machine learning: a perspective on combining mechanistic modelling with machine learning for systems biology**. *Frontiers in Systems Biology* 2024, **4**, 1407994, https://doi.org/10.3389/fsysb.2024.1407994.

62. Roohani YH, Hua TJ, Tung P-Y, Bounds LR, Yu FB, Dobin A, *et al*.: **Virtual cell challenge: toward a turing test for the virtual cell**. *Cell* 2025, **188**:3370–3374, https://doi.org/10.1016/j.cell.2025.06.008.