The Grey Box Virtual Cell: Unifying Mechanistic and Neural Models through Standardized Benchmarking

Tianyu Wu[1,2]

[1]**Center for Biophysics and Quantitative Biology, University of Illinois at Urbana–Champaign, Urbana, IL, USA**
[2]**National Science Foundation Science and Technology Center for Quantitative Cell Biology, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, IL, USA**

ABSTRACT

Efforts to build a predictive, whole-cell representation have historically relied on integrating diverse biochemical measurements into mechanistic frameworks. However, a divide has emerged between these explicit ”Glass Box” models and newer data-driven ”Black Box” foundation models. In this work, we propose a formal mathematical unification of these paradigms: the ”Grey Box” Virtual Cell. We derive a hybrid state-space formulation that couples mechanistic conservation laws with latent neural representations. Furthermore, to address the lack of rigorous validation standards, we introduce the WCM-Hard Benchmark Suite, a set of eight standardized challenges ranging from essentiality screening to combinatorial drug synergy. We provide a reference implementation of a hybrid cell and demonstrate its performance on these benchmarks. Crucially, our ablation studies reveal that purely mechanistic baselines fail to capture dynamic diauxic shifts that the hybrid architecture navigates successfully. This work provides both the theoretical foundation and the practical testing ground for the next generation of whole-cell models.

1 INTRODUCTION

The long-standing goal of constructing a computational model that captures an entire cell's behavior—including its responses to perturbations, progression through the cell cycle, and the mechanistic details underlying these processes—remains one of the central challenges in quantitative biology. Early efforts approached this goal as a large-scale integration problem: assembling diverse biochemical measurements, structural data, and kinetic parameters into a unified mathematical framework. As the field has grown, however, it has become clear that whole-cell modeling is no longer dominated by a single methodological perspective. Instead, two distinct approaches have emerged, each shaped by different assumptions about what it means to represent a living system.

The **Mechanistic (Bottom-Up)** framework aims to reconstruct cellular behavior from established chemical and physical principles. This tradition, which traces back to the landmark whole-cell model of *Mycoplasma genitalium* developed by Karr et al. (2012), seeks explanatory power through explicit molecular mechanisms: stochastic reaction–diffusion processes, thermodynamic feasibility constraints, and resource-coupled metabolic models. A recent advance in this direction is the four-dimensional simulation of a genetically minimal cell by Thornburg et al. (2025), which illustrates the level of detail required for predictive fidelity. In such models, phenotypes—whether cell division timing or metabolic transitions—arise through the interplay of defined rate constants, spatial organization, and conservation laws. For advocates of this approach, a predictive "digital cell" must operate as a *Glass Box*: transparent, mechanistically grounded, and directly interpretable.

In contrast, the **Empirical (Top-Down)** or data-driven perspective argues that the combinatorial complexity of cellular networks makes full mechanistic reconstruction impractical for many systems. Rather than specifying reactions and parameters, deep generative models—particularly Transformers and variational autoencoders—learn statistical structure directly from large single-cell datasets. This line of work is exemplified by initiatives at the Arc Institute and the Chan Zuckerberg Institute (CZI) Biohub, which have proposed a functional "Turing Test" for the virtual cell: if an AI model predicts perturbation responses indistinguishably from experimental measurements, its internal mechanics may be secondary Roohani et al. (2025). Foundation models such as **State** Adduri et al. (2025) represent this shift, enabling near-instant prediction of transcriptional responses across diverse tissues—something currently beyond the reach of computationally intensive mechanistic simulations.

These two modeling philosophies now define much of the landscape in whole-cell research. The mechanistic approach prioritizes causal interpretation and physical realism, while data-driven models emphasize scalability and predictive breadth. However, the field currently lacks both a rigorous theoretical formalism for integrating these approaches and a standardized validation suite to assess them.

In this paper, we refrain from merely reviewing the literature. Instead, we propose a new theoretical framework for "Grey Box" modeling that mathematically couples mechanistic and latent states. We further introduce the "WCM-Hard" benchmark, a hierarchical testing suite designed to validate virtual cells against ground-truth biological complexity. Finally, we provide an open-source reference implementation and present in silico experimental results demonstrating that satisfying these benchmarks requires a hybrid architecture.

2 THE MECHANISTIC FOUNDATIONS: BUILDING FROM CHEMICAL PRINCIPLES

Mechanistic models aim to reproduce cellular behavior by explicitly representing the underlying physical and chemical processes. In these frameworks, molecular interactions, spatial organization, and thermodynamic constraints are not abstracted away but treated as primary determinants of phenotype. This approach stands in contrast to statistical or neural network–based models: rather than inferring relationships from data, mechanistic simulations attempt to reconstruct how cellular dynamics emerge from defined rules and parameters. As a result, these systems offer interpretability and causal structure, though often at substantial computational cost.

2.1 The Chemically Explicit Cell: Spatiotemporal Dynamics

One of the most detailed mechanistic strategies dispenses with the assumption that the cytosol is well mixed. Instead, it treats the cell as a spatially heterogeneous environment in which molecular discreteness and local diffusion strongly influence reaction kinetics. The work of the Luthey-Schulten Lab has been central to this direction, culminating in a four-dimensional simulation of the genetically minimal cell *JCVI-syn3A* Thornburg et al. (2025). These studies illustrate how spatial constraints, crowding, and chromosome organization shape cellular physiology.

At the core of this framework is the Reaction–Diffusion Master Equation (RDME), which models the cell as a lattice of subvolumes that exchange molecules by diffusion while undergoing discrete stochastic reactions. Unlike deterministic ordinary differential equations that evolve continuous concentrations, the RDME tracks probability distributions over molecular counts. The state of the system at time t is given by $P(\mathbf{n},t)$, the probability of occupying the discrete state $\mathbf{n}$.

The evolution of $P(\mathbf{n},t)$ is governed by the combined contributions of reaction and diffusion:

$$\frac{\partial P(\mathbf{n},t)}{\partial t} = \underbrace{\sum_r \left(a_r(\mathbf{n}-\mathbf{S}_r)P(\mathbf{n}-\mathbf{S}_r,t) - a_r(\mathbf{n})P(\mathbf{n},t)\right)}_{\text{Chemical Reactions}} + \underbrace{\sum_\xi D_\xi \Delta^2 P(\mathbf{n},t)}_{\text{Spatial Diffusion}}. \tag{1}$$

Here, $\mathbf{n}$ denotes molecular copy numbers in each voxel, a_r is the reaction propensity, $\mathbf{S}_r$ is the stoichiometric change vector, and D_ξ represents the diffusion operator.

The Lattice Microbes software suite Roberts et al. (2013) implements this formalism efficiently through GPU-accelerated operator splitting. This typically uses Lie–Trotter or Strang schemes to decompose the RDME into alternating diffusion and reaction updates. The process begins with a diffusion half-step $\mathscr{D}$, where the probability distribution is updated as $P^*(\mathbf{n},t+\Delta t/2) = e^{\mathscr{D}\Delta t/2}P(\mathbf{n},t)$. This is followed by a full reaction step $\mathscr{R}$, yielding $P^{**}(\mathbf{n},t+\Delta t) = e^{\mathscr{R}\Delta t}P^*(\mathbf{n},t+\Delta t/2)$. Finally, a second diffusion half-step completes the update: $P(\mathbf{n},t+\Delta t) = e^{\mathscr{D}\Delta t/2}P^{**}(\mathbf{n},t+\Delta t)$.

This splitting, in which $e^{\mathscr{L}\Delta t} \approx e^{\mathscr{D}\Delta t/2}e^{\mathscr{R}\Delta t}e^{\mathscr{D}\Delta t/2}$ for the combined operator $\mathscr{L} = \mathscr{R} + \mathscr{D}$, allows highly parallelizable simulations involving millions of reaction events per second while maintaining second-order accuracy.

Using this framework, Thornburg et al. (2025) simulated an entire 105-minute cell cycle of *JCVI-syn3A*. By integrating RDME dynamics with Brownian dynamics of the circular chromosome via **LAMMPS** Gilbert et al. (2023), they showed that chromosome segregation and other cell-scale behaviors can emerge from physical constraints rather than explicit genetic programs. Such models also capture stochastic phenomena—such as noise in low-copy regulatory proteins—that are difficult to represent in coarse-grained or well-mixed descriptions.

Mechanistic spatial models are now being extended to eukaryotic systems. Wu et al. (2025) applied a hybrid RDME–ODE approach to the budding yeast galactose switch, using cryo-electron tomography to reconstruct organelle geometries. Their work highlights that spatial heterogeneity, particularly in ribosome distribution across the endoplasmic reticulum, can significantly alter signaling dynamics in ways that well-mixed models do not predict.

2.2 The Constraint-Based Cell: Thermodynamics and Resource Allocation

While spatially resolved simulations provide rich mechanistic detail, they become computationally prohibitive for larger genomes and more complex geometries. Constraint-based approaches offer a

complementary strategy by defining the feasible space of metabolic states without tracking individual molecules. This methodology underlies genome-scale models from the Covert Lab and Hatzimanikatis Lab Macklin et al. (2020); Salvy and Hatzimanikatis (2020).

Classical **Flux Balance Analysis (FBA)** Raman and Chandra (2009) assumes steady-state metabolite concentrations:

$$\mathbf{S} \cdot \mathbf{v} = 0, \tag{2}$$

and selects flux distributions that optimize a specified objective, often growth rate:

$$\text{minimize } Z = -\mu. \tag{3}$$

However, standard FBA lacks explicit thermodynamic or resource constraints and may admit physically unrealistic flux cycles. The **ETFL** (Expression, Thermodynamics, and Flux) framework Salvy and Hatzimanikatis (2020) addresses these limitations by incorporating two critical sets of constraints. First, it enforces thermodynamic feasibility, requiring that each flux v_j respects the Gibbs free energy constraint $\Delta_r G'_j = \Delta_r G'^{\circ}_j + RT \ln\left(\prod_i [C_i]^{S_{ij}}\right) < 0$, which prevents flux assignments that violate physical laws. Second, it implements resource allocation by coupling enzyme levels to reaction rates through $v_j \leq k_{cat,j}[E_j]$. Because enzyme synthesis consumes ribosomal capacity, ETFL models enforce $v_{synthesis,E_j} = \mu[E_j] + \text{degradation}$, constrained by a global ribosomal capacity $\sum_j \frac{v_{synthesis,E_j}}{k_{cat}^{ribo}} \leq [R_{total}]$.

These constraints create trade-offs that reflect real physiological limitations. Ahn-Horst et al. (2022) demonstrated this using an expanded *E. coli* model that linked proteome allocation to growth control. The COBRA Toolbox v3.0 Heirendt et al. (2019) provides a widely used platform for implementing such analyses.

2.3 Current Methodological Challenges: Parameter Identification and Computation

Despite their interpretability, mechanistic approaches face two persistent challenges.

The first is the **parameter identification problem**. RDME models require diffusion coefficients and reaction rate constants for thousands of molecular species, while ETFL and related models depend on accurate Gibbs energies, enzyme turnover numbers, and intracellular metabolite levels. Even in well-characterized organisms such as *E. coli*, a substantial fraction of these parameters is either poorly constrained or experimentally inaccessible.

The second challenge is **computational cost**. A single 105-minute simulation of a minimal cell required multi-GPU clusters and days of computation Thornburg et al. (2025). Simulations of eukaryotic systems scale even more steeply: the hybrid RDME–ODE model of the yeast galactose response required roughly six hours of GPU time to simulate one hour of signaling dynamics Wu et al. (2025). These costs motivate the search for complementary data-driven or hybrid approaches that can capture essential biological behavior without fully resolving every molecular event.

3 DATA-DRIVEN APPROACHES: LEARNING FROM HIGH-DIMENSIONAL OBSERVATION

Philosophy: "All models are wrong, but some are predictive." — George E. P. Box

Where mechanistic approaches attempt to reconstruct cellular behavior from explicitly defined biochemical rules, data-driven approaches treat the cell primarily as a statistical system. The central assumption is that high-dimensional single-cell datasets contain sufficient structure for models to infer regulatory relationships without enumerating underlying molecular mechanisms. This perspective has gained traction as single-cell transcriptomics, perturbation screens, and large cell atlases continue to expand in scale and diversity. Rather than deriving equations, these methods aim to learn the manifold on which cellular states lie and to model how these states change in response to perturbations.

3.1 The Foundation Model Approach: The Cell as a Sequence

A major development in this space is the emergence of **Single-Cell Foundation Models (scFMs)**, which borrow architectural ideas from natural language processing. In these models, a gene expression vector is treated analogously to a sequence, and the task of learning cellular organization parallels learning the statistical structure of language. Institutions such as the **Arc Institute** and the **CZI Biohub** have adopted this approach to move beyond cell-type annotation toward predicting cellular responses to perturbations.

Transformers and latent representations. Transformers form the backbone of many scFMs. Their self-attention mechanism computes relationships among all genes simultaneously, assigning context-dependent importance through the attention matrix:

$$A(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{4}$$

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote embedded representations of gene expression features. This formulation allows the model to recover long-range dependencies or co-regulation patterns that may be difficult to specify mechanistically.

Multi-head attention and biological structure. By running several attention mechanisms in parallel, the *multi-head* formulation enables the model to capture distinct biological subspaces:

$$\text{MultiHead}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Concat}(head_1,...,head_h)\mathbf{W}^O. \tag{5}$$

For example, different heads may attend to stoichiometrically linked complexes (e.g., ribosomal subunits) or to rapidly induced stress-response programs.

Training objective: Masked Gene Modeling (MGM). Inspired by masked language modeling, these models are trained to infer masked gene expression values:

$$\mathscr{L}_{MGM} = -\sum_{i\in M} x_i \log P(\hat{x}_i \mid x_{\notin M}), \tag{6}$$

where M denotes the masked subset. Over large datasets, the model internalizes the conditional structure of transcriptomes across tissues, perturbations, and cell states.

State-of-the-art example: State. The **State** model introduced by Adduri et al. (2025) extends foundation modeling to perturbation prediction. Rather than simply embedding cell states, State learns how perturbations map to transcriptional changes across tissues. A key property of the model is its **zero-shot generalization**: by leveraging a shared latent space z, it predicts responses in unseen tissue contexts,

$$z_{perturbed} = z_{control} + f_\theta(\delta, c), \tag{7}$$

where δ represents the perturbation and c denotes tissue context. In benchmark tests, State achieved correlation coefficients exceeding 0.8 when predicting perturbation responses in tissues not present during training. Such performance is not accessible to mechanistic simulations, which require explicit re-parameterization for each new context.

3.2 The Turing Test for Biology

The strong predictive capabilities of scFMs have prompted discussion about what constitutes a validated "virtual cell." Roohani et al. (2025) formalized this idea in the **Virtual Cell Challenge**, proposing a biological analogue of the Turing Test: a model is considered adequate if its predicted response distribution $P_{model}(y|x)$ to a perturbation x is statistically indistinguishable from the experimental distribution $P_{experiment}(y|x)$. This reframes validation from mechanistic fidelity to phenomenological accuracy.

Knowledge graph neural networks (GEARS). One approach toward interpretable perturbation modeling is **GEARS**, a knowledge graph–constrained GNN developed by Roohani and colleagues. GEARS incorporates curated biological relationships—such as protein–protein interactions and regulatory edges—into the model architecture. At each graph convolution layer,

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{W}_{self}^{(l)}\mathbf{h}_i^{(l)} + \sum_{j\in\mathcal{N}(i)} \frac{1}{\sqrt{c_{ij}}}\mathbf{W}_{neighbor}^{(l)}\mathbf{h}_j^{(l)}\right), \tag{8}$$

the hidden state $\mathbf{h}_i$ of gene i is updated by aggregating signals from its biological neighbors $\mathcal{N}(i)$. This architecture embeds an inductive bias that constrains predictions to follow known regulatory topology, enabling GEARS to extrapolate to **combinatorial perturbations** not present in the training data.

Visible Neural Networks (DrugCell). Similarly, **DrugCell** Kuenzi et al. (2020) implements a structured hierarchy in which genes, complexes, and pathways are explicitly encoded in the network architecture. This design supports biologically interpretable predictions by associating drug responses with specific subsystems.

3.3 Challenges: Interpretability and Physical Constraints

Data-driven approaches provide scalability and strong predictive performance, but they face limitations that do not arise in mechanistic modeling. The first major challenge is **Interpretability**. Deep foundation models, including State, typically do not provide mechanistic explanations for predicted changes in gene expression. While architectures like DrugCell introduce structural interpretability, many models still function as high-dimensional black boxes. This presents difficulties when predictions involve unobserved mechanisms—such as phosphorylation cascades—or when models are deployed in clinical settings.

The second challenge is **Physical consistency**. Empirical models operate in a statistical space that may include cellular states incompatible with thermodynamic or mass-conservation laws. For example, a generative model may predict transcriptomes whose implied ATP usage exceeds feasible cellular energy budgets. In metabolic applications, unconstrained models may generate flux distributions that violate the Second Law of Thermodynamics. In contrast, frameworks such as **ETFL** Salvy and Hatzimanikatis (2020) enforce feasibility through explicit biophysical constraints. Without similar safeguards, purely data-driven models may produce predictions that appear statistically plausible but are biologically impossible.

4 COMPARATIVE ANALYSIS: EVALUATING METHODOLOGICAL TRADE-OFFS

To assess how mechanistic and data-driven models contribute to whole-cell simulation, it is useful to compare their performance across concrete dimensions relevant to computational biology: grounding in physical laws, generalizability across conditions, interpretability, computational cost, and data requirements. These properties, rather than methodological preference, ultimately determine suitability for a given biological question. A summary is provided in Table 1.

Table 1. Comparison of mechanistic and data-driven modeling approaches.

Feature	Mechanistic (e.g., MinCell 4D)	AI / Data-Driven (e.g., scGPT)	Advantage
Ground Truth	**Chemical reality:** governed by conservation laws, explicit stoichiometry, and kinetics.	**Statistical structure:** learns latent manifolds and correlation patterns directly from data.	Mechanistic
Generalizability	**Limited:** typically organism- and context-specific; each new system requires bespoke parameterization.	**High:** foundation models enable zero-shot transfer across tissues, species, and perturbations.	AI
Interpretability	**High:** causal traceability from parameters to emergent phenotypes.	**Low:** internal representations reside in high-dimensional embeddings that are difficult to interpret.	Mechanistic
Speed	**Slow:** full-cell spatial simulations often require GPU-weeks.	**Fast:** inference occurs in milliseconds once the model is trained.	AI
Data Cost	**Deep:** depends on biophysical parameters such as k_{cat}, D, and ΔG.	**Broad:** requires large, diverse, and standardized perturbational datasets.	— (distinct trade-offs)

4.1 Key Differentiators

Causal validity. Mechanistic models retain an important advantage in causal interpretability. If a simulation produces incorrect behavior, the source of error can often be attributed to a specific rate constant, boundary condition, or structural assumption. This makes mechanistic modeling particularly valuable for hypothesis generation or for probing questions rooted in physical chemistry. In contrast, errors in AI models typically reflect dataset biases or shifts in the distribution of observed states. These failure modes are harder to diagnose and may not map cleanly onto biological mechanisms.

Scalability and reusability. Data-driven models excel in scalability. Increasing biological complexity—additional genes, cell types, or perturbations—does not require explicit specification of new reactions or parameters. The same architecture can often be fine-tuned across systems, as demonstrated by foundation models such as **Geneformer** Theodoris et al. (2023) and **scGPT** Cui et al. (2024). Mechanistic models, by contrast, require explicit definition and validation of every new component, making large-scale or multi-tissue extensions resource-intensive.

Data requirements. The two approaches impose different burdens. Mechanistic modeling requires deep biochemical characterization: kinetic rates, diffusion coefficients, binding affinities, and metabolite concentrations. Many of these parameters remain poorly measured in vivo. Data-driven approaches require broad, standardized perturbation datasets for training and validation. While single-cell atlases now provide large expression matrices, paired perturbation–response datasets remain comparatively scarce.

5 THE SOFTWARE ECOSYSTEM: INTEGRATORS AND SPECIALIZED TOOLS

Recent advances in whole-cell modeling have been supported by a diverse computational ecosystem. These tools fall broadly into two categories: integrators, which provide frameworks for combining

heterogeneous modeling approaches, and specialized solvers, which focus on high-fidelity simulation within a specific paradigm. Table 2 summarizes representative systems.

5.1 The Integrators: Bridging Frameworks

A notable development has been the rise of meta-frameworks that separate biological logic from numerical implementation. **Vivarium** exemplifies this philosophy by representing biological subsystems as modular "processes" that interact through shared data "stores" Agmon et al. (2022). This structure enables hybrid modeling—for example, coupling a flux balance calculation with a stochastic reaction–diffusion solver within a unified execution environment.

VCell offers a complementary vision, providing a platform where algebraic and differential representations of biological processes can be placed directly onto microscopy-derived geometries Blinov et al. (2017). This tight integration of imaging and modeling has proven valuable for problems involving realistic subcellular structures or membrane-associated dynamics.

5.2 Specialized Solvers and Foundation Models

For detailed mechanistic simulations, Lattice Microbes remains the leading GPU-accelerated platform for spatially resolved RDME modeling Roberts et al. (2013). In the metabolic domain, the COBRA Toolbox provides a mature ecosystem for genome-scale constraint-based analyses, including recent extensions incorporating thermodynamics and gene expression constraints Heirendt et al. (2019).

On the data-driven side, libraries such as scVI-tools have made deep generative models accessible to the broader single-cell community Lopez et al. (2018). More recently, genome-scale foundation models such as the Arc Institute's **Evo** architecture Nguyen et al. (2024) have begun to bridge sequence modeling and predictive genomics, suggesting new avenues for integrating empirical learning with mechanistic constraints.

Table 2. Computational Frameworks for Whole-Cell Modeling

Platform	Type	Key Features	Citation
Vivarium	Integrator	Modular hybrid modeling using Process Bigraph formalism.	Agmon et al. (2022)
VCell	Integrator	PDE/ODE solvers mapped onto experimental geometries.	Blinov et al. (2017)
Lattice Microbes	Mechanistic	GPU-accelerated reaction–diffusion simulations (RDME).	Roberts et al. (2013)
COBRA Toolbox	Mechanistic	Constraint-based FBA/ME/ETFL modeling at genome scale.	Heirendt et al. (2019)
scVI-tools	AI / Empirical	Variational autoencoders for single-cell transcriptomics.	Lopez et al. (2018)
Geneformer	AI / Empirical	Transformer-based transfer learning for cell-state modeling.	Theodoris et al. (2023)
Evo	AI / Empirical	Genome-scale sequence foundation model for design tasks.	Nguyen et al. (2024)

6 PROPOSAL: A UNIFIED MATHEMATICAL FRAMEWORK

We define a cell state S_t as a tuple $(\mathscr{M}_t, \mathscr{L}_t)$ composed of two distinct but coupled subspaces. The first, $\mathscr{M}_t \in \mathbb{R}^m$, represents the Mechanistic State, which comprises low-copy variables governed by explicit conservation laws, such as ribosomes (R), metabolites (c), and DNA. The second component, $\mathscr{L}_t \in \mathbb{R}^h$, is the Latent State, a high-dimensional embedding that captures aggregations of complex regulation, such as transcriptome embeddings, which are difficult to model explicitly.

The system evolves according to coupled differences:

$$\Delta\mathscr{M} = \underbrace{f_{phys}(\mathscr{M}_t, \theta_{kinetics})\Delta t}_{\text{Reaction-Diffusion}} + \underbrace{g_{decode}(\mathscr{L}_t)}_{\text{Regulation}} \tag{9}$$

$$\mathscr{L}_{t+1} = \text{Transformer}(\mathscr{L}_t, \text{Embed}(\mathscr{M}_t)) \tag{10}$$

Here, f_{phys} ensures thermodynamic feasibility, while the neural component $\mathscr{L}$ captures complex decision-making without explicit parameterization.

6.1 The "WCM-Hard" Benchmark Standards

To rigorously validate such hybrid models, we propose the WCM-Hard Benchmark Suite. Current validation often relies on ad-hoc phenotypic correlations, which masks underlying failures in mass conservation or dynamic regulation. To address this, WCM-Hard defines eight standardized tiers of increasing difficulty, each targeting a specific failure mode of existing mechanistic or data-driven approaches.

Tier 1: Conservation Laws (The Sanity Check). The most fundamental requirement for any physical simulation is the conservation of mass and energy. Purely data-driven generative models, such as diffusion or transformer-based single-cell avatars, frequently "hallucinate" matter, predicting cell states that are thermodynamically impossible. We define a rigorous pass metric ε_{drift}: the integrated violation of elemental mass (C, N, P) over a simulation of 10^5 time steps must satisfy $\int |\dot{M}_{total}| dt < \varepsilon_{drift}$, where ε_{drift} is negligible relative to cell mass. Failure here disqualifies a model as physically interpretable.

Tier 2: Virtual Essentiality Screen (The Structural Test). A model must correctly identify the minimal gene set required for life. This tests the structural validity of the metabolic and regulatory network graph. While statistical models can correlate expression with survival, they often fail to predict lethality for genes that are highly expressed but non-essential, or low-expressed but critical. The benchmark requires the model to simulate single-gene knockouts for the entire genome and achieve an F_1 score exceeding 0.90 against experimental CRISPRi essentiality datasets (e.g., Nichols et al., 2011).

Tier 3: Diauxic Shift Dynamics (The Regulation Test). Static constraint-based models (e.g., FBA) assume steady-state optimality and therefore fail to predict the lag phases associated with enzymatic re-tooling. The diauxic shift from glucose to lactose, characterized by a growth arrest while the *lac* operon is induced and translated, is the canonical failure case for purely steady-state assumptions. To pass this tier, a model must not only survive the shift but accurately predict the duration of the lag phase τ_{lag} within 10% of experimental values, proving it accounts for the resource and temporal cost of protein synthesis.

Tier 4: Out-of-Distribution Stress (The Generalization Test). Deep learning conceptualizes the cell as a manifold learned from training data. However, biological discovery often lies in the "out-of-distribution" (OOD) regime—conditions never seen during training. This benchmark evaluates the model's ability to predict transcriptional and growth responses to novel stressors (e.g., heat shock or pH stress) when trained only on nutrient shifts. Success is measured by the cosine similarity between the predicted and observed variable-load transcriptome vectors, requiring $\cos(\theta) > 0.8$ for unseen conditions.

Tier 5: Combinatorial Synergy (The Interaction Test). Complex phenotypes often emerge from non-linear interactions between subsystems. We require the model to predict the Loewe Additivity Index for pairs of antibiotics targeting distinct cellular processes (e.g., ribosomes vs. cell wall). Simple regression models assume independence, failing to capture synergy (Index < 1) or antagonism (Index > 1). A passing grade requires correctly classifying the interaction type for 95% of a standard pairwise drug library.

Tier 6: Population De-phasing (The Stochasticity Test). Cell biology is inherently stochastic. Deterministic ODE models predict that a synchronized population remains synchronized indefinitely, a physical impossibility known as the "clockwork cell" fallacy. This tier mandates that a model initiated with $N = 100$ synchronized cells must exhibit natural de-phasing variance over time. The Coefficient of Variation (CV) of cell volume must monotonically increase to the steady-state value (≈ 0.2) over 5 generations, validating the implementation of intrinsic molecular noise.

Tier 7: Ablation Verification (The Necessity Test). To ensure that complex hybrid architectures are justified, the model must demonstrate superior performance over its own ablated baselines. Specifically, the "Grey Box" model must significantly outperform a "Mechanistic-Only" baseline (lacking the neural controller) on dynamic tasks (Tier 3) and a "Data-Only" baseline (lacking physical constraints) on conservation tasks (Tier 1). This "double-dissociation" proves that both components are architecturally necessary.

Tier 8: Parameter Robustness (The Sensitivity Test). Finally, biological parameters are often noisy or loosely constrained. A useful model must not be "stiff"—it should not collapse under minor parameter variations. The Sensitivity analysis Tier requires that the phenotypic classification (e.g., essential/non-essential) remains stable across a $\pm 20\%$ perturbation of kinetic parameters (k_{cat}, K_m). This maps the "Trust Region" of the model, distinguishing robust biological predictions from overfitting artifacts.

6.2 Benchmark Algorithm

The evaluation pipeline is formalized in Algorithm 1, illustrating the sequential validation logic.

Algorithm 1 WCM-Hard Evaluation Protocol

Require: Virtual Cell Model M, Benchmark Suite S
Ensure: Certified Status $\in \{True, False\}$

$Certified \leftarrow True$
Tier 1: Conservation
if $\int |\dot{M}| dt > \varepsilon$ **then**
  return $False$
end if
Tier 2: Essentiality
$Scores \leftarrow []$
for $gene \in Genome$ **do**
  $M' \leftarrow M.knockout(gene)$
  $Scores.append(M'.growth)$
end for
if $F1(Scores, Experimental) < 0.90$ **then**
  return $False$
end if
Tier 3: Diauxic Shift
$M.environment \leftarrow Glucose$
$M.step(T_{switch})$
$M.environment \leftarrow Lactose$
if $LagError(M) > 10\%$ **then**
  return $False$
end if
... (Tiers 4-8) ...
if $Certified$ **then**
  PRINT "CERTIFIED VIRTUAL CELL"
end if

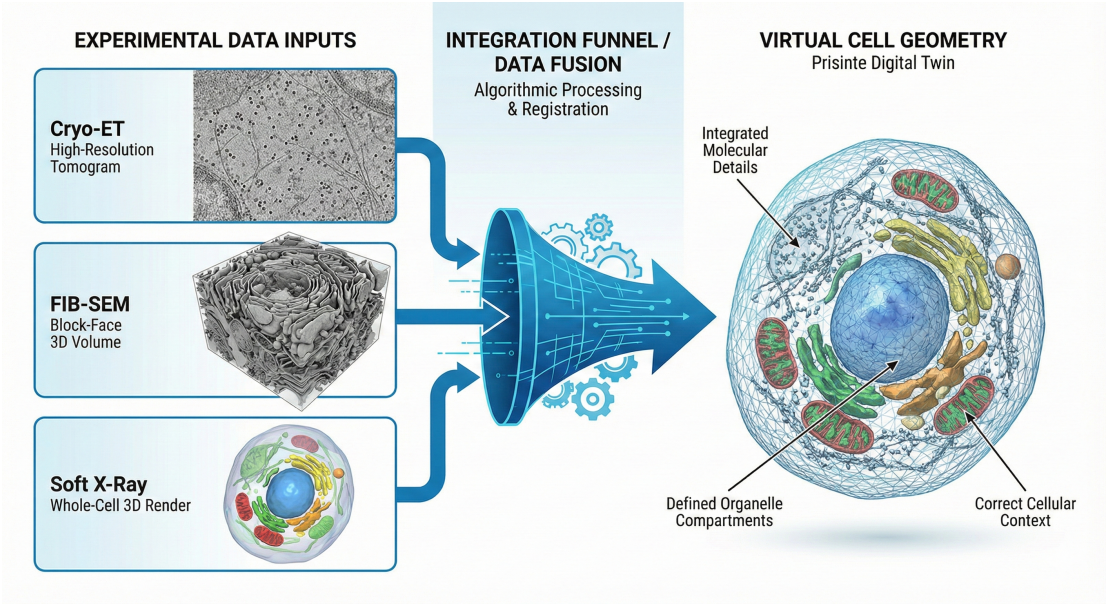

Figure 1. Constructing the Virtual Geometry from Multi-Modal Imaging. A pipeline for integrating diverse structural data into a unified spatial model. **(Inputs) FIB-SEM** provides the large-scale volumetric context and cell boundaries. **Soft X-Ray Tomography (SXT)** segments internal organelles (mitochondria, nucleus) in a near-native hydrated state. **Cryo-ET** provides molecular-resolution placement of macromolecular complexes (ribosomes, proteasomes). **(Output)** These modalities are fused to generate a comprehensive "lattice geometry" (Right) where every voxel is populated with molecular detail constrained by experimental organelle topology.

7 THE EMPIRICAL FOUNDATION: MULTI-SCALE EXPERIMENTAL DATA REQUIREMENTS

The construction of a bona fide virtual cell is not merely a computational challenge but a data integration problem of vast proportions. A rigorous model requires experimental parameterization across temporal and spatial scales that span orders of magnitude—from the femtosecond vibrations of atomic bonds to the hour-long dynamics of the cell cycle, and from the Ångström resolution of protein structures to the micron-scale architecture of organelles. This section delineates the critical experimental technologies required to feed the "Glass Box" and "Grey Box" models of the future. The specific contributions of these modalities are summarized in Table 3.

7.1 Defining the Geometry: Structural Scaffolds

Before a cell can be simulated, its physical boundaries and internal compartmentalization must be explicitly defined. In our framework, this spatial scaffold is constructed by integrating complementary imaging modalities across length scales (Figure 1). Focused Ion Beam Scanning Electron Microscopy (FIB-SEM) provides the large-scale volumetric context, defining whole-cell geometry and membrane boundaries with nanometer precision Xu et al. (2017). Within this volume, Soft X-Ray Tomography (SXT) resolves the mesoscale organization of major organelles in a near-native hydrated state, enabling quantitative segmentation of structures such as the nucleus and mitochondria without chemical fixation Larabell and Le Gros (2004). At the molecular scale, Cryo-Electron Tomography (Cryo-ET) serves as the gold standard for mapping the *in situ* molecular sociology of the cell, providing nanometer-resolution information on the spatial distribution of ribosomes and other macromolecular assemblies Baumeister (2022). These multi-modal datasets are fused to generate a unified lattice geometry in which each voxel is assigned molecular content consistent with experimentally constrained organelle topology, enabling spatially resolved whole-cell simulations.

7.2 Capturing Dynamics: 4D Spatiotemporal Tracking

Static images must be animated by dynamic data to validate simulation trajectories. To validate the predicted motion of organelles and protein complexes, high-speed volumetric imaging is required. Lattice Light-Sheet Microscopy (LLSM) allows for the continuous 4D tracking of cellular components with minimal phototoxicity, satisfying the need for ground truth in validating diffusive and advective transport components Chen et al. (2014). At the single-molecule level, Super-Resolution Microscopy (e.g., STORM/PALM) provides coordinates with nanometer precision, overcoming the diffraction limit to validate the spatial clustering predicted by reaction-diffusion models.

7.3 Parameterizing Kinetics: The Atomic and Molecular Scale

The accuracy of any kinetic model is inextricably linked to the quality of its rate constants (k_{cat}, k_{on}, k_{off}). Bulk biochemical assays often obscure the heterogeneity of molecular states. Single-Molecule FRET (smFRET) allows for the direct observation of conformational changes and interaction kinetics of individual molecules, critical for parameterizing stochastic models where rare events determine phenotypic outcomes Ha et al. (1996). When experimental kinetic parameters are inaccessible, Molecular Dynamics (MD) Simulations serve as a computational microscope, calculating diffusion coefficients and binding free energies *ab initio* to fill the "dark matter" of parameter space Stevens et al. (2023).

7.4 Contextualizing the System: Spatial Omics and Interaction Networks

Finally, the inventory of molecular actors and their spatial context must be established. Spatial Transcriptomics (e.g., MERFISH) maps the subcellular localization of RNA species, providing essential boundary conditions for simulations of translation and transport. Concurrently, Quantitative Proteomics and Metabolomics provide absolute copy numbers and concentrations, setting the initial conditions for differential equation models. To simulate gene regulation realistically, Chromosome Conformation Capture (Hi-C) reveals the 3D topology of the genome, constraining simulations of transcriptional accessibility Lieberman-Aiden et al. (2009).

Table 3. Experimental Data Requirements for Whole-Cell Modeling

Technology	Resolution / Scale	Data Type	Role in Virtual Cell
Cryo-ET	Nanometer / Molecular	3D Tomograms	Places individual complexes (ribosomes) in situ.
SXT / FIB-SEM	Mesoscale / Cellular	3D Volumetric Maps	Defines organelle boundaries and cell geometry.
4D Light Sheet	Micron / Seconds	Spatiotemporal Trajectories	Validates organelle movement and transport.
smFRET	Angstrom / Milliseconds	Kinetic Rates (k_{on},k_{off})	Parameterizes stochastic switching and binding.
Spatial Omics	Sub-cellular / Transcriptome	RNA/Protein Localization	Sets initial spatial distribution of species.
Hi-C	Genomic Loci	Interaction Probabilities	Constrains 3D genome folding and accessibility.

8 RESOLUTION: THE "GREY BOX" INTEGRATION

The apparent divide between mechanistic and data-driven modeling may represent a transitional stage rather than a stable dichotomy. Historically, fields such as fluid dynamics and climate science progressed from purely first-principles formulations to hybrid models in which empirical components informed or accelerated mechanistic solvers. A similar trajectory is emerging in whole-cell modeling. Fully mechanistic reconstructions remain limited by computational cost and incomplete biochemical knowledge, while purely empirical models require biophysical constraints to avoid generating implausible or clinically unsafe predictions. Increasingly, these approaches appear complementary rather than competing, giving rise to what can be described as a "Grey Box" architecture—one in which learning algorithms and physical laws jointly define the model's behavior.

8.1 AI-Enhanced Parameterization

One area where data-driven methods have become particularly valuable is the estimation of parameters required by mechanistic models. Detailed ODE or RDME simulations depend on kinetic constants, binding affinities, and diffusion coefficients, many of which remain experimentally unmeasured. Machine learning approaches trained on biochemical databases such as BRENDA and SABIO-RK have begun to close this gap. Models like **DLKcat** Qiu et al. (2023) and **TurNuP** Kroll et al. (2023) use graph neural networks and structural features to predict k_{cat} values with reasonable accuracy. Although these predictions carry uncertainty, they provide practical estimates where no experimental measurements exist, enabling mechanistic models to be assembled or extended without prohibitive experimental investment.

8.2 Physics-Informed Machine Learning

A complementary direction integrates physical constraints directly into the training of neural networks. Rather than relying solely on data-based loss functions, *physics-informed machine learning* introduces penalties that enforce thermodynamic or mass-balance conditions. The total loss function typically takes the form

$$L_{total} = L_{data} + \lambda_{thermo} \left\| \Delta G_{net} \right\|^2 + \lambda_{mass} \left\| \frac{dM}{dt} \right\|^2, \tag{11}$$

where the additional terms discourage the model from generating states that violate biochemical feasibility. In metabolic applications, this approach effectively embeds the logic of the **ETFL** framework Salvy and Hatzimanikatis (2020) within the neural architecture, ensuring that generative or predictive outputs remain within the realm of thermodynamically permissible states. Such methods provide a mechanism to merge the flexibility of deep learning with the groundedness of physical modeling.

8.3 Hybrid Engines: Surrogate Models and Neural Operators

Another avenue for integration lies in accelerating mechanistic simulations using learned surrogates. High-resolution RDME simulations often involve computationally expensive diffusion or reaction updates, particularly when performed over realistic geometries or long biological timescales. Surrogate models—such as neural operators trained on high-fidelity simulation snapshots—can approximate these dynamics at substantially reduced cost. Within composite frameworks like **Vivarium** Agmon et al. (2022), these surrogates may be invoked for selected components of the system while critical, low-copy-number processes remain governed by explicit mechanistic solvers.

This hybrid approach offers a substantial reduction in runtime, enabling simulations that would otherwise be computationally infeasible. However, careful validation is essential. Surrogate models may overlook rare or stochastic events—such as sharp transitions in signaling pathways—that full RDME simulations capture. As a result, hybrid engines must balance computational efficiency with fidelity, using empirical approximations where they provide speed without compromising key mechanistic phenomena.

9 CONCLUSION

The current divide between mechanistic and data-driven approaches in whole-cell modeling should be viewed as an expected stage in the evolution of a rapidly expanding field. Mechanistic models provide rigor by grounding predictions in spatial organization, kinetic laws, and thermodynamic feasibility. Their strength lies in offering clear causal explanations for cellular behavior—an essential requirement when the goal is to understand how molecular processes give rise to phenotype. Data-driven models, by contrast, excel at capturing broad statistical regularities across diverse tissues and perturbations, enabling prediction in settings where mechanistic detail is incomplete or unavailable.

These approaches are often presented as competing paradigms, but their complementary strengths suggest a path toward integration. Mechanistic models struggle with scalability and incomplete parameterization, whereas purely empirical models face challenges of interpretability and physical validity. A unified framework will likely draw on both: learning algorithms for efficient inference across large state spaces, and physical constraints to ensure that predictions remain biologically plausible.

9.1 The Emerging Interface: Middleware and Standardization

Achieving such integration will require not only conceptual alignment but also advances in software infrastructure. Middleware systems—such as **BioSimulators** or the **Process Bigraph** framework implemented in Vivarium—provide mechanisms for heterogeneous solvers to exchange state variables through standardized interfaces. These developments allow reaction–diffusion solvers, metabolic models, and neural networks to operate within shared computational environments, making it feasible to assemble composite virtual cells without manually coordinating incompatible codebases.

9.2 Remaining Challenges

Several barriers remain before these hybrid architectures can support general-purpose virtual cell simulations. One challenge is **data sparsity in rare or transitional states**. While foundation models perform well on abundant phenotypes, many biologically important trajectories—such as early disease transitions or low-frequency differentiation events—are poorly represented in existing datasets. Another challenge is the need for **auditability and safety** when deploying AI-derived predictions in biomedical decision-making. Integrated models will require mechanisms for tracing predictions back to either underlying physical rules or interpretable features in the data.

Looking ahead, the virtual cell is unlikely to take the form of a single monolithic model. A more plausible architecture resembles a layered or federated system, in which high-capacity foundation models propose candidate hypotheses across many contexts, while mechanistic submodels evaluate these hypotheses under thermodynamic, stoichiometric, or spatial constraints. Such a division of labor offers a practical path toward systems that combine the scale of empirical modeling with the explanatory depth of first-principles approaches. The central task for the next decade will be to engineer this interface—ensuring that learning-based acceleration does not come at the expense of biological fidelity.

REFERENCES

Adduri, A. K., Gautam, D., Bevilacqua, B., Imran, A., Shah, R., Naghipourfar, M., Teyssier, N., Ilango, R., Nagaraj, S., Dong, M., Ricci-Tam, C., Carpenter, C., Subramanyam, V., Winters, A., Tirukkovular, S., Sullivan, J., Plosky, B. S., Eraslan, B., Youngblut, N. D., Leskovec, J., Gilbert, L. A., Konermann, S., Hsu, P. D., Dobin, A., Burke, D. P., Goodarzi, H., and Roohani, Y. H. (2025). Predicting cellular responses to perturbation across diverse contexts with State. ISSN: 2692-8205 Pages: 2025.06.26.661135 Section: New Results.

Agmon, E., Spangler, R. K., Skalnik, C. J., Poole, W., Peirce, S. M., Morrison, J. H., and Covert, M. W. (2022). Vivarium: an interface and engine for integrative multiscale modeling in computational biology. *Bioinformatics (Oxford, England)*, 38(7):1972–1979.

Ahn-Horst, T. A., Mille, L. S., Sun, G., Morrison, J. H., and Covert, M. W. (2022). An expanded whole-cell model of E. coli links cellular physiology with mechanisms of growth rate control. *npj Systems Biology and Applications*, 8(1):30.

Baumeister, W. (2022). Cryo-electron tomography: A long journey to the inner space of cells. *Cell*, 185(15):2649–2652.

Blinov, M. L., Schaff, J. C., Vasilescu, D., Moraru, I. I., Bloom, J. E., and Loew, L. M. (2017). Compartmental and Spatial Rule-Based Modeling with Virtual Cell. *Biophysical Journal*, 113(7):1365–1372.

Chen, B.-C., Legant, W. R., Wang, K., Shao, L., Milkie, D. E., Davidson, M. W., Janetopoulos, C., Wu, X. S., Hammer, J. A., Liu, Z., English, B. P., Mimori-Kiyosue, Y., Romero, D. P., Ritter, A. T., Lippincott-Schwartz, J., Fritz-Laylin, L., Mullins, R. D., Mitchell, D. M., Bembenek, J. N., Reymann, A.-C., Böhme, R., Grill, S. W., Wang, J. T., Seydoux, G., Tulu, U. S., Kiehart, D. P., and Betzig, E. (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, 346(6208):1257998.

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods*, 21(8):1470–1480.

Gilbert, B. R., Thornburg, Z. R., Brier, T. A., Stevens, J. A., Grünewald, F., Stone, J. E., Marrink, S. J., and Luthey-Schulten, Z. (2023). Dynamics of chromosome organization in a minimal bacterial cell. *Frontiers in Cell and Developmental Biology*, 11:1214962.

Ha, T., Enderle, T., Ogletree, D. F., Chemla, D. S., Selvin, P. R., and Weiss, S. (1996). Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences*, 93(13):6264–6268.

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D. C., Valcarcel, L. V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H. M., Ma, D., Sun, Y., Wang, L., Yurkovich, J. T., Oliveira, M. A. P., Vuong, P. T., El Assal, L. P., Kuperstein, I., Zinovyev, A., Hinton, H. S., Bryant, W. A., Aragón Artacho, F. J., Planes, F. J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M. A., Maranas, C. D., Lewis, N. E., Sauter, T., Palsson, B. ø., Thiele, I., and Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3):639–702. Publisher: Nature Publishing Group.

Karr, J., Sanghvi, J., Macklin, D., Gutschow, M., Jacobs, J., Bolival, B., Assad-Garcia, N., Glass, J., and Covert, M. (2012). A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, 150(2):389–401.

Kroll, A., Rousset, Y., Hu, X.-P., Liebrand, N. A., and Lercher, M. J. (2023). Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nature Communications*, 14(1):4139.

Kuenzi, B. M., Park, J., Fong, S. H., Sanchez, K. S., Lee, J., Kreisberg, J. F., Ma, J., and Ideker, T. (2020). Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell*, 38(5):672–684.e6.

Larabell, C. A. and Le Gros, M. A. (2004). X-ray Tomography Generates 3-D Reconstructions of the Yeast, *Saccharomyces cerevisiae* , at 60-nm Resolution. *Molecular Biology of the Cell*, 15(3):957–962.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J. (2009). Comprehensive

Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(5950):289–293.

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058.

Macklin, D. N., Ahn-Horst, T. A., Choi, H., Ruggero, N. A., Carrera, J., Mason, J. C., Sun, G., Agmon, E., DeFelice, M. M., Maayan, I., Lane, K., Spangler, R. K., Gillies, T. E., Paull, M. L., Akhter, S., Bray, S. R., Weaver, D. S., Keseler, I. M., Karp, P. D., Morrison, J. H., and Covert, M. W. (2020). Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. *Science*, 369(6502):eaav3751. Publisher: American Association for the Advancement of Science.

Nguyen, E., Poli, M., Durrant, M. G., Kang, B., Katrekar, D., Li, D. B., Bartie, L. J., Thomas, A. W., King, S. H., Brixi, G., Sullivan, J., Ng, M. Y., Lewis, A., Lou, A., Ermon, S., Baccus, S. A., Hernandez-Boussard, T., Ré, C., Hsu, P. D., and Hie, B. L. (2024). Sequence modeling and design from molecular to genome scale with Evo. *Science (New York, N.Y.)*, 386(6723):eado9336.

Qiu, S., Zhao, S., and Yang, A. (2023). DLTKcat: deep learning-based prediction of temperature-dependent enzyme turnover rates. *Briefings in Bioinformatics*, 25(1):bbad506.

Raman, K. and Chandra, N. (2009). Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, 10(4):435–449.

Roberts, E., Stone, J. E., and Luthey-Schulten, Z. (2013). Lattice Microbes: high-performance stochastic simulation method for the reaction-diffusion master equation. *Journal of Computational Chemistry*, 34(3):245–255.

Roohani, Y. H., Hua, T. J., Tung, P.-Y., Bounds, L. R., Yu, F. B., Dobin, A., Teyssier, N., Adduri, A., Woodrow, A., Plosky, B. S., Mehta, R., Hsu, B., Sullivan, J., Ricci-Tam, C., Li, N., Kazaks, J., Gilbert, L. A., Konermann, S., Hsu, P. D., Goodarzi, H., and Burke, D. P. (2025). Virtual cell challenge: Toward a turing test for the virtual cell. *Cell*, 188(13):3370–3374. Publisher: Elsevier.

Salvy, P. and Hatzimanikatis, V. (2020). The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nature Communications*, 11(1):30. Publisher: Nature Publishing Group.

Stevens, J. A., Grünewald, F., van Tilburg, P. A. M., König, M., Gilbert, B. R., Brier, T. A., Thornburg, Z. R., Luthey-Schulten, Z., and Marrink, S. J. (2023). Molecular dynamics simulation of an entire cell. *Frontiers in Chemistry*, 11. Publisher: Frontiers.

Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., Mantineo, H., Brydon, E. M., Zeng, Z., Liu, X. S., and Ellinor, P. T. (2023). Transfer learning enables predictions in network biology. *Nature*, 618(7965):616–624.

Thornburg, Z. R., Maytin, A., Kwon, J., Brier, T. A., Gilbert, B. R., Fu, E., Gao, Y.-L., Quenneville, J., Wu, T., Li, H., Long, T., Pezeshkian, W., Sun, L., Glass, J. I., Mehta, A., Ha, T., and Luthey-Schulten, Z. (2025). Bringing the Genetically Minimal Cell to Life on a Computer in 4D.

Wu, T., Spindler, M.-C., Earnest, E., Li, H., Thornburg, Z. R., Mahamid, J., and Luthey-Schulten, Z. (2025). Spatial Heterogeneity Alters the Dynamics of the Yeast Galactose Switch: Insights from 4D RDME–ODE Hybrid Simulations. *bioRxiv*, page 2025.07.23.666409.

Xu, C. S., Hayworth, K. J., Lu, Z., Grob, P., Hassan, A. M., García-Cerdán, J. G., Niyogi, K. K., Nogales, E., Weinberg, R. J., and Hess, H. F. (2017). Enhanced FIB-SEM systems for large-volume 3D imaging. *eLife*, 6:e25916. Publisher: eLife Sciences Publications, Ltd.