Biotechnology Advances 92 (2026) 108962

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Review article

The role of computational cellular models in industrial bioprocesses: From genetic engineering to plant design

Sergio Garcia

School of Science and Technology, IE University, Paseo de la Castellana, 259E, 28046 Madrid, Spain

ARTICLE INFO

Keywords:
Cellular modeling
Genome-scale models
Machine learning
Hybrid modeling
Multiscale modeling
Techno-economic analysis
Sustainable biomanufacturing

ABSTRACT

Engineering cellular metabolism has enabled bioprocesses to manufacture sustainable fuels, materials, foods, commodity chemicals, and pharmaceuticals. Yet much of this potential remains unrealized because cellular behavior is complex and decisions must be coordinated across scales from the cell to the plant. Computational modeling therefore becomes essential to navigate this design space and to turn biological capability into reliable, scalable processes. In this review we examine the main cellular modeling approaches, including mechanistic, statistical and hybrid models. We cover classic macroscopic models through genome-scale models incorporating protein synthesis, regulation, and kinetics, together with statistical approaches that learn from databases, high-throughput experimentation, and online sensing. We compare model families by assumptions, data and curation needs, and computational tractability, and map how they guide genetic edits, medium and feeding choices, reactor design and operation, and early linkage to techno-economic and life cycle assessments. Three themes emerge from this synthesis: fitness to purpose matters more than novelty because simple kinetic models remain central for screening and control, while advanced cellular models reveal physiological limits that prioritize what to engineer; value grows when predictions are carried forward to process and plant metrics so economic and environmental trade offs are visible early; and finally, tractability governs multiscale integration which motivates model reduction, precomputation, and fast surrogates. As data pipelines and automation mature, cellular models serve as a shared language across scales, translating biological insight into strain design, feed formulations, control policies, and plant designs that are economically viable and environmentally responsible.

1. Introduction

There is a renewed interest in bioprocessing and industrial fermentation as manufacturing pivots toward sustainable fuels (Benavides et al., 2024; Cui et al., 2024), materials (Samadhiya et al., 2022; Varghese et al., 2022; Benavides et al., 2024), foods (Eastham and Leman, 2024; Goodwin et al., 2024), commodity chemicals (Kim et al., 2023; Cowan et al., 2023), and pharmaceuticals (Love et al., 2024; Vo and Trinh, 2025). This shift is enabled by advances in synthetic biology and metabolic engineering that reprogram cellular metabolism to perform targeted conversions. Progress, however, is often hindered by gaps in common standards and interoperable data, slow and costly process characterization with limited automation and real-time measurements, challenges in transferring processes across scales and sites, and supply-chain and feedstock variability (Asin-Garcia et al., 2025). To address these constraints, data analysis and simulation models must play a larger role throughout bioprocess design.

Many industrial bioprocesses depend on an engineered cell factory for converting raw substrates into desired products. Microbial workhorses like bacteria and yeasts underpin sustainable routes to fuels, materials, and chemicals, complemented by filamentous fungi for enzyme and organic-acid production and by mammalian cell lines for therapeutic proteins (Cho et al., 2022; Futing et al., 2021; Gonzalez-Hernandez and Perre, 2024). Achieving economic and sustainable performance depends on coordinated design across three levels: the cell strain (genetic and regulatory features controlling metabolism), the bioreactor (vessel geometry and operational policies for feeding, aeration, mixing, and temperature), and the overall plant (integration of fermentation with purification, utilities, and supply chains). Because these decisions couple phenomena from intracellular flux allocation to reactor dynamics and plant logistics, mathematical models have become essential tools. In brief, the modeling toolbox has progressed from empirical growth laws and lumped balances to constraint-based genome-scale formalisms and dynamic, resource-aware descriptions of metabolism, now increasingly paired with hybrid and digital-twin approaches that combine first-principles structure with machine learning

E-mail address: sergio.garciamanogil@ie.edu.

https://doi.org/10.1016/j.biotechadv.2026.108962
Received 15 January 2026; Received in revised form 29 June 2026; Accepted 1 July 2026
Available online 11 July 2026

List of abbreviations

AI	artificial intelligence
ANN	artificial neural network
CBM	constraint-based metabolic model
CFD	computational fluid dynamics
CHO	Chinese hamster ovary
dFBA	dynamic FBA
ETFL	Expression and Thermodynamics Flux
FBA	flux balance analysis
GECKO	Genome-scale model to account for Enzyme Constraints, using Kinetics and Omics
GEM	genome-scale metabolic model
HPC	high-performance computing
LCA	life cycle assessment
LLM	large language model
mAb	monoclonal antibody
ME	metabolism–expression
MILP	mixed integer linear program
ML	machine learning
MPC	model predictive control
MSP	minimum selling price
ODE	ordinary differential equation
pFBA	parsimonious flux balance analysis
RSM	response-surface methodology
RTM	reactive transport model
TEA	techno-economic analysis

to support monitoring, optimization, and control (Turanli et al., 2024; Yatipanthalawa and Gras, 2024; Park et al., 2021; Sokolov et al., 2021; Narayanan et al., 2023; Noor et al., 2025). Despite this progress, practitioners still lack a concise, practice-oriented synthesis with the most recent advances that compares model families on assumptions, data and curation effort, tractability, and extrapolation, while showing how to operationalize them for medium and feeding design, reactor operation, and plant-level decisions, motivating the present review.

This review captures recent advances in cellular modeling and their application across the major aspects of bioprocess design. First, we synthesize the modeling landscape, from lumped stoichiometric balances and unstructured kinetics to state-of-the-art mechanistic, statistical, and hybrid models, presented in a comparable format that clarifies assumptions, data and curation requirements, computational tractability, and predictive scope to aid model selection. Second, we show how these models inform practice: prioritizing strain edits; tailoring medium composition and feeding to align intracellular demands with reactor constraints; guiding bioreactor design, monitoring, and control; and linking cell-level predictions to plant-level analyses so that financial costs and environmental consequences are evaluated early. Our aim is to provide an up-to-date, practical guide that connects cellular physiology simulations to process performance and plant objectives.

2. Cellular models

Computational representations of the cell are the analytical backbone of modern bioprocess development: they translate *genotype* and *culture conditions* into quantitative *phenotype* predictions that inform every engineering layer, from pathway editing and medium formulation to bioreactor scale-up and whole-plant integration (Fig. 1). The available modeling spectrum spans overall stoichiometric mass balances, through dynamic kinetic and proteome-constrained frameworks, to data-driven or hybrid surrogates that blend first-principles with machine learning (ML). Each tier entails a trade-off between the effort and data needed for reconstruction and the depth explanatory power and reliability of its forecasts. The subsections that follow survey these approaches in order of increasing mechanistic detail, highlighting their mathematical structure, data requirements, and key advantages and limitations.

2.1. Lumped overall stoichiometric models

Lumped overall stoichiometric models describe microbial or chemical conversion processes at steady state using only elemental and degree-of-reduction (i.e., electron equivalents) balances, while neglecting intracellular intermediates as well as kinetic and regulatory detail (Bailey and Ollis, 1986; Stephanopoulos et al., 1998). These macroscopic descriptions constitute the simplest class of process models in biochemical engineering and are commonly used for yield analysis, process benchmarking, and estimation of theoretical conversion limits.

A generic overall reaction for aerobic biomass formation on a single carbon source can be written as

$$s_{S,X}\,C_aH_bO_c + s_{O_2,X}\,O_2 + s_{N,X}\,NH_3 \longrightarrow s_X\,CH_xO_yN_z + s_{CO_2,X}\,CO_2 + s_{H_2O,X}\,H_2O,$$

where $C_aH_bO_c$ denotes a generic organic substrate, $CH_xO_yN_z$ represents biomass with an empirical elemental composition, and $s_{i,X}$ are stoichiometric coefficients specifying the molar amounts of each species participating in the overall biomass-formation reaction. These coefficients are obtained by enforcing elemental balances on C, H, O, and N and, when required, degree-of-reduction consistency. For growth on glucose, $C_aH_bO_c = C_6H_{12}O_6$, while typical biomass compositions (Stephanopoulos et al., 1998) satisfy $x \approx 1.8$, $y \approx 0.5$, and $z \approx 0.2$.

Overall product formation from the same substrate can be expressed analogously as

$$s_{S,P}\,C_aH_bO_c \longrightarrow s_P\,P + s_{CO_2,P}\,CO_2 + s_{H_2O,P}\,H_2O,$$

where P denotes the product of interest and $s_{i,P}$ are stoichiometric coefficients associated with the product-formation reaction, determined from elemental and degree-of-reduction balances. Additional reactions of this type may be introduced to represent multiple substrates or products within the same lumped framework.

Yield coefficients follow directly from the overall stoichiometry and quantify biomass or product formation on a molar substrate basis. The biomass and product yields on substrate are defined as

$$Y_{X/S} = \frac{s_X}{s_{S,X}}, \qquad Y_{P/S} = \frac{s_P}{s_{S,P}}.$$

In practice, constructing an overall stoichiometric model typically requires on the order of hours to days when elemental biomass composition, substrate formulas, and representative yield coefficients are already available for a well-characterized host organism, and up to weeks for newly sequenced organisms for which elemental composition and yield data must first be experimentally determined. Once the required data are available, model construction proceeds in two steps: (i) formulation of overall reactions for biomass and, if relevant, major products; and (ii) solution of elemental and, when redox fidelity is important, degree-of-reduction balances, typically anchored by at least one measured yield coefficient. When only the theoretical maximum product yield is sought, step (i) relies solely on pathway stoichiometry and step (ii) reduces to verifying that the pathway equation is mass- and charge-balanced, eliminating the need for experimental yield data.

Once built, such models solve analytically at negligible computational cost, making them ideal for rapid techno-economic or life-cycle screening and as mass-balanced boundary conditions in fluid dynamics simulations. Their limitations are the flip side of their simplicity: no temporal dynamics, no regulation, and no by-products beyond those explicitly included. Even so, they provide indispensable mass-consistent anchors that guide and validate more detailed kinetic or genome-scale models.

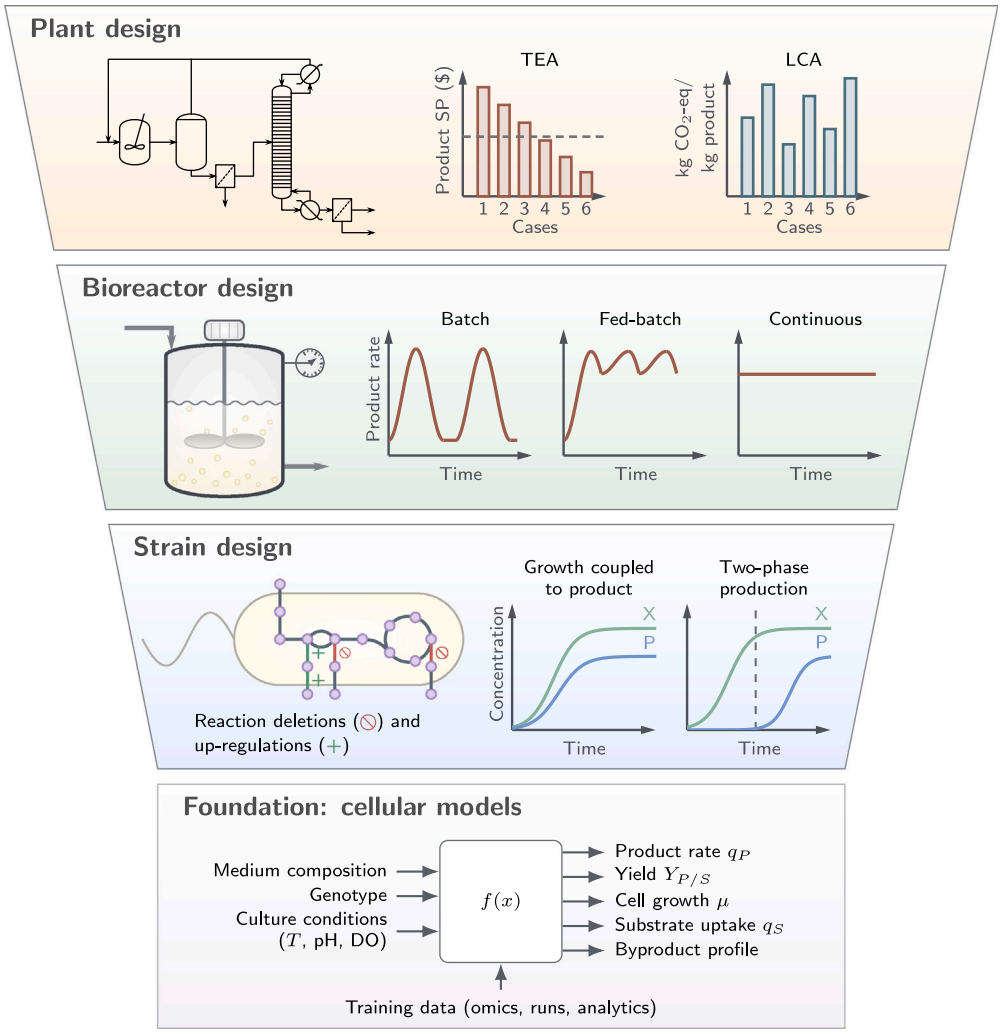

Fig. 1. Cellular modeling across design scales. At the cellular level, strain engineering tunes how metabolic fluxes are directed to produce the target molecule. At the bioreactor scale, design choices such as feed strategy, mixing, and aeration determine how much product can be made and how efficiently. At the plant scale, the entire process is evaluated for economic and environmental viability by linking strain and bioreactor performance with other unit operations across the production chain. The bottom layer summarizes the generic cellular-model inputs and process-facing outputs that support these applications; later sections show that specific decisions require different levels of resolution in these outputs, with more advanced cellular models needed when flux allocation, dynamic switching, resource limits, product quality, or other physiological details can change the design choice.

2.2. Unstructured kinetic models

Unstructured (macroscopic) kinetic models describe the time evolution of bulk variables, including substrate S, biomass X, product P, and occasionally by-products and multiple substrates, through ordinary differential equations (ODEs) that capture dominant empirical trends while treating the cell as a "black box". For a batch culture, the canonical Monod law (Eq. (1)),

$$\mu = \mu_{\max} \frac{[S]}{K_S + [S]}, \tag{1}$$

relates the specific growth rate μ to the substrate concentration $[S]$ in the growth medium (Monod, 1949). In this model $\mu_{\max}$ is the maximum specific growth rate and K_S is the Monod (half-saturation) constant, defined as the substrate concentration at which $\mu = \mu_{\max}/2$. Monod's equation is often coupled with the biomass balance (Eq. (2)), where [X] is biomass concentration, and an uptake constraint (Eq. (3)), where $Y_{X/S}$ is biomass yield from substrate. Together, these three equations predict substrate depletion and cell-mass rise in batch and fed-batch cultures.

$$\frac{d[X]}{dt} = \mu[X] \tag{2}$$

$$\frac{d[S]}{dt} = -\frac{1}{Y_{X/S}} \mu[X], \tag{3}$$

Product formation is often represented by the Luedeking–Piret relationship (Eq. (4)),

$$\frac{d[P]}{dt} = \alpha \frac{d[X]}{dt} + \beta[X], \tag{4}$$

where α and β weight growth-associated and non-growth-associated synthesis, respectively (Luedeking and Piret, 1959). Numerous alternative rate laws (e.g., for substrate inhibition, oxygen limitation, or mixed feeds) are cataloged in Table 1 of Du et al. (2022).

The data needed to calibrate an unstructured kinetic model includes time-resolved biomass, substrate and product profiles from one to three well-instrumented batch or fed-batch runs. Additional data like byproduct, O_2, and CO_2 concentrations can be collected, extending the model equations beyond what is discussed here (e.g. adding a maintenance demand for substrate uptake (Eq. (3)), a cell death for biomass change (Eq. (2)), or more equations to represent mass balance for additional metabolites). Non-linear regression of the Monod–Luedeking–Piret parameters $\{\mu_{\max}, K_S, Y_{X/S}, \alpha, \beta\}$ including data curation can take around 2 days from gathering existing data, while

generating the laboratory data typically adds three to six weeks (Singh et al., 2017). The fitted low-order ODE system then reproduces lag, exponential growth, substrate depletion and product formation with modest computational cost, making it a practical tool for seed-train sizing, feed-profile screening and early techno-economic or life-cycle assessments.

The simplicity that makes these models attractive also limits them: the lumped parameters must be re-identified whenever strain, medium or operating mode changes; they extrapolate poorly to unseen genetic or environmental perturbations; parameter sloppiness (i.e. multiple parameter combinations yield indistinguishable model predictions and individual parameters become poorly identifiable) arises under noisy data; and scale-dependent mixing or mass-transfer effects are absorbed empirically, restricting portability across bioreactor sizes. Even so, unstructured kinetics remain a work-horse for early process development and a convenient benchmark for more detailed hybrid or machine-learning surrogates.

2.3. Constraint-based metabolic models

CBMs extend stoichiometric bookkeeping to account for intracellular metabolic reactions, often at the genome-scale in which case they are named genome-scale metabolic models (GEMs). Mathematically, this information is organized as a sparse stoichiometric matrix $S \in \mathbb{R}^{|I|\times|J|}$, whose entries S_{ij} denote the signed stoichiometric coefficient of metabolite i in reaction j. The sets of all metabolites and reactions in the network are I and J, respectively. Assuming intracellular quasi-steady state, metabolic fluxes for each reaction v_j in mmol/gCDW/s are given by mass conservation (Eq. (5)) and physico-biochemical limits (e.g. substrate uptake rate or reaction reversibility) captured by linear bounds (Eq. (6)).

$$\sum_{j\in J} S_{ij}\, v_j = 0 \qquad \forall\, i \in I, j \in J \tag{5}$$

$$v_j^{\min} \le v_j \le v_j^{\max} \qquad \forall\, j \in J \tag{6}$$

When only stoichiometric constraints and irreversibility ($v_j \ge 0$ for irreversible reactions) are imposed, the feasible flux space forms a polyhedral cone, commonly called the *flux cone*, because any non-negative scalar multiple of a feasible flux vector is itself feasible. Adding the condition-specific upper and lower bounds of Eq. (6) intersects this cone with a bounded box, producing the convex polytope

$$\mathcal{F}(\mathbf{b}) = \left\{ \{v_j\}_{j\in J} : \sum_{j\in J} S_{ij}\, v_j = 0\ \forall i \in I, \quad v_j^{min}(\mathbf{b}) \le v_j \le v_j^{max}(\mathbf{b})\ \forall j \in J \right\}, \tag{7}$$

where $\mathbf{b}$ collects the experimental conditions encoded in the bounds: measured substrate uptake rates, oxygen availability, gene knockouts ($v_j^{min} = v_j^{max} = 0$), or flux intervals from ^{13}C fluxomics. Tightening or relaxing $\mathbf{b}$ reshapes $\mathcal{F}$ without changing any model parameter, so the same network represents multiple experimental scenarios.

The polytope $\mathcal{F}(\mathbf{b})$ is not a specific prediction of fluxes, but a space of possible fluxes that can be explored in several complementary ways. Most commonly, flux balance analysis (FBA) selects a single point by solving the linear program:

$$\{v_j^*\}_{j\in J} = \arg \max_{\{v_j\}\in\mathcal{F}(\mathbf{b})} \sum_{j\in J} c_j\, v_j, \tag{8}$$

where c_j is the objective coefficient for reaction j; typically $c_j = 1$ for the biomass reaction and zero elsewhere (Orth et al., 2010). A common variant is parsimonious FBA (pFBA), which further minimizes $\sum_{j\in J} |v_j|$ among all solutions that achieve the same optimal objective, favoring enzyme-economical predictions (Lewis et al., 2010). These single-objective methods return one flux distribution – a single point in $\mathcal{F}(\mathbf{b})$ – and therefore cannot capture the full range of metabolic states the network can support. Two complementary approaches explore this broader capability: random sampling draws unbiased flux distributions from the interior of $\mathcal{F}(\mathbf{b})$ to characterize its geometry without imposing any objective (Gelbach et al., 2024). Usually understanding the extremes is more relevant, so flux variability analysis (FVA) is used to probe the limits of the polytope by solving, for each reaction $j \in J$, a minimization and a maximization:

$$v_j^- = \min\ v_j, \qquad v_j^+ = \max\ v_j \qquad \text{s.t. } \{v_j\}_{j\in J} \in \mathcal{F}(\mathbf{b}), \tag{9}$$

where $\mathcal{F}(\mathbf{b})$ may be augmented with any user-defined conditions—for example, enforcing a minimum growth rate $v_{\text{biomass}} \ge \mu_{\min}$ to ensure biological viability while probing each reaction's allowable range. The result is the interval collection $\{[v_j^-, v_j^+]\}_{j\in J}$: reactions with $v_j^- = v_j^+$ carry a uniquely determined flux under the imposed constraints, while those with a wide interval are metabolically flexible (Kenefake et al., 2022).

The strategies above cover the most widely used ways to extract information from $\mathcal{F}(\mathbf{b})$. Many variants built on the same mathematical structure have been developed for specific applications, including context-specific network reconstruction from omics data, for example integrating transcriptomics or proteomics to restrict the active flux space to condition-relevant subnetworks (Machado and Herrgård, 2014; Opdam et al., 2017), and metabolic engineering target identification, which we discuss in Section 3.1.

Building a constraint-based model begins with a high-quality genome annotation that is mapped to mass-balanced reactions via gene–protein–reaction rules; a strain-specific biomass equation is then formulated, and iterative gap-filling plus *in silico* gene-essentiality tests close remaining network holes and benchmark predictive skill (Thiele and Palsson, 2010). Kinetic parameters are not required; instead, condition-specific uptake/secretion bounds, growth yields, and optional multi-omics layers (RNA-seq, proteomics, ^{13}C fluxes) are used to tighten flux limits or tailor objectives. Community toolkits such as *memote* add > 150 automated checks for stoichiometry, annotation coverage, and simulation fidelity to streamline reuse and updates (Lieven et al., 2020).

In terms of effort, starting from an existing draft for a model organism (for example *E. coli*, *S. cerevisiae*, or CHO) typically requires ~3–4 months of curation (gap-filling, biomass calibration, validation), whereas a newly sequenced non-model microbe demands ~6–12 months for end-to-end annotation, reconstruction, and testing (Thiele and Palsson, 2010; Lieven et al., 2020). Fuelled by automated pipelines and inexpensive sequencing, the catalog of GEMs has expanded rapidly: a 2019 survey counted reconstructions for 6239 bacterial, archaeal, and eukaryotic species (Gu et al., 2019). The current catalog spans canonical industrial workhorses, namely *Escherichia coli* (Monk et al., 2017), *Saccharomyces cerevisiae* (Lu et al., 2019), *Bacillus subtilis* (Massaiu et al., 2019), *Corynebacterium glutamicum* (Zhang et al., 2017), *Komagataella phaffii* (formerly *Pichia pastoris*) (Tomàs-Gamisans et al., 2016), and Chinese-hamster ovary cells (Hefzi et al., 2016). The catalog also covers hosts of expanding industrial interest: the thermophilic cellulolytic bacterium *Clostridium thermocellum* (Garcia et al., 2020), the acidophile *Acidithiobacillus Ameehan* (Wu et al., 2023), the oleaginous yeast *Yarrowia lipolytica* (Kerkhoven et al., 2016), the solvent-tolerant soil bacterium *Pseudomonas putida* (Nogales et al., 2020), and the filamentous fungus *Aspergillus niger* (Brandl et al., 2018). Automated pipelines can now assemble draft GEMs at population scale by templated reconstruction from genome annotations, automated gap-filling, and standardized quality control; yielding hundreds of thousands of models with minimal manual intervention (Heinken and Thiele, 2025). These drafts are excellent for comparative and hypothesis-generating studies but, because biomass composition, transport specificity, gene-protein-reaction rules, and condition-specific bounds are generic or missing, they should be treated as screening-grade starting points that still require organism- and process-specific curation for engineering use.

The size of full GEMs hampers certain analyses so systematic reduction algorithms such as redGEM (Ataman and Hatzimanikatis, 2017) or FASTCORE (Klamt et al., 2014) derive compact "core" networks while retaining key physiological functions. These core models can serve as basis for kinetic models, metabolic flux analysis using isotopic tracers for precise flux quantification, and elementary mode analysis EMA) (Trinh et al., 2009). Elementary modes are the smallest, non-decomposable steady-state flux vectors that satisfy Eq. (5); every feasible flux distribution is a non-negative weighted sum of these modes (Schilling et al., 2000). Because the number of modes grows combinatorially with network size, full enumeration is generally practical only for core or reduced models (Terzer and Stelling, 2008). Once the mode set is available, however, it provides an ultrafast surrogate model: metrics such as specific product yield, ATP cost, or oxygen requirement can be obtained by a single matrix–vector multiplication of objective weights with the pre-computed mode matrix, bypassing linear programming altogether.

Leveraging linear mass-balance constraints, CBMs scale to networks of thousands of reactions and solve in milliseconds, making them ideal for high-throughput *in-silico* nutrient-essentiality screens, strain-design knockouts, media tailoring, metabolic-engineering target ranking, and even plant-wide process optimization, as we will cover later in this review. Their scope, however, is bounded by the quasi-steady-state assumption, which overlooks transient metabolite pools and regulatory kinetics; and by the existence of multiple alternate optima, quantified by FVA (Eq. (9)), that require careful interpretation of any single optimal solution (Kenefake et al., 2022; Gelbach et al., 2024). Thus, while CBMs powerfully connect genome content to process design with minimal parameterization, their predictions must ultimately be tempered by dynamic modeling and experimental validation.

2.4. Proteome-constrained metabolic models

Where classical GEMs assume unlimited catalytic capacity, *proteome-constrained* models, such as the metabolism–expression (ME) formulation (Lerman et al., 2012; O'Brien et al., 2013), variants of ME like the Expression and Thermodynamics Flux (ETFL) formulation (Salvy and Hatzimanikatis, 2020), and enzyme-constrained models like the Genome-scale model to account for Enzyme Constraints, using Kinetics and Omics (GECKO) framework (Sánchez et al., 2017), embed transcription, translation, and enzyme-allocation limits into the stoichiometric framework.

Denoting v_j as the flux of metabolic reaction $j \in J$, e_k as the concentration of catalyst $k \in K$ (enzyme, ribosome, or protein complex), and $e_{\sigma(j)}$ as the abundance of the catalyst assigned to reaction j by the mapping $\sigma : J \to K$ (needed because several reactions can share the same enzyme and, conversely, one reaction may require a multisubunit complex), the general proteome-constrained framework invokes three linked constraints: the metabolite steady-state balance (Eq. (10)), identical to the one used in CBMs; the catalytic-capacity limit (Eq. (11)), that couples each flux to its catalytic capacity, and the global protein-mass budget (Eq. (12)).

$$\sum_{j \in J} S_{ij} v_j = 0 \qquad \forall i \in I \tag{10}$$

$$v_j \le k_j^{\text{cat}} e_{\sigma(j)} \qquad \forall j \in J \tag{11}$$

$$\sum_{k \in K} \text{MW}_k e_k \le \rho_{\text{prot}} \tag{12}$$

All variants share the core idea that cellular resources, rather than pure stoichiometry, bound achievable phenotypes; the differences lie in how rigorously enzyme synthesis and turnover are represented.

Analogous to the flux polytope $\mathcal{F}(\mathbf{b})$ defined for CBMs (Eq. (7)), proteome-constrained models define a feasible set in a joint flux-and-enzyme space. For enzyme-constrained models such as GECKO, all three constraints in Eqs. (10)–(12) are linear, so the feasible set remains a convex polytope in the combined (v, e) space and the alternate-optima and variability-analysis concepts of CBMs carry over in principle. The practical complication is that k_{cat} values span several orders of magnitude, contributing to numerical ill-conditioning; in full ME models, which couple flux and protein-synthesis variables, this ill-conditioning is severe enough to require specialized quad-precision solvers such as *solveME* and the DQQ workflow (Yang et al., 2016; Ma et al., 2017), whereas GECKO-style ecModels remain ordinary linear programs and solve without special numerical treatment. Full ME models track enzyme synthesis explicitly and include a dilution-by-growth term μe_k; since both μ and e_k are variables, their product is bilinear and the feasible set is no longer a simple polytope. In practice this is addressed by solving a sequence of linear programs at fixed μ, treating the model as a parametric family of polytopes across growth rates (Lloyd et al., 2018). ETFL goes further by incorporating thermodynamic constraints via reaction Gibbs energies and intracellular metabolite concentrations, which enter as logarithmic terms and push the problem into nonlinear or mixed-integer programming territory (Salvy and Hatzimanikatis, 2020).

Beyond a curated GEM, proteome-constrained models (GECKO, ME, ETFL) require enzyme turnover numbers k_{cat}, protein molecular weights, multi-subunit complex stoichiometries, and either absolute protein abundances or a total proteome (or ribosome) budget; full ME/ETFL variants also need transcription/translation rates and protein degradation half-lives. Databases such as BRENDA, SABIO-RK, and UniProt supply many k_{cat} values, with machine-learning imputation used to fill gaps (Li et al., 2022b; Zhai et al., 2025). Several toolchains help implement this: *COBRAme* streamlines full ME reconstructions (e.g., *E. coli*, *Thermotoga maritima*) (Lloyd et al., 2018; Yang et al., 2016), while the GECKO toolbox inserts the catalytic-capacity and proteome-budget constraints of Eqs. (11)–(12) onto existing GEMs (Sánchez et al., 2017); readers seeking step-by-step construction guidance are referred to the dedicated protocols and worked examples in those references and in Chen et al. (2024). In terms of effort, given that a GEM is available, for a proteomics-rich host (e.g., yeast) an ecModel via GECKO typically takes 4–8 months of curation (data collation, mapping, quality control), whereas assembling a full ME model or porting either approach to a sparsely characterized organism generally requires 12–24 months for kinetic curation, numerical tuning, and validation (Sánchez et al., 2017; Lloyd et al., 2018; Yang et al., 2016).

Coupling each metabolic flux to an enzyme variable inflates model size by one–two orders of magnitude, producing ill-conditioned mixed linear/convex programs; quad-precision pipelines such as *solveME* and the DQQ scaling workflow cut solve times from days to minutes (Yang et al., 2016; Ma et al., 2017), while condensed ME formulations and GECKO's linear relaxation routinely converge in ~10–60 s on commodity hardware (Lloyd et al., 2018; Sánchez et al., 2017), yet thousands of sequential solves, as required for dynamic simulations, are still computationally taxing. When their computational cost is not an issue, proteome-constrained models mechanistically link genotype, resource allocation, and flux, predicting growth–yield trade-offs, overflow metabolism, stress responses, and the protein cost of heterologous pathways without fitted kinetics, thereby refining strain-design rankings and pinpointing medium supplements that relieve catalytic bottlenecks (Dinh et al., 2018). Their accuracy, however, hinges on comprehensive k_{cat} datasets and precise complex stoichiometries (i.e., subunit composition and cofactor content of each protein assembly that catalyses a reaction), both scarce for non-model organisms. Overall, incomplete data, poor conditioning, or unmodeled regulatory layers (e.g. transcription-factor cascades) can erode predictive power and hinder applications.

2.5. Large-scale kinetic metabolic models

Kinetic models elevate the metabolite mass balance (Eq. (5)) from a linear flux-balance to a nonlinear, time-dependent description. Defining

$C_i(t)$ as the concentration of metabolite $i \in I$ at time t, $v_j\big(C(t),k\big)$ as the rate of reaction $j \in J$ with kinetic parameter vector k, and S_{ij} as the stoichiometric matrix used previously for CBMs, intracellular dynamics follow metabolite mass balance (Eq. (13)).

$$\frac{dC_i}{dt} = \sum_{j \in J} S_{ij}\, v_j\big(C(t),k\big), \qquad \forall i \in I, \tag{13}$$

Typical rate laws are reversible Michaelis–Menten, convenience kinetics, or modular network-embedded formulas that embed thermodynamic reversibility and allosteric regulation (Islam et al., 2021). Algebraic Haldane relationships couple the forward and reverse constants to the reaction equilibrium constant, thereby reducing the free parameter count.

A very common choice for $v_j\big(C(t),k\big)$ in Eq. (13) is the reversible Michaelis–Menten rate law (Islam et al., 2021). For a single-substrate reversible reaction $A \rightleftharpoons B$, this takes the form

$$v_j\big(C(t),k\big) = \frac{V_{\max,f}\,\dfrac{[A]}{K_{m,A}} - V_{\max,r}\,\dfrac{[B]}{K_{m,B}}}{1 + \dfrac{[A]}{K_{m,A}} + \dfrac{[B]}{K_{m,B}}}, \tag{14}$$

where $V_{\max,f}$ and $V_{\max,r}$ are the forward and reverse maximum reaction rates – the maximum rates achievable at saturating substrate and product concentrations, respectively – and $K_{m,A}$, $K_{m,B}$ are the Michaelis constants for substrate and product. In practice, many kinetic models do not track enzyme concentration e_j explicitly; instead, each product $k_{\text{cat}}\, e_j$ is collapsed into a single apparent parameter ($V_{\max,f}$ or $V_{\max,r}$), which is what gets measured from enzyme assays or fitted from flux and metabolomics data (Liebermeister et al., 2010; Gopalakrishnan et al., 2020; Hu et al., 2024). At face value, Eq. (14) therefore contains four independent parameters per reaction ($V_{\max,f}$, $V_{\max,r}$, $K_{m,A}$, $K_{m,B}$). The Haldane constraint reduces this by one: because the net rate must vanish at thermodynamic equilibrium ($v_j = 0$ when $[B]/[A] = K_{eq}$), the parameters must satisfy

$$K_{eq} = \frac{V_{\max,f}\,K_{m,B}}{V_{\max,r}\,K_{m,A}}, \qquad K_{eq} = \exp\!\left(-\frac{\Delta_r G^\circ}{RT}\right), \tag{15}$$

where $\Delta_r G^\circ$ is the standard Gibbs energy of reaction, R the gas constant, and T absolute temperature. Since $\Delta_r G^\circ$ can be obtained from thermodynamic databases such as eQuilibrator (Beber et al., 2022), Eq. (15) fixes the ratio $V_{\max,f}/V_{\max,r}$ before any fitting is done, guaranteeing thermodynamic consistency and lowering the number of free parameters that must be inferred from experimental data. More complex rate laws for multi-substrate reactions (convenience kinetics, modular rate laws) apply the same Haldane principle with additional Michaelis terms.

In practice, large-scale kinetic reconstruction starts from a curated network (typically 100–500 reactions), selects a rate-law template such as Eq. (14), and then populates parameters by (i) mining databases such as BRENDA and SABIO-RK for K_m, k_{cat}, and K_i; (ii) enforcing the Haldane constraint (Eq. (15)) using $\Delta_r G^\circ$ from eQuilibrator (Beber et al., 2022) together with steady-state flux targets from ^{13}C fluxomics; and (iii) calibrating against mutant flux maps, time-series metabolomics, and (when available) proteomics. Frameworks such as K-FIT (Gopalakrishnan et al., 2020), KETCHUP (Hu et al., 2024), and RENAISSANCE (Choudhury et al., 2024) systematize these steps and conclude with ensemble approaches that sample thousands of parameter sets consistent with all constraints to address under-identifiability. Emerging deep-learning tools such as CatPred further impute missing k_{cat} or K_m values from sequence and substrate features (Boorla and Maranas, 2025). For organisms with an existing kinetic atlas like *E. coli*, parameter balancing and ensemble generation typically require 6–12 months (Khodayari and Maranas, 2016); for non-model strains with little information available, the end-to-end effort, including data generation and curation, could extend to 24–36 months.

Kinetic models capture transients such as substrate pulses, feast–famine cycles and genetic switches, quantify enzyme saturation and control coefficients, and thereby enable dynamic feed optimization and identification of rate-limiting steps that steady-state frameworks miss. Their chief drawbacks are the thousands of poorly identifiable kinetic constants, which necessitate ensemble sampling and sensitivity analysis, and stiff equations that still solve in minutes even with quad-precision and model-reduction tricks. Generative ML pipelines such as KETCHUP and CatPred are narrowing the parameter gap, positioning these models as dynamic complements to GEMs for next-generation strain and process optimization.

2.6. Statistical models

In contrast to mechanistic approaches, statistical models specify a likelihood $p(y \mid X, \theta)$ that relates observed responses y to input features X through unknown parameters θ, and in Bayesian settings, a prior $p(\theta)$ that encodes beliefs about the values of θ. Here, $X \in \mathbb{R}^{N \times p}$ is a design matrix of N experiments and p input variables (e.g., medium components, temperature, pH), $y \in \mathbb{R}^N$ collects the measured outcomes (e.g., growth rate, titer), and θ are the model parameters to be inferred. Common statistical modeling tasks in bioprocessing include classification of culture phases from on-line signals using logistic regression or support-vector machines (Nikzad-Langerodi et al., 2017), clustering to group media formulations by phenotype (Grzesik and Warth, 2021), dimensionality reduction to compress (multi-)omics via principal component analysis (Meng et al., 2016), and time–series/state–space analysis with Kalman filters for soft sensing and drift detection (Yousefi-Darani et al., 2021; Duarte et al., 2021). A very common case is supervised regression, where the goal is to map experimental settings to a continuous response via

$$y = f_\theta(X) + \varepsilon, \tag{16}$$

with f_θ a parametrized function (e.g., linear or polynomial regression, Gaussian process, random forest, neural network) and ε an experimental noise term. Parameter fitting uses least squares or machine-learning loss functions on a training set $D = \{(X_n, y_n)\}_{n=1}^N$ after which f_θ predicts unseen conditions in milliseconds.

Statistical models require only labeled examples and no explicit stoichiometry or pathway knowledge. A coarse planning heuristic is to target on the order of 10–30 observations per varying input variable, for the 10–50 factors typical of medium-design studies, this rule of thumb implies 10^2–10^3 unique conditions. Recent cases align with this order of magnitude (Hashizume et al., 2023), while others can reduce the sampling needs through adaptive sampling (Narayanan et al., 2025). When curated data already exist, cleaning and fitting a random-forest, Gaussian-process, or neural-network regressor typically takes 1–2 weeks; generating a fresh high-throughput screening data set adds ~1–3 months for experimental execution and data wrangling.

These models are inexpensive to train/evaluate, capture nonlinear interactions, and readily fuse heterogeneous features (sequence descriptors, media composition, on-line signals), but they extrapolate poorly beyond the training domain, cannot enforce mass or energy conservation, and offer limited mechanistic insight. Consequently, they are best suited to medium optimization, bioprocess control, and as components in hybrid workflows, as discussed later in this review.

2.7. Hybrid models

Hybrid models cover a heterogeneous "catch-all" space that mixes the mechanistic formalisms reviewed above with statistical or rule-based layers, or couples metabolism to additional cellular scales such as transcriptional regulation, protein secretion, or plasmid burden. A validated GEM or kinetic core supplies mass-balance constraints, after which high-throughput data (e.g., fluxomics, time-series titers,

transcriptomics) train the statistical block to learn residuals, regulatory switches, or uncertainty envelopes.

Integration of ML with constraint-based models generally follows the three-way scheme cataloged by Sahu et al. (2021): *(i) Kinetically-augmented CBMs* enrich flux-balance cores with explicit rate laws, thermodynamic bounds or enzyme budgets (here we considered this an established modeling approach discussed under the section Proteome-constrained metabolic models); *(ii) Logic/event layers* superimpose discrete regulation on continuous fluxes, often via Petri nets (i.e., a graph-based formalism that tracks tokens to capture concurrent discrete state changes) or their Flexible Net derivatives; and *(iii) Data-driven surrogates* embed physics-informed neural networks or Gaussian processes inside the optimization loop so model predictions comply with mass balance constraints (e.g., Eq. (5)). Recent work illustrates each category: Artificial Metabolic Networks differentiate FBA through a neural layer (Faure et al., 2023) (surrogate); Flexible-Net conversions of proteome-constrained GEMs add discrete regulation (Lázaro et al., 2022) (logic/event); genome-scale hybrid-cybernetic models to optimize product yield (kinetically-augmented CBMs) (Luo et al., 2023); and an uncertainty-aware dynamic FBA for CHO cultures blends enzyme-constrained kinetics with a macro-kinetic shell (Pennington et al., 2024) (kinetically-augmented CBMs). NEXT-FBA trained artificial neural networks (ANNs) on Chinese hamster ovary (CHO) exometabolomic data paired with ^{13}C fluxomics to predict upper and lower bounds for central-carbon reactions in a CBM, identifying asparagine catabolism as the dominant source of ammonia accumulation across three feeding strategies in CHO culture (Morrissey et al., 2025).

Some recent hybrids depart from the taxonomy either by forgoing ML or by coupling metabolism to additional cellular scales. COSMIC-dFBA toggles cellular objectives in perfusion bioreactors via a phase ML classifier (Gopalakrishnan et al., 2024), while kilometre-scale reactive-transport models use neural surrogates of FBA to inject metabolic realism into environmental partial differential equations (Song et al., 2025). A Bayesian extension of an enzyme-constrained yeast GEM estimates temperature-dependent catalytic parameters on-the-fly, producing an uncertainty-aware model that pinpoints thermotolerance bottlenecks (Li et al., 2021). Multi-scale hybrids integrate transcriptional-regulatory network (e.g., OptRAM (Shen et al., 2019), etiBsu1209 (Bi et al., 2023)), secretion pathways (e.g., pcSecYeast (Li et al., 2022a)), or plasmid-expression burdens (e.g., rETFL (Oftadeh and Hatzimanikatis, 2024)). Whole-cell models, which couple metabolism with DNA replication, transcription and translation, remain largely exploratory but showcase the ultimate hybrid ambition (Macklin et al., 2014; Thornburg et al., 2022; Macklin et al., 2020).

In terms of data requirements, hybrid models require all inputs required for the mechanistic core (CBM, ecModel, or kinetics) plus time-series or on-line data streams to train the ML or logic layer (e.g. residuals, phase classifiers, surrogate objectives). Adding an ML layer to an existing core model typically requires 1–3 months (feature engineering, cross-validation, integration) (Doyle et al., 2023; Faure et al., 2023).

2.8. Summary of model construction requirements and expected AI acceleration

The model families reviewed above differ markedly in data, curation effort, and simulation time. Increasing mechanistic detail shifts workflows from simple conversion chemistry or culture trajectories toward curated reaction networks, catalytic parameters, omics constraints, and condition-specific validation. Table 1 summarizes these trade-offs, while Table 2 expands the compressed "Data requirements" column into evidence types such as conversion chemistry, network structure, catalytic capacity, culture time series, omics constraints, and condition-response records. Hybrid models are included in Table 1 because they are a recurring modeling architecture, but they are not assigned a standalone column in Table 2, since their data requirements are the union of the chosen mechanistic, machine-learning, rule-based, or control components rather than a stable checklist.

Large language models (LLMs) and agentic artificial intelligence (AI) are expected to accelerate model curation mainly when the limiting work is digital, through two channels: information curation, such as extracting and structuring evidence from literature, databases, and model annotations; and software automation, such as writing parsers, model-conversion scripts, database queries, quality-control wrappers, and workflow code that curators have often developed manually. This view is supported by software-development studies showing task-dependent gains, with time saved in code writing partly offset by the need to review, correct, and maintain AI-generated output (Vella and Blincoe, 2026; Becker et al., 2025; Xu et al., 2025). It is also consistent with LLM-assisted data-extraction and biocuration studies, which support source-linked extraction as curator assistance rather than autonomous curation (Schmidt et al., 2024; Caufield et al., 2024). For CBMs, likely targets include reaction-evidence triage, gene-protein-reaction rules, biomass and transport definitions, gap-filling review, and *memote*-style quality control; for protein-constrained and large-scale kinetic models, they include enzyme-reaction mapping, k_{cat} or kinetic-constant selection, complex stoichiometry, omics integration, rate-law or model-code assembly, numerical troubleshooting, and parameter-uncertainty checks (Thiele and Palsson, 2010; Lieven et al., 2020; Karp et al., 2018; Sánchez et al., 2017; Lloyd et al., 2018; Li et al., 2022b; Khodayari and Maranas, 2016; Choudhury et al., 2024). Existing tools such as GECKO, machine-learning k_{cat} prediction, automated draft reconstruction, and generative kinetic parameterization provide structured workflows that agents can run, audit, and document (Sánchez et al., 2017; Li et al., 2022b; Heinken and Thiele, 2025; Choudhury et al., 2024).

For these exposed digital components, we estimate that domain-specific LLM/agent workflows could reduce curator desk-time by roughly 50%–70% in data-rich CBM, protein-constrained, and large-scale kinetic model-curation workflows. This range is intended as a scenario estimate, consistent with task-dependent gains reported for code generation and structured information-retrieval settings, while allowing for review, correction, and maintenance overhead (Vella and Blincoe, 2026; Becker et al., 2025). It applies to digital curation tasks, not to total project time or validated industrial deployment. Acceleration should plateau when the limiting work is new phenotyping, proteomics, fluxomics, metabolomics, kinetic assays, expert biological adjudication, parameter uncertainty, or independent validation. These experimental bottlenecks may also begin to shift as agentic AI interfaces with automated data-generation platforms, including industrial biofoundries such as Ginkgo Bioworks' cell-engineering platform and high-throughput screening workflows, translating model-curation hypotheses into assay designs and standardized data returns. However, these technologies remain early, tied to particular organisms, assay formats, automation hardware, and measurement modalities; it remains to be seen how broadly they can cover the phenotyping, omics, flux, and kinetic assays needed for model-specific validation (Ginkgo Bioworks Holdings, Inc., 2026).

3. Strain design and medium optimization

Before a process ever reaches the bioreactor, its performance is largely fixed at the culture scale, where the strain's genotype and the medium's composition together set metabolic capacity, regulatory balance, and overall cell health. Recent work in strain design and medium optimization has relied significantly on constraint based models and optimization methods. This allows us to describe both as the generic mathematical optimization problem shown in Fig. 2. In this formulation, the decision variables represent either genetic or medium manipulations, a cellular model maps those decisions to physiological responses, and the objective maximizes classic performance metrics (e.g., titer, rate, yield, robustness, growth-coupled production,

Table 1
Comparison of major cellular modeling paradigms. In the lower bound of curation time all experimental data is available, while in the upper bound most experimental data needs to be generated. These curation times are estimated for the effort of a single researcher. In practice, several modeling tasks, such as experimental characterization or pathway-specific curation, can be carried out in parallel by a team, thereby reducing the overall wall-clock time required for model assembly.

Model class	Brief description	Data requirements	Curation effort	Main limitations
Stoichiometric models	Pseudo-reaction(s) linking substrate to biomass and/or product; closed-form algebra; used for consistency checks and fluid dynamics or plant-level simulations.	Main metabolic pathways or measured biomass composition and product or biomass yields.	0.5 days–2 weeks	No time dimension; ignores regulation and by-products.
Unstructured kinetic models	≤10 ODEs for biomass, key substrates and products; capture lag, growth, and metabolic shifts for feed-rate design.	One or few cultures with overall rates (biomass, substrate, product vs. time).	2 days–6 weeks	Valid only near calibration window; no intracellular insight.
Constraint-based metabolic models (CBM)	Genome-scale reaction network that predicts feasible fluxes, including metabolite uptake, intracellular conversion, metabolite secretion, and biomass formation.	Annotated genome, biomass composition, uptake bounds; optional omics and gene essentiality.	3–12 months	Needs objective choice; lacks kinetics/regulation.
Proteome-constrained models	Add enzyme capacity or full expression machinery to CBM; allocates a finite proteome budget to metabolic fluxes.	Same as CBM inputs plus k_{cat} values and protein features (e.g. complex stoichiometry, abundance, degradation rates, etc.).	4–24 months (from existing CBM)	Sparse kinetic data; ill-conditioned mathematics; long runtimes.
Large-scale kinetic models	Hundreds of ODEs with parametrized rate laws	Fluxomics, metabolomics, enzyme kinetic constants (K_m, k_{cat}, K_i), and reaction Gibbs energies $\Delta_r G'^{\circ}$.	6–36 months	Kinetic parameter identifiability; stiff numerics; minutes per simulation.
Statistical models	ML regressors (e.g., random forest, Gaussian process, artificial neural network) linking inputs like culture conditions to target phenotypes like growth or titer; no mechanistic terms.	Condition and phenotype data generated through a high throughput screening platform (generally 100–1000 observations).	1 week–3 months	Poor extrapolation; limited interpretability; limited to scenarios where sufficient experimental data can be generated; retraining needed after changes.
Hybrid models	Mechanistic core (e.g. unstructured kinetic, CBM) plus ML or rule layer for bounds, objective, or regulation.	Mechanistic data and on-line or historical process data.	1–3 months (from existing mechanistic model)	Higher complexity; risk of hidden-constraint violation; data-hungry ML part.

or medium efficiency). The advantage of using mathematical optimization frameworks is that many algorithms exist to solve this type of problems. In addition to these approaches, we will also cover medium optimization through the use of statistical and machine learning models either independently or alongside mechanistic models.

3.1. Strain design

Over the past two decades metabolic-engineering research has advanced from "one-reaction-at-a-time" heuristics to formal optimization frameworks that recommend complete edit sets *in silico* before a single colony is picked. Classical tools such as *OptKnock* showed that a genome-scale stoichiometric model, cast as a mixed integer linear program (MILP) optimization problem, can couple product synthesis to growth and thereby steer laboratory evolution (Burgard et al., 2003). Their enduring appeal lies in mechanistic transparency: each reaction bound is hard-wired to a gene, so a simulated knock-out translates cleanly into a wet-lab deletion. These early successes entrenched constraint-based model (CBM) analysis as the default language for strain design and opened the door to larger genomes, richer objectives and multi-objective formulations.

Growth-coupling design remains pivotal for adaptive laboratory evolution (G. Wang et al., 2022): *OptCouple* broadens the search to include medium additives alongside gene edits (Jensen et al., 2019), whereas *OptEnvelope* seeks the minimal set of modifications that enforces coupling (Motamedian et al., 2023). By merging growth-coupling principles with the modular-platform concept, *ModCell2* employs multi-objective optimization to create a single chassis whose growth remains coupled to a suite of exchangeable product modules (Garcia and Trinh, 2019). A goal-attainment formulation later ensured coupling to every specified product (Garcia and Trinh, 2020), and high-performance extensions now scale the search to hundreds of target modules (Garcia and Trinh, 2021). Statistical post-processing is also emerging, CFSA (Comparative Flux Sampling Analysis) flags reactions whose flux distributions shift between growth and production optima, revealing non-obvious deletion or amplification targets (van Rosmalen et al., 2024).

As genome-scale kinetic reconstructions have matured, kinetic-model–driven strain design has gained relevance. The earliest "extended" GEMs enriched stoichiometric backbones with kinetic rate laws, letting design algorithms optimize both flux and catalytic control. Hybrid schemes such as *k-OptForce* (Chowdhury et al., 2014) embed elementary kinetics into a constraint-based core, extending earlier OptForce-based strain designs (Xu et al., 2011), so the solver can tune rate-limiting steps alongside steady-state fluxes, while the genome-scale *NOMAD* workflow applies the same idea to propose near–wild-type interventions that boost anthranilate titers (Narayanan et al., 2024). Large kinetic reconstructions strengthen this trend: $k_{ecoli457}$

Table 2
Data taxonomy for cellular model requirements. Legend: R = normally required; O = optional or decision-specific; – = not normally needed. Clarifications: [a]For unstructured kinetic models, full elemental/redox characterization is not always required. [b]For protein-constrained models absolute proteomics are useful but not strictly required. [c]For protein-constrained models, R applies to ME/ETFL or regulatory resource-allocation variants; GECKO/ecModels generally do not require transcriptomic or regulatory data. [d]For CBM, protein-constrained, and statistical models, R indicates data normally needed for credible predictive use or evaluation, not necessarily for minimal model assembly.

Data item	Typical examples	Model					
		Stoich.	Unstr. kin.	CBM	Prot. constr.	Large-scale kin.	Stat.
Fundamental biochemical and model-structure data							
Overall conversion chemistry	Substrate/product/by-product formulas; biomass composition; measured or theoretical yields; mass and redox closure.	R	R[a]	R	R	R	–
Network reconstruction	Genome annotation; reaction stoichiometry; reversibility; transporters; GPRs; biomass reaction.	–	–	R	R	R	–
Catalytic capacity limits	Reaction-enzyme mapping; flux-capacity constants used to cap enzyme usage, such as k_{cat} or enzyme-level V_{max}.	–	–	–	R	R	–
Protein allocation constraints	Protein molecular weights; complex stoichiometry; enzyme/proteome budget; expression or dilution parameters.	–	–	–	R	–	–
Mechanistic kinetic/thermodynamic parameters	Rate-law and thermodynamic parameters beyond capacity limits: K_m, K_i, K_{eq} or $\Delta_r G'^{\circ}$; regulation; initial metabolite concentrations.	–	–	–	O	R	–
Fundamental experimental physiology and omics data							
Culture setup, trajectories, and rates	Mode, feed, volume, sampling times; biomass, substrate, product, off-gas, pH/DO, uptake/secretion; perturbation or scale-transfer runs.	O	R	O	O	R	O
Flux and metabolite constraints	^{13}C fluxes; metabolomics; measured intracellular or exchange flux states.	–	O	O	O	R	O
Protein pool and abundance data	Total protein/enzyme pool; protein fractions; absolute or relative proteomics; enzyme abundance.	–	–	–	R[b]	O	O
Expression/regulatory data	RNA-seq or expression ratios; transcription-factor states; regulatory rules; stress-response markers.	–	–	O	R[c]	O	O
Experimental design, phenotype, and genotype records[d]	Media, strain, genotype, or process-condition variables paired with growth/no-growth, secretion, essentiality, titer, rate, yield, or quality outcomes.	O	O	R	R	O	R

STRAIN AND MEDIUM DESIGN OPTIMIZATION PROBLEM

- *Objective:* maximize a chosen strain-level performance metric, e.g., product formation rate, product-to-substrate yield, or a weighted combination of several phenotypes.
- *State variables:* all metabolic fluxes, including those that account for growth rate, product formation, or medium utilization.
- *Design variables:* genetic manipulations or medium alterations. Often represented with binary variables.
- *Constraints:*
  - **Stoichiometric consistency**: intracellular mass balances from the genome-scale model must be satisfied.
  - **Genetic edit budget**: only a limited number of knock-outs, knock-ins, or over-expressions may be applied per design cycle.
  - **Viable growth**: the engineered strain must retain at least a specified minimum growth rate.
  - **Physiological bounds**: each reaction flux must stay within plausible thermodynamic or capacity limits; key uptakes (e.g. glucose, O_2) cannot exceed experimental maxima.

Fig. 2. Verbal form of the general strain and medium design mathematical optimization problem used in the following sections. Optimization is a common framework for design problems because it uses the predictive potential of cellular models to compare candidate genetic or medium interventions against explicit objectives and constraints. The schematic highlights the objectives, state and design variables, and types of constraints used in contemporary modeling studies.

ranked 320 engineered *E. coli* strains across 24 products with $r = 0.84$ to experiment (Khodayari and Maranas, 2016); a 775-reaction map of *Pseudomonas putida* pinpointed ATP-buffering control points (Tokić et al., 2020); and a proteome-constrained kinetic atlas of *S. cerevisiae* reproduced multi-omics shifts over dozens of conditions, enabling *in-silico* knockout screens for higher fluxes (Choudhury et al., 2024). Kinetic analyses also reveal redox bottlenecks: NADPH-limited isobutanol pathways benefited from cofactor-swapped enzymes and transhydrogenase overexpression predicted *in silico* and later validated experimentally (Millard et al., 2021; Almquist et al., 2014).

Beyond pure kinetics, newer models layer additional biological constraints onto GEMs. Regulatory hybrids such as *OptRAM* (Shen et al., 2019) and *BeReTa/h-BeReTa* (Kim et al., 2017; Koduru et al., 2018) couple transcription-factor logic to metabolism, allowing simultaneous TF and metabolic edits and reducing false-positive targets in *E. coli*, *S. cerevisiae*, and *B. subtilis*, the latter via the thermodynamic-regulatory *etiBsu1209* model, which doubled menaquinone-7 titers (Bi et al., 2023). Protein-constrained ME and ecGEM variants filter designs by enzyme capacity: rescoring 2 600 knockouts with iJL1678b-ME yielded 42 growth-coupled *E. coli* strains robust to k_{cat} uncertainty (Dinh et al., 2018), while *pcSecYeast* guided CRISPRi/a screens that doubled α-amylase titers (Li et al., 2022a). Similar capacity limits have uncovered edits for L-lysine in *E. coli* (Ye et al., 2020), poly-γ-glutamate in *B. subtilis* (Massaiu et al., 2019), and a 70-fold heme boost in yeast (Ishchuk et al., 2022). Finally, Bayesian pipelines have fitted temperature-dependent kinetics to build *etcGEMs*, rationalizing thermotolerance limits and pointing toward thermophile-friendly interventions (Li et al., 2021). Together, these kinetic, regulatory, and proteome-aware extensions provide a richer mechanistic lens than flux balance alone, narrowing experimental effort to the most promising, and physiologically feasible, engineering strategies.

We covered examples where machine learning augmented mechanistic frameworks, however recent work also explores purely data-driven pipelines. The Bayesian recommender *ART* learns a genotype–phenotype function directly from screening data and, in retrospective tests, proposed edits that raised titers faster than any model-based search (Radivojević et al., 2020). The multi-agent reinforcement-learning platform *MARL* takes an even more agnostic stance, treating each gene edit as an action in a black-box environment and iteratively refining policies that maximize production (Sabzevari et al., 2022). Such engines excel when mechanism is poorly understood, yet they falter with sparse or biased data and still await broad *in-vivo* validation. Network-based data-driven strategies have also proven effective in non-model hosts: in *Yarrowia lipolytica*, co-expression connectivity metrics derived from genome-wide expression data ranked tolerance genes whose experimental knockout and overexpression confirmed predicted targets for high ionic-liquid resistance (Walker et al., 2022). For a deeper discussion of how fully statistical methods complement mechanistic models, and the challenges of merging the two, see the recent review by Lu et al. (2024).

Despite steady algorithmic progress, experimental validation remains the main bottleneck: only a handful of kinetic or machine-learning designs have been confirmed *in vivo*, and even constraint-based successes often stop at computational proof of concept. Closing this gap will require tighter Design–Build–Test–Learn loops and community benchmarks that report titers, rates, and yields on standardized media. Looking forward, machine learning is poised to parameterize and automate ever more complex models using high-throughput data to populate kinetic constants, regulatory logic, or even proteome constraints, thereby adding mechanistic insight without manual curation (Narayanan et al., 2024; Radivojević et al., 2020; Sabzevari et al., 2022). At the same time, simpler flux models and growth-coupling heuristics will remain valuable wherever metabolic complexity is not the primary bottleneck, offering transparent links from optimization variables to gene edits and media tweaks (Khodayari and Maranas, 2016; Jensen et al., 2019; Garcia and Trinh, 2021). Choosing the right level of abstraction and validating the resulting designs experimentally, will therefore be central to the next wave of strain-engineering advances.

3.2. Medium optimization

Medium optimization aims to alter medium composition and concentration of ingredients to direct flux toward target product. It can be pursued either statically, by fixing a recipe before inoculation, or dynamically, by adjusting feeds on the fly, dynamic strategies are covered in more detail later in the bioreactor operation section. Studies in static medium optimization over the past decade have taken two distinct approaches, one based mechanistic models the other in purely statistical models.

Constraint based genome-scale models have been the basis of mechanistic approaches, recent examples include: *ToMI-FBA* added a "total-membrane-influx" constraint and correctly predicted that L-valine or late cellobiose pulses would lift alcohol titers in *Bacillus* and *Clostridium* fermentations (Nazem-Bokaee and Senger, 2015). The animal-cell counterpart, *Nutrition-GEM*, balanced elemental uptake in a CHO GEM, replaced glucose with cheaper carbon sources, and cut feed cost 12% while protecting antibody output (Weston and Thiele, 2023). The same element-budget logic doubled glutathione in *S. cerevisiae* (Schmacht et al., 2017), dynamic GEM simulations of *E. coli* flagged nitrogen limitation and guided an Asn/Gln/Arg co-feed that doubled single-chain variable fragment (scFv) titers (Behravan et al., 2022), and a constraint-based medium redesign for an enterobactin pathway pinpointed two missing amino acids that boosted siderophore yields (Swayambhu et al., 2020). As mentioned before, some constraint-based methods can even combine genetic manipulations with medium manipulation (Jensen et al., 2019). When stoichiometry or elemental balance dominates, these CBM approaches successfully identify the few uptake fluxes that matter most.

Unlike genetic manipulations, medium optimization lends itself to true high-throughput screening because microplate-scale cultivation makes it trivial to reshuffle dozens of nutrients across hundreds of wells. Singh et al. (2017) provides an in-depth review of fermentation medium-optimization strategies, from response-surface methodology (RSM) designs to evolutionary heuristics, concluding that statistically grounded and heuristic searches can efficiently navigate multi-factor media spaces when paired with high-throughput experimentation, while also noting limits around nonlinear interactions and transferability. A classical example of such a statistical approach is RSM, which fits the quadratic model

$$\hat{y} = \beta_0 + \sum_i \beta_i x_i + \sum_i \beta_{ii} x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j \tag{17}$$

to a small structured experiment in which k factors (e.g. glucose, nitrogen source, and pH, each coded to $[-1, +1]$) are varied across a fraction of their possible combinations and then locates the optimum analytically. The linear terms capture main effects, the quadratic terms curvature, and the cross-terms pairwise nutrient interactions. RSM is effective when the response has a single smooth optimum and interactions are limited to pairs, but is generally outperformed by flexible learners when the landscape is irregular or higher-order interactions dominate, as shown for fungal melanin where a neural network surpassed an RSM design (Saber et al., 2023). More recently, primary studies have leveraged microplate-scale screening to exploit that throughput: active-learning loops identified NaCl as a key lever for flaviolin in *P. putida* (Zournas et al., 2025), ML-assisted surveys disentangled distinct nutrient drivers for 4-aminophenylalanine versus tyrosine in engineered *E. coli* (Aida et al., 2023), and an "optimum seeker" tuned serum and 29 components to enhance HeLa-S3 growth signals (Hashizume et al., 2023). Complementary examples span straightforward regressions for elicitor timing in *Corylus avellana* (Farhadi et al.,

2020), GRNN-based confirmation under altered carbon supply (Salehi et al., 2021), Bayesian and single-cycle deep networks rapidly selecting carbon blends for recombinant protein in *E. coli* (Yoshida et al., 2023; Watanabe et al., 2024), and gradient-boosted active-learning schemes that favor *L. plantarum* over *E. coli* (Zhang et al., 2023). Even when executed in micro-bioreactor platforms (e.g., Ambr 15), the main lever in these studies remains medium composition rather than dynamic operation or scale transfer, as illustrated by DOE-guided simplification for *Neisseriaceae* outer-membrane vesicles (Baccante et al., 2023). In sum, microplate-scale reformulation makes medium composition the primary high-throughput lever for optimization, allowing ML models to efficiently map and exploit nutrient main effects and interactions to deliver reproducible, scalable gains without genetic edits or changes in operation mode.

Mechanistic CBM approaches excel at exposing elemental bottlenecks and prescribing biochemically transparent fixes, but they depend on painstaking model curation, steady-state assumptions, and a clear cellular objective, which can limit transferability across strains or dynamic phases (Nazem-Bokaee and Senger, 2015; Weston and Thiele, 2023). Conversely, the surge of microplate robotics has made it far easier to shuffle media recipes than to rebuild genomes, so data-driven statistical and machine-learning algorithms now dominate high-throughput screens, discovering nutrient interactions that elude human intuition, yet their black-box predictions can extrapolate poorly outside the measured design space and give little mechanistic insight (Hashizume et al., 2023; Aida et al., 2023). The most promising future direction therefore lies in hybridizing the two: use stoichiometric or elemental-balance constraints to prune the search space and regularize models, while letting active-learning ensembles explore the remaining degrees of freedom. Early examples, such as coupling *Nutrition-GEM* element budgets with Bayesian optimization for CHO feeds (Weston and Thiele, 2023), or interpreting random-forest feature importance through pathway context in engineered *E. coli* (Aida et al., 2023), hint that such hybrids can combine interpretability with predictive power, overcome each method's blind spots, and accelerate medium design toward truly adaptive, strain-aware recipes.

4. Integrating cellular models with bioreactor design and operation

Bioreactor design and operation entail choosing vessel geometry, establishing mixing and aeration capacity, and orchestrating time-varying inputs (e.g., feeds, temperature, pH, and dissolved oxygen) so cells convert substrates into product safely, efficiently, and at scale. In this section we focus on mechanistic descriptions and their hybrids with data-driven layers because these are the tools that explicitly link intracellular or population-level kinetics to reactor levers and constraints. We do not cover purely or mostly statistical/ML approaches in depth, as several recent reviews already synthesize that landscape: Helleckes et al. (2023) survey how ML has been deployed across strain engineering, process optimization, scale-up, monitoring, and control; Khanal et al. (2023) frame "smart bioprocesses" as the fusion of online data acquisition, hybrid modeling, and real-time decision-making while emphasizing validation and data governance for deployment; Agharafeie et al. (2023) map the progression from shallow neural network hybrids to deeper hybrid architectures for soft sensing, open/closed-loop control, and digital twins; and Wang et al. (2025) consolidate ML-first workflows for media and feed optimization that combine factor screening, feature-importance ranking, surrogate-model optimization, and staged experimental refinement. In parallel, a critical perspective on soft sensors highlights the practical hurdles that determine whether data-driven estimators can be trusted for batch phase control, sensor fault detection, and data synchronization, underscoring the need to anchor data-driven layers to conservation-law models and robust process analytical technology streams (Brunner et al., 2021).

Guided by that context, we organize the remainder of this section into four parts: first, unstructured kinetic models and their hybrids that guide quick feed and temperature strategies; second, advanced cellular models (CBM, proteome-constrained, and large-scale kinetic) and their hybrids that link intracellular fluxes to process levers; third, a subsection on embedding any of these cellular models within computational-fluid-dynamics simulations to resolve mixing heterogeneity; and fourth, a focused discussion of model tractability for repeated-call multiscale workflows. Together the tiers trace a path from the fastest, least detailed tools to the most mechanistic, showing where each fits into modern bioprocess development.

4.1. Unstructured kinetic models and hybrids

Unstructured kinetic models still dominate early-stage bioreactor studies because they couple seamlessly to the vessel mass-balance ODEs, run in milliseconds and can therefore be embedded in optimizer loops, Monte-Carlo robustness scans or even on-line soft-sensor frameworks. In practice they supply two main services: *(i)* fast *in-silico* screening of operating policies such as feed-rate trajectories or temperature schedules, and *(ii)* quick-to-tune models that support day-to-day decisions when data or time are scarce. For example, purely *in-silico* multi-objective optimization raised predicted specific productivity of hybridoma cultures by reshaping the glucose feed (Maria, 2020), while a similar model, enriched with elementary modes, identified a robust amino-acid recipe for CHO perfusion under parameter uncertainty (Wang et al., 2024). When wet-lab validation is performed, modest parameter sets still capture key trends: a two-stage temperature profile for *Lactiplantibacillus plantarum* boosted lactate titer and productivity (Sriphochanart and Skolpap, 2022); implementing a sucrose-limited feeding strategy increased BY-2 biomass yield (Nausch et al., 2023); similarly control of glucose supply on fed-batch increased succinate titer (Escanciano et al., 2023); and an exponential rapeseed-oil feed tripled MEL biosurfactant output in *Moesziomyces aphidis* (Beck et al., 2022). Even temperature-shift strategies for CHO fed-batch processes were narrowed to half the usual experimental workload once a six-parameter kinetic model was in place (Z. Wang et al., 2022). More recent examples have combined unstructured kinetics with additional layers, Pinto et al. incorporated model predictive control (MPC) and neural networks to improve fed-batch productivity of monoclonal antibody (mAb) (Pinto et al., 2023).

These examples, summarized in Table 3, illustrate how a "cheap" model can still pay rich dividends when interfaced with controllable levers in the bioreactor.

Looking forward, such simplicity is unlikely to disappear in the short term; instead, simple kinetics will serve as the lightweight layer in multi-fidelity digital twins, providing rapid predictions for most of the design space while higher-order (hybrid or GEM-based) surrogates are called only for critical gradients or quality-by-design assessments. Their speed also makes them ideal candidates for real-time model-predictive control on standard industrial hardware and for rapid uncertainty propagation when process sensors inject new data every few minutes.

4.2. Advanced cellular models and hybrids

Advanced cellular models for bioreactor process design span a spectrum of mechanistic depth, from constraint-based abstractions to resource-aware and fully kinetic descriptions (Table 4). Within this spectrum, CBMs remain the bioprocess workhorse because they deliver genome-scale flux insight at modest curation effort and computational cost. CBMs have been used to understand, design, and control bioprocesses.

In terms of explanatory use, a recent review by Park et al. (2024a) covered in depth the many explanatory applications of CBMs, focuses on CHO cell cultivation, from diagnosing pathway bottlenecks and predicting uptake/secretion phenotypes to guiding media and feed

Table 3
Recent studies that employ simple kinetic formulations for bioreactor design and operation.

Method[a]	Organism	Product	Lever	Key outcome	Ref.
Unstructured kinetic + multi-objective optimization	Hybridoma cells	mAb	Fed-batch glucose feed profile	Predicted higher specific productivity.[b]	Maria (2020)
Unstructured kinetic + elementary modes	CHO	mAb	Amino-acid formulation (perfusion)	Medium predicted to maximize titer under uncertainty.[b]	Wang et al. (2024)
Kinetic model (temperature shift)	CHO	mAb	32 to 28 °C shift and feed	Highest titer with 50% fewer runs.	Z. Wang et al. (2022)
Unstructured kinetic	*L. plantarum*	L-lactate	Temperature profile and feed	Titer 1.5-fold higher	Sriphochanart and Skolpap (2022)
Unstructured kinetic	Tobacco BY-2 cells	Biomass	Sucrose concentration	Reduced sucrose 30%, biomass increased 25%.	Nausch et al. (2023)
Unstructured kinetics	*Moesziomyces aphidis*	MEL biosurfactant	Exponential rapeseed-oil feed	Tripled biomass; 50 g L^{-1} MEL.	Beck et al. (2022)
Unstructured kinetic	*Actinobacillus succinogenes*	Succinic acid	Inoculum, yeast extract, agitation, CO_2 flow	Model predicted fed-batch conditions with higher succinate titer and selectivity.[b]	Escanciano et al. (2023)
Unstructured kinetic + MPC	CHO	mAb	Real-time nutrient feed	Specific productivity up more than 20%.	Pinto et al. (2023)

[a] These approaches combine the indicated method with basic reactor mass balance models.
[b] Purely *in silico* studies (no wet-lab validation).

composition, cell-line engineering targets, integrating multi-omics layers, and product-quality attributes when secretory and glycosylation pathways are coupled to the metabolic backbone. A more recent primary research example is the use of parsimonious flux balance analysis (pFBA) to rationalize pH-dependent glycoform shifts in mAb production (Reddy et al., 2025).

CBMs play an important role as design tools. dFBA couples reactor-scale mass balance ODEs with repeated FBA optimizations over time under a quasi-steady-state intracellular assumption (Mahadevan et al., 2002). Numerically, however, dFBA requires care because non-unique FBA optima, switches in active constraints, and substrate-depletion events can make the exchange fluxes supplied to the reactor ODEs discontinuous or infeasible. Reliable implementations address these issues with lexicographic tie-breaking, event handling, and differential–algebraic or embedded-optimization formulations (Höffner et al., 2013; Gomez et al., 2014; Harwood et al., 2016); when repeated solves dominate cost, validated surrogates can also be used within a defined state domain Song et al. (2025). Many dFBA variants are now routinely applied in bioprocess development: they have guided fed-batch shikimate production (Kuriya and Araki, 2020), two-stage polyhydroxybutyrate synthesis (Lasry Testa et al., 2022), and multi-stage mAb production in CHO perfusion processes using state-informed constraints (Gopalakrishnan et al., 2024). Specific tools like OptMSP have been conceived for the design of multi-stage processes (e.g., in the growth stage biomass is formed and in the non-growth stage product is formed) coupling a GEM with ODEs (Bauer and Klamt, 2024); another approach besides using ML (Gopalakrishnan et al., 2024) or ODEs (Bauer and Klamt, 2024) to account for different cellular states, are Flexible-Net Petri nets, which were used to embed both GEM and reactor balances in a single MILP, reproducing historical citramalate benchmarks (Lázaro et al., 2022).

Within design applications, feeding protocols are a primary lever for aligning cellular metabolism with process objectives. In practice, models are used to explore substrate compositions and scheduling (e.g., phase-specific amino acids, lipids, and trace nutrients; pulse versus continuous additions) under typical bioreactor constraints such as oxygen transfer and osmolality, with distinct growth- and production-phase targets. Reported process-facing outcomes include *in silico* reductions in medium cost via GEMs-driven media optimization for continuous culture (Pérez-Fernández et al., 2021), and performance gains from CHO model-guided feed tailoring (Park et al., 2024a).

Finally, as control tools, CBMs have been embedded within nonlinear MPC (a receding-horizon scheme that forecasts system trajectories, optimizes the next control move to stay on target, and updates as new measurements arrive) for yeast ethanol production (Chang et al., 2016) and to propose glucose-feed policies that increased CHO cell mAb titers *in silico* (Monteiro et al., 2023). Collectively, these studies illustrate a progression from design-time screening of feed composition and timing to run-time supervisory control, with models serving both to reduce medium cost and to navigate the productivity–byproduct trade-offs typical of high-density cultures.

Beyond classical CBMs, several theoretical frameworks now model metabolism dynamically with explicit resource allocation or detailed kinetics. Representative examples include the dynamicME method, which uses an *E. coli* ME-model to describe batch and fed batch scenarios (Yang et al., 2019), the regulatory dynamic enzyme cost FBA formulation (r-deFBA) that augments dFBA with enzyme capacity and transcriptional regulation to describe batch fermentations (Liu and Bockmayr, 2020), and large scale kinetic reconstructions used to interrogate mechanisms such as acetate overflow in *E. coli* and amino acid limitations in CHO cells (Millard et al., 2021; Robitaille et al., 2015). At present these methods are at an initial formulation and explanatory stage, and there are no published demonstrations of bioreactor design or feedback control driven by these models; nevertheless, their ability to couple metabolism, regulation, and process perturbations indicates clear potential for future design optimization and MPC in industrial settings.

4.3. Coupling cellular models to computational fluid dynamics and reactive-transport models

Bioreactor design and operation hinge on predicting how spatial transport of momentum and species shapes cell physiology, productivity, and robustness; to that end, we use computational fluid dynamics (CFD) to denote models that solve the Navier–Stokes equations on a three-dimensional mesh to obtain local velocity, pressure, and scalar fields, revealing heterogeneities (e.g., mixing deficits, oxygen-starvation pockets, and pH microenvironments) that well-mixed models cannot capture (Rathore et al., 2016). As a computationally cheaper alternative, we refer to one-dimensional reactive transport models (RTMs), where ‘‘one-dimensional’’ means the reactor is reduced to a

Table 4
Advanced cellular models applied to bioreactor design and operation.

Method[a]	Organism	Product	Lever	Key outcome	Ref.
CBM	*E. coli*	Citramalate	Dilution rate and substrate concentration	Predicted optimal combination; matched data[b]	Lázaro et al. (2022)
CBM	CHO	Biomass	Amino-acid profile (chemostat)	Predicted 25% medium-cost saving[b]	Pérez-Fernández et al. (2021)
dFBA + ML (*COSMIC-dFBA*)	CHO	mAb	Perfusion nutrient feed	Forecasted cell density and size and metabolite consumption and secretion better than alternative methods[b]	Gopalakrishnan et al. (2024)
dFBA	*E. coli*	Shikimic acid	Genotype evaluation	High-producer reached 84% of simulated maximum	Kuriya and Araki (2020)
CBM + ODEs (*OptMSP*)	*E. coli*	Lactate	Stage sequencing	Volumetric productivity maximized[b]	Bauer and Klamt (2024)
dFBA	*Synechocystis* sp.	PHB	Nitrogen-starved production	Predicted 4.8 g L^{-1} PHB for two stage process[b]	Lasry Testa et al. (2022)
dFBA + non-linear MPC	*S. cerevisiae*	Ethanol	Glucose and oxygen feeds	Ethanol titer +15%	Chang et al. (2016)
CBM + kinetic model + nonlinear MPC	CHO	mAb	Glucose feed concentration and volumetric feed/bleed rates	Predicted up to 95% higher mAb titer versus an emulated baseline[b]	Monteiro et al. (2023)

[a] These approaches combine the indicated method with basic reactor mass balance models.
[b] Purely *in silico* studies.

single spatial coordinate (e.g., axial height in a column or a representative radial path) with cross-sectional variations averaged out; RTMs solve advection–dispersion–reaction balances for species along that coordinate while *prescribing* hydrodynamics via correlations (e.g., dispersion coefficients, mass-transfer rates) rather than solving the momentum equations. In reactor design, this makes RTMs well suited for rapid screening of operating windows and for estimating axial profiles (e.g. substrates, dissolved oxygen, pH, products), albeit with fidelity limited by the one dimensional averaging and the accuracy of the underlying correlations.

A recent review by Singh et al. (2024) catalogs coupling strategies between CFD and cellular reaction kinetics, from loose one-way links (passing mass-transfer coefficients) to fully integrated transport-kinetics solvers. An audit of their post-2015 corpus indicates that the majority of studies (roughly 80%) embed unstructured Monod-type or Luedeking–Piret rate expressions in each CFD cell, with the remaining studies using overall stoichiometric models or population balance models (i.e., common chemical engineering models that describe particles, in this case cells, per volume with birth/death and advection–diffusion–reaction terms). These models are used because they are simple to construct and, most importantly, very inexpensive to evaluate.

To date, we are not aware of published CFD applications that embed advanced cellular models such as CBMs directly; however, several couplings of CBMs with RTMs have been demonstrated. Syngas fermentation is a particularly relevant testbed, with commercial deployments motivating reactor-scale analysis (e.g., industrial acetogen processes by LanzaTech), and a one-dimensional multiphase convection–dispersion RTM coupled to a genome-scale CBM for *C. ljungdahlii* showed that very high CO transfer and increased H_2 uptake substantially boost ethanol titer and shift selectivity away from acetate, highlighting mass transfer as a dominant design lever (Gomez et al., 2015). In that study, computational cost remained manageable, however many studies coupling CBMs with RTMs have encountered computational challenges that prompted acceleration strategies. The main approach to contain the prohibitive cost of per-cell CBM solves has been *fast surrogates*: off-line look-up tables of FBA solutions interpolated during transport integration (Scheibe et al., 2009), direct (on-the-fly) enrichment that grows the table as new regimes are explored (Fang et al., 2011), and, most recently, ANNs that emulate FBA end-to-end and deliver orders-of-magnitude speedups while improving numerical robustness (Song et al., 2025). Closely related to the same syngas bubble-column application, Siebler et al. (2020) combined a fast 1-D RTM (using unstructured kinetics) for design screening with targeted 3-D CFD to verify hydrodynamics, thereby addressing computational burden; this hybrid RTM–CFD staging offers a practical route to eventually couple more expensive CBMs to CFD in large-scale reactor studies.

4.4. Model tractability for multiscale integration

In the context of multi-scale integration, model tractability means that a cellular model can be evaluated repeatedly and robustly inside a larger process calculation. A single FBA solve can be fast, but dFBA, RTM, CFD, optimization, or MPC workflows may request growth, uptake, and secretion rates across many time points, grid cells, parameter samples, or control iterations. The practical question is therefore whether repeated evaluations can deliver the needed rates within the decision time budget while remaining stable near substrate depletion, metabolic switching, or changing constraints.

For accuracy–speed trade-offs in metabolic simulations, the relevant surrogates are cheaper evaluators for repeated FBA calls in dynamic or spatial workflows. Early RTM couplings used precomputed or reused FBA solutions: Scheibe et al. (2009) interpolated from 1000 FBA solutions over different medium conditions, whereas Fang et al. (2011) stored FBA solutions and reused them through a binary-tree search. These studies demonstrated how repeated linear-programming calls could be avoided, but they did not run the direct FBA-in-the-loop baseline needed for a clean speed-up or accuracy-loss comparison. Song et al. (2025) provides the clearest direct benchmark: their ANN surrogate for FBA in *S. oneidensis* matched FBA targets with correlations above 0.99, reproduced substrate switching, and reduced batch dFBA runtime to 0.1–0.2% of the linear-programming runtime, equivalent to roughly a 500–1000-fold speed-up. In the 1-D RTM, it was about three orders of magnitude faster for 10–30 grids; at 100 grids, the ANN-based model solved in minutes on a desktop, whereas the linear-programming

RTM did not generate stable solutions. Thus, no material accuracy loss was reported within the sampled substrate and oxygen domain, but extrapolation outside that domain remains the main limitation. These three examples remain ad hoc implementations: no general tool currently converts an FBA model into a validated surrogate and embeds it in dFBA, RTM, or CFD.

When access to high-performance computing (HPC) is available, the key question is whether parallelization overhead is small relative to each cellular-model call. Independent FBA sweeps can be distributed efficiently when each task is large enough, but dispatching many small state queries from a tightly coupled dFBA, RTM, CFD, optimization, or MPC loop can make overhead comparable to the solve being accelerated. As an order-of-magnitude calculation, the overhead is $N\tau$, where N is the number of grid-cell/time-step queries and τ is per-query overhead. Dask, a common Python library for parallel computing, reports about 200 μs–1 ms of overhead per task and warns that graphs with millions of tasks can incur minutes-to-hours of overhead (Dask Development Team, 2026). Thus, for $N = 10^6$, $\tau = 1$ ms gives 10^3 s, or 16.7 min, while a sensitivity case of $\tau = 10$ ms gives 10^4 s, or 2.8 h, before any linear programs are solved. Parallel computing therefore does not remove the need for validated surrogates; it makes them especially useful when the upper-level workflow requires many fast, local evaluations so computation can be spent on finer grids, uncertainty samples, or longer control horizons rather than repeated cellular solves.

5. Linking cellular models to TEA and LCA

Techno-economic analysis (TEA) converts mass- and energy-balance data from process simulations into financial metrics such as minimum selling price (MSP), providing a fast screen for commercial plausibility and R&D prioritization of bioconversion processes (Scown et al., 2021; Poddar and Scown, 2025). Life cycle assessment (LCA) translates the same inventories into cradle-to-grave indicators to benchmark environmental performance, with a common example being CO_2 equivalents emitted per kg of product, which is especially valuable at early development stages when design choices are still fluid (Wowra et al., 2023). Used iteratively, TEA and LCA establish the economic and sustainability "gates" that new biotechnologies should clear before scale-up (Scown et al., 2021; Wowra et al., 2023).

At the plant level, MSP is commonly defined as the product price that makes the project net present value equal to zero,

$$\mathrm{NPV}(\mathrm{MSP}) = 0. \tag{18}$$

Net present value converts future revenues and costs to today's value using an assumed discount rate, i.e. the chosen rate used to account for the time value of money, and is expressed in currency units. Eq. (18) therefore means that product revenues and coproduct credits balance capital, operating, financing, and other project costs over the project lifetime after all cash flows are expressed on the same present-value basis. In early steady-state screening TEA, where annual production and costs are approximated by representative yearly values, this discounted-cash-flow calculation is often summarized by the conceptual annualized approximation

$$\mathrm{MSP} \approx \frac{\text{net annualized plant cost}}{\text{annual target-product output}} = \frac{a_{\mathrm{cap}} C_{\mathrm{CAPEX}} + C_{\mathrm{OPEX}} - R_{\mathrm{coproduct}}}{m_{P,\mathrm{yr}}}, \tag{19}$$

where the left fraction makes clear that MSP has units of price, for example USD kg^{-1} product when annual costs are expressed in USD yr^{-1} and production in kg yr^{-1}. In the expanded form, C_{CAPEX} is total capital investment (USD), a_{cap} is a capital charge factor that converts upfront capital into an equivalent annual cost (yr^{-1}), C_{OPEX} is annual operating cost (USD yr^{-1}), $R_{\mathrm{coproduct}}$ is annual coproduct revenue or credit (USD yr^{-1}), and $m_{P,\mathrm{yr}}$ is annual production of the target product (kg yr^{-1}). Analogously, LCA converts a process inventory into impact indicators by multiplying inventory flows by characterization factors,

$$\mathrm{Impact}_k = \frac{\text{total impact in category } k}{\text{functional-unit product mass}} = \frac{1}{m_{P,\mathrm{FU}}} \sum_{\ell} \mathrm{CF}_{k\ell} I_{\ell}, \tag{20}$$

where k denotes an impact category, such as climate change, eutrophication, acidification, or water use. I_{ℓ} is the amount of inventory flow ℓ crossing the modeled system boundary after upstream or background processes have been accounted for; examples include kg glucose or other feedstock, kWh electricity, MJ steam, kg CO_2 emitted, m^3 wastewater, or kg nitrogen discharged. $\mathrm{CF}_{k\ell}$ has units of impact in category k per unit of I_{ℓ}, so the final result has units such as kg CO_2-equivalent kg^{-1} product for climate change. $m_{P,\mathrm{FU}}$ is the product mass used as the functional-unit denominator, often 1 kg of product.

Eqs. (19) and (20) also clarify the boundary between cellular models and plant-level accounting. Cellular and fermentation models (i.e. models that integrate cellular behavior with bioreactor operation) provide biological inputs to the process model, such as product yield on substrate $Y_{P/S}$ (kg product kg^{-1} substrate), by-product or coproduct yield $Y_{j/S}$ for side product j (kg side product kg^{-1} substrate), product titer C_P (kg m^{-3}), and productivity r_P (kg m^{-3} h^{-1}). The process model then translates these inputs into the quantities used by TEA/LCA: feedstock flow F_S (kg substrate yr^{-1}), annual target-product output $m_{P,\mathrm{yr}}$, annual side-product flow $m_{j,\mathrm{yr}}$ (kg yr^{-1}), equipment sizes that contribute to C_{CAPEX}, operating costs C_{OPEX}, coproduct credits $R_{\mathrm{coproduct}}$, and inventory flows I_{ℓ}.

In conventional bioprocess TEA/LCA, this fermentation block is usually kept simple. It may be an overall stoichiometric equation, a set of fixed titer-rate-yield assumptions, or a simple unstructured kinetic model summarized into titer, yield, productivity, and cycle time for the plant model (Baral and Shah, 2016; Kadhum et al., 2018; Balina et al., 2023; Davis and Dempsey, 2023). For example, if annual product output is fixed, $F_S \approx m_{P,\mathrm{yr}}/Y_{P/S}$, so improving $Y_{P/S}$ lowers feedstock cost in C_{OPEX} and reduces feed-related inventory flows I_{ℓ}. Similarly, $m_{j,\mathrm{yr}} \approx Y_{j/S} F_S$, so side products become coproduct revenue if they are recovered and sold, or waste-treatment and LCA burdens if they are not. This coarse representation is transparent and computationally cheap, which is valuable when feedstock-price uncertainty or product recovery and purification assumptions dominate the result (Scown et al., 2021; Poddar and Scown, 2025).

The stoichiometric and unstructured kinetic approaches cannot ask what happens when a new strain, variable feedstock, co-feed, inhibitor, or oxygen-limited regime changes the cell's yield, coproduct pattern, nutrient demand, oxygen demand, off-gas composition, titer, or productivity. One exception is that simple kinetic models can represent time courses and inhibition, but only within their calibration range. More detailed cellular models become useful when those physiological changes are large enough to alter process inputs to TEA/LCA. A small but growing set of studies makes this link more explicit by coupling CBMs to process and TEA/LCA models. The purpose is mainly to compare biologically feasible conversion routes and then ask whether the route changes feed use, coproducts, separations, utilities, emissions, or product cost. Akbari and Barton (2019) recast an algal CBM as constraints inside a plant-level optimization problem, allowing metabolic routes, separations, and utilities to be selected together. Other studies use a CBM-derived flux map as a smaller surrogate for the full metabolic model. In this context, the CBM is solved first for selected substrate-to-product routes, and the resulting uptake rates, secretion rates, and coproduct patterns are passed to the process model as simplified reaction blocks. This keeps the process optimization computationally tractable because equipment sizing, cost estimation, and LCA calculations do not need to solve the full CBM at every iteration. For example, Van der Hauwaert et al. (2023) used flux-map reaction blocks in an optimization model that evaluated MSP and midpoint LCA indicators, and later extended this logic to food-waste valorization with equipment

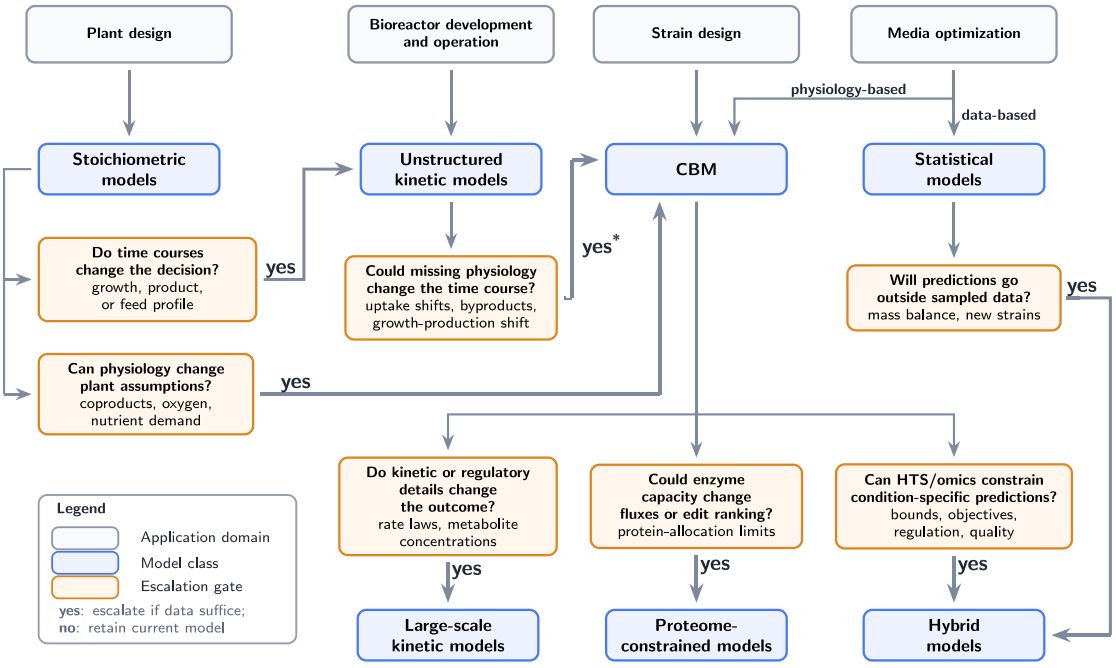

Fig. 3. Fitness-to-purpose model selection. Practical summary of the application patterns found in this review. When several positive paths are plausible, the branch should be selected according to the dominant bottleneck. Escalation should also account for data availability and the computational time budget. * denotes using dFBA or another dynamic CBM coupling to retain time-course detail.

sizing and cost correlations (Van der Hauwaert et al., 2024). Baral et al. (2024) coupled a microbial CBM with a biorefinery process model for bisabolene production and iterated between flux predictions and equipment sizing; the result was not a direct cellular prediction of price, but a process-level assessment of whether microbial lipids and biogenic CO_2 upgrading could approach market competitiveness under specific yield and H_2 cost assumptions.

In current practice, bioprocess TEA/LCA still relies mainly on lumped stoichiometries, fixed process-performance assumptions, or simple unstructured kinetics. The recent CBM-linked examples therefore point to an interesting but still underexplored direction: integrating advanced cellular models into TEA/LCA, using reduced surrogates when needed for computational tractability, to test whether physiology changes cost or impact conclusions.

6. Conclusions and outlook

The evidence reviewed here points to a practical rule for industrial bioprocess modeling: cellular models should be selected by the decision they can improve, not by their mechanistic sophistication alone. Simple stoichiometric and unstructured kinetic models remain valuable when the task is rapid screening, control-oriented simulation, or mass-balanced plant assessment, whereas CBMs, proteome-constrained models, kinetic models, and hybrids are most useful when intracellular physiology can change strain priorities, feed design, byproduct formation, product quality, oxygen or nutrient demand, or downstream burden. A pragmatic multi-fidelity workflow therefore starts with the simplest model that can support the decision, escalates only when missing physiology could alter the outcome, keeps repeated simulations tractable through reduction, precomputation, or validated surrogates, and carries cellular predictions into TEA and LCA early enough for economic and environmental constraints to feed back into strain and process targets (Fig. 3).

The modeling families reviewed here are already used across industrial bioprocessing, but at different levels of maturity. Stoichiometric balances, unstructured kinetics, statistical models, and process-model hybrids are established tools for process development, monitoring, optimization, and automation. Future development should focus especially on advanced cellular models, including CBMs and their resource-aware, kinetic, and hybrid extensions, because they offer the clearest route to stronger predictive power when decisions depend on the mechanistic link between strain genotype, medium composition, flux redistribution, and process performance. Public company-linked examples already show model-guided workflows supporting strain design for metabolite production, optimization of gas fermentation, filamentous fungal production hosts, amino-acid production, CHO media and process development, and mechanistic understanding of dairy starter-strain metabolism (Yim et al., 2011; Heffernan et al., 2020; Liew et al., 2022; Pel et al., 2007; Vongsangnak et al., 2008; Schwentner et al., 2019; Park et al., 2024b; Rau et al., 2022). Much additional industrial use is likely internal, because strain history, proprietary media, feedstock specifications, omics and enzyme measurements, batch records, product-quality assays, and validation results are commercially sensitive and platform-specific. This limits transferability, benchmarking, and external reuse, while routine deployment also requires independent validation, model maintenance as strains, raw materials, sensors, media, and equipment change, and integration with laboratory information, manufacturing execution, online sensing, control, and quality-system platforms.

Broader use of advanced cellular models will require workflows that reduce dependence on specialized in-house curation while connecting data generation, model updating, and validation in the same loop. Existing model-quality checks, automated draft-reconstruction pipelines, reaction and metabolite databases, enzyme-parameter resources, public omics repositories, and machine-learning tools for missing kinetic parameters already lower the digital curation barrier (Lieven et al., 2020; Heinken and Thiele, 2025; Li et al., 2022b; Zhai et al., 2025; Boorla and Maranas, 2025). The next step is to extend automation beyond early draft assembly: fermentation phenotypes, omics measurements, medium recipes, strain variants, enzyme parameters, and batch context still need more standardized descriptions and more automated tools

that translate experimental conditions into model configurations, validation datasets, and auditable curation decisions. In the near term, LLM-assisted and agentic AI workflows are likely to help bridge this data-to-curation gap by accelerating literature triage, parameter extraction, model-code generation, consistency checks, and experiment prioritization, while still requiring expert review and independent validation (Schmidt et al., 2024; Caufield et al., 2024). As curation, data-integration, and simulation-time bottlenecks are reduced, advanced cellular models are more likely to add value in multiscale workflows where genotype-, medium-, and flux-dependent constraints alter yield, productivity, coproduct formation, oxygen and nutrient demand, or downstream burden. Realizing this value will also require shared data schemas, versioned assumptions, collaboration platforms, and multidisciplinary teams linking strain, fermentation, downstream, automation/data, and TEA/LCA expertise, so that cellular predictions can be translated into process and plant decision variables—and bioprocess development can move from empirical trial-and-error toward validated predict-and-design workflows.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work the author used ChatGPT to assist with grammar and style suggestions and LaTeX programming. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Funding

This work was supported by the IE University Tenure-Track Faculty Seed Fund.

Declaration of competing interest

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

Data availability

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

References

Agharafeie, R., Ramos, J.R.C., Mendes, J.M., Oliveira, R., 2023. From shallow to deep bioprocess hybrid modeling: Advances and future perspectives. Fermentation 9 (10), 922. http://dx.doi.org/10.3390/fermentation9100922.

Aida, H., Uchida, K., Nagai, M., Hashizume, T., Masuo, S., Takaya, N., Ying, B.-W., 2023. Machine learning-assisted medium optimization revealed the discriminated strategies for improved production of the foreign and native metabolites. Comput. Struct. Biotechnol. J. 21, 2654–2663. http://dx.doi.org/10.1016/j.csbj.2023.04.020.

Akbari, A., Barton, P.I., 2019. Integrating genome-scale and superstructure optimization models in techno-economic studies of biorefineries. Processes 7 (5), 286. http://dx.doi.org/10.3390/pr7050286.

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., Jirstrand, M., 2014. Kinetic models in industrial biotechnology – improving cell factory performance. Metab. Eng. 24, 38–60. http://dx.doi.org/10.1016/j.ymben.2014.03.007.

Asin-Garcia, E., Fawcett, J.D., Batianis, C., Martins dos Santos, V.A., 2025. A snapshot of biomanufacturing and the need for enabling research infrastructure. Trends Biotechnol. 43 (5), 1000–1014. http://dx.doi.org/10.1016/j.tibtech.2024.10.014, URL: https://www.sciencedirect.com/science/article/pii/S0167779924003068.

Ataman, M., Hatzimanikatis, V., 2017. redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. PLoS Comput. Biol. 13 (7), e1005444. http://dx.doi.org/10.1371/journal.pcbi.1005444.

Baccante, A., Petruccelli, P., Saudino, G., Ragnoni, E., Johansson, E., Di Cioccio, V., Mazarakis, K., 2023. Optimization of a bacterial cultivation medium via a design-of-experiment approach in a sartorius ambr® 15 fermentation micro-bioreactor system. Fermentation 9 (12), 1002. http://dx.doi.org/10.3390/fermentation9121002.

Bailey, J.E., Ollis, D.F., 1986. Biochemical Engineering Fundamentals. McGraw-Hill.

Balina, K., Soloha, R., Suleiko, A., Dubencovs, K., Liepins, J., Dace, E., 2023. Prospective life cycle assessment of microbial sophorolipid fermentation. Fermentation 9 (9), 839. http://dx.doi.org/10.3390/fermentation9090839.

Baral, N.R., Banerjee, D., Mukhopadhyay, A., Simmons, B.A., Singer, S.W., Scown, C.D., 2024. Integration of genome-scale metabolic model with biorefinery process model reveals market-competitive carbon-negative sustainable aviation fuel utilizing microbial cell mass lipids and biogenic CO2. BioResources 19 (3), 4056–4086. http://dx.doi.org/10.15376/biores.19.3.4056-4086.

Baral, N.R., Shah, A., 2016. Techno-economic analysis of cellulosic butanol production from corn stover through acetone–butanol–ethanol fermentation. Energy & Fuels 30 (7), 5779–5790. http://dx.doi.org/10.1021/acs.energyfuels.6b00819.

Bauer, J., Klamt, S., 2024. OptMSP: A toolbox for designing optimal multi-stage (bio)processes. J. Biotech. 383, 94–102. http://dx.doi.org/10.1016/j.jbiotec.2024.01.009.

Beber, M.E., Gollub, M.G., Mozaffari, D., Shebek, K.M., Flamholz, A.I., Milo, R., Noor, E., 2022. eQuilibrator 3.0: a database solution for thermodynamic constant estimation. Nucleic Acids Res. 50 (D1), D603–D609. http://dx.doi.org/10.1093/nar/gkab1106.

Beck, A., Vogt, F., Hägele, L., Rupp, S., Zibek, S., 2022. Optimization and kinetic modeling of a fed-batch fermentation for mannosylerythritol lipids (MEL) production with *Moesziomyces aphidis*. Front. Bioeng. Biotechnol. 10, 913362. http://dx.doi.org/10.3389/fbioe.2022.913362.

Becker, J., Rush, N., Barnes, E., Rein, D., 2025. Measuring the impact of early-2025 AI on experienced open-source developer productivity. arXiv:2507.09089.

Behravan, A., Hashemi, A., Marashi, S.-A., 2022. A constraint-based modeling approach to reach an improved chemically defined minimal medium for recombinant antiEpEX-scFv production by *Escherichia coli*. Biochem. Eng. J. 179, 108339. http://dx.doi.org/10.1016/j.bej.2022.108339.

Benavides, P.T., Balchandani, S., Gracida-Alvarez, U.R., 2024. Environmental analysis of biotechnologies for biofuels, bioplastics, and bioproducts: a greenhouse gas (GHG) emissions review. Biotechnol. Environ. 1 (1), 10. http://dx.doi.org/10.1186/s44314-024-00010-5.

Bi, X., Cheng, Y., Xu, X., Lv, X., Liu, Y., Li, J., Du, G., Chen, J., Ledesma-Amaro, R., Liu, L., 2023. etiBsu1209: A comprehensive multiscale metabolic model for Bacillus subtilis. Biotechnol. Bioeng. 120 (6), 1623–1639. http://dx.doi.org/10.1002/bit.28355.

Boorla, V.S., Maranas, C.D., 2025. CatPred: a deep-learning framework for in-vitro enzyme kinetic parameter prediction. Nat. Commun. 16, 2072.

Brandl, J., Aguilar-Pontes, M.V., Schäpe, P., Noerregaard, A., Arvas, M., Ram, A.F.J., Meyer, V., Tsang, A., de Vries, R.P., Andersen, M.R., 2018. A community-driven reconstruction of the Aspergillus niger metabolic network. Fungal Biology Biotechnol. 5, 16. http://dx.doi.org/10.1186/s40694-018-0060-7.

Brunner, V., Siegl, M., Geier, D., Becker, T., 2021. Challenges in the development of soft sensors for bioprocesses: A critical review. Front. Bioeng. Biotechnol. 9, 722202. http://dx.doi.org/10.3389/fbioe.2021.722202.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84 (6), 647–657.

Caufield, H., Kroll, C., O'Neil, S.T., Reese, J.T., Joachimiak, M.P., Hegde, H., Harris, N.L., Krishnamurthy, M., McLaughlin, J.A., Smedley, D., Haendel, M.A., Robinson, P.N., Mungall, C.J., 2024. CurateGPT: A flexible language-model assisted biocuration tool. arXiv:2411.00046.

Chang, L., Liu, X., Henson, M.A., 2016. Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. J. Process Control 42, 137–149. http://dx.doi.org/10.1016/j.jprocont.2016.04.012.

Chen, Y., Gustafsson, J., Tafur Rangel, A., Anton, M., Domenzain, I., Kittikunapong, C., Li, F., Yuan, L., Nielsen, J., Kerkhoven, E.J., 2024. Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO toolbox 3.0. Nat. Protoc. 19, 629–667. http://dx.doi.org/10.1038/s41596-023-00931-7.

Cho, J.S., Kim, G.B., Eun, H., Moon, C.W., Lee, S.Y., 2022. Designing microbial cell factories for the production of chemicals. JACS Au 2, 1781–1799. http://dx.doi.org/10.1021/jacsau.2c00344.

Choudhury, S., Narayanan, B., Moret, M., Hatzimanikatis, V., Miskovic, L., 2024. Generative machine learning produces kinetic models that accurately characterise intracellular metabolic states. Nat. Catal. 7, 1086–1098. http://dx.doi.org/10.1038/s41929-024-01220-6.

Chowdhury, A., Zomorrodi, A.R., Maranas, C.D., 2014. k-OptForce: integrating kinetics with flux balance analysis for strain design. PLoS Comput. Biol. 10 (2), e1003487.

Cowan, A.E., Klass, S.H., Winegar, P.H., Keasling, J.D., 2023. Microbial production of fuels, commodity chemicals, and materials from sustainable sources of carbon and energy. Curr. Opin. Syst. Biology 36, 100482. http://dx.doi.org/10.1016/j.coisb.2023.100482.

Cui, Y., Xu, Y., Zhang, X., Li, Z., 2024. Sustainable aviation fuel: Biomass fostered future aviation. Innov. Energy 1 (1), 100007. http://dx.doi.org/10.59717/j.xinn-energy.2024.100007, URL: https://www.the-innovation.org/data/article/energy/preview/pdf/XINNENERGY-2024-0010.pdf.

Dask Development Team, 2026. Dask best practices. (Accessed 19 June 2026), https://docs.dask.org/en/stable/best-practices.html.

Davis, R., Dempsey, J., 2023. Techno-Economic Case Study: Fermentation Cost Impacts from Selected Critical Material Attributes. Technical Report, National Renewable Energy Laboratory, http://dx.doi.org/10.2172/2530727, URL: https://www.osti.gov/biblio/2530727.

Dinh, H.V., King, Z.A., Palsson, B.O., Feist, A.M., 2018. Identification of growth-coupled production strains considering protein costs and kinetic variability. Metab. Eng. Commun. 7, e00080. http://dx.doi.org/10.1016/j.mec.2018.e00080, URL: https://www.sciencedirect.com/science/article/pii/S2214030118300014.

Doyle, K., Tsopanoglou, A., Fejér, A., Glennon, B., del Val, I.J., 2023. Automated assembly of hybrid dynamic models for CHO cell culture processes. Biochem. Eng. J. 191, 108763. http://dx.doi.org/10.1016/j.bej.2022.108763.

Du, Y.-H., Zhang, Z.-H., Guo, Y.-R., Li, N., Yang, W.-D., 2022. Optimization and scale-up of fermentation processes driven by models. Bioengineering 9 (9), 473. http://dx.doi.org/10.3390/bioengineering9090473.

Duarte, R., Jorge, J., Ferreira, P., Neves, J., 2021. Sensor fusion in bioprocesses—On the estimation of biomass formation using the extended and unscented Kalman filters. Sensors 21 (7), 2344. http://dx.doi.org/10.3390/s21072344.

Eastham, J.L., Leman, A.R., 2024. Precision fermentation for food proteins: ingredient innovations, bioprocess considerations, and outlook—A mini-review. Curr. Opin. Food Sci. 58, 101194. http://dx.doi.org/10.1016/j.cofs.2024.101194.

Escanciano, I.A., Ladero, M., Santos, V.E., Blanco, A., 2023. Development of a simple and robust kinetic model for the production of succinic acid from glucose depending on different operating conditions. Fermentation 9 (3), 155. http://dx.doi.org/10.3390/fermentation9030155.

Fang, Y., Scheibe, T.D., Mahadevan, R., Garg, S., Long, P.E., Lovley, D.R., 2011. Direct coupling of a genome-scale microbial *In Silico* model and a groundwater reactive transport model. J. Contam. Hydrol. 122 (1–4), 96–103. http://dx.doi.org/10.1016/j.jconhyd.2010.11.007.

Farhadi, S., Salehi, M., Moieni, A., Safaie, N., Sabet, M.S., 2020. Modeling of paclitaxel biosynthesis elicitation in *Corylus avellana* cell culture using adaptive neuro-fuzzy inference system–genetic algorithm (ANFIS–GA) and multiple regression methods. PLoS One 15 (8), e0237478. http://dx.doi.org/10.1371/journal.pone.0237478.

Faure, L., Mollet, B., Liebermeister, W., Faulon, J.-L., 2023. A neural–mechanistic hybrid approach improving the predictive power of genome-scale metabolic models. Nat. Commun. 14, 4669. http://dx.doi.org/10.1038/s41467-023-40380-0.

Futing, P., Barthel, L., Cairns, T.C., Briesen, H., Schmideder, S., 2021. Filamentous fungal applications in biotechnology: a combined bibliometric and patentometric assessment. Fungal Biol. Biotechnol. 8, 23. http://dx.doi.org/10.1186/s40694-021-00131-6.

Garcia, S., Thompson, R.A., Giannone, R.J., Dash, S., Maranas, C.D., Trinh, C.T., 2020. Development of a genome-scale metabolic model of clostridium thermocellum and its applications for integration of multi-omics datasets and computational strain design. Front. Bioeng. Biotechnol. 8, 772. http://dx.doi.org/10.3389/fbioe.2020.00772, URL: https://www.frontiersin.org/article/10.3389/fbioe.2020.00772.

Garcia, S., Trinh, C.T., 2019. Multiobjective strain design: A framework for modular cell engineering. Metab. Eng. 51, http://dx.doi.org/10.1016/j.ymben.2018.09.003.

Garcia, S., Trinh, C.T., 2020. Harnessing natural modularity of metabolism with goal attainment optimization to design a modular chassis cell for production of diverse chemicals. ACS Synth. Biol. 9 (7), 1665–1681. http://dx.doi.org/10.1021/acssynbio.9b00518, arXiv:https://doi.org/10.1021/acssynbio.9b00518, PMID: 32470305.

Garcia, S., Trinh, C.T., 2021. Computational design and analysis of modular cells for large libraries of exchangeable product synthesis modules. Metab. Eng. 67, 453–463.

Gelbach, P.E., Cetin, H., Finley, S.D., 2024. Flux sampling in genome-scale metabolic modeling of microbial communities. BMC Bioinformatics 25 (1), 45. http://dx.doi.org/10.1186/s12859-024-05655-3.

Ginkgo Bioworks Holdings, Inc., 2026. Annual report on form 10-k for the fiscal year ended december 31, 2025. Filed February 26, 2026, https://www.sec.gov/Archives/edgar/data/1830214/000162828026012346/dna-20251231.htm.

Gomez, J.A., Höffner, K., Barton, P.I., 2014. DFBAlab: A fast and reliable MATLAB code for dynamic flux balance analysis. BMC Bioinformatics 15 (1), 409. http://dx.doi.org/10.1186/s12859-014-0409-8.

Gomez, J.A., Höffner, K., Barton, P.I., Henson, M.A., 2015. Metabolic modeling of synthesis gas fermentation in bubble column reactors. Biotechnol. Biofuels 8, 89. http://dx.doi.org/10.1186/s13068-015-0272-5.

Gonzalez-Hernandez, Y., Perre, P., 2024. Building blocks needed for mechanistic modeling of bioprocesses: A critical review based on protein production by CHO cells. Metab. Eng. Commun. 18, e00232. http://dx.doi.org/10.1016/j.mec.2024.00232.

Goodwin, C.M., Aimutis, W.R., Shirwaiker, R.A., 2024. A scoping review of cultivated meat techno-economic analyses to inform future research directions for scaled-up manufacturing. Nat. Food 5, 901–910. http://dx.doi.org/10.1038/s43016-024-01061-3, URL: https://www.nature.com/articles/s43016-024-01061-3.pdf.

Gopalakrishnan, S., Dash, S., Maranas, C.D., 2020. K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data. Metab. Eng. 61, 197–205.

Gopalakrishnan, S., Johnson, W., Valderrama-Gomez, M.A., Icten, E., Tat, J., Ingram, M., Fung Shek, C., Chan, P.K., Schlegel, F., Rolandi, P., Kontoravdi, C., Lewis, N.E., 2024. COSMIC-dFBA: A novel multi-scale hybrid framework for bioprocess modeling. Metab. Eng. 82, 183–192. http://dx.doi.org/10.1016/j.ymben.2024.02.012, URL: https://www.sciencedirect.com/science/article/pii/S1096717624000284.

Grzesik, P., Warth, S.C., 2021. One-time optimization of advanced t cell culture media using a machine learning pipeline. Front. Bioeng. Biotechnol. 9, 614324. http://dx.doi.org/10.3389/fbioe.2021.614324.

Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., Lee, S.Y., 2019. Current status and applications of genome-scale metabolic models. Genome Biology 20, 121. http://dx.doi.org/10.1186/s13059-019-1730-3.

Harwood, S.M., Höffner, K., Barton, P.I., 2016. Efficient solution of ordinary differential equations with a parametric lexicographic linear program embedded. Numer. Math. 133 (4), 623–653. http://dx.doi.org/10.1007/s00211-015-0760-3.

Hashizume, T., Ozawa, Y., Ying, B.-W., 2023. Employing active learning in the optimization of culture medium for mammalian cells. Npj Syst. Biol. Appl. 9, 20. http://dx.doi.org/10.1038/s41540-023-00284-7.

Van der Hauwaert, L., Regueira, A., Mauricio-Iglesias, M., 2023. Integration of metabolic models in biorefinery designs using superstructure optimisation. In: Computer Aided Chemical Engineering. vol. 52, Elsevier, pp. 2249–2254. http://dx.doi.org/10.1016/B978-0-443-15274-0.50358-9.

Van der Hauwaert, L., Regueira, A., Mauricio-Iglesias, M., 2024. Advancing biorefinery design through the integration of metabolic models. J. Clean. Prod. 465, 142793. http://dx.doi.org/10.1016/j.jclepro.2024.142793.

Heffernan, J.K., Valgepea, K., de Souza Pinto Lemgruber, R., Casini, I., Plan, M., Tappel, R., Simpson, S.D., Köpke, M., Nielsen, L.K., Marcellin, E., 2020. Enhancing CO2-valorization using Clostridium autoethanogenum for sustainable fuel and chemicals production. Front. Bioeng. Biotechnol. 8, 204. http://dx.doi.org/10.3389/fbioe.2020.00204.

Hefzi, H., Ang, K.S., Hanscho, M., et al., 2016. A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. Cell Syst. 3, 434–443. http://dx.doi.org/10.1016/j.cels.2016.10.016.

Heinken, A., Thiele, I., 2025. APOLLO: A genome-scale metabolic reconstruction resource of 247,092 diverse human microbes spanning multiple continents, age groups, and body sites. Cell Syst. 16 (2), 101196. http://dx.doi.org/10.1016/j.cels.2025.101196.

Helleckes, L.M., Hemmerich, J., Wiechert, W., von Lieres, E., Grunberger, A., 2023. Machine learning in bioprocess development: From promise to practice. Trends Biotechnol. 41 (6), 739–752. http://dx.doi.org/10.1016/j.tibtech.2022.10.010.

Höffner, K., Harwood, S.M., Barton, P.I., 2013. A reliable simulator for dynamic flux balance analysis. Biotechnol. Bioeng. 110 (3), 792–802. http://dx.doi.org/10.1002/bit.24748.

Hu, M., Suthers, P.F., Maranas, C.D., 2024. KETCHUP: Parameterizing large-scale kinetic models using multiple datasets with different reference states. Metab. Eng. 82, 123–133.

Ishchuk, O., Domenzain, I., Sánchez, B., Muñiz Paredes, F., Martínez, J., Nielsen, J., Petranovic, D., 2022. Genome-scale modeling drives 70-fold improvement of intracellular heme production in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 119, e2108245119. http://dx.doi.org/10.1073/pnas.2108245119.

Islam, M.M., Schroeder, W.L., Saha, R., 2021. Kinetic modeling of metabolism: Present and future. Curr. Opin. Syst. Biol. 26, 72–78. http://dx.doi.org/10.1016/j.coisb.2021.04.003, URL: https://www.sciencedirect.com/science/article/pii/S2452310021000147.

Jensen, K., Broeken, V., Hansen, A.S.L., Sonnenschein, N., Herrgård, M.J., 2019. OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. Metab. Eng. Commun. 8, e00087. http://dx.doi.org/10.1016/j.mec.2019.e00087.

Kadhum, H.J., Rajendran, K., Murthy, G.S., 2018. Optimization of surfactant addition in cellulosic ethanol process using integrated techno-economic and life cycle assessment for bioprocess design. ACS Sustain. Chem. Eng. 6 (11), 13687–13695. http://dx.doi.org/10.1021/acssuschemeng.8b00387.

Karp, P.D., Weaver, D., Latendresse, M., 2018. How accurate is automated gap filling of metabolic models? BMC Syst. Biol. 12, 73. http://dx.doi.org/10.1186/s12918-018-0593-7.

Kenefake, D., Armingol, E., Lewis, N.E., Pistikopoulos, E.N., 2022. An improved algorithm for flux variability analysis. BMC Bioinformatics 23, 550. http://dx.doi.org/10.1186/s12859-022-05089-9.

Kerkhoven, E.J., Pomraning, K.R., Baker, S.E., Nielsen, J., 2016. Regulation of amino-acid metabolism controls flux to lipid accumulation in Yarrowia lipolytica. Npj Syst. Biol. Appl. 2 (1), 16005. http://dx.doi.org/10.1038/npjsba.2016.5.

Khanal, S.K., Tarafdar, A., You, S., 2023. Artificial intelligence and machine learning for smart bioprocesses. Bioresour. Technol. 375, 128826. http://dx.doi.org/10.1016/j.biortech.2023.128826.

Khodayari, A., Maranas, C.D., 2016. A genome-scale escherichia coli kinetic metabolic model *k*-ecoli457 satisfying flux data for multiple mutant strains. Nat. Commun. 7 (13806), http://dx.doi.org/10.1038/ncomms13806.

Kim, J.Y., Ahn, Y.-J., Lee, J.A., Lee, S.Y., 2023. Recent advances in the production of platform chemicals using metabolically engineered microorganisms. Curr. Opin. Green Sustain. Chem. 40, 100777. http://dx.doi.org/10.1016/j.cogsc.2023.100777.

Kim, M., Sun, G., Lee, D.-Y., Kim, B.-G., 2017. Bereta: A systematic method for identifying target transcriptional regulators to enhance microbial production of chemicals. Bioinformatics 33 (1), 87–94. http://dx.doi.org/10.1093/bioinformatics/btw557, URL: https://doi.org/10.1093/bioinformatics/btw557.

Klamt, S., Krömer, J.O., Davidsen, P., Savoglidis, G., Ruiz, J., 2014. Integrative model reduction to speed up metabolic network analysis and design. Microb. Cell Factories 13, 135.

Koduru, L., Lakshmanan, M., Lee, D.-Y., 2018. In silico model-guided identification of transcriptional regulator targets for efficient strain design. Microb. Cell Factories 17, 167. http://dx.doi.org/10.1186/s12934-018-1015-7.

Kuriya, Y., Araki, M., 2020. Dynamic flux balance analysis to evaluate the strain production performance on shikimic acid production in *Escherichia coli*. Metabolites 10 (5), 198. http://dx.doi.org/10.3390/metabo10050198.

Lasry Testa, R.D., Delpino, C., Estrada, V.G., Díaz, M.S., 2022. Development of *in silico* strategies to photoautotrophically produce poly-β-hydroxybutyrate (PHB) by cyanobacteria. Algal Res. 62, 102621. http://dx.doi.org/10.1016/j.algal.2021.102621, Article 102621.

Lázaro, J., Jansen, G., Yang, Y., Torres-Acosta, M.A., Lye, G., Oliver, S.G., Júlvez, J., 2022. Combination of genome-scale models and bioreactor dynamics to optimize the production of commodity chemicals. Front. Mol. Biosci. 9, 855735. http://dx.doi.org/10.3389/fmolb.2022.855735.

Lerman, J.A., Hyduke, D.R., Latif, H., Portnoy, V.A., Lewis, N.E., Orth, J.D., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Zengler, K., Palsson, B.O., 2012. In silico method for modeling metabolism and gene product expression at genome scale. Nat. Commun. 3, 929, URL: https://www.nature.com/articles/ncomms192.pdf.

Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., Weitz, K.K., Eils, R., König, R., Smith, R.D., Palsson, B.Ø., 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6, 390. http://dx.doi.org/10.1038/msb.2010.47.

Li, F., Chen, Y., Qi, Q., Wang, Y., Yuan, L., Huang, M., Elesemani, I.E., Feizi, A., Kerkhoven, E.J., Nielsen, J., 2022a. Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints. Nat. Commun. 13, 2969. http://dx.doi.org/10.1038/s41467-022-30689-7.

Li, G., Hu, Y., Zrimec, J., Luo, H., Wang, H., Zelezniak, A., Ji, B., Nielsen, J., 2021. Bayesian genome-scale modelling identifies thermal determinants of yeast metabolism. Nat. Commun. 12, 190. http://dx.doi.org/10.1038/s41467-020-20338-2.

Li, F., Yuan, L., Lu, H., Li, G., Chen, Y., Engqvist, M.K.M., Kerkhoven, E.J., Nielsen, J., 2022b. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. Nat. Catal. 5 (8), 662–672. http://dx.doi.org/10.1038/s41929-022-00798-z.

Liebermeister, W., Uhlendorf, J., Klipp, E., 2010. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. Bioinformatics 26 (12), 1528–1534. http://dx.doi.org/10.1093/bioinformatics/btq141.

Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., et al., 2020. MEMOTE for standardized genome-scale metabolic model testing. Nature Biotechnol. 38 (3), 272–276.

Liew, F.E., Nogle, R., Abdalla, T., Rasor, B.J., Canter, C., Jensen, R.O., Wang, L., Strutz, J., Chirania, P., De Tissera, S., Mueller, A.P., Ruan, Z., Gao, A., Tran, L., Engle, N.L., Bromley, J.C., Daniell, J., Conrado, R., Tschaplinski, T.J., Giannone, R.J., Hettich, R.L., Karim, A.S., Simpson, S.D., Brown, S.D., Leang, C., Jewett, M.C., Köpke, M., 2022. Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. Nature Biotechnol. 40 (3), 335–344. http://dx.doi.org/10.1038/s41587-021-01195-w.

Liu, L., Bockmayr, A., 2020. Regulatory dynamic enzyme-cost flux balance analysis: a unifying framework for constraint-based modeling. J. Theoret. Biol. 501, 110317. http://dx.doi.org/10.1016/j.jtbi.2020.110317.

Lloyd, C.J., Ebrahim, A., Yang, L., King, Z.A., Catoiu, E., O'Brien, E.J., Liu, J.K., Palsson, B.O., 2018. COBRAme: A computational framework for genome-scale models of metabolism and gene expression. PLoS Comput. Biol. 14 (7), e1006302. http://dx.doi.org/10.1371/journal.pcbi.1006302.

Love, K.R., Martin, S.E., Morrison, D.G., Crowell, L.E., 2024. Perfusion fermentation sets a path to democratize biomanufacturing. Trends Biotechnol. 42 (12), 1819–1821. http://dx.doi.org/10.1016/j.tibtech.2024.08.003.

Lu, H., Li, F., Liang, Z., Ebrahim, A., Palsson, B.O., et al., 2019. Yeast8 bridges the gap between genotype and phenotype in Saccharomyces cerevisiae. Nat. Commun. 10, 3586. http://dx.doi.org/10.1038/s41467-019-11581-3.

Lu, H., Xiao, L., Liao, W., Yan, X., Nielsen, J., 2024. Cell factory design with advanced metabolic modelling empowered by artificial intelligence. Metab. Eng. 85, 61–72.

Luedeking, R., Piret, E.L., 1959. A kinetic study of the lactic acid fermentation: Batch process at controlled pH. J. Biochem. Microbiol. Technol. Eng. 1, 393–412.

Luo, H., Li, P., Ji, B., Nielsen, J., 2023. Modeling the metabolic dynamics at the genome scale by optimized yield analysis. Metab. Eng. 75, 119–130. http://dx.doi.org/10.1016/j.ymben.2022.12.001.

Ma, D., Yang, L., Palsson, B.O., Saunders, M.A., 2017. Reliable and efficient solution of genome-scale models of metabolism and macromolecular expression. Sci. Rep. 7, 40863. http://dx.doi.org/10.1038/srep40863.

Machado, D., Herrgård, M., 2014. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput. Biol. 10 (4), e1003580. http://dx.doi.org/10.1371/journal.pcbi.1003580.

Macklin, D.N., Ahn-Horst, T.A., Choi, H., Ruggero, N.A., Carrera, J., Mason, J.C., Sun, G., Agmon, E., DeFelice, M.M., Maayan, I., Lane, K., Spangler, R.K., Gillies, T.E., Paull, M.L., Akhter, S., Bray, S.R., Weaver, D.S., Keseler, I.M., Karp, P.D., Morrison, J.H., Covert, M.W., 2020. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. Science 369 (6502), eaav3751. http://dx.doi.org/10.1126/science.aav3751, URL: https://www.science.org/doi/abs/10.1126/science.aav3751, arXiv:https://www.science.org/doi/pdf/10.1126/science.aav3751.

Macklin, D.N., Ruggero, N.A., Covert, M.W., 2014. The future of whole-cell modeling. Curr. Opin. Biotechnol. 28, 111–115. http://dx.doi.org/10.1016/j.copbio.2014.01.012.

Mahadevan, R., Edwards, J.S., Doyle, F.J., 2002. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys. J. 83 (3), 1331–1340. http://dx.doi.org/10.1016/S0006-3495(02)73903-9.

Maria, G., 2020. Model-based optimization of a fed-batch bioreactor for mab production using a hybridoma cell culture. Molecules 25 (23), 5648. http://dx.doi.org/10.3390/molecules25235648.

Massaiu, I., Pasotti, L., Sonnenschein, N., Rama, E., Cavaletti, M., Magni, P., Calvio, C., Herrgård, M., 2019. Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains. Microb. Cell Factories 18, 3. http://dx.doi.org/10.1186/s12934-018-1052-2.

Meng, C., Zeleznik, O.A., Thallinger, G.G., Kuster, B., Gholami, A.M., Culhane, A.C., 2016. Dimension reduction techniques for the integrative analysis of multi-omics data. Brief. Bioinform. 17 (4), 628–641. http://dx.doi.org/10.1093/bib/bbv108.

Millard, P., Enjalbert, B., Uttenweiler-Joseph, S., Portais, J.-C., Létisse, F., 2021. Control and regulation of acetate overflow in *Escherichia coli*. ELife 10, e63661. http://dx.doi.org/10.7554/eLife.63661.

Monk, J.M., Lloyd, C.J., Brunk, E., et al., 2017. iML1515, a knowledgebase that quantitatively maps Escherichia coli metabolism. Nature Biotechnol. 35, 904–908. http://dx.doi.org/10.1038/nbt.3956.

Monod, J., 1949. The growth of bacterial cultures. Annu. Rev. Microbiol. 3, 371–394.

Monteiro, M., Fadda, S., Kontoravdi, C., 2023. Towards advanced bioprocess optimization: A multiscale modelling approach. Comput. Struct. Biotechnol. J. 21, 3639–3655. http://dx.doi.org/10.1016/j.csbj.2023.07.003.

Morrissey, J., Barberi, G., Strain, B., Facco, P., Kontoravdi, C., 2025. NEXT-FBA: A hybrid stoichiometric/data-driven approach to improve intracellular flux predictions. Metab. Eng. 91, 130–144. http://dx.doi.org/10.1016/j.ymben.2025.03.010.

Motamedian, E., Berzins, K., Muiznieks, R., Stalidzans, E., 2023. OptEnvelope: A target point guided method for growth-coupled production using knockouts. PLOS ONE 18 (11), e0294313. http://dx.doi.org/10.1371/journal.pone.0294313.

Narayanan, H., Hinckley, J.A., Barry, R., Dang, B., Wolffe, L.A., Atari, A., Tseng, Y., Love, J.C., 2025. Accelerating cell culture media development using Bayesian optimization-based iterative experimental design. Nat. Commun. 16, 6055. http://dx.doi.org/10.1038/s41467-025-61113-5.

Narayanan, H., von Stosch, M., Feidl, F., Sokolov, M., Morbidelli, M., Butté, A., 2023. Hybrid modeling for biopharmaceutical processes: advantages, opportunities, and implementation. Front. Chem. Eng. 5, 1157889. http://dx.doi.org/10.3389/fceng.2023.1157889.

Narayanan, B., Weilandt, D., Masid, M., Miskovic, L., Hatzimanikatis, V., 2024. Rational strain design with minimal phenotype perturbation. Nat. Commun. 15, 723. http://dx.doi.org/10.1038/s41467-024-44831-0.

Nausch, H., Baldan, M., Teichert, K., Lutz, J., Claussen, C., Bortz, M., Buyel, J.F., 2023. Simulation and optimization of nutrient uptake and biomass formation using a multi-parameter monod-type model of tobacco BY-2 cell suspension cultures in a stirred-tank bioreactor. Front. Plant Sci. 14, 1183254. http://dx.doi.org/10.3389/fpls.2023.1183254.

Nazem-Bokaee, H., Senger, R.S., 2015. ToMI-FBA: A genome-scale metabolic flux based algorithm to select optimum hosts and media formulations for expressing pathways of interest. AIMS Bioeng. 2 (4), 335–374. http://dx.doi.org/10.3934/bioeng.2015.4.335.

Nikzad-Langerodi, R., Lughofer, E., Saminger-Platz, S., Zahel, T., Sagmeister, P., Herwig, C., 2017. Automatic feed phase identification in multivariate bioprocess profiles by sequential binary classification. Anal. Chim. Acta 982, 48–61. http://dx.doi.org/10.1016/j.aca.2017.05.034.

Nogales, J., Mueller, J., Gudmundsson, S., Canalejo, F.J., Duque, E., Monk, J., Feist, A.M., Ramos, J.L., Niu, W., Palsson, B.O., 2020. High-quality genome-scale metabolic modelling of Pseudomonas putida highlights its broad metabolic capabilities. Environ. Microbiol. 22 (1), 255–269. http://dx.doi.org/10.1111/1462-2920.14843.

Noor, M.S., Ferdous, S., Salehi, R., Gates, H., Dey, S., Raghunath, V.S., Zargar, M.R., Chowdhury, R., 2025. Next-generation metabolic models informed by biomolecular simulations. Curr. Opin. Biotechnol. 92, 103259. http://dx.doi.org/10.1016/j.copbio.2024.103259.

O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B.Ø., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biology 9, 693.

Oftadeh, O., Hatzimanikatis, V., 2024. Genome-scale models of metabolism and expression predict the metabolic burden of recombinant protein expression. Metab. Eng. 84, 109–116.

Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D.C., Lewis, N.E., 2017. A systematic evaluation of methods for tailoring genome-scale metabolic models. Cell Syst. 4 (3), 318–329. http://dx.doi.org/10.1016/j.cels.2017.01.010.

Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? Nature Biotechnol. 28, 245–248. http://dx.doi.org/10.1038/nbt.1614.

Park, S.-Y., Choi, D.-H., Song, J., Lakshmanan, M., Richelle, A., Yoon, S., Kontoravdi, C., Lewis, N.E., Lee, D.-Y., 2024a. Driving towards digital biomanufacturing by CHO genome-scale models. Trends Biotechnol. 42 (9), 1192–1203. http://dx.doi.org/10.1016/j.tibtech.2024.03.001.

Park, S.Y., Park, C.H., Choi, D.-H., Hong, J.K., Lee, D.-Y., 2021. Bioprocess digital twins of mammalian cell culture for advanced biomanufacturing. Curr. Opin. Chem. Eng. 33, 100702. http://dx.doi.org/10.1016/j.coche.2021.100702.

Park, S.-Y., Song, J., Choi, D.-H., Park, U., Cho, H., Hong, B.H., Silberberg, Y.R., Lee, D.-Y., 2024b. Exploring metabolic effects of dipeptide feed media on CHO cell cultures by in silico model-guided flux analysis. Appl. Microbiol. Biotechnol. 108 (1), 123. http://dx.doi.org/10.1007/s00253-023-12997-0.

Pel, H.J., de Winde, J.H., Archer, D.B., Dyer, P.S., Hofmann, G., Schaap, P.J., Turner, G., de Vries, R.P., Albang, R., Albermann, K., Andersen, M.R., Bendtsen, J.D., Benen, J.A.E., van den Berg, M., Breestraat, S., Caddick, M.X., Contreras, R., Cornell, M., Coutinho, P.M., Danchin, E.G.J., Debets, A.J.M., Dekker, P., van Dijck, P.W.M., van Dijk, A., Dijkhuizen, L., Driessen, A.J.M., d'Enfert, C., Geysens, S., Goosen, C., Groot, G.S.P., de Groot, P.W.J., Guillemette, T., Henrissat, B., Herweijer, M., van den Hombergh, J.P.T.W., van den Hondel, C.A.M.J.J., van der Heijden, R.T.J.M., van der Kaaij, R.M., Klis, F.M., Kools, H.J., Kubicek, C.P., van Kuyk, P.A., Lauber, J., Lu, X., van der Maarel, M.J.E.C., Meulenberg, R., Menke, H., Mortimer, M.A., Nielsen, J., Oliver, S.G., Olsthoorn, M., Pal, K., van Peij, N.N.M.E., Ram, A.F.J., Rinas, U., Roubos, J.A., Sagt, C.M.J., Schmoll, M., Sun, J., Ussery, D., Varga, J., Vervecken, W., van de Vondervoort, P.J.J., Wedler, H., Wösten, H.A.B., Zeng, A.-P., van Ooyen, A.J.J., Visser, J., Stam, H., 2007. Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88. Nature Biotechnol. 25 (2), 221–231. http://dx.doi.org/10.1038/nbt1282.

Pennington, O., Espinel Ríos, S., Torres Sebastian, M., Dickson, A., Zhang, D., 2024. A multiscale hybrid modelling methodology for cell cultures enabled by enzyme-constrained dynamic metabolic flux analysis under uncertainty. Metab. Eng. 86, 274–287. http://dx.doi.org/10.1016/j.ymben.2024.10.013.

Pérez-Fernández, B.A., Fernandez-de Cossio-Diaz, J., Boggiano, T., León, K., Mulet, R., 2021. In-silico media optimization for continuous cultures using genome scale metabolic networks: The case of CHO-K1. Biotechnol. Bioeng. 118 (5), 1884–1897. http://dx.doi.org/10.1002/bit.27704.

Pinto, J., Ramos, J.R.C., Costa, R.S., Rossell, S., Dumas, P., Oliveira, R., 2023. Hybrid deep modeling of a CHO-K1 fed-batch process: Combining first-principles with deep neural networks. Front. Bioeng. Biotechnol. 11, 1237963. http://dx.doi.org/10.3389/fbioe.2023.1237963.

Poddar, T.K., Scown, C.D., 2025. Technoeconomic analysis for near-term scale-up of bioprocesses. Curr. Opin. Biotechnol. 92, 103258. http://dx.doi.org/10.1016/j.copbio.2025.103258.

Radivojević, T., Costello, Z., Workman, K., Garcia Martin, H., 2020. A machine learning automated recommendation tool for synthetic biology. Nat. Commun. 11, 4879. http://dx.doi.org/10.1038/s41467-020-18008-4.

Rathore, A.S., Shekhawat, L.K., Loomba, V., 2016. Computational fluid dynamics for bioreactor design. In: Bioreactors: Design, Operation and Novel Applications. Springer, pp. 295–322. http://dx.doi.org/10.1007/978-3-319-26769-3_12.

Rau, M.H., Gaspar, P., Jensen, M.L., Geppel, A., Neves, A.R., Zeidan, A.A., 2022. Genome-scale metabolic modeling combined with transcriptome profiling provides mechanistic understanding of Streptococcus thermophilus CH8 metabolism. Appl. Environ. Microbiol. 88 (16), e00780–22. http://dx.doi.org/10.1128/aem.00780-22.

Reddy, J.V., Singh, S.K., Leibiger, T., Lee, K.H., Ierapetritou, M., Papoutsakis, E.T., 2025. Flux balance analysis and peptide mapping elucidate the impact of bioreactor pH on Chinese hamster ovary (CHO) cell metabolism and N-linked glycosylation in the Fab and Fc regions of the produced IgG. Metab. Eng. 87, 37–48. http://dx.doi.org/10.1016/j.ymben.2024.11.005.

Robitaille, J., Chen, J., Jolicoeur, M., 2015. A single dynamic metabolic model can describe mab-producing CHO cell batch and fed-batch cultures on different culture media. PLoS One 10 (9), e0136815. http://dx.doi.org/10.1371/journal.pone.0136815.

van Rosmalen, R.P., Moreno-Paz, S., Duman-Özdamar, Z.E., Suárez-Díez, M., 2024. CFSA: Comparative flux sampling analysis as a guide for strain design. Metab. Eng. Commun. 19, e00244. http://dx.doi.org/10.1016/j.mec.2024.e00244.

Saber, W.E.I.A., Ghoniem, A.A., Al-Otibi, F.O., El-Hersh, M.S., Eldadamony, N.M., Menaa, F., Elattar, K.M., 2023. A comparative study using response surface methodology and artificial neural network towards optimized production of melanin by *Aureobasidium pullulans* AKW. Sci. Rep. 13, 13545. http://dx.doi.org/10.1038/s41598-023-40549-z.

Sabzevari, M., Szedmak, S., Penttilä, M., Jouhten, P., Rousu, J., 2022. Strain design optimization using reinforcement learning. PLoS Comput. Biol. 18 (6), e1010177. http://dx.doi.org/10.1371/journal.pcbi.1010177.

Sahu, A., Blatke, M.-A., Szymanski, J.J., Topfer, N., 2021. Advances in flux balance analysis by integrating machine learning and mechanism-based models. Comput. Struct. Biotechnol. J. 19, 4626–4640. http://dx.doi.org/10.1016/j.csbj.2021.08.004, URL: https://www.sciencedirect.com/science/article/pii/S2001037021003354.

Salehi, M., Farhadi, S., Moieni, A., Safaie, N., Hesami, M., 2021. A hybrid model based on general regression neural network and fruit fly optimization algorithm for forecasting and optimizing paclitaxel biosynthesis in *Corylus avellana* cell culture. Plant Methods 17, 13. http://dx.doi.org/10.1186/s13007-021-00714-9.

Salvy, P., Hatzimanikatis, V., 2020. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. Nat. Commun. 11 (1), 30. http://dx.doi.org/10.1038/s41467-019-13818-7.

Samadhiya, K., Sangtani, R., Nogueira, R., Bala, K., 2022. Insightful advancement and opportunities for microbial bioplastic production. Front. Microbiol. 12, 674864.

Sánchez, B.J., Zhang, C.C., Nilsson, A., et al., 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. 13, 935. http://dx.doi.org/10.15252/msb.20167411.

Scheibe, T.D., Fang, Y., Murray, A.R., Brooks, A.E., Roden, E.E., Zachara, J.M., 2009. Coupling a genome-scale metabolic model with a reactive transport model to describe in situ uranium bioremediation. Microb. Biotechnol. 2 (2), 274–286. http://dx.doi.org/10.1111/j.1751-7915.2009.00087.x.

Schilling, C.H., Letscher, D., Palsson, B.Ø., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theoret. Biol. 203, 229–248.

Schmacht, M., Lorenz, E., Stahl, U., Senz, M., 2017. Medium optimization based on yeast's elemental composition for glutathione production in *Saccharomyces cerevisiae*. J. Biosci. Bioeng. 123 (5), 555–561. http://dx.doi.org/10.1016/j.jbiosc.2016.12.011.

Schmidt, L., Hair, K., Graziosi, S., Campbell, F., Kapp, C., Khanteymoori, A., Craig, D., Engelbert, M., Thomas, J., 2024. Exploring the use of a large language model for data extraction in systematic reviews: a rapid feasibility study. arXiv:2405.14445.

Schwentner, A., Feith, A., Münch, E., Stiefelmaier, J., Lauer, I., Favilli, L., Massner, C., Öhrlein, J., Grund, B., Hüser, A., Takors, R., Blombach, B., 2019. Modular systems metabolic engineering enables balancing of relevant pathways for L-histidine production with Corynebacterium glutamicum. Biotechnol. Biofuels 12, 65. http://dx.doi.org/10.1186/s13068-019-1410-2.

Scown, C.D., Baral, N.R., Yang, M., Vora, N., Huntington, T., 2021. Technoeconomic analysis for biofuels and bioproducts. Curr. Opin. Biotechnol. 67, 58–64. http://dx.doi.org/10.1016/j.copbio.2021.01.002.

Shen, F., Sun, R., Yao, J., Li, J., Liu, Q., Price, N.D., Liu, C., Wang, Z., 2019. OptRAM: In-silico strain design via integrative regulatory–metabolic network modeling. PLoS Comput. Biol. 15 (3), e1006835. http://dx.doi.org/10.1371/journal.pcbi.1006835.

Siebler, F., Lapin, A., Takors, R., 2020. Synergistically applying one-dimensional modeling and CFD for designing industrial-scale bubble column syngas bioreactors. Eng. Life Sci. 20, e1900132. http://dx.doi.org/10.1002/elsc.1900132.

Singh, V., Haque, S., Niwas, R., Srivastava, A., Pasupuleti, M., Tripathi, C.K.M., 2017. Strategies for fermentation medium optimization: An in-depth review. Front. Microbiol. 7, 2087. http://dx.doi.org/10.3389/fmicb.2016.02087.

Singh, V.K., Jiménez del Val, I., Glassey, J., Kavousi, F., 2024. Integration approaches to model bioreactor hydrodynamics and cellular kinetics for advancing bioprocess optimization. Bioengineering 11 (6), 546. http://dx.doi.org/10.3390/bioengineering11060546.

Sokolov, M., von Stosch, M., Narayanan, H., Feidl, F., Butté, A., 2021. Hybrid modeling—A key enabler towards realizing digital twins in biopharma? Curr. Opin. Chem. Eng. 34, 100715. http://dx.doi.org/10.1016/j.coche.2021.100715.

Song, H.-S., Ahamed, F., Lee, J.-Y.e., 2025. Coupling flux balance analysis with reactive transport modelling through machine learning for rapid and stable simulation of microbial metabolic switching. Sci. Rep. 15, 6042. http://dx.doi.org/10.1038/s41598-025-89997-9.

Sriphochanart, W., Skolpap, W., 2022. Temperature shift and feeding strategies for improving L-lactic acid production by Lactiplantibacillus plantarum in batch and fed-batch cultures. Process Biochem. 113, 11–21. http://dx.doi.org/10.1016/j.procbio.2021.12.006.

Stephanopoulos, G., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering: Principles and Methodologies. Academic Press.

Swayambhu, G., Moscatello, N., Atilla-Gokcumen, G.E., Pfeifer, B.A., 2020. Flux balance analysis for media optimization and genetic targets to improve heterologous siderophore production. IScience 23, 101016. http://dx.doi.org/10.1016/j.isci.2020.101016.

Terzer, M., Stelling, J., 2008. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24 (19), 2229–2235. http://dx.doi.org/10.1093/bioinformatics/btn401.

Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5 (1), 93–121. http://dx.doi.org/10.1038/nprot.2009.203.

Thornburg, Z.R., Bianchi, D.M., Brier, T.A., Gilbert, B.R., Earnest, T.M., Melo, M.C.R., Safronova, N., Sáenz, J.P., Cook, A.T., Wise, K.S., Hutchison, I., Smith, H.O., Glass, J.I., Luthey-Schulten, Z., 2022. Fundamental behaviors emerge from simulations of a living minimal cell. Cell 185 (2), 345–360.e28. http://dx.doi.org/10.1016/j.cell.2021.12.025.

Tokić, M., Hatzimanikatis, V., Misković, L., 2020. Large-scale kinetic metabolic models of Pseudomonas putida KT2440 for consistent design of metabolic engineering strategies. Biotechnol. Biofuels 13, 33. http://dx.doi.org/10.1186/s13068-020-01667-1.

Tomàs-Gamisans, M., Ferrer, P., Albiol, J., 2016. Integration and validation of the genome-scale metabolic models of Pichia pastoris: a comprehensive update of protein glycosylation pathways, lipid and energy metabolism. PLoS One 11 (1), e0148031. http://dx.doi.org/10.1371/journal.pone.0148031.

Trinh, C.T., Wlaschin, C.P., Srienc, F., 2009. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl. Microbiol. Biotechnol. 81, 813–826, URL: https://link.springer.com/article/10.1007/s00253-008-1770-8.

Turanli, B., Gulfidan, G., Onluturk Aydogan, O., Kula, C., Selvaraj, G., Arga, K.Y., 2024. Genome-scale metabolic models in translational medicine: the current status and potential of machine learning in improving the effectiveness of the models. Mol. Omics 20, 560–576. http://dx.doi.org/10.1039/D3MO00152K.

Varghese, S., Dhanraj, N., Rebello, S., Sindhu, R., Binod, P., Pandey, A., Jisha, M., Awasthi, M.K., 2022. Leads and hurdles to sustainable microbial bioplastic production. Chemosphere 305, 135390.

Vella, A., Blincoe, K., 2026. The impact of AI coding assistants on software engineering: A longitudinal study. arXiv:2605.23135.

Vo, D.-K., Trinh, K.T.L., 2025. Molecular farming for immunization: Current advances and future prospects in plant-produced vaccines. Vaccines 13 (2), 191. http://dx.doi.org/10.3390/vaccines13020191, URL: https://www.mdpi.com/2076-393X/13/2/191.

Vongsangnak, W., Olsen, P., Hansen, K., Krogsgaard, S., Nielsen, J., 2008. Improved annotation through genome-scale metabolic modeling of Aspergillus oryzae. BMC Genomics 9, 245. http://dx.doi.org/10.1186/1471-2164-9-245.

Walker, C., Ryu, S., Garcia, S., Dooley, D., Mendoza, B., Trinh, C.T., 2022. Gene coexpression connectivity predicts gene targets underlying high ionic-liquid tolerance in Yarrowia lipolytica. Msystems 7 (4), e00348–22.

Wang, G., Li, Q., Zhang, Z., Yin, X., Wang, B., Yang, X., 2022. Recent progress in adaptive laboratory evolution of industrial microorganisms. J. Ind. Microbiol. Biotechnol. 50 (1), kuac023. http://dx.doi.org/10.1093/jimb/kuac023.

Wang, Y., Pasquini, M., Chotteau, V., Hjalmarsson, H., Jacobsen, E.W., 2024. Iterative learning robust optimization with application to medium optimization of CHO cell cultivation in continuous monoclonal antibody production. J. Process Control 137, 103196. http://dx.doi.org/10.1016/j.jprocont.2023.11.005.

Wang, Z., Wang, C., Chen, G., 2022. Kinetic modeling: A tool for temperature shift and feeding optimization in cell culture process development. Protein Express. Purif. 198, 106130. http://dx.doi.org/10.1016/j.pep.2022.106130.

Wang, Z., Zeng, D., Zhu, Y., Zhou, M., Kondo, A., Hasunuma, T., Zhao, X., 2025. Fermentation design and process optimization strategy based on machine learning. BioDesign Res. 7 (1), 25–34, URL: https://www.sciencedirect.com/science/article/pii/S2693125725000032.

Watanabe, K., Chiou, T.-Y., Konishi, M., 2024. Optimization of medium components for protein production by *Escherichia coli* with a high-throughput pipeline that uses a deep neural network. J. Biosci. Bioeng. 137 (4), 304–312. http://dx.doi.org/10.1016/j.jbiosc.2024.01.005.

Weston, B.R., Thiele, I., 2023. A nutrition algorithm to optimize feed and medium composition using genome-scale metabolic models. Metab. Eng. 76, 167–178. http://dx.doi.org/10.1016/j.ymben.2023.01.010.

Wowra, K., Hegel, E., Scharf, A., Grünberger, A., Rosenthal, K., 2023. Estimating environmental impacts of early-stage bioprocesses. Trends Biotechnol. 41 (9), 1199–1212.

Wu, P., Yuan, Q., Cheng, T., Han, Y., Zhao, W., Liao, X., Wang, L., Cai, J., He, Q., Guo, Y., et al., 2023. Genome sequencing and metabolic network reconstruction of a novel sulfur-oxidizing bacterium acidithiobacillus ameehan. Front. Microbiol. 14, 1277847.

Xu, F., Medappa, P.K., Tunc, M.M., Vroegindeweij, M., Fransoo, J.C., 2025. AI-assisted programming decreases the productivity of experienced developers by increasing the technical debt and maintenance burden. arXiv:2510.10165.

Xu, P., Ranganathan, S., Fowler, Z.L., Maranas, C.D., Koffas, M.A.G., 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. 13 (5), 578–587. http://dx.doi.org/10.1016/j.ymben.2011.06.008.

Yang, L., Ebrahim, A., Lloyd, C.J., Saunders, M.A., Palsson, B.Ø., 2019. DynamicME: dynamic simulation and refinement of integrated models of metabolism and protein expression. BMC Syst. Biol. 13, 2. http://dx.doi.org/10.1186/s12918-018-0675-6.

Yang, L., Ma, D., Ebrahim, A., Lloyd, C.J., Saunders, M.A., Palsson, B.O., 2016. solveME: fast and reliable solution of nonlinear ME models. BMC Bioinformatics 17, 391. http://dx.doi.org/10.1186/s12859-016-1240-1, URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1240-1.

Yatipanthalawa, B.S., Gras, S.L., 2024. Predictive models for upstream mammalian cell culture development: A review. Digit. Chem. Eng. 10, 100137. http://dx.doi.org/10.1016/j.dche.2023.100137.

Ye, C., Luo, Q., Guo, L., Gao, C., Xu, N., Zhang, L., Liu, L., Chen, X., 2020. Improving lysine production through construction of an *Escherichia coli* enzyme-constrained model. Biotechnol. Bioeng. 117 (11), 3533–3544. http://dx.doi.org/10.1002/bit.27485.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B., Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Van Dien, S., 2011. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat. Chem. Biol. 7 (7), 445–452. http://dx.doi.org/10.1038/nchembio.580.

Yoshida, K., Watanabe, K., Chiou, T.-Y., Konishi, M., 2023. High throughput optimization of medium composition for *Escherichia coli* protein expression using deep learning and Bayesian optimization. J. Biosci. Bioeng. 135 (2), 127–133. http://dx.doi.org/10.1016/j.jbiosc.2022.12.004.

Yousefi-Darani, A., Paquet-Durand, O., Hitzmann, B., 2021. The Kalman filter for the supervision of cultivation processes. In: Herwig, C., Pörtner, R., Möller, J. (Eds.), Digital Twins: Applications To the Design and Optimization of Bioprocesses. Springer International Publishing, Cham, pp. 95–125. http://dx.doi.org/10.1007/10_2020_145.

Zhai, J., Qi, X., Cai, L., Liu, Y., Tang, H., Xie, L., Wang, J., 2025. Nnkcat: deep neural network to predict catalytic constants (kcat) by integrating protein sequence and substrate structure with enhanced data imbalance handling. Brief. Bioinform. 26 (3), bbaf212. http://dx.doi.org/10.1093/bib/bbaf212.

Zhang, S., Aida, H., Ying, B.-W., 2023. Employing active learning in medium optimization for selective bacterial culture. Appl. Microbiol. 3 (4), 1355–1369. http://dx.doi.org/10.3390/applmicrobiol3040091.

Zhang, Y., Cai, J., Shang, X., Wang, B., Liu, S., Chai, X., Tan, T., Zhang, Y., Wen, T., 2017. A new genome-scale metabolic model of Corynebacterium glutamicum and its application. Biotechnol. Biofuels 10, 169. http://dx.doi.org/10.1186/s13068-017-0856-3.

Zournas, A., Incha, M.R., Radivojevic, T., Blay, V., Martí, J.M., Costello, Z., Thompson, M.G., Hillson, N.J., Keasling, J.D., García Martin, H., 2025. Machine learning-led semi-automated medium optimization reveals salt as key for flaviolin production in *Pseudomonas putida*. Commun. Biol. 8, 630. http://dx.doi.org/10.1038/s42003-025-08039-2.