Cross-kingdom virtual cells as an Artificial Intellegence-based architecture for predictive biological modeling: animal, plant and microbial systems

Zhasmin Abdilda[1], Aisha Yergali[2], Zhaniya Baizhan[3], Yermekova Alua[4], Aisha Zharkynkyzy[5], Mariam Tutayeva[6], Aida Akhmetova[7], Merey Dauytkali[8], Yeraly Gainulla[9], Miras Zharas[10]

[1]Nazarbayev Intellectual School of Physics and Mathematics, Almaty 050040, Kazakhstan

[2]Lyceum #163, Almaty 050000, Kazakhstan

[3]Taraz Grammar school #40, Taraz 080000, Kazakhstan

[4]Nurorda School, Almaty 050038, Kazakhstan

[5]Nazarbayev Intellectual School of Physics and Mathematics, Taldykorgan 040000, Kazakhstan

[6]Nazarbayev Intellectual School of Astana, Astana 010000, Kazakhstan

[7]Department of Mathematics, Nazarbayev University, Astana 010000, Kazakhstan

[8]Nazarbayev Intellectual School of Chemistry and Biology in Almaty, Almaty 050006, Kazakhstan

[9]Aqbobek International School, Aktobe 030000, Kazakhstan

[10]Salisbury School, 251 Canaan Rd, Salisbury, CT 06068, United States of America

Supervision and Correspondence: Professor Yingqiu Xie[a]

[a]Department of Biology, School of Sciences and Humanities, Nazarbayev University, Astana 010000, Kazakhstan

Email

[1]abdilda_zh0804@fmalm.nis.edu.kz

[2]yergaliaisha2008@gmail.com

[3]zhaniya.baizhan@gmail.com

[4]alua_yermekova@mail.ru

[5]aishazharqynqyzy@gmail.com

[6]mariamtutayeva@gmail.com

[7]aidaakhmetovaxx@gmail.com

[8]dmerey08@mail.ru

[9]eralygainulla@gmail.com

[10]maratuly.miras.zharas@gmail.com

[a]yingqui.xie@nu.edu.kz

Introduction

Problem of traditional drug testing

Drug development remains a high-risk, high-cost endeavor, with attrition rates in clinical trials persistently exceeding 85-90 % across therapeutic areas [7]. In oncology, success rates can be as low as 3-5 %, reflecting poor translatability of preclinical results to human outcomes [7]. Conventional animal models, though

historically indispensable, frequently fail to replicate human physiology and disease complexity, leading to inaccurate predictions of efficacy and safety [5]. These translational gaps contribute to 12-15 years of development timelines and costs averaging $2-2.5 billion per approved drug [7]. In addition, regulatory, ethical, and reproducibility concerns are driving the search for alternative models that better capture human biology without relying heavily on in vivo animal studies.

Potential of virtual organism models

Advances in induced pluripotent stem cells (iPSC) and organoid technologies are redefining the preclinical landscape. Human iPSC-derived organoids can self-organize into physiologically relevant 3D architectures that mimic native tissue functions, enabling modeling of organ-level drug responses [6]. This capability supports personalized medicine approaches where patient-specific cells are used to predict individual drug responses. Parallel developments in AI-driven virtual modeling, including multi-omics integration, spatial transcriptomics alignment [4], and *in silico* patient simulations, are accelerating the design-build-test-learn (DBTL) cycle for therapeutic discovery [2].

Importantly, regulatory shifts such as the FDA Modernization Act 2.0 recognize “new approach methods” (NAMs) as viable alternatives to animal testing, opening a pathway for AI-enhanced virtual models to play a central role in preclinical decision-making [7].

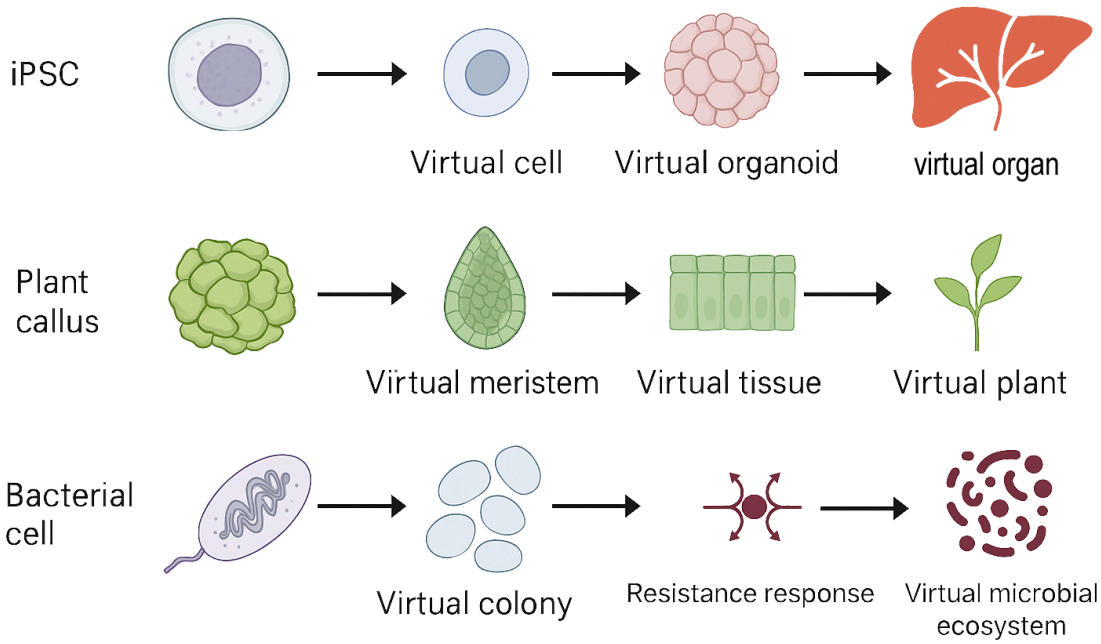
From Animal to Non-Animal: Redefining the Stem Cell Analogy

While iPSC and embryonic stem cell (ESC) systems have revolutionized modeling in human and animal biology, analogous regenerative and self-organizing systems exist in other kingdoms. This work proposes extending the virtual cell framework, which was traditionally centered on animal stem cells, to include plant callus cells and bacterial colonies as functionally analogous platforms.

Animal iPSC: capabilities and pipeline

iPSCs are pluripotent, capable of differentiating into virtually any somatic lineage. In vitro, they can be directed toward specific tissues or combined into organoids that replicate aspects of organ physiology [6]. These models support drug toxicity screening, disease modeling, and personalized therapy prediction. Integration with AI, spatial transcriptomics, and predictive modeling tools [4] enables simulation of developmental trajectories and response profiling at single-cell resolution.

Why this analogy matters for digital modeling



This cross-kingdom perspective is not merely theoretical. In human health, AI-enhanced modeling of iPSC directly addresses the reproducibility crisis in preclinical research and reduces reliance on animal testing. In infectious disease management, AI-driven methods such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and gradient-boosted decision trees already enable rapid pathogen detection, resistance prediction, and antibiotic treatment optimization, cutting inappropriate prescriptions by up to 67 % in urinary tract infection cases [1].

Plant callus cells as functional analogs

In agriculture and environmental biotechnology, plant virtual cells could transform agrochemical safety testing. Currently, greenhouse and field trials are expensive and

slow, and many candidate compounds fail late in development, defined as a pattern analogous to drug discovery bottlenecks [2]. AI-guided virtual plant models could simulate hormone-driven developmental responses (e.g., auxin/cytokinin balance) to predict pesticide efficacy and toxicity before costly real-world trials.

Plant callus cells are totipotent: any differentiated plant cell can revert to a pluripotent-like state and regenerate an entire organism under the influence of phytohormones such as auxin and cytokinin. This mirrors the role of morphogen signaling (e.g., BMP4, Activin A) in iPSC reprogramming. As in animal systems, callus formation follows distinct molecular pathways that can be predicted, manipulated, and optimized, making them viable candidates for AI-driven virtual plant cell modeling. In sum, such models could accelerate agrochemical testing by forecasting toxicity and growth effects prior to greenhouse or field trials.

Bacterial cells as self-organizing, programmable systems

In microbiology, bacterial virtual colonies could function as predictive testbeds for antibiotic resistance evolution and biofilm-associated treatment failure — problems that conventional in vitro assays poorly capture. Although bacteria lack traditional multicellular differentiation, they demonstrate remarkable programmability. Through engineered gene circuits, phase variation, and biofilm formation, bacterial colonies can display coordinated behaviors with emergent properties. Biofilms, in particular, are clinically significant responsible for up to 65 % of human infections and displaying antibiotic resistance 10-1000 times greater than planktonic cells [1]. AI-powered genomic and phenotypic prediction pipelines already achieve > 90 %

accuracy in resistance classification [2] and resistance for pathogens like *Staphylococcus aureus* and *Pseudomonas aeruginosa* [2], highlighting the readiness of bacterial systems for integration into a virtual colony modeling framework.

Therefore, by extending the virtual cell concept beyond animal paradigms, we can address multiple high-stakes bottlenecks across health, agriculture, and biotechnology.

Limitations of existing models and the novelty of this work

These challenges extend beyond human therapeutics. Agrochemical development faces a severe efficiency problem: large numbers of pesticide candidates fail in late-stage field testing due to unforeseen crop toxicity or insufficient pest control [2]. The lack of predictive, mechanistically accurate in silico plant models prolongs time-to-market and increases development costs.

In microbiology, antimicrobial resistance (AMR) continues to outpace the development of new antibiotics. Globally, AMR was directly responsible for 1.27 million deaths and associated with 4.95 million deaths in 2019, and is projected to cause up to 10 million annual deaths by 2050 without intervention [1]. In addition, up to 65 % of all human infections involve biofilms, which drastically reduce antibiotic efficacy [1].

AI models have demonstrated tangible clinical utility, for example, real-time prescription optimization in sepsis that aligns with clinical practice, and

genomic-driven resistance prediction with > 90 % accuracy [2]. However, these tools remain siloed from plant and bacterial virtual modeling pipelines.

Our proposed framework unifies these disparate efforts into a cross-kingdom AI-based virtual cell architecture, enabling reproducible, scalable, and mechanistically grounded predictive modeling across biomedical, agrochemical, and environmental domains.

Background

Today, biological modeling has increasingly relied on artificial intelligence tools that can analyze complicated data and predict cellular behavior[8]. However, well before AI, scientists used rule-based simulators grounded in mechanistic principles, wherein a user would define particular biological rules, such as chemical reactions, diffusion, or physical constraints, and the system would simulate the resulting dynamics[9].The use of these transparent and interpretable tools facilitates hypothesis testing, generates detailed, quantitative predictions, and allows exploration of all "what-if" scenarios[10]. On the other hand, their obvious disadvantage is that the approach requires clearly designed input parameters and a very good understanding of the biology of the system under consideration[9].They may also have difficulty scaling in large, data-rich environments that AI-based models excel in, learning patterns directly from the data[11]. To illustrate the strength and weaknesses of this methodology, the paper will focus on three commonly used platforms: Virtual Cell

(VCell), PhysiCell, and CompuCell3D. They embody an almost complete spectrum of ways to model cells and tissues-from molecular-level[12].

The Virtual Cell (VCell) platform serves perfectly in simulating the impact of spatial organization on intracellular processes[14]. Users can specify reaction networks within biological compartments, say the cytoplasm, nucleus, or extracellular space[14]. These configurations are then automatically translated by the system into mathematical models using ordinary or partial differential equations (ODEs/PDEs). With options for deterministic and stochastic solvers, and several inbuilt geometries of cells for 2D/3D modeling (one can import cell geometries from microscopy), VCell facilitates spatially-resolved simulations aligned with imaging data and structurally analyzable with regard to spatial effects on signal transduction and regulatory dynamics solvers[14]. One of the unique capabilities of VCell is integrated rule-based modeling, where the user is allowed to specify reaction rules (using BioNetGen language-BNGL) without needing to list every reaction, thus enabling network-free simulation of combinatorially large systems. VCell, in fact, incorporates the NFSim network-free simulator to accommodate large rule-based models with a high number of potential species configurations[15].To summarize, VCell is designed to span multiple modeling formalisms: deterministic, stochastic, spatial, non-spatial, and rule-based, all accessible in a single platform[12].

Inside, VCell uses a tiered approach: a BioModel defines the species, compartments, reactions (or rules), while specific Applications inside the BioModel define the simulation type (ODE, PDE, deterministic, or stochastic), geometry, and solver settings[14]. From the biological model definition, VCell automatically

generates the necessary math equations (such as systems of ODEs or PDEs), freeing users from the task of manual derivation of equations. It shares with the user a graphical user interface and database-backed environment in which one builds models by drawing compartments and reactions; the math that bog-they-minds may be inspected and even directly edited[14].

An open-source, physics-based framework for simulating multicellular systems in both 2-D and 3-D environments is PhysiCell[17]. It is designed to scale computationally and efficiently; thus, simulations of 10^4-10^6 interacting cells and dozens of diffusing biochemical substrates can run smoothly on a desktop, high-performance cluster, or cloud infrastructure (PhysiCell, n.d.). PhysiCell synergizes tightly with BioFVM: a parallelized solver for swiftly computing the diffusion, decay, and uptake of multiple substrates (such as oxygen, glucose, drugs, signaling molecules) that take place discreetly in 3D space-however, the direct coupling of extracellular gradients with cellular responses also happens-migration, growth arrest, or apoptosis[19]. Whereas lattice models force cells into fixed positions on the lattice, PhysiCell operates off-lattice, with cells assigned continuous spatial coordinates as autonomous agents. These positions and mechanical interactions vary dynamically through physical laws, rendering discretization artifacts largely moot. PhysiCell includes some prebuilt agents for major cellular mechanics and behavior submodels that are biologically realistic[17].

Each cell is internally represented by a hierarchical Phenotype data structure, which stores variables like cell cycle phase, death state, volume, motility parameters, and secretion rates. Instead of manually coding these behaviors, users define

environmental dependencies (e.g., "low oxygen increases necrosis") that dynamically update phenotype variables during simulation[20].This allows users to focus on modeling how microenvironmental conditions influence standard cellular behaviors, rather than re-implementing those behaviors themselves[20].PhysiCell also supports a human-readable, rule-based modeling language, which allows users to specify such logical or quantitative rules as “TGF-β reduces motility” or “oxygen below 5 mmHg increases apoptosis” that are automatically compiled into executable code at runtime [17].This promotes reproducibility and interpretability by reducing the need for low-level programming and enabling annotated model components.

The framework is implemented in standard C++ with minimal external dependencies, making it portable across platforms and easy to maintain [20]. It supports multithreaded execution via OpenMP, allowing simulations to scale linearly with the number of cells. PhysiCell’s modular design enables seamless integration with other modeling tools: for example, Boolean regulatory networks via PhysiBoSS, extracellular matrix modeling via PhysiMeSS, and intracellular signaling pathways using ODE solvers like libRoadRunner. These features make it suitable for multi-scale, multi-physics modeling across diverse biological systems[20].

PhysiCell is developed by a collaborative community of researchers, with continuous improvements enhancing its biological realism, computational efficiency, and interoperability with emerging modeling standardshttps[17]. While the platform was initially designed for cancer modeling, its architecture is general-purpose and has since been adapted to simulate a wide variety of biological processes[17]. Its

diffusion solvers and phenotype modules are broadly applicable, enabling its use in fields such as tissue engineering, angiogenesis, microbial ecology, and immune system dynamics.

Thanks to its modular architecture and core functionality, PhysiCell allows users to develop custom libraries for simulating physiological systems similar to how Microvessel Chaste was built upon the Chaste simulator, which is also based on biophysical principles [12]. Furthermore, users can define new types of substrates (e.g., extracellular matrix with zero diffusion), add custom cell types (e.g., fibroblasts with high motility and matrix-modifying behavior), and construct entire cellular systems, such as vascular networks that secrete oxygen and respond to gradients of angiogenic growth factors [21].

CompuCell3D is an intuitive and flexible modeling environment designed for building in silico virtual tissue simulations without requiring extensive programming expertise. Its scriptable architecture allows for rapid development and sharing of models across a wide range of multiscale, multicellular biological problems [22]. Written in C++ and equipped with Python bindings for model and simulation development, CompuCell3D uses the Cellular Potts Model (CPM) to simulate cell behavior, including shape changes, adhesion, and movement within tissues[23].

Application areas of CompuCell3D include angiogenesis, bacterial colony growth, cancer, developmental biology, tissue engineering, immune responses, evolutionary mechanisms, toxicology, and even modeling of non-cellular soft materials[22]. These

domains significantly overlap with those of PhysiCell, particularly in modeling cell–cell interactions and tissue behavior under dynamic microenvironmental conditions[12].

CompuCell3D models are configured through XML files (CC3DML), where CPM and PDE settings are defined, and Python scripts known as *stepables* execute custom logic during simulations. Low-level calculations are optimized in C++ for speed, while high-level behaviors are accessible through a rich Python API[23]. Plugins further expand functionality, enabling processes such as chemotaxis, external force application, cell elongation, and adhesion manipulation[22].

scVI

Single-cell Variational Inference, or scVI, is a deep generative model that uses a variational autoencoder (VAE) on single-cell RNA-seq count data [28]. It is trained using a probabilistic model (usually a negative-binomial likelihood) and a KL-divergence term in the VAE loss on the raw gene expression count matrix (raw UMI counts, not log-transformed). In actual use, scVI is frequently configured by first registering the raw counts in an AnnData object (scvi-tools), after which a low-dimensional latent embedding is learned that allows for the modeling and "denoising" of technical variation (batch, library size). Differential expression, clustering, and dimensionality reduction can all be done with the learned latent space. When it came to integrating intricate datasets and eliminating batch effects while maintaining biological variability, scVI excelled in benchmark studies. For instance, Lopez et al. (2018) demonstrated that scVI performed well on tasks such

as DE analysis, batch correction, visualization, and clustering[28]. Due to its generative nature, scVI can also be used to sample from the latent space in order to create new "virtual" cells (for example, generating gene-expression profiles for fictitious cell states) or interpolate between conditions, both of which are useful for exploratory analysis [31]. To summarize, scVI employs raw count data to train a VAE with a reconstruction loss (negative-binomial or ZINB) plus KL divergence. This results in a probabilistic latent embedding that corrects batch effects and lowers dimensionality.

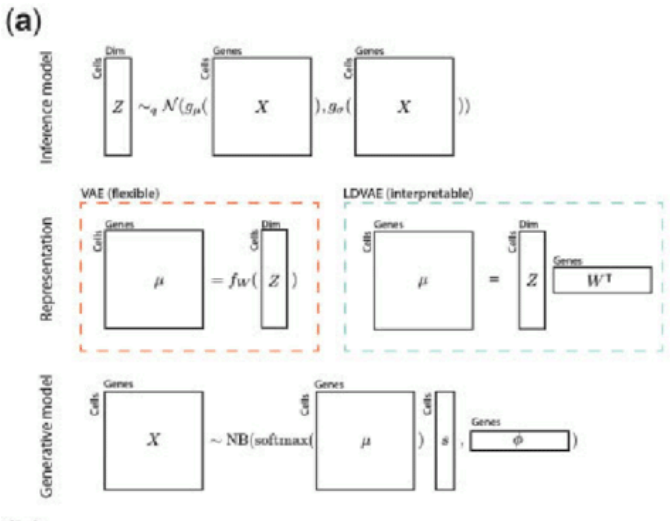


Figure 1. An outline of the general design of two different representation models for scVI autoencoders. This diagram shows how raw counts are embedded into a latent space and the encoder/decoder structure.

In terms of data, batch labels and the unnormalized count matrix or raw UMI counts are used to train scVI. The model employs an evidence lower bound (ELBO) objective (reconstruction error + KL divergence) and assumes a discrete count distribution (Poisson or negative-binomial). Frequently, the model handles

library-size normalization without the need for a previous log-transformation. After training, scVI yields a latent embedding that can be used for clustering, differential expression testing, and visualization (such as UMAP). According to studies, it successfully eliminates technical batch effects. By decoding to expression space and sampling from the learned latent prior, it can also produce new cells. These "in silico" cells can be used to enhance datasets or investigate unobserved cell states.

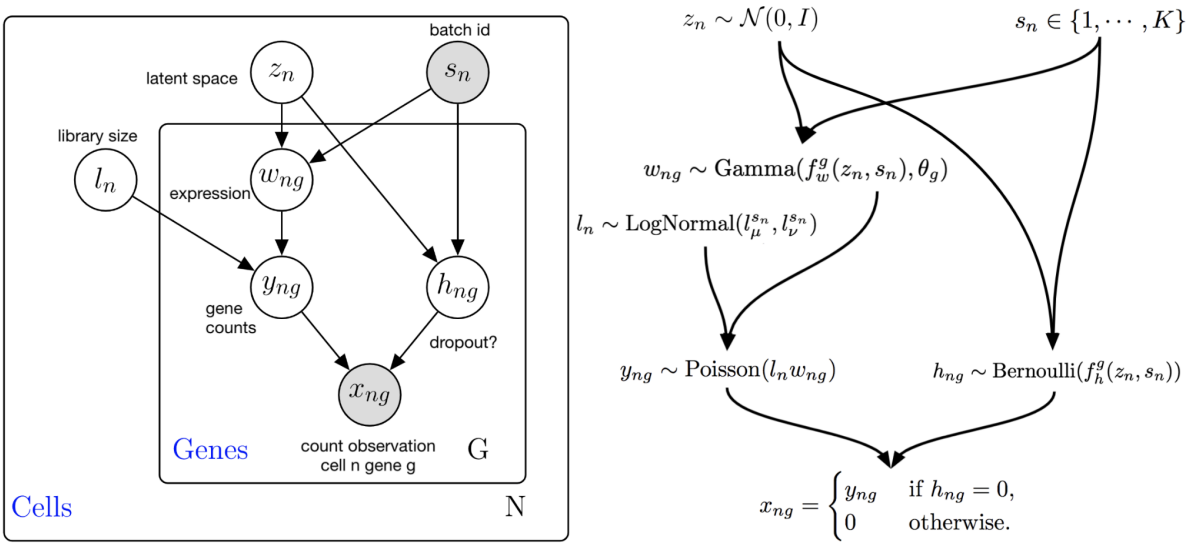


Figure 2. The diagram illustrates how the model produces normalized representations and how batch annotations are incorporated during training (The scvi-tools development team, 2025 [33]).

scGen

Single-cell perturbation prediction, or scGen, is a VAE-based model created especially to forecast how single-cell gene expression will alter in response to perturbations (such as medication therapy or genetic knockouts) [29]. By learning distinct latent representations for every condition and cell type during training, it

integrates a standard VAE with latent vector arithmetic, ensuring that the effect of a perturbation is represented by a constant vector shift in latent space [29]. In reality, scGen learns to encode each cell into a latent space conditioned on treatment by using as inputs the raw gene expression counts along with condition labels (such as control vs. treated). By appending the acquired "perturbation vector" to the latent representation of a control cell and then decoding back to expression space, scGen can forecast an invisible condition after training. Accordingly, scGen can replicate the effects of treatment on different cell types and even species. Lotfollahi et al. (2019) showed that scGen captures biological response signatures by accurately modeling drug or infection responses across cell types and studies[29]. The ability of scGen to forecast how a novel drug or stimulus would change the transcriptome of cells not visible in the training data is a crucial downstream application in silico drug testing. When annotated data were available, scGen (with cell-type labels provided) performed better than many other models in benchmarks for predicting perturbations [31].

scGen is usually trained using paired conditions on the same raw scRNA-seq counts as scVI. Each cell's stimulus or condition is listed on the training label. It employs a conditional VAE approach, which is essentially a CVAE, along with KL divergence and reconstruction loss, which is typically negative-binomial or MSE on log-expression. In essence, scGen learns an embedded condition vector: the model adds or subtracts perturbation effects using vector arithmetic after embedding cells and covariates (conditions) into the latent space. scGen can transfer a perturbation from one cell type or study to another after training. For instance, by extrapolating

from data on another cell type, it can mimic the appearance of an iPSC-derived neuron under a specific medication. It is helpful in silico combinatorial effect and dose-response prediction. The performance of scGen is on par with or better than other integration techniques when cell-type labels are known, according to nature.com. Generative investigation of perturbation effects is made possible by scGen's generation of "predicted" single-cell profiles under novel circumstances through the decoding of latent vectors.

DeepCell

Van Valen Lab's DeepCell framework is a collection of deep learning tools that were first created for single-cell analysis using images. Using convolutional neural networks trained on high-dimensional microscopy data (multi-channel images, time-series) to carry out tasks like cell segmentation, classification, tracking, and phenotyping is the main concept. Van Valen et al. (2016) demonstrated, for instance, that supervised CNNs are capable of accurately segmenting the cytoplasm and cell nuclei of both mammalian and bacterial cells, even differentiating between distinct cell types within the same image[34]. DeepCell models heavily rely on data augmentation and are trained on labeled image patches (such as microscopy images with annotated cell masks). Rotating and flipping images greatly enhanced segmentation in Van Valen's study (augmented training sets of 200-400k patches from a few hundred cells). The network learns to produce instance segmentations or cell masks, and the loss is usually a pixel-wise segmentation loss, cross-entropy or Dice. DeepCell has recently branched out into multi-task networks such as

combining segmentation and classification and cloud-scalable tools including DeepCell Kiosk. For example, in highly multiplexed images, the CelloType model [32] jointly segments cells and classifies cell types using a transformer-based CNN. In general, DeepCell's ecosystem comprises models for:

- Delineating cell and nuclear boundaries in 2D/3D microscopy with performance at or above human level is known as segmentation [26].
- Cell tracking is the process of connecting cells over time to create lineages.
- Determining cell types or states based on image features (e.g. via clustering or CNN classifiers) known as phenotyping.

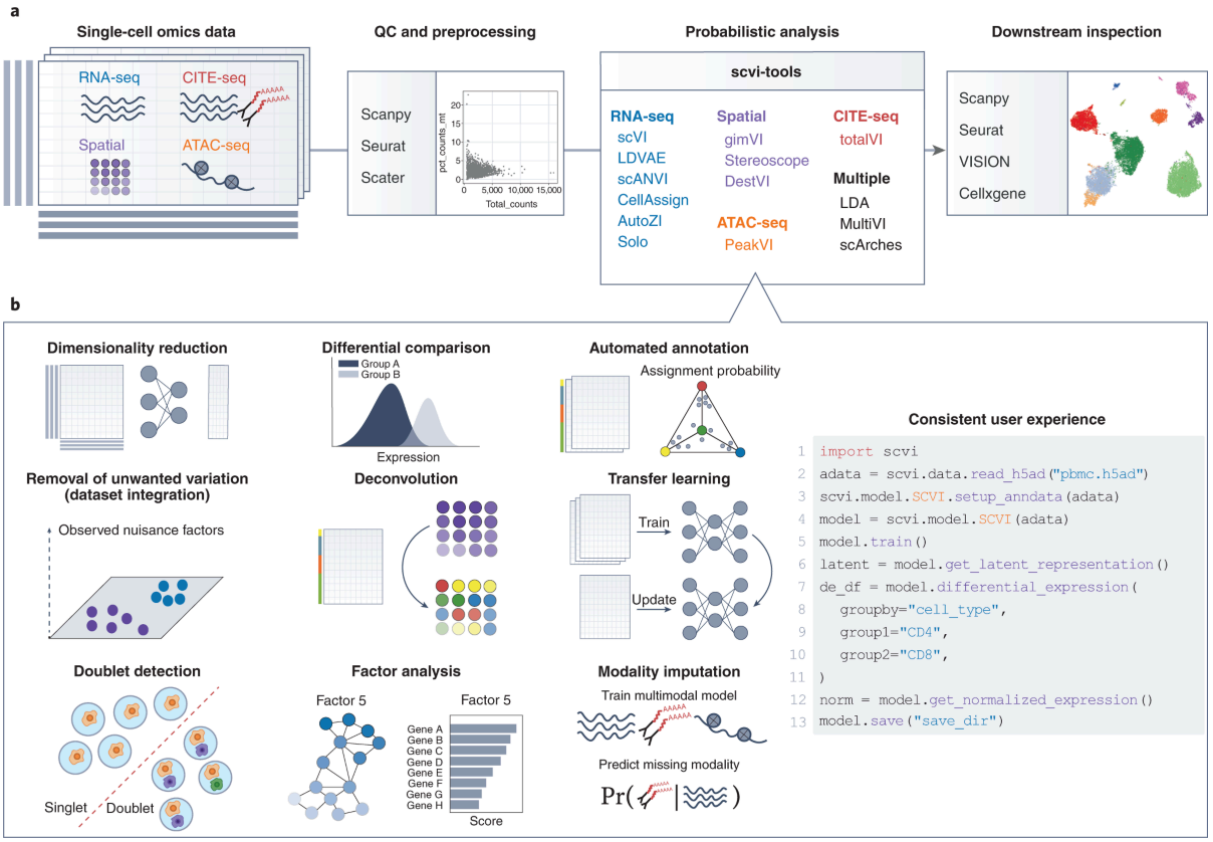


Figure 3. DeepCell multi-modal pipeline overview[25].

Despite its imaging focus, DeepCell is multimodal in nature; the lab also investigates the relationship between morphology and molecular states. For instance, the goal of DeepCell's GenAI platform is to combine gene expression and cell morphology, or "multiomics"[27]. Therefore, DeepCell offers a "digital cell" framework in which transcriptomic readouts are linked to image data (perhaps from live, label-free imaging). Applications of DeepCell models include automatically profiling cells in tissues such as distinguishing immune cells from tumor cells in cancer images and to use cloud deployment to speed up extensive image analysis. Raw pixel data from multi-channel microscopy is used in DeepCell networks. Fluorescence or phase-contrast images with manually annotated segmentation masks (nucleus, cytoplasm, whole cell) are common training data sources. To increase robustness, training uses data augmentation (rotations by 90° increments and reflections). In order to predict segmentation maps, models frequently employ CNN architectures such as U-Net or others with multi-resolution features that have been trained using supervised loss (cross-entropy or Dice). Dropout can be applied to fully connected layers, but given the paucity of manually annotated data, heavy augmentation is crucial.

Different data formats are used by these AI models. Single-cell RNA count matrices are consumed by scVI and scGen. Importantly, they anticipate receiving input from discourse.scverse.org in the form of raw UMI counts, not log- or library-size-normalized counts. Both are VAE-based; they use the standard ELBO loss function, which combines a KL-divergence and a reconstruction term. For counts, the reconstruction term usually takes a negative-binomial (or zero-inflated

Poisson) distribution. Any normalization is learned internally during training, negating the need for external feature scaling. Expression counts are generally not subjected to data augmentation in the conventional sense (such as permuting or adding noise), though certain VAE variants may employ subsampling or dropout masking.

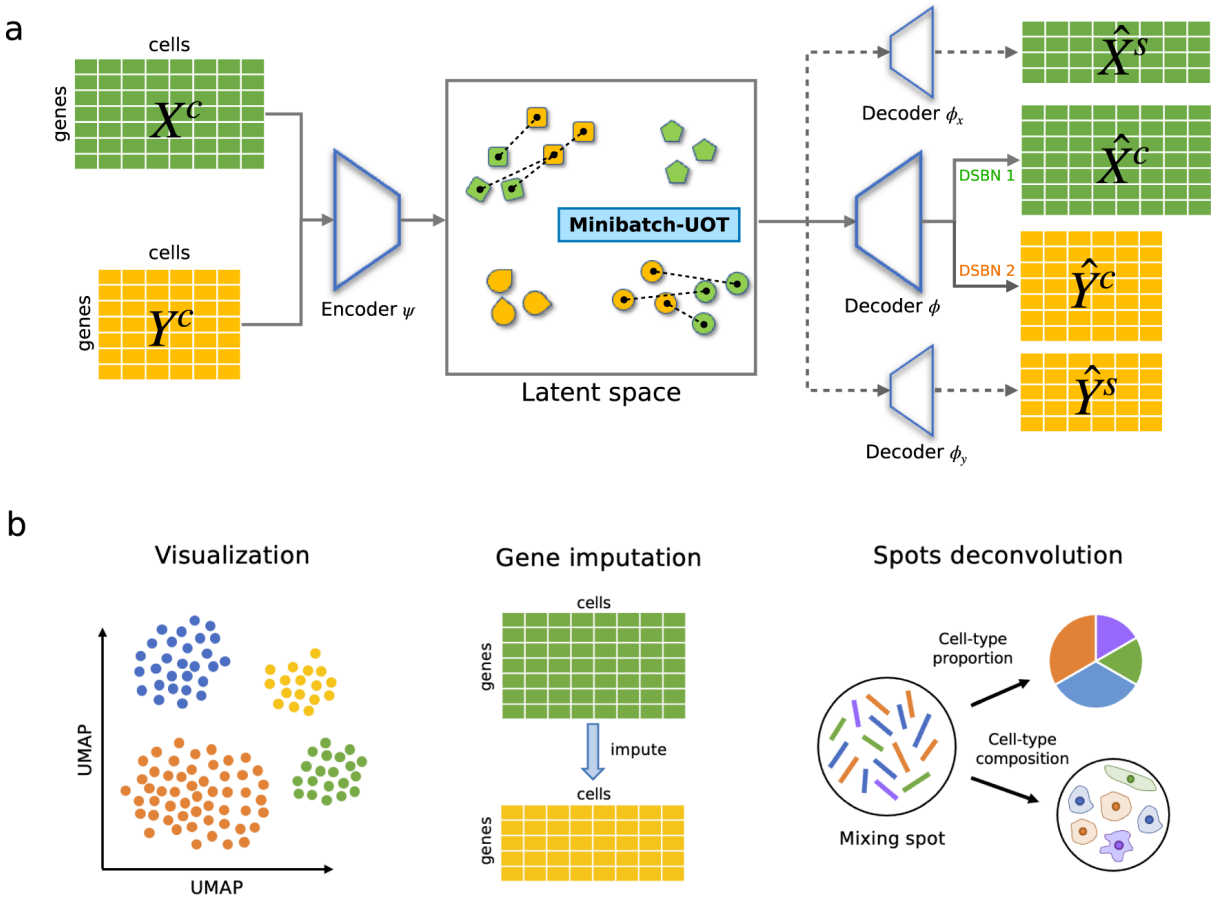


Figure 4. U-Net / VAE latent embedding comparison [24]. Latent space separation is shown graphically in this diagram, which is useful for describing cell simulation or cluster separation.

DeepCell models, on the other hand, are trained using image data. Multi-channel microscopy images (e.g., one channel of nuclear stain and one channel of phase

contrast) or even dozens of fluorescence channels in multiplexed assays are examples of input [30,31]. Pixel-level segmentations or object masks serve as ground truth labels. Extensive image augmentation is used during training. Van Valen et al. (2016) demonstrated that flipping and randomly rotating patches by 0°, 90°, 180°, and 270° greatly enhances segmentation performance[34]. Usually, the output mask suffers a supervised segmentation loss (such as the Dice coefficient or cross-entropy). Dropout is used to regularize sparingly, usually only in fully-connected layers.

DeepCell networks classify and segment images with human-level accuracy by identifying cells and subcellular structures [34]. For instance, DeepCell can identify each cell in multiplexed imaging or spatial transcriptomics and then pair it with marker expression to assign a cell-type label. Cell morphology, local cell density, and spatial context can all be precisely quantified thanks to the segmentation output. Cells by phenotype can also be directly classified from images using different DeepCell models (or branches of a network). Other analyses may use these image features as inputs (e.g., predicting gene expression from image or vice versa). As generative VAEs, scVI and scGen are able to create new data. By decoding and sampling latent vectors, scVI can produce "virtual" cells. When given unperturbed inputs, scGen specifically produces perturbed cells. The goal of DeepCell's most recent GenAI models is to create artificial cell images in novel settings (such as how a cell would appear following medication treatment). This suggests that multi-modal digital cells will be possible in the future when models are able to integrate modalities, predicting an image from expression or vice versa.

In conclusion, deep learning is used by contemporary AI-based models such as scVI, scGen, and DeepCell to learn data-driven cell models without the need for hardwired biology. They optimize variational or deep convolutional networks (with losses like KL divergence or cross-entropy) using raw high-dimensional inputs (RNA counts or images). These techniques have empirically demonstrated excellent performance in actual single-cell studies. For instance, DeepCell models have been utilized for high-throughput tissue imaging, and scVI/scGen has been applied to human iPSC differentiation data and other scRNA-seq tasks. Their accuracy and usefulness are validated by benchmarking across studies. These tools are essential parts of the contemporary single-cell toolkit because they work together to enable strong downstream analyses, such as creating new cellular profiles, integrating datasets, and categorizing cell phenotypes.

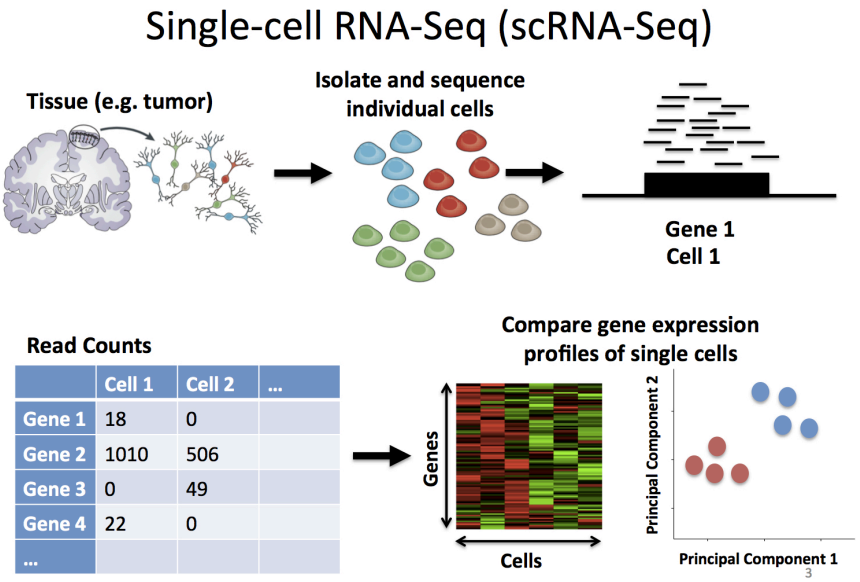
For our research, scVI should be used for batch-corrected latent embedding and dataset integration in our study on virtual organism construction using iPSC data; scGen is best suited for simulating transcriptomic responses to perturbations, such as virtual drug testing; and DeepCell can be integrated during tissue-level modeling and spatial reconstruction. These tools work together to create a complementary AI toolkit that includes spatial morphology, perturbation response, and gene expression dynamics. This toolkit is essential for simulating the transformation of a single cell into a functional virtual tissue or organ.

Main body

Digital Representation of the Cell: From Expression to Model in Animal cell

The transformation of a biological cell into a computational model, here referred to as a virtual cell, represents a cornerstone in systems biology, computational modeling, and AI-driven drug discovery. A virtual cell is more than a static abstraction of cellular traits; it is a dynamic, data-driven construct capable of simulating gene expression changes, regulatory shifts, and phenotypic responses in reaction to stimuli such as transcription factors, small molecules, or environmental stressors. This subsection details the theoretical and technical underpinnings of digital cell modeling, tracing the evolution from raw single-cell omics to trainable generative models.

Biological Data Foundations: From scRNA-seq to GRNs and Perturbations



The foundation of any digital cell model begins with single-cell RNA sequencing (scRNA-seq), a technology that has revolutionized our understanding of cell heterogeneity, differentiation landscapes, and context-specific gene expression [35]. Through scRNA-seq, it is possible to decompose tissue samples into individual transcriptomes, offering a high-dimensional representation of gene activity at the single-cell level. Recent projects such as the Tabula Muris Senis and Human Cell Atlas have provided expansive scRNA-seq datasets, which now serve as key training corpora for generative AI [36,37].

Beyond expression, gene regulatory networks (GRNs) offer a structured view of how cellular transcriptional output is orchestrated. These networks model the directed interactions between transcription factors (TFs) and their target genes, often reconstructed using statistical inference, Bayesian methods, or perturbation-based

learning algorithms [38]. Importantly, GRNs provide interpretability, ensuring that models do not merely correlate but embed causality in gene activity [39].

Another key layer includes drug perturbation signatures, as popularized by the LINCS L1000 database and the Connectivity Map, which systematically profile transcriptomic responses to thousands of chemical and genetic perturbations [40]. These datasets provide functional input–output mappings critical for simulating virtual interventions.

From Data to Representation: Vectorization and Latent Spaces

Once biological data are collected, the next challenge lies in embedding them into a mathematical space suitable for simulation. In the simplest form, a cell can be represented as a vector of gene expression values or activity levels of GRN nodes. However, such high-dimensional representations are often noisy and redundant. As such, dimensionality reduction is crucial, not only to reduce computational complexity but to extract biologically meaningful features.

Variational autoencoders (VAEs) and their probabilistic derivatives (e.g., scVI) have become the gold standard for this task [28]. These models compress high-dimensional expression profiles into a latent space, which is a continuous, low-dimensional embedding that captures the essence of cellular identity. This space is not only useful for clustering or visualization but forms the backbone for simulation, enabling interpolation between cellular states or projection into hypothetical conditions [41, 42]

In this framework, each cell is an embedding in latent space, where neighboring points reflect biological similarity, and directions correspond to biological processes, such as cell cycle, activation, or differentiation. The latent variables can thus be interpreted as axes of variation shaped by biological programs.

Modeling Transitions: Dynamic Behavior and Perturbation Response

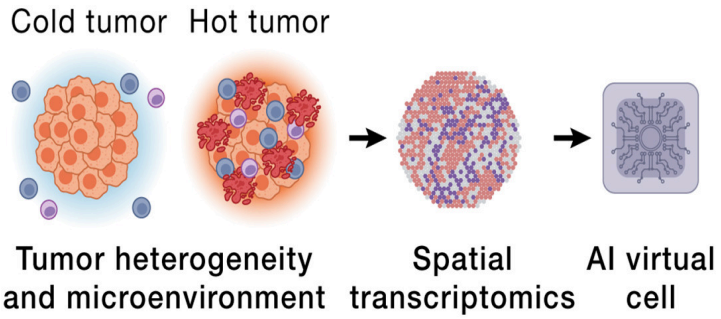
A true virtual cell must not only represent a static identity but simulate how that identity evolves in response to changes. This capability emerges in dynamic generative models, such as the Compositional Perturbation Autoencoder (CPA) and scGen, which allow conditional simulation of transcriptomic outcomes [29,30]. CPA extends the VAE framework by conditioning both encoder and decoder on covariates such as perturbation type, dose, or time point, thus enabling extrapolative generalization across unseen conditions.

For example, given a latent embedding of a CD34+ hematopoietic stem cell, CPA can predict its transcriptional state after exposure to interferon-gamma, even if that specific combination was never observed in training. Such capabilities are particularly valuable in virtual drug testing, where one wishes to simulate the effect of a compound on diverse cell types without empirical experimentation.

Such trajectory inference models as Monocle, PAGA, or RNA velocity further augment simulation by reconstructing pseudo-temporal orderings of cells based on scRNA-seq snapshots [43, 44]. These models offer vector fields over latent space,

suggesting how cells move from progenitor to terminal states. When embedded into virtual cells, these dynamics enable the simulation of developmental trajectories, disease progression, or reprogramming under transcription factor induction.

Integrating Structure and Interpretability



To ensure interpretability and biological realism, newer models integrate mechanistic priors, notably, GRNs or biochemical kinetics, into deep learning frameworks. For instance, DeepGRN and GeneCircuit models encode prior knowledge into neural network architectures, ensuring that learned embeddings reflect known TF-gene relationships [45,46]. This approach addresses one of the core challenges of black-box deep learning: biological plausibility.

Additionally, physics-informed machine learning and differentiable simulators such as Differentiable Cell (DiffC) attempt to integrate spatial constraints, diffusion, or feedback loops into cell-level simulations [47]. These hybrid approaches offer the benefits of both worlds: data-driven accuracy and mechanistic transparency.

Virtual Animal Cell

Input Data

Simulating organoid growth in a three-dimensional environment involves capturing both intracellular dynamics and cell–microenvironment interactions. To accurately reproduce morphogenesis, differentiation, and cellular self-organization, hybrid approaches leverage biophysical models guided by AI methods. These models rely on several key data sources: multi-cell single-cell RNA sequencing (scRNA-seq) for defining cell types and transcriptional states; precise spatial growth conditions (e.g. Matrigel stiffness, extracellular matrix composition, biochemical gradients); and imaging data (high-resolution 3D/4D microscopy of organoids) to calibrate and validate morphology, cell arrangement, and dynamic behaviors [48]. Of these input modalities, as discussed earlier, scRNA-seq serves as a key element throughout the virtual modeling pipeline – from lineage validation to AI-driven cell behavior modeling. To clarify, single-cell transcriptomics provides a comprehensive map of the cell states and differentiation trajectories within the organoid, ensuring that each simulated cell type behaves in accordance with its real-world counterpart. In practice, scRNA-seq data can guide the assignment of cell phenotypes in agent-based models or serve as training data for AI algorithms that predict cell fate decisions. Moreover, realistic spatial context is crucial; factors such as matrix stiffness and nutrient gradients influence how organoids develop structures like lumens or vascular networks, so these conditions must be encoded in the simulation. Concurrently, time-lapse imaging data offers ground-truth for growth patterns and

organoid morphology changes over time, helping to tune model parameters (e.g. cell proliferation rates, movement) so that in silico organoids mirror in vitro observations.

AI and Hybrid Modeling Approaches

A hybrid modeling approach combines data-driven AI techniques with mechanistic, physics-based simulations to capture the complexity of 3D organoid development [49]. On the AI side, generative models like 3D Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) have been proposed to synthesize realistic 3D forms. For example, a 3D-GAN can be trained on volumetric organoid imaging data to generate new plausible organoid morphologies under different conditions, allowing exploration of how changes in growth factors might alter organoid shape. The application of GANs/VAEs to de novo 3D morphogenesis is still nascent – AI-driven generation of complex organoid structures remains an emerging research area with limited training data, so currently AI is rarely used directly to create novel 3D morphogenetic predictions [50]. Nevertheless, such generative models hint at the potential for AI to suggest novel organoid architectures beyond what has been experimentally observed.

Meanwhile, deep learning is making an impact in organoid image analysis and morphology quantification. Tools like OrganoID [51] provide a versatile deep learning platform that can automatically recognize and track individual organoids in microscopy images, measuring features such as organoid count, size, and shape over time. OrganoID demonstrated >95% agreement with manual counts of pancreatic cancer organoids and ~97% for organoid size, without parameter

adjustments, and it maintained single-organoid tracking accuracy above 89% over four days. Such AI-driven image analysis platforms accelerate high-throughput phenotypic measurements: for instance, OrganoID can trace exact organoid shapes and monitor changes (e.g. in circularity, solidity, eccentricity) under different drug treatments automatically [51].This enables direct validation of simulation outputs – one can compare simulated organoid sizes or shapes with real image-derived metrics.

Another example, sometimes termed "OrgaNet," refers to deep neural networks designed for organoid morphological analysis, including segmentation of organoid structures and even cell nuclei within them. These networks (often based on 3D U-Net or similar architectures) can convert raw imaging data into quantitative morphological descriptors that feed into or benchmark the simulations [52].

On the mechanistic side, agent-based modeling frameworks simulate each cell as an independent “agent” following biological rules. Platforms such as PhysiCell and CellModeller allow researchers to embed cells in a 3D space where they grow, divide, move, and interact according to biophysical laws. For example, PhysiCell [53] is an open-source simulator that can model thousands of interacting cells, linking cell behaviors (division, death, secretion, etc.) to diffusible signals and mechanics in the microenvironment [53]. On the mechanistic side, agent-based modeling frameworks simulate each cell as an independent “agent” following biological rules. Platforms such as PhysiCell and CellModeller allow researchers to embed cells in a 3D environment where they grow, divide, move, and interact according to biophysical laws. For example, PhysiCell [53] is an open-source simulator that can model

thousands or even millions of interacting cells, linking cell behaviors (division, death, secretion, etc.) to diffusible signals and mechanics in the microenvironment. Ghaffarizadeh et al. [53] describe PhysiCell as a 3-D agent-based framework with built-in submodels for cell cycle, apoptosis, mechanics, and coupling to diffusion solvers for nutrients and signals. Using such a framework, one can simulate an organoid starting from a few stem cells that proliferate and self-organize: cells experience forces like cell–cell adhesion and pressure from confined growth, consume nutrients, and respond to signaling gradients – leading to emergent structures reminiscent of real organoids. For instance, agent-based models naturally capture phenomena like hollow lumen formation at the core (via cell polarization and apoptosis of interior cells) and spatial metabolic gradients (e.g. hypoxic center vs. oxygenated periphery in larger organoids) when appropriate rules are included. Experimental observations show that cells losing contact with a matrix or nutrient supply often undergo apoptosis and accumulate in the organoid lumen [54], and diffusion limits cause intrinsic hypoxic cores in organoids lacking vasculature [55]. By calibrating the model with experimental data (e.g. nutrient diffusion lengths or cell-death rates), these behaviors – luminal clearing and spatial differentiation – emerge in simulations, matching real organoid morphology.

Crucially, hybrid models can integrate AI with these mechanistic simulations, using machine learning to discover or optimize certain rules. For example, an evolutionary algorithm or other optimization method might tune dozens of model parameters so that the simulation outcomes (organoid size distribution, cell type proportions, etc.) match experimental observations. In practice, researchers have begun to combine data-driven components with physics-based models: Camacho-Gomez et al.[49]

present a framework where a neural network is trained on image-derived metrics to regulate cell “decision-making” within an agent-based organoid model [49]. In their approach, the simulation calls a deep learning model to decide which cell behavior (divide, differentiate, etc.) should occur, and a genetic algorithm optimizes the neural network so that the simulated pattern fits the observed morphogenesis [49]. This kind of AI-in-the-loop strategy shows how machine learning can augment traditional simulations – for instance, by predicting cell fate outcomes from gene expression or microenvironment data, and feeding those predictions into the next time-step of the simulation. Similarly, metaheuristic algorithms have been used to automatically tune agent-based model parameters for better agreement with data [56], illustrating the power of AI to handle complex parameter spaces that are intractable by manual adjustment. Notably, graph-based AI approaches have potential to further enrich these simulations. Graph Neural Networks (GNNs) can represent each cell as a node in a graph, with edges embodying cell–cell interactions (such as physical contacts or signaling proximity). Recent studies demonstrate that GNN models can learn the rules of cell fate coordination from live tissue imaging data [57].

For example, Yamamoto et al. [57] showed that an interpretable GNN, given spatiotemporal cell-tracking data from a live epithelium, could predict a cell’s fate (e.g. division or differentiation) based on its neighbors and reveal distinct “neighbor interaction” rules governing multicellular dynamics.

In an organoid context, one could likewise use a GNN to model how signals from neighboring cells influence a cell’s behavior. During a simulation, the GNN would dynamically update each cell’s state by passing “messages” along the cell interaction network (mimicking juxtacrine signaling, contact inhibition, etc.), providing an efficient

approximation of complex cell–cell communication. This approach is especially attractive for capturing emergent patterning – for example, how a small cluster of organizer cells can influence the spatial arrangement or differentiation of surrounding cells. While the incorporation of GNNs into organoid simulations is still largely theoretical, the concept is supported by the versatility of the GNN framework: it can infer general cell interaction rules from data without prior knowledge of the underlying signaling pathways [57]. By ensuring that local neighbor effects are learned from real data, GNN-augmented simulations could achieve more realistic collective behavior than using physics-based rules alone.

Limitations and Outlook

It is important to acknowledge the current limitations in 3D organoid modeling with AI. First, the use of AI for generating de novo 3D morphogenesis is still in its infancy. Most AI applications in biology have focused on 2D images or sequence data; applying AI to drive 3D shape formation (as in growing an organoid purely in silico) faces challenges due to the complexity of spatial data and the lack of large training datasets [58]. In other words, AI is rarely used directly to create novel 3D morphogenetic predictions today – this is a frontier that demands new approaches and substantially more data.

Second, there is a paucity of high-resolution 3D data at single-cell resolution to inform and validate these models. Many studies still rely on endpoint measurements of organoid size or gene expression, or on 2D cross-sections, which fail to capture the full 3D cellular architecture [59]. The scarcity of volumetric, single-cell-resolved

datasets means that models might oversimplify cell–cell interactions or miss subtle phenotypic heterogeneity. Advances in light-sheet microscopy and cleared-tissue imaging are starting to fill this gap, but data volume and analysis remain bottlenecks. Third, validation of simulation results is difficult without robust biomarkers and standardized metrics. For example, if a model predicts a certain spatial pattern of differentiation inside an organoid, do we have validated biomarkers or tracers to detect that pattern in real experiments? The field currently lacks universally accepted quantitative benchmarks for organoid morphology beyond generic ones (size, circularity, histology). Developing validation biomarkers and assays (such as specific immunostains for predicted cell states or functional readouts like calcium oscillation patterns for cardiomyocytes) is crucial to test model predictions. Without such benchmarks, it's hard to say whether a given simulation is "correct" or biologically relevant. In summary, 3D organoid modeling requires a hybrid approach that combines the strengths of AI and traditional physics-based methods. Neither approach alone is sufficient: purely data-driven AI might not capture physical constraints (like diffusion limits or mechanics), whereas purely mechanistic models may not fully leverage complex datasets or uncover hidden patterns. A synergy of the two can compensate for each other's weaknesses – for instance, AI can rapidly optimize parameters or suggest network interactions, while mechanistic models ensure adherence to biophysical laws. The consensus emerging in this interdisciplinary field is that we need to integrate AI with physical and mechanical modeling to realistically simulate organoids. This could mean AI-derived rules plugged into agent-based simulations, or simulators generating synthetic data for training AI models – likely both in iterative cycles. Finally, it's worth noting that the

concepts in 3D organoid modeling are influencing other domains of biology as well. Plant biology is seeing efforts to model 3D development of structures like root systems and shoot apical meristems. Here, researchers combine biomechanics (cell wall expansion, turgor pressure) with gene-regulatory network models to simulate patterning in a growing root or the phyllotaxis in a shoot. While these plant models don't yet heavily use AI, the frameworks being developed for organoids could conceivably be applied (for example, using machine learning to optimize a plant root growth model against observed root architectures). Likewise, in microbiology, agent-based modeling has been used for years to simulate bacterial colony growth – for instance, BacSim [60] was an early individual-based model capturing how E. coli cells grow and compete in a colony. Going forward, modern AI techniques like GNNs could enhance bacterial colony simulations by learning interaction rules between bacteria or predicting colony morphologies under various conditions (nutrient levels, agar stiffness, etc.). These examples in plants and microbes echo the theme that understanding complex 3D biological phenomena benefits from hybrid modeling. In all cases – whether organoids, plant meristems, or bacterial biofilms – the integration of AI with biophysical simulations offers a promising route to unravel emergent behaviors in three dimensions.

Organ Integration

Functional architecture is a way of organizing and modeling organs and tissues. It describes a set of system functions, their interactions and interrelations, and how

they are implemented in the system components. They transmit signals, coordinate processes, and exchange substances.

This paper will consider an approach to the virtual integration of tissues and organs in order to create functional organs using AI modeling. This paper also considers AI systems and simulators (BioDynaMo, OpenCMISS, DeepCell) to link organelle data with the functioning of the organs as a whole [61]. For a complete demonstration, three organs were chosen: the liver, heart, and brain.

As input, organoid components characteristic of each of the organs are used. For the liver (hepatoid organoids): organoids are assembled from the main parenchymal cells of the liver and with the help of auxiliary cells such as vascular endothelial cells (form the vascular network) and cholangiocytes (epithelial cells of the bile ducts). For the heart (cardioid organoids): cardiomyocytes (contractile muscle cells) and vascular endothelial cells. Cardioids are 3D modules of the heart for the brain, self-organizing from pluripotent cells and their derivatives [66]. For the brain (cerebroid organoids). Cerebroid organoids are 3D structures derived from pluripotent stem cells that resemble the developing human brain. These organoids contain types of nerve cells such as neurons, astrocytes (a type of cell that supports neurons), microglia (the brain's innate immune cells), and oligodendrocytes.

Organoid-Based Brain Modeling

Modeling neural structures. Cerebral organoids (3D brain organoids) reproduce various aspects of human brain function, including neurogenesis and cortical areas. Using AI and simulations, researchers are trying to recreate the functional architecture of the brain based on data from brain organoids. One suitable AI tool is BioDynaMo [61,63]. For example, a new computer model of the development and growth of neurons in our brain was built in the Journal of Mathematical Biology 2024 [63]. The simulations used neurons in the hippocampus, a critical brain region responsible for memory storage. The research team used a method called “approximate Bayesian inference” (ABS) that successfully simulated the growth patterns of real brain neurons [63]. This shows that such AI tools are able to simulate neurogenesis and the network of connections in mini-brains, which in turn can help understand brain development and various neuron-related diseases. Analysis of images and signals from brain organoids[63]. Often, the analysis of large and complex data sets taken from organoids can be ineffective and corrupted. Modern AI methods offer a promising solution for efficient information extraction, making a forecast based on various types of data. In particular, algorithms based on convolutional neural networks are able to segment the nucleus and cells in 3D images of organelles, tracking their growth and changes. For example, the DeepCell system was developed so that automatic recognition of microscopic images of cells completely coincides with the segmentation of human tissues [61]. AI tools such as DeepCell allow you to accurately analyze the architecture and composition of the mini-brain, tracking the dynamics of its further changes [61].

Case study: In the Journal of Mathematical Biology 2024, scientists combined the BioDynaMo simulator with experimental data on the growth of neurons from organoids [63]. As a result, the simulation reliably reproduced the branching of hippocampal neurons, similar to in vitro observations [63]. This study shows that the combination of realistic modeling and AI optimization helps to create a functional architecture of a part of the brain, in other words, a digital analogue of the neural network of the organoid, which in turn opens up innovations to the possibility of virtual testing of hypotheses in the field of neuroscience, without direct experiments with tissues, which is very important.

Modeling the heart based on organoids

The heart is an organ with a pronounced multi-scale nature, pumping blood throughout the body. The heart maintains blood circulation and works as a pump, thereby electrical impulses at the level of ion channels cause cell contraction for this function. To reproduce such a functional architecture, computed modules are required. One such tool is the OpenCMISS library [64] - it stores a large part of the information on multiphysics modeling in biomedicine. OpenCMISS was used to build three-dimensional models of the heart, where cellular electrophysiological parameters are integrated into the model of tissue and ventricles [64]. Such simulations help to predict the propagation of waves and contraction of the heart in norm and pathology.

Scientists have collected a library of 230 different cardioids of geometric configurations (rectangles with different sides of the ratios, circles of different diameters, etc.) [66]. Using methods called "clustering and nonlinear dimensionality reduction", the organoids were immediately grouped by similarity [66]. As it turned out, geometry significantly affected the function. Machine learning made it possible to identify variables and determine the optimal forms of the organoid to achieve the specified properties [66]. Thus, AI can help with the creation of a mini heart with the desired functionality and architecture. An example of AI in cardiology is the Living Heart project [61]. In this project, the model solves 30 million equations in a real anatomical frame of the heart and can predict any changes. In subsequent works, this work is supplemented by AI modules so that it is possible to use clinical data of ECG and MRI of the heart of patients in order to personalize the simulation [61]. In general, the modern strategy has two complementary lines: 1) Physically based simulators [64], 2) AI analysis of cardioid data [66]. Such a combination will help to accurately create a "virtual heart" on which in the future it will be possible to try various drugs and treatments without risking the patient's health.

Liver modeling based on organoids

Design of a functional analog of the liver considering processors such as: blood flow in sinusoids, metabolite transport and cellular signaling networks of hepatocytes. In the work npj Digital Medicine, 2024, a model of a virtual human liver lobule is provided [65]. The model considers metabolic zoning, distinguishing enzyme activities between the periphery and the center of the lobule, which is very important

[65]. Thanks to this, the simulator was able to predict the zonal nature of drug toxicity, consistent with clinical observations. This is a kind of "digital twin" showing personalized damage in different patients at a given dose of the drug [65]. Now they have begun to try to create 3D liver models based on medical images, where each lobule is described by a set of diffusion reaction equations. Thus, computer simulators of the liver allow you to create a functional architecture of the organ [65].

AI for liver organoid analysis

There is a developed model DILITracer, which uses a convolutional transformer trained on 700 thousand cell images, to analyze liver organoids [67].

Organ on a chip

The liver very rarely acts in isolation, it is most often connected to other organs. Modern approaches try to combine organoids of different types into multi-organ chips and use AI to analyze interactions [61]. An example of a study is the work of Nature Communications 2022 [68]. In this study, a liver organoid on a chip was exposed to an anti-cancer drug, while simultaneously observing its effect on heart cells. Thus, thanks to the data, it was possible to build a model of the liver-heart connection.

Limitations of existing approaches

Lack and quality of data. In order to use AI modeling, a lot of extensive and reliable experimental data is required. But obtaining such data for organoids is very difficult due to biological variability [61,66]. Even with the same protocol, two different

organelles can differ greatly in composition and architecture. In addition, many processes are simply difficult to observe since datasets are simply not available [61].

Ethical and methodological issues. The use of AI in biology and treatment often prompts the question: can we fully trust the model when making decisions, especially when it comes to a clinic [61]. Full certification of models is needed, while the area is practically not formalized by law [61].

Summary

AI can effectively model the functional architecture of organs such as the liver [65,67], heart [61,64,66], and brain [61,63] based on organoid components. However, the reliability and clinical applicability of such modules are entirely dependent on careful validation [61]. Without proper validation, these modules remain limited in their exploration and predictions.

Virtual Organism (Animal cell)

Currently, one of the most promising areas is the creation of virtual organisms. These are full-fledged modules that recreate the interactions of various organs and systems of the body. These virtually developed modules help to study biological processes, test new drugs and medications without harming human health. Models of virtual organisms are computer representations of living systems that allow you to take into account both physiological and molecular aspects of the functioning of the body. Such modules allow you to predict the effectiveness of

therapy, assess risks and also develop personalized approaches in medicine. Here we will consider modern achievements in this field and also limitations.

Input data and models

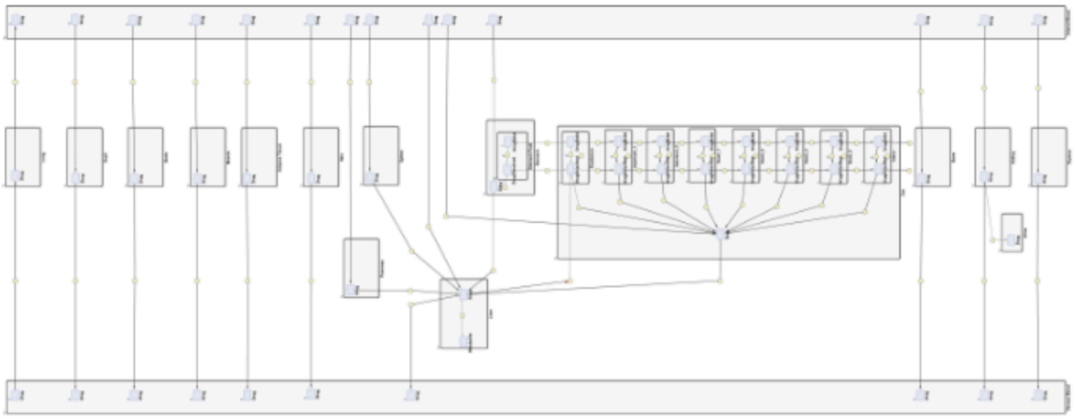
In order to assemble an accurate model of a virtual organism that will function correctly, you need to collect all the necessary data from various levels of organization of biological systems and integrate them [70].

Models and their applications

HumMod- models the physiology of the human body. It includes more than 9,000 variables that describe various physiological processes in the body, allowing you to study the effects of diseases on the body and its changes in functioning. WholeBodyPK, OpenCOR, SimBody – pharmacokinetics and dynamics [69].

WholeBodyPK- model, which includes an extended version of the compartmental absorption and transit (CAT) model of gastrointestinal absorption, additionally including enterohepatic recirculation, analyzes pharmacokinetics, which in turn allows tracking how drugs are metabolized in the human body [77].

OpenCOR is an open source environment for organizing, editing, simulating, and analyzing models encoded in CellML format [71].



SimBody is a dynamic engine that solves problems that arise in biomedical research. It is also used for scientific and engineering modeling of mechanical systems, including biomechanical structures [72].

Possibility of using multi-agent RL to coordinate the behavior of organs

The use of multi-agent learning methods, which include 3 structured levels: agent activation, task planning, and trajectory perception, is a promising direction. For example, the study "VIKI-R: Coordinating Embodied Multi-Agent Cooperation via Reinforcement Learning" demonstrates how MARL can be used to coordinate the actions of complex systems [73].

Detoxification

The kidneys and liver are among the most important parts of the body for detoxification. With a decrease in kidney function, the liver is damaged. Because of

this, modeling the interactions between the kidney and liver is important. Research shows that the gut microbiota and its metabolites can influence kidney and liver health [74,76].

Neurohumoral Loops

The microbiota-gut-brain axis describes the interactions between the gut and the brain through neural, immune, and endocrine pathways. Modeling this axis allows us to study microbial metabolites, immune responses, and neural activity, providing valuable insights into the communication between the gut and the brain [75].

Immune-Liver-Tumor systems

Modeling the interactions between the immune system, liver, and tumors is important for understanding cancer better and developing treatments for it. Research shows that the interactions between these systems can significantly affect the effectiveness of therapy [78].

Limitations

Lack of datasets: The lack of comprehensive data on various parts of human and other organisms' physiology hinders accurate modeling [70].

The difficulty of simulating chronic processes: Modeling long-term chronic diseases such as cancer or diabetes is particularly challenging because these processes develop gradually and can extend over time, and their impact on the body can manifest itself at different levels. This requires the creation of more complex and adaptive models that can analyze long-term changes in the body.

(2) Virtual Plant Cell

Architecture and Simulation

Callus Induction and Differentiation Modeling

The induction of totipotent callus represents a uniquely plant-specific route to pluripotency, in which fully differentiated somatic cells are experimentally reprogrammed to regain a proliferative, undifferentiated state with the capacity to regenerate entire organs or even whole plants. This transformation is mediated primarily by the interplay between auxin and cytokinin, which function not simply as mitogenic signals but as master regulators of developmental reprogramming, controlling cell identity, chromatin accessibility, and spatial polarity establishment [81]. Auxin perception via TIR1/AFB-AUX/IAA-ARF signaling modules initiates transcriptional programs that destabilize differentiated cell states, promote cell cycle re-entry, and establish polarity cues, while cytokinin perception through CRE1/AHK

histidine kinase receptors and type-B ARR transcription factors stimulates meristematic cell proliferation and patterning. The relative auxin-to-cytokinin ratio operates as a developmental bifurcation parameter: elevated auxin favors root-meristem-like identity, whereas elevated cytokinin biases toward shoot meristem specification. In contrast to mammalian systems, where pluripotency arises intrinsically during early embryogenesis, plant totipotency is not a default developmental state but rather an induced condition, achieved through exogenous hormonal regimes that reactivate latent morphogenetic programs otherwise inaccessible in differentiated tissues. This biochemical reprogramming is underpinned by genome-wide transcriptional rewiring, including the activation of pluripotency-associated transcription factors (e.g., WUSCHEL, BABY BOOM), epigenetic remodeling such as histone acetylation and DNA demethylation, and the reorganization of intercellular signaling networks, including plasmodesmatal transport and hormonal flux redistribution. Together, these molecular and cellular processes collectively reinstate developmental plasticity, rendering the callus competent to initiate organogenesis or somatic embryogenesis under appropriate inductive cues [81].

Hormonal Control as a Computational Problem

Auxin perception through the TIR1/AFB–AUX/IAA–ARF signaling cascade destabilizes repressors of auxin-responsive genes, thereby initiating transcriptional programs that promote cell cycle re-entry, dedifferentiation, and the establishment of new polarity axes [81]. Cytokinin is perceived via CRE1/AHK histidine kinase

receptors, which activate type-B ARR transcription factors to stimulate meristematic cell proliferation, regulate organ boundary formation, and maintain stem cell activity [81]. The interplay between these pathways is not merely additive but functions as a developmental bifurcation system in which the auxin-to-cytokinin ratio governs trajectory selection: elevated auxin biases callus differentiation toward root-meristem-like states, whereas elevated cytokinin promotes shoot meristem specification [79,81].

From a computational perspective, such ratio-dependent fate decisions can be formalized as a multi-stable dynamical system in which hormone concentrations act as control parameters [79]. In this framework, auxin and cytokinin distributions form spatially heterogeneous fields shaped by active transport (e.g., PIN-mediated auxin efflux), passive diffusion, and localized biosynthesis. The resulting gradients establish positional information that is interpreted by cells to determine developmental outcomes. Multi-scale models, such as those described by Band et al. [79], capture this coupling between molecular signaling and tissue-level geometry, linking intracellular regulatory states to emergent organogenic patterns. These models are particularly relevant for callus systems, where auxin/cytokinin feedback loops and spatial transport processes jointly define the accessible differentiation trajectories.

Trajectory Inference in Plant Systems

Simulating callus differentiation in silico requires reconstructing developmental trajectories from high-dimensional single-cell transcriptomic profiles, while explicitly

integrating hormone distribution patterns as dynamic covariates. Unlike in mammalian models, where trajectory inference can often proceed without spatial constraints, plant systems demand the incorporation of spatial hormone gradients and the mechanical context imposed by rigid cell walls. This constraint fundamentally alters both the computational representation of developmental state spaces and the mathematical modeling of state transitions [79,82].

In plant-specific adaptations, pseudotemporal trajectory mapping must be extended to account for directional auxin fluxes, driven by polarized transport through PIN-FORMED (PIN) efflux carriers, and cytokinin redistribution through vascular networks such as phloem and xylem conduits. These transport processes create persistent morphogen fields whose local maxima and minima act as positional cues for cell fate transitions. Multi-scale models integrate these spatial gradients from the subcellular scale (PIN localization) to tissue-scale vascular architecture, enabling a quantitative link between hormone topology and differentiation outcomes [79].

Plant single-cell transcriptomic datasets, such as those collated in the Plant Single Cell Atlas, provide the necessary resolution to couple gene expression dynamics with hormone signaling domains [82]. By embedding these profiles in low-dimensional pseudotemporal manifolds, via methods such as Monocle or diffusion pseudotime, while annotating each cell with local hormone concentrations, it becomes possible to predict divergence points where callus cells commit toward root- or shoot-like lineages. This integration produces not just lineage graphs but spatially anchored fate maps, in which each branch point is contextualized by the surrounding morphogen landscape [82].

Frameworks like VirtualRoot operationalize this integration by simulating auxin and cytokinin fields under realistic transport constraints, providing in silico morphogen distributions that can be aligned with experimental single-cell trajectories. Such models can thereby test whether predicted branch points coincide with experimentally observed fate transitions, allowing iterative refinement of both transport parameters and trajectory inference algorithms [79]. This convergence of spatial hormone modeling with high-resolution transcriptomic pseudotime reconstruction offers a pathway to predictive, mechanistically grounded simulations of plant callus differentiation that are directly translatable to experimental regeneration systems [79,82].

Graph Neural Networks for Fixed-Topology Systems

In multicellular plant tissues, the spatial arrangement of cells is constrained by the rigid extracellular matrix and stabilized by cell wall connections, resulting in a relatively immutable, planar lattice of cell-cell contacts. This structural immobility, coupled with symplastic connectivity through plasmodesmata, defines a fixed-topology communication network that lends itself naturally to graph-based computational formalisms [80]. In such a framework, each cell is represented as a discrete node characterized by multi-modal feature vectors encompassing its transcriptomic state, local morphogen profile, mechanical stress parameters, and positional metadata. Edges encode stable, anatomically defined adjacency relations, which inherently capture both physical proximity and the potential for direct molecular exchange.

Graph Neural Networks (GNNs) exploit this representation by performing iterative neighborhood aggregation, whereby each node updates its state as a function of both its intrinsic features and the aggregated signals from its immediate neighbors. In plant developmental contexts, this enables the emergence of biologically interpretable patterns such as localized auxin enrichment zones, which frequently coincide with pre-organogenic centers, and cytokinin depletion domains that are associated with root-like differentiation foci [80]. Unlike statistical correlation models, which treat each cell as an independent observation, GNNs preserve the explicit spatial dependencies encoded in tissue topology, allowing predictions to account for the positional constraints of plant morphogenesis.

Recent methodological advances integrate spatial transcriptomics with GNN-based architectures to directly couple molecular state variation with tissue geometry. These models leverage high-resolution single-cell or subcellular transcriptomic maps, aligning them with reconstructed tissue graphs to infer context-dependent cell fate trajectories. The fixed-topology assumption simplifies graph construction, enabling stable node–edge mappings that remain valid across developmental timepoints. In the context of callus differentiation, such models can resolve fine-scale fate specification domains, detect shifts in hormone-driven patterning boundaries, and predict lineage commitment events with greater spatial fidelity than non-graph-based approaches. This integration of spatial omics data with GNN-driven inference thus provides a powerful computational paradigm for mechanistically grounded modeling of plant tissue differentiation under experimentally controlled hormonal regimes [80].

Rule-Based Hormone Diffusion Models

Rule-based simulation offers a mechanistically interpretable approach to modeling hormone transport during callus differentiation, formalizing auxin and cytokinin dynamics based on transport laws derived from empirical observations. In such models, auxin movement is characterized by passive diffusion across cell walls and active efflux via polarized PIN-FORMED (PIN) proteins, whose localization responds to local auxin flux, establishing a positive feedback mechanism known as canalization. These rules produce emergent patterns where auxin flux becomes self-reinforcing along discrete strand-like pathways, consistent with vascular strand formation during organogenesis [84].

Similarly, cytokinin dynamics can be encoded through rules governing synthesis, diffusion, degradation, and inhibitory cross-regulation, allowing the model to reproduce threshold-mediated suppression of proliferative signaling within callus tissues. When cytokinin concentration surpasses specified thresholds, the model applies regulatory feedback to modulate cell division rates and tissue expansion, emulating observed inhibitory effects on undifferentiated growth.

Crucially, rule-based models support hybridization with machine learning frameworks to form physics-informed neural networks. In such architectures, the rule-based component enforces physically plausible hormone transport dynamics, while a neural network component ingests such additional data as gene expression profiles, mechanical stress features, or cell shape metrics to adaptively refine spatial hormone distribution and fate prediction performance.

For example, simplified lattice-based auxin transport models, parameterized by cell geometry and PIN polarity rules, can be embedded within GNN architectures: nodes represent cells, edges denote adjacency, and rules dictate local hormone updates. The neural component updates latent cell states conditioned on rule-based hormone patterns, enabling simulations in which experimentally tunable parameters, like PIN localization bias or degradation rates, are modulated to explore alternative differentiation outcomes. This synergy preserves interpretability while enhancing predictive flexibility.

Generative Modeling for Hormone Perturbation

Generative neural network frameworks, such as scGen, can be tailored to plant single-cell datasets to predict transcriptional responses to hormone perturbations without direct experimental measurements for each condition. In this approach, a model is trained to capture latent transformations between baseline (pre-treatment) and perturbed (post-treatment) cellular states, using single-cell RNA-seq data as input. By embedding cells into a shared latent space, the model infers a “perturbation vector” that encodes the transcriptional shift induced by specific hormone regimes.

For instance, in Arabidopsis thaliana root callus differentiation, transcriptomic profiles under varied auxin/cytokinin ratios provide paired datasets from which scGen can learn both hormone-specific and generalizable transcriptional transitions [83]. Once trained, the model can apply these learned perturbation vectors to unseen cell populations, effectively generating in silico predictions for novel hormone combinations not present in the training set.

This capability is particularly valuable for auxin/cytokinin optimization in regeneration protocols. Instead of exhaustively testing every hormone ratio experimentally, the trained generative model can simulate hypothetical treatments, predict lineage-specific transcriptional trajectories, and rank hormone cocktails for their likelihood to induce desired tissue fates (e.g., shoot meristem specification vs. root identity). Importantly, the model output can be cross-referenced with known transcriptional markers from datasets such as Zhang et al. (2019), which provide high-resolution single-cell maps of Arabidopsis root developmental stages [83].

Such in silico screening could substantially reduce experimental trial-and-error in plant biotechnology by narrowing down candidate hormone regimes before wet-lab validation. In agricultural contexts, this approach could accelerate regeneration system design for species with recalcitrant tissue culture responses, improving both efficiency and reproducibility.

Reinforcement Learning for Sequential Induction Strategies

Sequential optimization of plant regeneration protocols can be naturally formulated as a Markov Decision Process (MDP), where each state corresponds to the multidimensional cellular context defined by its transcriptomic signature and local hormonal environment, and each action represents a discrete modification to auxin/cytokinin concentrations or other culture parameters (e.g., photoperiod, nutrient composition). The reward is quantified as the proximity of the resulting cellular state to a predefined target fate, such as shoot meristem initiation or root primordium formation.

This conceptual framing draws on the methodology established by Sootla et al. (2013), who demonstrated that reinforcement learning (RL) algorithms can effectively control genetic toggle-switch networks without explicit mechanistic equations [85]. In their work, fitted Q-iteration was used to identify optimal sequences of control inputs that transition the system between distinct stable states, even under stochastic fluctuations. Translating this approach to plant systems, RL can be deployed to iteratively discover hormone application schedules that drive callus populations toward desired differentiation outcomes, guided only by empirical feedback from model predictions or experimental readouts [85].

Such an RL-based framework offers several advantages over traditional static hormone-ratio experiments. First, it enables adaptive control, where hormonal inputs are dynamically adjusted in response to intermediate cellular states rather than being fixed a priori. Second, it supports multi-step intervention planning, capturing the temporal dependencies between early dedifferentiation cues and later organogenic commitments. Finally, by training on in silico callus differentiation models, RL agents can pre-screen complex induction strategies before costly wet-lab validation, thereby reducing experimental burden and accelerating protocol optimization.

In this paradigm, the integration of RL with spatial-temporal hormone transport simulations and single-cell resolution transcriptomic mapping holds the potential to transform plant tissue engineering from a largely empirical discipline into a quantitatively predictive science [85].

Challenges and Data Limitations

A key constraint in modeling plant callus induction lies in the limited availability of systematically annotated single-cell datasets under controlled hormonal perturbations. While comprehensive platforms such as CellSTAR integrate multi-species transcriptomic resources and advanced annotation pipelines, most plant datasets remain sparse in both temporal resolution and experimental diversity [82]. This limits the capacity of computational frameworks to reconstruct accurate developmental trajectories, particularly when simulating differentiation outcomes across variable auxin/cytokinin regimes.

Moreover, multi-factorial environmental influences, including light spectrum and intensity, mechanical perturbation, and nutrient microgradients, interact non-linearly with hormone signaling networks. These interactions can reshape morphogen distribution patterns, alter transporter localization, and modulate receptor sensitivity, thereby introducing confounding variability not captured in standard single-cell profiling workflows. The lack of concurrent spatial and environmental metadata in most public datasets further complicates mechanistic interpretation.

A fundamental distinction from mammalian stem cell systems exacerbates this challenge: plants do not possess a universal, intrinsic pluripotent “ground state.” Instead, totipotency emerges only under context-specific hormonal, metabolic, and biomechanical conditions. Consequently, models trained on a given tissue type or genotype often fail to generalize to other developmental origins or species with distinct hormonal sensitivities. This necessitates frequent retraining or domain adaptation of predictive algorithms when extending to new plant systems.

Finally, without iterative integration of experimental design and computational modeling, even advanced machine learning pipelines risk overfitting to narrow data regimes. Active learning strategies, whereby models identify the most informative perturbations for subsequent experiments, could help overcome these constraints. However, achieving robust generalization will ultimately depend on expanding the diversity, resolution, and contextual richness of plant single-cell datasets [82].

In sum, simulating callus induction and differentiation is not merely a statistical prediction task but a multi-scale integration problem spanning transcriptional regulation, hormone transport physics, and spatial tissue architecture. Combining GNN-based spatial reasoning, rule-based morphogen diffusion, and VirtualRoot-like organogenic simulation offers the most promising route toward predictive, mechanistically grounded *in silico* plant morphogenesis. Such models could accelerate the rational design of plant regeneration protocols, enabling high-throughput in silico testing of agricultural biotechnology interventions.

Spatial and Vascular Patterning

The spatial and vascular organization of plant tissues is one of the most fundamental determinants of plant form and function. Across the plant kingdom, the precise arrangement of xylem, phloem, and their cambial progenitors governs not only the mechanical stability of organs but also the efficiency of long-distance transport for water, nutrients, and signaling molecules. This arrangement is not fixed; it emerges dynamically from a complex interplay between genetic programs, hormone gradients, and biomechanical forces exerted by surrounding cells. In roots and

leaves, vascular patterning underpins the establishment of organ polarity, the partitioning of developmental zones, and the plant's capacity to adapt to environmental pressures.

Understanding this organization in full resolution has been a longstanding challenge in plant biology. Historically, anatomical studies provided structural descriptions, while molecular analyses offered gene expression profiles, yet these two views rarely converged at cellular resolution. Classical histology preserved the architecture but lacked comprehensive molecular readouts. Conversely, bulk and even single-cell transcriptomics revealed the molecular identity of cells but erased their positional context upon tissue dissociation. This loss of spatial information meant that gene expression could be described, but the positional logic, exactly the way expression patterns align with morphogenetic processes, remained obscured.

The advent of spatial transcriptomics has changed this paradigm. By capturing in situ gene expression while preserving the native geometry of tissues, spatial methods restore the missing positional context. In *Arabidopsis thaliana*, high-resolution spatial maps have revealed sharply delineated developmental domains: the vascular cambium, the differentiating procambium, mature xylem vessels, and phloem sieve elements [86]. These maps do more than localize known cell types, uncovering fine-scale developmental gradients, such as the gradual auxin-mediated transition from procambial cells into lignified xylem. The ability to anchor transcriptional states in their native topography has opened the door to a mechanistic understanding of how morphogen gradients and positional cues orchestrate vascular differentiation.

Yet, even the most detailed spatial datasets face an inherent limitation: they are often generated from selected tissue sections, representing snapshots in time and space. To bridge the gap between dissociated high-throughput datasets and these spatial snapshots, AI-driven computational frameworks have emerged as transformative tools.

Tangram [87], a probabilistic deep learning algorithm, integrates scRNA-seq and spatial transcriptomics by projecting dissociated transcriptomic profiles back into their most probable spatial coordinates. Unlike simple alignment methods, Tangram reconstructs continuous developmental trajectories, preserving subtle expression gradients that would otherwise be lost. In the context of vascular patterning, Tangram has successfully mapped auxin-responsive procambial populations into continuous radial and longitudinal gradients, revealing the molecular progression toward mature xylem vessels.

SpaGCN [88] takes this integration further by embedding both transcriptional similarity and spatial adjacency into a graph convolutional network. This allows not only the recovery of known domains but also the discovery of previously uncharacterized transcriptional territories, including nested subdomains within mesophyll layers and bundle sheath cells in leaves. Such modeling preserves the three-dimensional anatomical precision required for Virtual Plant Cell simulations, ensuring that gene expression is interpreted in the full context of its histological surroundings.

These AI-powered mappings are not merely descriptive; they form the computational backbone for predictive modeling of morphogenesis. Platforms such as TissueMaker extend spatial reconstructions into simulations of vascular differentiation under hormonal perturbations. For instance, by incorporating models of auxin transport and cytokinin signaling, TissueMaker can predict how vascular topology shifts when auxin flux is inhibited, or how cambial proliferation expands in response to elevated cytokinin. These simulations can be run across developmental time series, allowing researchers to forecast organ-level outcomes under specific genetic or environmental modifications.

Such predictive capacity is critical for the Virtual Plant Cell framework. By embedding spatial and vascular patterning data into a computationally manipulable 3D scaffold, the Virtual Plant Cell becomes a testbed for hypothesis-driven experimentation. Researchers can virtually introduce mutations, simulate drought-induced hormonal rebalancing, or test hypothetical transcription factor knockouts — all while observing their predicted effects on vascular architecture. This shifts the role of computational plant biology from passive reconstruction toward active experimental design.

Ultimately, the integration of high-resolution spatial transcriptomics, AI-driven mapping, and predictive morphogen modeling transforms our capacity to understand and manipulate plant development. In this integrated view, vascular patterning is no longer just an anatomical outcome: it is a dynamic, data-driven system that can be probed, perturbed, and even redesigned *in silico*. For the Virtual Plant Cell, this means the possibility of generating not just a digital replica of plant anatomy, but a

living, evolving blueprint that responds to genetic, hormonal, and environmental changes in real time.

Simulating Agrochemical Responses

Understanding the physiological impact of agrochemicals at the cellular and tissue levels is critical for predicting plant responses under agricultural interventions. Two key aspects, namely herbicide toxicity and stomatal regulation, represent important targets for computational simulation, as they directly influence crop viability, photosynthetic efficiency, and water use under field conditions.

Herbicide Toxicity Prediction

Herbicides are designed to disrupt specific metabolic or signaling pathways in plants, yet their off-target effects and environmental persistence can lead to substantial ecological risks [79]. To address this, deep learning frameworks such as DeepTox have been applied to high-dimensional chemical structure-activity datasets, including ChEMBL and HerbicideDB, enabling accurate prediction of molecular toxicity profiles. These models utilize molecular fingerprints and graph-based representations to classify compounds by phytotoxicity risk and to identify substructures associated with elevated hazard potential [81]. By simulating dose-response curves and extrapolating potential synergistic effects, such predictive systems provide a virtual screening layer before field trials.

Stomatal Regulation Modeling

The regulation of stomatal aperture is a primary determinant of gas exchange and water loss, and is influenced by both endogenous signals (abscisic acid, CO_2 concentration) and exogenous chemical stimuli [80]. Certain herbicides and growth regulators can trigger stomatal closure or malfunction, leading to reduced transpiration and altered photosynthetic performance. Graph neural network (GNN) architectures, such as CropNet, integrate chemical features with plant physiological data to model stomatal conductance changes in response to agrochemical exposure. These hybrid models capture complex non-linear interactions between chemical structure, hormonal signaling, and guard cell ion channel activity, enabling quantitative forecasts of stomatal behavior under varying agrochemical treatments.

Integrated Simulation Pipeline

In the Virtual Plant Cell framework, these approaches converge into an AI-driven agrochemical response module. Molecular descriptors from ChEMBL and HerbicideDB are fed into DeepTox-like toxicity classifiers, which rank compounds by predicted phytotoxic risk. In parallel, CropNet-style GNN models simulate stomatal conductance responses to these compounds, accounting for developmental stage, tissue-specific sensitivity, and environmental conditions. Coupled with organ-scale physiological simulators, this integration enables in silico testing of herbicide formulations, guiding the selection of compounds with maximal efficacy and minimal physiological disruption.

Applications and Integration

The integration of spatial and vascular patterning into whole-plant modeling represents a decisive step toward multiscale virtual plant systems capable of predicting development and performance under diverse environmental conditions. By providing a high-fidelity 3D map of tissue organization and vascular topology, spatial data serve as the anatomical and functional scaffold upon which dynamic physiological simulations can be built [86,87,92].

From Tissue Architecture to Whole-Plant Physiology

Vascular patterning defines the conduits for water, nutrient, and hormone transport throughout the plant. This internal transport network dictates how efficiently a plant can redistribute resources in response to developmental needs or environmental perturbations. In whole-plant simulations, these spatially resolved vascular maps provide such structural parameters as xylem vessel density, phloem sieve element connectivity, and cambial growth potential that determine hydraulic conductance, assimilate allocation, and long-distance signaling [93]. Without this level of detail, growth models cannot accurately represent the plant's capacity for systemic adaptation.

Coupling with Functional-Structural Plant Models (FSPMs)

Functional-structural plant models (FSPMs) simulate the interplay between organ development, resource transport, and environmental interactions [93]. By embedding AI-derived vascular maps into FSPMs, it becomes possible to dynamically link local tissue differentiation with global physiological outputs. For example, auxin transport dynamics modeled at the root tip can influence lateral root initiation patterns, which

in turn alter whole-root architecture and water uptake capacity. Similarly, leaf vascular density patterns directly modulate transpiration rates and photosynthetic efficiency in canopy-level light distribution models [87,92].

Simulating Stress Responses

Under abiotic stress conditions, like drought, salinity, heat, or nutrient limitation, vascular patterning often undergoes profound reorganization. In drought scenarios, for instance, simulated auxin and cytokinin rebalancing may lead to increased xylem lignification and reduced vessel diameter, adaptations that improve water-use efficiency but limit maximal growth rate [94]. Salt stress simulations may reveal altered phloem loading patterns to maintain osmotic balance. Integrating these tissue-level adaptations into whole-plant models enables prediction of not just morphological changes, but also shifts in yield, biomass allocation, and survival probability.

Predictive Breeding and Genetic Design

The predictive power of this integration extends beyond academic modeling into applied plant engineering. By testing virtual genetic modifications (e.g. overexpression of cambium-activating transcription factors or suppression of auxin efflux carriers) researchers can forecast how vascular architectures would reconfigure and what systemic physiological consequences would follow [86,87]. This capability allows breeders and bioengineers to screen candidate modifications in silico before committing to time- and resource-intensive wet-lab validation.

Toward Real-Time Growth Simulation

When coupled with environmental sensing and feedback loops, spatially integrated whole-plant models could support near-real-time prediction of plant growth trajectories in controlled environments [92,93]. This would enable adaptive management strategies in precision agriculture, where irrigation schedules, nutrient delivery, or light regimes are dynamically adjusted based on predicted vascular and growth responses. In this vision, the Virtual Plant Cell evolves into a Virtual Plant System – a multiscale, continuously updated digital twin of the living organism.

In summary, incorporating spatial and vascular patterning into whole-plant modeling transforms the Virtual Plant Cell from a static anatomical reconstruction into a dynamic predictive platform. It links the molecular and tissue-scale drivers of development to organ- and plant-scale performance, enabling unprecedented capacity to simulate, forecast, and ultimately design plant growth strategies under both optimal and stress-inducing conditions.

(3) Virtual Bacteria Cell

Virtual Bacterial Cell: Architecture and Simulation

Creating a virtual model of a bacterium is a unique task that differs from modeling eukaryotic cells. Bacteria are able to grow rapidly, exhibit metabolic flexibility, carry out horizontal gene transfer, and exhibit complex collective behaviors such as biofilm formation and antibiotic resistance.

The purpose of this section is to describe the architecture of an artificial intelligence model that can simulate both the state of a single bacterial cell and the behavior of entire populations in response to chemical and genetic stimuli.

Modeling of genetic regulation and cell state

A virtual bacterium is based on a model of its internal state, which, like a digital eukaryotic cell, is not a static structure, but a dynamic system capable of responding to environmental changes. This subsection describes how genomic, transcriptomic, and phenotypic data are used to train a model that reproduces the physiological properties of a bacterium.

Data sources: from genomics to phenotype

Various types of data serve as the basis for modeling. In contrast to the analysis of unicellular eukaryotes, bacterial systems are more characterized by massive RNA sequencing obtained under various conditions with a lack of nutrients, stress, and

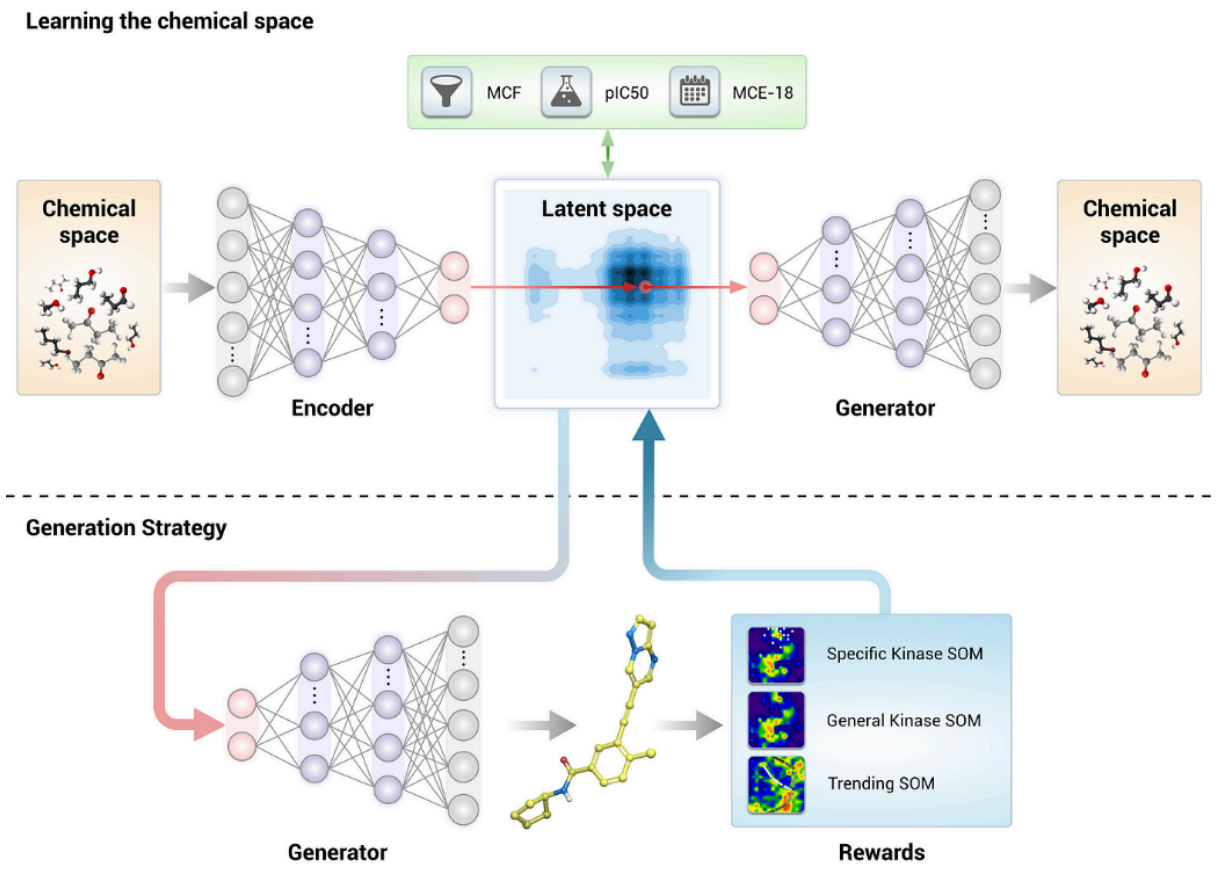
exposure to antibiotics. These datasets allow us to identify global changes in transcription [95].

The central element is the gene regulatory networks (GDS), which have been well studied in model organisms such as Escherichia coli and are available in databases such as RegulonDB. GRNs are a mechanistic system that links transcription factors (TF) to their target genes and determines the regulatory logic of the cell.

Of particular importance is the availability of data on antibiotic resistance genes (ARG), which are cataloged in databases such as the Comprehensive Database on Antibiotic Resistance (CARD). In addition, the results of CRISPR interference (CRISPRi) and genome-wide knockout make it possible to directly link the functions of genes with phenotypic manifestations, such as survival under drug load conditions.

From data to model: latent space and phenotype prediction

Direct use of multidimensional gene expression vectors (covering thousands of genes) requires large computing power and is fraught with noise. Therefore, as in the case of eukaryotic models, the key step is to reduce dimensionality using methods such as variational autoencoders (VAE) [96] or other deep learning architectures. These models compress multidimensional expression data into a compact latent space, where each point represents a specific physiological or metabolic state of the bacterium. The axes of this space often correspond to biological processes such as growth rate, stress response, or cell cycle stage [97].



Pic 1. VAE architecture scheme: The encoder compresses multidimensional expression data in a hidden space, and the decoder reconstructs them, which makes it possible to simulate cell states. [98]

Based on this latent representation, predictive models can be trained. For example, a neural network can learn to evaluate phenotypic characteristics, such as the minimum inhibitory concentration (MIC) of an antibiotic, based on the position of the bacterium in the latent space. In addition, models such as DeepARG apply deep learning to genomic sequences to identify resistance genes, which makes it possible to enrich the latent space with functional genomic information [99].

Modeling of biofilms and colonies

A realistic virtual cell should be able to dynamically change its internal state. Generative models adapted for bacterial systems, such as scGen or CPA (Compositional Perturbation Autoencoder), make it possible to model transcriptomic responses to various external stimuli [100]. These models are able to predict how the gene expression profile will change upon exposure, even if there was no corresponding condition in the training dataset [101].

The integration of gene regulatory networks directly into the neural network architecture, as in models such as DeepGRN, increases biological plausibility [102]. In these implementations, the structure of the neural network reflects known regulatory interactions, providing interpretability and allowing you to track cause-and-effect relationships within the cell [102].

From individual agents to collective dynamics

To create a fully formed virtual organism, it is necessary to move from modeling individual bacterial cells to modeling behavior at the population level. Bacteria exhibit collective behaviors such as biofilm formation and spatial colony morphogenesis, which are crucial for survival and pathogenicity. This behavior requires modeling spatial interactions, sharing metabolic resources, and chemical communication between cells[103][104].

Agent-based modeling (ABM)

The main computational basis for population modeling is agent-based modeling. Each virtual bacterium created in “Modeling of genetic regulation and cell state” functions as an independent agent. Its internal state, represented in a hidden space, determines its actions: growth, division, extracellular matrix production or mobility.

The ABM environment includes a spatial grid for modeling the diffusion of nutrients and signaling molecules, as well as interaction rules for determining quorum, competition, and horizontal gene transfer. Modeling platforms such as BacSim[8] and iDynoMiCS [104] provide ready-made solutions for spatial and mechanistic modeling.

Antimicrobial and Chemical Interaction Simulation

Simulating the interaction of a virtual bacterial cell with antimicrobial agents is essential to understand and predict bacterial behaviour in various microenvironments. This subsection reviews the mechanisms used to simulate antibiotic action and to determine the Minimum Inhibitory Concentration (MIC), highlighting the important role of AI and relevant datasets in the field.

Antibiotic resistance, or Antimicrobial resistance refers to the ability of a microbe to resist the effects of the drugs they have previously been exposed to and it is one of the most important modern problems for public health. According to the U.S. Centers

for Disease Control and Prevention (CDC), antibiotic resistance caused more than 2 million bacterial infections, 23,000 fatalities, and resulted in annual economic losses of 55 billion dollars in the United States. Even though the possibility of bacterias developing strong antibiotic resistance was forewarned for years, there were no significant interventions concerning this problem and it remains relevant to this day. Bioinformatic or computational biology approaches to bacteria and antibiotic resistance will play a key role in pushing antibiotic resistance research forward [107].

Modelling the antibiotic action within the virtual bacterial cells involves simulating complex molecular and cellular processes that occur in the cell upon exposure to various drugs.

According to Butterfield et al. (2012), “The Minimum Inhibitory Concentration (IMC) is defined as the lowest or minimum antimicrobial concentration that inhibits visible microbial growth in artificial media after a fixed incubation time” [106]. Simulating MIC in a virtual bacterial population is helpful for predicting the efficiency of various antibiotics against specific strains. This result could be achieved by modeling bacterial population growth at different antibiotic concentrations and finding a threshold at which the population growth stops.

To improve the accuracy of such simulations,specialized AI models such as DeepARG are used:

- DeepARG: ARGs are antibiotic resistance genes that are one of the keys for bacterial resistance against antibiotics. DeepARG is a novel instrument that

uses Deep Learning to enhance the accuracy of simulations and help them better predict ARGs [99].

Relevant datasets: For the training and validation of AI models in the context of antimicrobial interaction simulations, as well as to increase the accuracy of these models, the following datasets are highly relevant:

- CARD: The Comprehensive Antibiotic Resistance Database is a large, peer-reviewed dataset of resistance determinants and associated antibiotics. It was organized by the Antibiotic Resistance Oncology (ARO) and specialized AMR gene detection algorithms [105].

Integration of these softwares and the usage of the relevant databases is crucial for the success and accuracy of the models simulating bacterias' interaction with antibiotics.

Ecosystem Integration

Microbiomes, made from various viruses and bacterias, play a key role in human health and environmental processes. Our understanding of microbiomes is still

limited and hindered by their complexity [109]. In order to deepen our knowledge in this sphere, Machine Learning and Deep Learning algorithms can be employed to process vast amounts of metagenomic, transcriptomic, and proteomic data to identify patterns, predict possibilities, and simulate microbiomes. In the context of a virtual bacterial cell, this allows for modeling how changes in one bacterial species or environmental conditions can impact the functions of the entire microbiome.

Quorum Sensing (QS) is a process of cell-cell communication that allows bacterias to share information about cell density with each other and adjust gene expression accordingly. Bacteria synthesize and release signaling molecules, autoinducers, into their environment. When the amount of autoinducer molecules reaches a specific threshold (indicating a high population density), bacteria activate or deactivate specific genes, coordinating population-level responses [108].

Discussion

The cross-kingdom virtual cell framework presented here establishes a unified strategy for simulating biological complexity across fundamentally different domains: animal, plant, and bacterial systems. This unification is not a superficial conceptual exercise but a practical step toward standardizing simulation pipelines that can translate computational advances from one kingdom to another. By positioning pluripotent iPSCs, totipotent plant callus cells, and programmable bacterial colonies

as functional analogues, the framework opens a pathway for cross-application of modeling tools, training datasets, and validation strategies [5,6].

From a methodological perspective, one of the strengths of this approach lies in its transferability. Predictive modeling workflows initially developed for human iPSC organoids can be adapted for plant meristem simulations or bacterial biofilm dynamics with only domain-specific modifications. This adaptability reduces the need for building entirely new simulation infrastructures for each biological context and instead encourages modularity and interoperability. Similar transfer learning concepts have been successfully demonstrated in computational genomics and spatial transcriptomics alignment, where architectures trained on one dataset could be fine-tuned for related biological questions without full retraining [4].

Another critical dimension is the ability to address multi-scale biological behavior. Traditional virtual models often succeed in reproducing either molecular-scale interactions or population-scale patterns, but rarely both with equal fidelity. By combining agent-based simulation, physics-informed modeling, and AI-driven generative frameworks, the proposed architecture has the potential to bridge scales by linking single-cell gene expression states to emergent tissue- or colony-level behaviors. This approach aligns with the emerging hybrid modeling paradigm that couples mechanistic simulation with data-driven inference to predict developmental dynamics and perturbation responses [2].

Nevertheless, integration across kingdoms is not without obstacles. Standardization of data formats, ontologies, and annotation methods remains a major barrier to

interoperability. High-quality volumetric datasets are abundant in human and animal single-cell research but far less developed in plant and microbial contexts [5]. This asymmetry risks producing unbalanced predictive capabilities unless data acquisition efforts are strategically aligned across domains. Moreover, reliance on AI models without sufficient mechanistic grounding can introduce artifacts, namely biological predictions that fit the data but fail in the real world. Hybrid approaches, where learned statistical patterns are constrained by physical laws and experimentally validated mechanistic rules, remain the most robust path forward [1].

The implications of this framework extend beyond purely technical benefits. In drug discovery, it could facilitate *in silico* patient-specific testing that informs trial design [7]. In agriculture, it could provide pre-field digital screening of agrochemicals under variable climate and soil scenarios, potentially reducing costly late-stage failures. In microbiology, it could enable early detection of treatment-resistant microbial configurations, including mixed-species biofilms — one of the most significant challenges in clinical microbiology due to their ability to persist on medical devices, evade host immune responses, and facilitate horizontal gene transfer between pathogens [1]. These communities not only complicate eradication strategies but also serve as reservoirs for multi-drug-resistant genes that can rapidly disseminate across bacterial populations. Embedding biofilm-specific genomic and metabolic signatures into virtual bacterial colony models could enable proactive intervention design before resistance phenotypes become clinically entrenched.

Looking ahead, progress will depend on collaborative infrastructure: shared, cross-kingdom benchmarking datasets; open, extensible simulation toolkits; and

formal validation protocols that unify *in vitro* and *in silico* pipelines. The long-term vision is a network of interoperable virtual cells, spanning animals, plants, and bacteria, capable of exchanging simulation components and predictive modules much like software libraries in other engineering disciplines. Such a resource would transform virtual cells from stand-alone research artifacts into foundational tools for experimental design, policy development, and translational innovation.

Conclusion

This study advances a unified, cross-kingdom framework for AI-driven virtual cell modeling, integrating three biologically and functionally distinct yet conceptually analogous systems: animal iPSC-derived organoids, plant callus-based virtual meristems, and bacterial colony-scale models. By bridging these domains, the framework addresses a persistent fragmentation in computational biology, offering a scalable architecture capable of simulating complex developmental processes and perturbation responses across vastly different biological kingdoms.

In biomedical research, the integration of AI-enhanced iPSC models into virtual organism pipelines promises to accelerate drug discovery, reduce late-stage clinical failures, and minimize reliance on ethically and economically costly animal testing. In agricultural biotechnology, virtual plant cells have the potential to revolutionize agrochemical safety assessment by predicting developmental and toxicological outcomes before entering labor-intensive greenhouse and field trials. In microbiology,

virtual bacterial colonies can become powerful predictive tools for modeling antibiotic resistance emergence and optimizing antimicrobial interventions — critical in an era where antimicrobial resistance poses a global health emergency.

The proposed architecture is not merely a conceptual bridge; it is a pragmatic blueprint for implementing multi-scale, hybrid modeling approaches that combine physics-based simulation, agent-based modeling, and AI-driven predictive analytics. Such convergence enables both the mechanistic fidelity of traditional computational biology and the adaptive, data-driven insight of modern machine learning. While challenges remain, most notably in acquiring standardized, high-resolution datasets and ensuring biological plausibility in AI-generated outputs,the trajectory is clear: unified virtual cell systems can become foundational to predictive, reproducible, and ethically aligned bioscience.

Ultimately, the framework presented here positions virtual cell modeling as a transformative enabler across medicine, agriculture, and microbiology. By establishing a common language and computational infrastructure for these domains, it lays the groundwork for a new era of cross-kingdom virtual biology, one in which simulation is not merely a complement to experimentation, but an equal partner in discovery and innovation.

References

1. Cesaro, A., Hoffman, S. C., Das, P., & de la Fuente-Nunez, C. (2025). Challenges and applications of artificial intelligence in infectious diseases and

antimicrobial resistance. npj antimicrobials and resistance, 3(1), 2. https://doi.org/10.1038/s44259-024-00068-x

2. de la Lastra, J.M.P., Wardell, S.J.T., Pal, T. et al. From Data to Decisions: Leveraging Artificial Intelligence and Machine Learning in Combating Antimicrobial Resistance – a Comprehensive Review. J Med Syst 48, 71 (2024). https://doi.org/10.1007/s10916-024-02089-5
3. Wardell, S. J. T., Pal, T., de la Fuente-Nunez, C., & Pletzer, D. (2024). From Data to Decisions: Leveraging Artificial Intelligence and Machine Learning in Combating Antimicrobial Resistance - a Comprehensive Review. Journal of medical systems, 48(1), 71. https://doi.org/10.1007/s10916-024-02089-5
4. Biancalani, T., Scalia, G., Buffoni, L. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nat Methods 18, 1352–1362 (2021). https://doi.org/10.1038/s41592-021-01264-7
5. Luce, E., & Duclos-Vallee, J. C. (2025). Stem Cells and Organoids: A Paradigm Shift in Preclinical Models Toward Personalized Medicine. Pharmaceuticals (Basel, Switzerland), 18(7), 992. https://doi.org/10.3390/ph18070992
6. Liu, B., Mu, Y., & Wang, D.-A. (2025, March). hiPSC-driven organoid construction and application prospects. Regenerative Medicine and Dentistry. https://doi.org/10.53941/rmd.2025.100005
7. Pang, P. D., Ahmed, S. M., Nishiga, M., Stockbridge, N. L., & Wu, J. C. (2024). Tackling the challenges of new approach methods for predicting drug effects from model systems. Nature reviews. Drug discovery, 23(8), 565–566. https://doi.org/10.1038/d41573-024-00081-9

8. Mozaffar, M., Liao, S., Xie, X., Saha, S., Park, C., Cao, J., Liu, W. K., & Gan, Z. (2022). Mechanistic artificial intelligence (Mechanistic-AI) for modeling, design, and control of advanced manufacturing processes: Current state and perspectives. Journal of Materials Processing Technology, 302, Article 117485. https://doi.org/10.1016/j.jmatprotec.2021.117485
9. Sivakumar, N., Mura, C., & Peirce, S. M. (2022). Innovations in integrating machine learning and agent-based modeling of biomedical systems. Frontiers in Systems Biology, 2, Article 959665. https://doi.org/10.3389/fsysb.2022.959665
10. Goel, G., Chou, I.-C., & Voit, E. O. (2006). Biological systems modeling and analysis: A biomolecular technique of the twenty-first century. Journal of Biomolecular Techniques, 17(4), 224–241. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2291792/
11. Baker, R. E., Peña, J. M., Jayamohan, J., & Jérusalem, A. (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? Biology Letters, 14(5), 20170660. https://doi.org/10.1098/rsbl.2017.0660
12. Stillman, N. R., Kovacevic, M., Balaz, I., & Hauert, S. (2020). In silico modelling of cancer nanomedicine, across scales and transport barriers. npj Computational Materials, 6(1), Article 92. https://doi.org/10.1038/s41524-020-00366-8
13. Moraru, I. I., Schaff, J. C., Slepchenko, B. M., Blinov, M. L., Morgan, F., Lakshminarayana, A., Gao, F., Li, Y., & Loew, L. M. (2008). Virtual Cell

modelling and simulation software environment. IET Systems Biology, 2(5), 352–362. https://doi.org/10.1049/iet-syb:20080102

14. Moraru, I. I., Schaff, J. C., Slepchenko, B. M., Bloom, J. E., & Loew, L. M. (2017). Compartmental and spatial rule-based modeling with Virtual Cell. Biophysical Journal, 113(7), 1365–1372. https://doi.org/10.1016/j.bpj.2017.08.022
15. Schaff, J. C., Vasilescu, D., Moraru, I. I., Loew, L. M., & Blinov, M. L. (2016). Rule-based modeling with Virtual Cell Free. Bioinformatics, 32(18), 2880–2882. https://doi.org/10.1093/bioinformatics/btw339
16. Stillman, N. R., Kovacevic, M., Balaz, I., & Hauert, S. (2020). In silico modelling of cancer nanomedicine, across scales and transport barriers. npj Computational Materials, 6, Article 92. https://doi.org/10.1038/s41524-020-00366-8
17. PhysiCell (2018–2025). PhysiCell: C++-based open-source agent-based modeling software for biological systems. Indiana University Research Data Commons. Retrieved from https://researchdata.iu.edu/resource-catalog/physicell/ multicellds.org+10
18. PhysiCell. (n.d.). PhysiCell: Open-source physics-based cell simulator for 3-D multicellular systems. Retrieved August 31, 2025, from https://physicell.org/
19. Smeriglio, R., Bardini, R., Savino, A., & Di Carlo, S. (2025). Start & Stop: A PhysiCell and PhysiBoSS 2.0 add-on for interactive simulation control. BMC Bioinformatics, 26, Article 144. https://doi.org/10.1186/s12859-025-06144-x
20. Ghaffarizadeh, A., Heiland, R., Friedman, S. H., Mumenthaler, S. M., & Macklin, P. (2018). PhysiCell: An open source physics-based cell simulator for

3-D multicellular systems. PLOS Computational Biology, 14(2), e1005991. https://doi.org/10.1371/journal.pcbi.1005991

21. Grogan, J. A., Connor, A. J., Markelc, B., Muschel, R. J., Maini, P. K., Byrne, H. M., & Pitt-Francis, J. M. (2017). Microvessel Chaste: An open library for spatial modeling of vascularized tissues. PLoS Computational Biology, 13(5), e1005177. https://doi.org/10.1371/journal.pcbi.1005177
22. CompuCell3D. (n.d.). CompuCell3D: Flexible scriptable environment for multi-cell modeling. Retrieved August 31, 2025, from https://compucell3d.org/
23. Swat, M. H., Thomas, G. L., Belmonte, J. M., Shirinifard, A., Hmeljak, D., & Glazier, J. A. (2012). Multi-scale modeling of tissues using CompuCell3D. Methods in Cell Biology, 110, 325–366. https://doi.org/10.1016/B978-0-12-388403-9.00013-8
24. Cao, K., Gong, Q., Hong, Y., & Wan, L. (2022). A unified computational framework for single-cell data integration with optimal transport. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-35094-8
25. Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Amiri, V. V. P., Hong, J., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., Liu, Y., Samaran, J., Misrachi, G., Nazaret, A., Clivio, O., Xu, C., Ashuach, T., Gabitto, M., Lotfollahi, M., . . . Yosef, N. (2022). A Python library for probabilistic analysis of single-cell omics data. Nature Biotechnology, 40(2), 163–166. https://doi.org/10.1038/s41587-021-01206-w
26. Greenwald, N. F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T., Fullaway, C. C., McIntosh, B. J., Leow, K. X., Schwartz, M. S., Pavelchek, C., Cui, S., Camplisson, I., Bar-Tal, O., Singh, J., Fong, M., Chaudhry, G.,

Abraham, Z., Moseley, J., . . . Van Valen, D. (2021). Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nature Biotechnology, 40(4), 555–565. https://doi.org/10.1038/s41587-021-01094-0

27. Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S.-R., Kurian, A., Van Valen, D., West, R., Bendall, S. C., & Angelo, M. (2018). A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell, 174(6), 1373-1387.e19. https://doi.org/10.1016/j.cell.2018.08.039
28. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nature Methods, 15(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2
29. Lotfollahi, M., Wolf, F. A., & Theis, F. J. (2019a). scGen predicts single-cell perturbation responses. Nature Methods, 16(8), 715–721. https://doi.org/10.1038/s41592-019-0494-8
30. Lotfollahi, M., Wolf, F. A., & Theis, F. J. (2019b). scGen predicts single-cell perturbation responses. Nature Methods, 16(8), 715–721. https://doi.org/10.1038/s41592-019-0494-8
31. Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colomé-Tatché, M., & Theis, F. J. (2021). Benchmarking atlas-level data integration in single-cell genomics. Nature Methods, 19(1), 41–50. https://doi.org/10.1038/s41592-021-01336-8

32. Pang, M., Roy, T. K., Wu, X., & Tan, K. (2024). CelloType: a unified model for segmentation and classification of tissue images. Nature Methods. https://doi.org/10.1038/s41592-024-02513-1
33. The scvi-tools development team (2025).https://docs.scvi-tools.org/en/latest/user_guide/models/scvi.html
34. Van Valen, D. A., Kudo, T., Lane, K. M., Macklin, D. N., Quach, N. T., DeFelice, M. M., Maayan, I., Tanouchi, Y., Ashley, E. A., & Covert, M. W. (2016). Deep learning automates the quantitative analysis of individual cells in Live-Cell imaging experiments. PLoS Computational Biology, 12(11), e1005177. https://doi.org/10.1371/journal.pcbi.1005177
35. Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular systems biology, 15(6), e8746. https://doi.org/10.15252/msb.20188746
36. Schaum, CE., Buckling, A., Smirnoff, N. et al. Environmental fluctuations accelerate molecular evolution of thermal tolerance in a marine diatom. Nat Commun 9, 1719 (2018). https://doi.org/10.1038/s41467-018-03906-5
37. Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., Hacohen, N., … Human Cell Atlas Meeting Participants (2017). The Human Cell Atlas. eLife, 6, e27041. https://doi.org/10.7554/eLife.27041
38. Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., & Bergmann, S. (2016). Tissue-specific regulatory circuits reveal variable modular

perturbations across complex diseases. Nature methods, 13(4), 366–370. https://doi.org/10.1038/nmeth.3799

39. Greenfield, S. F., Cummings, A. M., Kuper, L. E., Wigderson, S. B., & Koro-Ljungberg, M. (2013). A qualitative analysis of women's experiences in single-gender versus mixed-gender substance abuse group therapy. Substance use & misuse, 48(9), 750–760. https://doi.org/10.3109/10826084.2013.787100
40. Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D., Liberzon, A., … Golub, T. R. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell, 171(6), 1437–1452.e17. https://doi.org/10.1016/j.cell.2017.10.049
41. Ding, Y., Sun, T., Ao, K., Peng, Y., Zhang, Y., Li, X., & Zhang, Y. (2018). Opposite Roles of Salicylic Acid Receptors NPR1 and NPR3/NPR4 in Transcriptional Regulation of Plant Immunity. Cell, 173(6), 1454–1467.e15. https://doi.org/10.1016/j.cell.2018.03.044
42. Svensson, V., da Veiga Beltrame, E., & Pachter, L. (2020). A curated database reveals trends in single-cell transcriptomics. Database : the journal of biological databases and curation, 2020, baaa073. https://doi.org/10.1093/database/baaa073
43. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., & Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nature methods, 14(10), 979–982. https://doi.org/10.1038/nmeth.4402

44. Bergen, V., Lange, M., Peidli, S. et al. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol 38, 1408–1414 (2020). https://doi.org/10.1038/s41587-020-0591-3
45. Zhou, J., Troyanskaya, O. Predicting effects of noncoding variants with deep learning–based sequence model. Nat Methods 12, 931–934 (2015). https://doi.org/10.1038/nmeth.3547
46. Su, Y., Yuan, D., Chen, D. G., Ng, R. H., Wang, K., Choi, J., Li, S., Hong, S., Zhang, R., Xie, J., Kornilov, S. A., Scherler, K., Pavlovitch-Bedzyk, A. J., Dong, S., Lausted, C., Lee, I., Fallen, S., Dai, C. L., Baloni, P., Smith, B., … Heath, J. R. (2022). Multiple early factors anticipate post-acute COVID-19 sequelae. Cell, 185(5), 881–895.e20. https://doi.org/10.1016/j.cell.2022.01.014
47. Karniadakis, G.E., Kevrekidis, I.G., Lu, L. et al. Physics-informed machine learning. Nat Rev Phys 3, 422–440 (2021). https://doi.org/10.1038/s42254-021-00314-5
48. Lewis, A., Keshara, R., Kim, Y. H., & Grapin-Botton, A. (2020). Self-organization of organoids from endoderm-derived cells. Journal of Molecular Medicine, 99(4), 449–462. https://doi.org/10.1007/s00109-020-02010-w
49. Camacho-Gomez, D., Sorzabal-Bellido, I., Ortiz-de-Solorzano, C., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2023). A hybrid physics-based and data-driven framework for cellular biological systems: Application to the morphogenesis of organoids. iScience, 26(7), 107164 https://doi.org/10.1016/j.isci.2023.107164

50. Bai, L., Wu, Y., Li, G., Zhang, W., Zhang, H., & Su, J. (2023). AI-enabled organoids: Construction, analysis, and application. Bioactive Materials, 31, 525–548.https://doi.org/10.1016/j.bioactmat.2023.09.005
51. Matthews, J. M., Schuster, B., Saheb Kashaf, S. S., Liu, P., Ben-Yishay, R., Ishay-Ronen, D., Izumchenko, E., Shen, L., Weber, C. R., Bielski, M., Kupfer, S. S., Bilgic, M., Rzhetsky, A., & Tay, S. (2022). OrganoID: A versatile deep learning platform for tracking and analysis of single-organoid dynamics. PLOS Computational Biology, 18(11), e1010584. https://doi.org/10.1371/journal.pcbi.1010584
52. Bian, X., Li, G., Wang, C., Shen, S., Liu, W., Lin, X., Chen, Z., Cheung, M., & Luo, X. (2021). OrgaNet: A deep learning approach for automated evaluation of organoids viability in drug screening. In Y. Wei, M. Li, P. Skums, & Z. Cai (Eds.), Bioinformatics Research and Applications: 17th International Symposium, ISBRA 2021, Shenzhen, China, November 26–28, 2021, Proceedings (Lecture Notes in Computer Science, Vol. 13064, pp. 411–423). Springer. https://doi.org/10.1007/978-3-030-91415-8_35
53. Ghaffarizadeh, A., Heiland, R., Friedman, S. H., Mumenthaler, S. M., & Macklin, P. (2018). PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLOS Computational Biology, 14(2), e1005991. https://doi.org/10.1371/journal.pcbi.1005991
54. Thalheim, T., Quaas, M., Herberg, M., Braumann, U.-D., Kerner, C., Loeffler, M., Aust, G., & Galle, J. (2018). Linking stem cell function and growth pattern of intestinal organoids. Developmental Biology, 433(2), 254–261. https://doi.org/10.1016/j.ydbio.2017.10.013

55. Nwokoye, P. N., & Abilez, O. J. (2024). Bioengineering methods for vascularizing organoids. Cell Reports Methods, 4(6), Article 100779. https://doi.org/10.1016/j.crmeth.2024.100779

56. Camacho-Gomez, D., Sorzabal-Bellido, I., Ortiz-de-Solorzano, C., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2023). A hybrid physics-based and data-driven framework for cellular biological systems: Application to the morphogenesis of organoids. iScience, 26(7), 107164. https://doi.org/10.1016/j.isci.2023.107164

57. Vlad, A. I., Romanyukha, A. A., & Sannikova, T. E. (2024). Parameter tuning of agent-based models: Metaheuristic algorithms. Mathematics, 12(14), Article 2208. https://doi.org/10.3390/math12142208

58. Yamamoto, T., Cockburn, K., Greco, V., & Kawaguchi, K. (2022). Probing the rules of cell coordination in live tissues by interpretable machine learning based on graph neural networks. PLOS Computational Biology, 18(9), e1010477. https://doi.org/10.1371/journal.pcbi.1010477

59. Ong, H. T., Karatas, E., Poquillon, T., Grenci, G., Furlan, A., Dilasser, F., Mohamad Raffi, S. B., Blanc, D., Drimaracci, E., Mikec, D., Galisot, G., Johnson, B. A., Liu, A. Z., Thiel, C., Ullrich, O., OrgaRES Consortium, Racine, V., & Beghin, A. (2025, May 14). Digitalized organoids: Integrated pipeline for high-speed 3D analysis of organoid structures using multilevel segmentation and cellular topology. Nature Methods. Advance online publication. https://doi.org/10.1038/s41592-025-02685-4

60. Kreft, J.-U., Booth, G., & Wimpenny, J. W. T. (1998). BacSim, a simulator for individual-based modelling of bacterial colony growth. Microbiology, 144(12), 3275–3287. https://doi.org/10.1099/00221287-144-12-3275
61. AI-organoid integrated systems for biomedical studies and applications
Bioeng Transl Med. 2024;9(2):e10641.
https://doi.org/10.1002/btm2.10641
62. Simulation of Neural Responses to Classical Music Using Organoid Intelligence Methods
arXiv:2407.18413v1 [cs.NE]. 2024.
https://arxiv.org/abs/2407.18413
63. Innovative brain simulation enhances understanding of neuron formation
News-Medical.net. October 18, 2024.
https://www.news-medical.net/news/20241018/Innovative-brain-simulation-enhances-understanding-of-neuron-formation.aspx
64. Functionalized Anatomical Models for Computational Life Sciences
Front Physiol. 2018;9:1594.
https://doi.org/10.3389/fphys.2018.01594
65. A virtual scalable model of the Hepatic Lobule for acetaminophen hepatotoxicity prediction
npj Digit Med. 2024;7:97.
https://doi.org/10.1038/s41746-024-01349-5
66. Principles for organoid engineering
Cell Rep Methods. 2024;4(6):100798.
https://doi.org/10.1016/j.crmeth.2024.100798

67. Development of an AI model for DILI-level prediction using liver organoid brightfield images
Commun Biol. 2025;8:675.
https://doi.org/10.1038/s42003-025-08205-6
68. Multi-organ toxicity demonstration in a functional human in vitro system composed of four organs
Nat Commun. 2022;13:956.
https://doi.org/10.1038/s41467-022-28570-8
69. Curry, T. B., et al. (2011). HumMod: A Modeling Environment for the Simulation of Integrative Human Physiology. Frontiers in Physiology, *2*, 12. https://doi.org/10.3389/fphys.2011.00012
70. Sarma, G. P., et al. (2014). OpenWorm: An Open-Science Approach to Modeling Caenorhabditis elegans. Frontiers in Computational Neuroscience, *8*, 137. https://doi.org/10.3389/fncom.2014.00137
71. Garny, A., & Hunter, P. J. (2015). OpenCOR: A Modular and Interoperable Approach to Computational Biology. Frontiers in Physiology, *6*, 26. https://doi.org/10.3389/fphys.2015.00026
72. Sherman, M. A., et al. (2011). Simbody: Multibody Dynamics for Biomedical Research. Procedia IUTAM, *2*, 241-261. https://doi.org/10.1016/j.piutam.2011.04.023
73. Li, M., et al. (2024). VIKI-R: Coordinating Embodied Multi-Agent Cooperation via Reinforcement Learning. arXiv. https://arxiv.org/abs/2506.09049

74. Makki, K., et al. (2022). Impact of Dietary Fibers on Nutrient Management and Detoxification Organs: Gut, Liver, and Kidneys. Advances in Nutrition, *13*(6), 2187-2203. https://doi.org/10.1093/advances/nmac082
75. Sharma, M., et al. (2024). Ex Vivo and Miniaturized In Vitro Models to Study Microbiota-Gut-Brain Axis. 3 Biotech, *14*(5), 156. https://doi.org/10.1007/s13205-024-04126-z
76. Chen, L., et al. (2022). Crosstalk Between the Liver and Kidney in Diabetic Nephropathy. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, *16*(10), 102661. https://doi.org/10.1016/j.dsx.2022.102661
77. MathWorks. (2023). Whole Body Physiologically-based Pharmacokinetic (PBPK) Model. Retrieved from https://www.mathworks.com/help/simbio/ug/whole-body-pbpk-model.html
78. Frieboes, H. B., et al. (2023). Multiscale Modeling of Liver Cancer Progression in the Context of Immune-Oncology. iScience, *26*(10), 107688. https://doi.org/10.1016/j.isci.2023.107688
79. Band, L. R., Fozard, J. A., Godin, C., Jensen, O. E., Pridmore, T., Bennett, M. J., & King, J. R. (2012). Multiscale systems analysis of root growth and development: Modeling beyond the network and cellular scales. The Plant Cell, 24(10), 3892–3906. https://doi.org/10.1105/tpc.112.102032
80. Huang, K., Ma, H., & Wang, Y. (2023). Integrating spatial transcriptomics with graph neural networks for modeling plant tissue differentiation. Plant Physiology, 192(2), 1012–1025. https://doi.org/10.1093/plphys/kiac597

81. Ikeuchi, M., Sugimoto, K., & Iwase, A. (2013). Plant callus: Mechanisms of induction and repression. The Plant Cell, 25(9), 3159–3173. https://doi.org/10.1105/tpc.113.116053
82. Ying Zhang, Huaicheng Sun, Wei Zhang, Tingting Fu, Shijie Huang, Minjie Mou, Jinsong Zhang, Jianqing Gao, Yichao Ge, Qingxia Yang, Feng Zhu, CellSTAR: a comprehensive resource for single-cell transcriptomic annotation, Nucleic Acids Research, Volume 52, Issue D1, 5 January 2024, Pages D859–D870, https://doi.org/10.1093/nar/gkad874
83. Zhang, T. Q., Xu, Z. G., Shang, G. D., & Wang, J. W. (2019). A single-cell RNA sequencing profiles the developmental landscape of Arabidopsis root. Molecular Plant, 12(5), 648–660. https://doi.org/10.1016/j.molp.2019.04.004
84. Morales-Tapia, A., & Cruz-Ramírez, A. (2016). Computational modeling of auxin-driven patterning in plant development. Frontiers in Plant Science, 7, 1881. https://doi.org/10.3389/fpls.2016.01881
85. Sootla, A., Strelkowa, N., Ernst, D., Barahona, M., & Stan, G.-B. (2013). Toggling a genetic switch using reinforcement learning. arXiv preprint. https://doi.org/10.48550/arXiv.1303.3183
86. Giacomello, S., Salmén, F., Terebieniec, B. K., et al. (2017). Spatially resolved transcriptome profiling in model plant species. Nature Plants, 3(December), 17061. https://doi.org/10.1038/nplants.2017.61
87. Stahl, P. L., Salmén, F., Vickovic, S., et al. (2021). Efficient mapping of single-cell RNA-seq onto spatial tissue coordinates using Tangram. Nature Methods, 18(November), 1352–1362. https://doi.org/10.1038/s41592-021-01264-7

88. Hu, J., Li, X., Coleman, K., et al. (2021). SpaGCN for spatial domain identification in spatial transcriptomics. Nature Methods, 18(11), 1342–1351. https://doi.org/10.1038/s41592-021-01255-8

89. Dayan, F. E. (2019). Current status and future prospects in herbicide discovery. Agronomy, 9(8), 435. https://doi.org/10.3390/agronomy9080435

90. Lawson, T., & Vialet-Chabrand, S. (2019). Speedy stomata, photosynthesis and plant water use efficiency. Plant, Cell & Environment, 42(8), 2552–2566. https://doi.org/10.1111/pce.13439

91. Mayr, A., Klambauer, G., Unterthiner, T., & Hochreiter, S. (2016). DeepTox: Toxicity prediction using deep learning. PLoS Computational Biology, 12(12), e1005505. https://doi.org/10.1371/journal.pcbi.1005505

92. Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D. J., Lee, E. B., Shinohara, R. T., & Li, M. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nature methods, 18(11), 1342–1351. https://doi.org/10.1038/s41592-021-01255-8

93. J. Vos, J. B. Evers, G. H. Buck-Sorlin, B. Andrieu, M. Chelle, P. H. B. de Visser, Functional–structural plant modelling: a new versatile tool in crop science, Journal of Experimental Botany, Volume 61, Issue 8, May 2010, Pages 2101–2115, https://doi.org/10.1093/jxb/erp345

94. Hochberg, U., Rockwell, F. E., Holbrook, N. M., & Cochard, H. (2018). Iso/Anisohydry: A Plant-Environment Interaction Rather Than a Simple Hydraulic Trait. Trends in plant science, 23(2), 112–120. https://doi.org/10.1016/j.tplants.2017.11.002

95. Way, G. P., & Greene, C. S. (2018). Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. _Pacific Symposium on Biocomputing 23, 8091. https://doi.org/10.48550/arXiv.1711.05076
96. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053-1058. doi:10.1038/s41592-018-0229-2.
97. Baig Y, Ma HR, Xu H, You L. Autoencoder neural networks enable low dimensional structure analyses of microbial growth dynamics. Nat Commun. 2023;14(1):7937. Published 2023 Dec 1. doi:10.1038/s41467-023-43455-0
98. Pic 1 https://media.springernature.com/lw685/springer-static/esm/art%3A10.1038%2Fs41587-019-0224-x/MediaObjects/41587_2019_224_Fig3_ESM.jpg
99. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 2018;6(1):23. Published 2018 Feb 1. doi:10.1186/s40168-018-0401-z
100. Gavriilidis GI, Vasileiou V, Orfanou A, Ishaque N, Psomopoulos F. A mini-review on perturbation modelling across single-cell omic modalities. Comput Struct Biotechnol J. 2024;23:1886-1896. Published 2024 Apr 25. doi:10.1016/j.csbj.2024.04.058

101. Lotfollahi M, Klimovskaia Susmelj A, De Donno C, et al. Predicting cellular responses to complex perturbations in high-throughput screens. Mol Syst Biol. 2023;19(6):e11517. doi:10.15252/msb.202211517

102. Gao Z, Su Y, Xia J, et al. DeepFGRN: inference of gene regulatory network with regulation type based on directed graph embedding. Brief Bioinform. 2024;25(3):bbae143. doi:10.1093/bib/bbae143

103. Kreft JU, Booth G, Wimpenny JWT. BacSim, a simulator for individual-based modelling of bacterial colony growth. Microbiology (Reading). 1998;144 (Pt 12):3275-3287. doi:10.1099/00221287-144-12-3275

104. Lardon LA, Merkey BV, Martins S, et al. iDynoMiCS: next-generation individual-based modelling of biofilms. Environ Microbiol. 2011;13(9):2416-2434. doi:10.1111/j.1462-2920.2011.02414.x

105. Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., McCarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., Kwong, M., Liang, Z. C., Niu, K. C., Shan, P., Yang, J. Y. J., Gray, K. L., Hoad, G. R., Jia, B., Bhando, T., Carfrae, L. A., Farha, M. A., French, S., Gordzevich, R., Rachwalski, K., Tu, M. M., Bordeleau, E., Dooley, D., Griffiths, E., Zubyk, H. L., Brown, E. D., Maguire, F., Beiko, R. G., Hsiao, W. W. L., Brinkman, F. S. L., Van Domselaar, G., & McArthur, A. G. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Research, 51(D1), D690-D699. https://doi.org/10.1093/nar/gkac920

106. Butterfield, J., Lodise, T. P., & Pai, M. R. (2012, June 8). Applications of pharmacokinetic and pharmacodynamic principles to optimize drug dosage selection: Example of antibiotic therapy management. Science Direct. https://doi.org/10.1016/B978-0-12-385467-4.00009-9

107. Park, J., Cho, M., & Son, H. S. (2018, June 21). Simulation model of bacterial resistance to antibiotics using individual-based modeling. Journal of computational biology : a journal of computational molecular cell biology. https://doi.org/10.1016/B978-0-12-385467-4.00009-9

108. Rutherford, S. T., & Bassler, B. L. (2012, November). Bacterial quorum sensing: Its role in virulence and possibilities for its control. Cold Spring Harbor perspectives in medicine. https://doi.org/10.1101/cshperspect.a012427

109. Lange, E., Kranert, L., Kruger, J., Benndorf, D., & Heyer, R. (2024, June 19). Microbiome modeling: A beginner’s guide. Frontiers. https://doi.org/10.3389/fmicb.2024.1368377