MDPI

Review

Genome-Scale Reconstruction of Microbial Dynamic Phenotype: Successes and Challenges

Nicolai S. Panikov [ID]

Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Ave., Boston, MA 02115, USA; n.panikov@northeastern.edu

Abstract: This review is a part of the SI 'Genome-Scale Modeling of Microorganisms in the Real World'. The goal of GEM is the accurate prediction of the phenotype from its respective genotype under specified environmental conditions. This review focuses on the dynamic phenotype; prediction of the real-life behaviors of microorganisms, such as cell proliferation, dormancy, and mortality; balanced and unbalanced growth; steady-state and transient processes; primary and secondary metabolism; stress responses; etc. Constraint-based metabolic reconstructions were successfully started two decades ago as FBA, followed by more advanced models, but this review starts from the earlier nongenomic predecessors to show that some GEMs inherited the outdated biokinetic frameworks compromising their performances. The most essential deficiencies are: (i) an inadequate account of environmental conditions, such as various degrees of nutrients limitation and other factors shaping phenotypes; (ii) a failure to simulate the adaptive changes of MMCC (MacroMolecular Cell Composition) in response to the fluctuating environment; (iii) the misinterpretation of the SGR (Specific Growth Rate) as either a fixed constant parameter of the model or independent factor affecting the conditional expression of macromolecules; (iv) neglecting stress resistance as an important objective function; and (v) inefficient experimental verification of GEM against simple growth (constant MMCC and SGR) data. Finally, we propose several ways to improve GEMs, such as replacing the outdated Monod equation with the SCM (Synthetic Chemostat Model) that establishes the quantitative relationships between primary and secondary metabolism, growth rate and stress resistance, process kinetics, and cell composition.

Keywords: growth kinetics; survival; death; substrate limitation; starvation; gene expression; conditional expression of macromolecules; protein allocation; batch culture; chemostat; cell cycle; metabolic network; pool; metabolic intermediates; cell composition; biomass brutto-formulae; kinetic order

1. Introduction

Everything should be made as simple as possible... but not simpler.

Albert Einstein

Presently, the NCBI lists more than 300,000 completed whole-genome projects. With ~7500 sequenced fungal species and more than 31,000 prokaryotic (archaeal and bacterial) OTUs, we can state that whole-genome sequence data are now available for nearly all cultivable microorganisms of medical, industrial, or environmental significance. Whole-genome projects are expected to increase over time [1] to further expand their coverage and as strain-specific resequencing, metagenomic surveys of communities, etc. Yet, let us face it, the main bottleneck now is not the number of sequenced genomes but, rather, our limited capacity to capitalize on the already accumulated genomic information [2]. Consequently, modern bioscience is confronted with the challenge to translate sequence data into solid biological knowledge, better industrial and healthcare solutions, and an improved mechanistic understanding of bioprocesses. This challenge can be met by system biology [3], using as research tools genome-scale metabolic reconstructions and genome-scale models (GEMs). The ambitious goal of GEM is the accurate prediction of the phenotype from the

respective genotype under specified environmental conditions. This undoubtedly would be a remarkable scientific breakthrough, bringing formerly descriptive biology to the top of natural sciences.

The most popular version of GEM, called FBA (Flux Balance Analyses), has been already established as a powerful tool in metabolic engineering and diverse applications [1,4–12]. It uses steady-state approximation and other simplifying assumptions that greatly improve the speed of computations with inevitable sacrifices, one of them being the static way of presenting metabolic reconstructions that reproduce a screenshot of cellular metabolism beyond the time axes. More advanced contemporary GEMs, such as dFBA, ME, and WC models, have been charged with the goal of the dynamic reconstruction of metabolism based on the simulation of natural regulatory mechanisms. The correct understanding of these mechanisms should allow the adequate reconstruction of biodynamic behaviors of microorganisms, from trivial binary division to a real-life microbial dynamic under specified environmental conditions: the balanced and unbalanced growth, steady-state and transient processes, survival and recovery from stresses, cell differentiation, biosynthesis of products, including secondary metabolites, etc. Taken together, the listed processes represent a dynamic phenotype [13] of microorganisms. However, modern dynamic GEMs are applied nearly exclusively to trivial data, such as exponential growth with constant SGR. It is well-justified at the initial stages of the methodology development; there is no doubt that new computational approaches should be tested against as simple of data as possible.

Today, genomic-scale system biology is mature enough to address the famous A. Einstein maxim (see epigraph) that gives sagacious advice to carefully balance between the simple and oversimplified. Too simple of data can hardly inspire generation of strong hypotheses crucial for development of systems biology, restrain practical GEM applications, and generally discourage research progress. This review illustrates typical nontrivial microbial behaviors that are commonly observed in laboratory experiments, industrial bioreactors, and in nature. The aim of this review is to show that complex biodynamics is not a confusing mixture of chaotic events but, rather, a highly reproducible phenomena that can be accurately recorded and interpreted using mechanistic mathematical models.

The review starts by defining the key terms and basic principles of growth theory to explain the distinction between simple and complex biodynamics of microorganisms. Next, we review the nongenomic predecessors of the currently available GEMs to show that some of them are built upon the outdated frameworks compromising their performances. Luckily, such deficiencies can be fixed in a straightforward way. Finally, we discuss the prospects for improvements of GEMs for the purpose of dynamic simulations. The technical details, including mathematical equations, their interpretation, alternative solutions, etc., the interested reader can find in the Appendix B.

2. Simple and Complex Growth: The Basic Principles

Here, we are going to justify the following basic principles used throughout the review:

1. The microbial specific growth rate (SGR) depends on environmental conditions and the macromolecular cell composition (MMCC). MMCC variation is the result of self-regulation, including differential gene expression in response to a changeable environment.
2. Variable MMCC is the main reason for the higher complexity of microbial growth kinetics vs. chemical and enzymatic reactions.
3. A nutrient supply is the primary environmental factor controlling microbial behavior; the effects of other modifying factors (T, pH, osmolarity, etc.) can be understood only in combination with the nutritional factor.
4. MMCC dependence on SGR is a common misconception; both complex variables depend on the environmental conditions.
5. A growth is called *simple* (syn: balanced and steady state) if it proceeds exponentially under steady environmental conditions with constant SGR and MMCC. Otherwise, the cell dynamics are *complex*.

2.1. MMCC and Growth Kinetics

Microbial growth under various tested conditions is usually recorded as a time series of cell mass concentration (x), nutrient substrates (s), and products (p). A specific growth rate (SGR) found by the numeric differentiation of growth curves (corrected for elimination in open systems, e.g., the dilution rate in a chemostat) is the most inclusive biokinetic variable. There are well-defined mass–balance relationships between SGR and other growth characteristics, e.g., the apparent yield, maintenance, respiration, substrate uptake, and products formation [14,15]. The listed biokinetic terms describe growth in a reductionistic 'chemical' way, but there are also remarkable changes in the quality of cells, e.g., size and shape, activities, stress resistance, productivity, etc., that, in the past, were combined under the vague term physiological state. Presently, the state of growing cells is surveyed by using omics molecular inventories on top of conventional chemical analyses of cells (elemental composition, total proteins, nucleic acids, lipids, etc.). About 10% of the cell dry mass is accounted for as a pool of small molecules using metabolomic analyses, and ~90% is represented by macromolecules evaluated by proteomic, glycomic, and lipidomic profiling. This review focuses on the macromolecular cell composition (MMCC) as a principal effector of growth kinetics; the metabolic pool is touched only from a biokinetic perspective (Appendix B.2.3).

Historically, the first growth-associated changes in MMCC were established ~60 years ago for the most abundant cellular constituent, the ribosomes making up to 50% of the total cell mass in fast-growing cells and dropping down to 5–10% under growth deceleration [16,17]. At the same time, the opposite trend was found for the storage polymers (glycogen in *E. coli*), which progressively decreased with the growth acceleration from 23 to 2% [18]. Presently, the changes in MMCC have been mostly followed by transcriptomic and proteomic techniques being referred to as 'condition-specific expression' [19–21]. A commonly accepted view is that the up- and downregulation of proteins is one of the most essential and fundamental qualities of life, the result of metabolic regulation and differential gene expression in response to environmental stimuli. However, this quality turned out to be difficult to incorporate into mechanistic growth models (see below).

2.2. Limited Applicability of Enzyme Kinetics to Microbial Growth

Enzymes are simpler than cells, and enzyme kinetics has a longer history of successful developments. It was the main reason for numerous attempts to adapt enzymological models to microbial growth. The obvious distinction between enzymes and growing microorganisms is that enzymes do not proliferate over time. However, there is a simple way to handle this complication by expressing microbially driven reaction rates per g cell biomass; then, the so-called specific rates (q_s and q_p) of cellular growth kinetics can be used as equivalents of the respective enzymatic rates, v [14]. For example, the Michaelis–Menten equation can be applied to the process of nutrient uptake by cells:

$$v = -\frac{ds}{dt} = V_{max}\frac{s}{K_s + s}; \ V_{max} = k_{cat}[E] \tag{1}$$

where $[E]$ is the concentration of the transmembrane transporter(s) responsible for uptake.

Dividing both parts of Equation (1) by cell mass x, we get the expression for SUR:

$$q_s = -\frac{1}{x}\frac{ds}{dt} = Q_s\frac{s}{K_s + s}; Q_s = \frac{V_{max}}{x} \tag{1a}$$

The mass–balance relationship between the consumed substrate $-ds$ and the produced biomass dx is set up by the yield, Y:

$$Y = -\frac{dx}{ds} = -\frac{1}{x}\frac{dx}{dt} : \frac{1}{x}\frac{ds}{dt} = \frac{\mu}{q_s} \therefore q_s = \frac{\mu}{Y} \tag{2}$$

The substitution of q_s in Equation (1a) for μ produces the well-known Monod equation:

$$\mu = Yq_s = \mu_m \frac{s}{K_s + s}; \; \mu_m = YQ_s \tag{3}$$

The Michaelis–Menten and Monod equations look similar, but there is a substantial difference. The enzymatic Equation (1) was deduced using the law of mass action and assuming the reversible formation of the enzyme–substrate complex as a unique mechanistic feature of enzymatic catalysis. It fits most experimental data, and the constants V_{max}, k_{cat}, and K_m can be found for individual enzymes from UniProt, BRENDA, or other reference sources. What about the Monod equation? Unfortunately, the parameters μ_m, Y and K_s cannot be found in the NCBI taxonomy browser or any other reference source as trustworthy reproducible constants. There are many reasons for that but from a kinetic standpoint, the most essential flaw of the Monod equation is that it tacitly assumes the constancy of enzyme concentration, the variable $[E]$ in Equation (1). This assumption is acceptable for a short-term response kinetic data obtained with either pure enzymes or intact cells but wrong as applied to microbial growth that is usually accompanied by significant MMCC changes. Figure 1 illustrates the phenomenon: the instant response of glucose-limited chemostat culture to a series of glucose pulses does fit Equation (1a) at each dilution rates D (Figure 1, Left). However, two parameters, Q_s and K_s were never remaining the same displaying a general trend to increase with D (Figure 1, Right). In Appendix B.3.2, we reproduced this pattern by simulating differential expression of low (L) and high (H) affinity glucose transporters. Every D-shift changes the ambient glucose concentration inducing the MMCC self-adjustment resulted in the optimized L:H ratio.
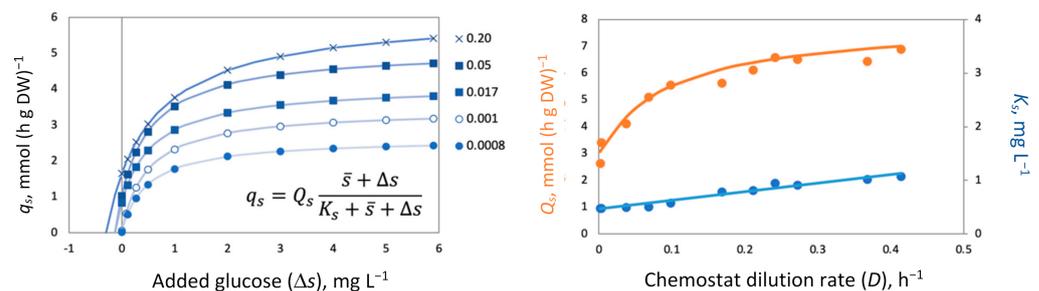

Figure 1. Failure of the Monod model to describe growth with variable MMCC. Yeasts *Schwanniomyces vanrijiae* were grown in a glucose-limited chemostat. **Left:** Short-term response of cells to glucose spike, Δs. The curves are fit to the modified Michaelis–Menten equation, where $\bar{s}$ is the steady-state glucose concentration in the chemostat. The legend shows the D and h^{-1}. **Right:** The pooled data on K_s and Q_s. Reprinted with permission from [22]. Copyright (1995) Springer Netherlands.

Other equations of enzyme kinetics (reviewed in [15]) are also expected to display a reasonable agreement with the short-term data on microbial populations but fail badly over longer periods, when microbial growth is accompanied with MMCC changes. Thus, it is the MMCC self-regulation that makes cell growth kinetics essentially more complex as compared with enzyme reactions.

2.3. Growth Conditions, the Hierarchy of Factors

The factors affecting growing cells can be physical (pressure, temperature, viscosity, stirring intensity, and mass transfer); chemical (impact of various chemical compounds); and biological (cell-to-cell interactions, differentiation into active and dormant, biofilm and planktonic forms, etc.). The biological factors are excluded from further discussion since they are parts of the dynamic phenotype to be the output rather than the input data in any growth model. The physical and chemical factors can be further split into (1) external independent variables, such as concentrations of nutrients or drugs, temperature, stirring rate, preset pH, etc., and (2) internal dependent variables, the environmental changes produced by growing cells, e.g., autoinhibitory products, metabolic acidification

or alkalinization, production of ROS or RNS, biogenic heat, etc. An account of the first group of factors as the instant microbial response is straightforward [14,15,23–25]. The long-term response simulation requires an explicit account of the variable MMCC (see below 'Synthetic chemostat model' and Appendix B.3.3). Modeling of the growth side effects is more challenging but still a doable task, for example, metabolic acidification can be simulated by coupling the ODEs for x and s with the uptake and release of the major ionized compounds (NH_4^+, HCO_3^-, and organic acids) affecting SID, the strong ion difference [26].

Most factors play a modifying role, i.e., they can accelerate or slow growth but never can start it without one essential primary factor that is the nutrient supply. The nutritional factor is multidimensional; it includes a spectrum of consumable substrates; the identity and concentration of the limiting substrate; and the regime of supply: single-term or continuous; the continuous regime can be steady or periodic (diurnal, sinusoidal, square wave); or randomly fluctuating. Sometimes, all factors are viewed as equally important, depending on local circumstances, e.g., the temperature regime is supposed to be a key factor in cold environments like polar sea and the tundra, water deficiency in deserts, nutrient supply in oligotrophic lakes, etc. [27]. However, the nutritional factor stands alone as the only one that plays the role of a consumable resource rather than the state of the environment (T, pH, red-ox, conductivity, etc.) associated with the other factors. Another unique quality of the nutrients especially important for system biology and GEMs is that nutrients participate in growth stoichiometry and are subjected to the mass conservation law. First, the consumed nutrients are reactants in those primary metabolic reactions that initiate the entire metabolic network, eventually producing new cell masses and products. Any quantitative (concentrations) or qualitative changes in nutrients (medium composition) affect the status of M- and E-matrixes in GEMs. Second, a consumable resource is always limited (actually or potentially) to be shared between various competitors: coexisting species in communities; subpopulations in pure culture (e.g., planktonic vs. attached, mutant vs. WT, prototroph vs. auxotroph, etc.); or alternative metabolic pathways at network branching points. The most delicate goal of any GEM is to mimic the natural regulatory mechanisms to recognize the optimal alternative metabolic pathway. Thus, the nutrient factor is more elaborate and more challenging in mathematical simulations, eventually giving a precious reward of more insightful models improving the understanding of dynamic systems. Third, the temperature and other modifying environmental factors are important; however, their exploration can be meaningful only in combination with the nutritional factors. The reason is that microbial stress responses to nonoptimal physicochemical factors such as heat shock, freezing, acidity, etc. include the expression of chaperones, porins, and other molecular structures minimizing the cellular damage. As any other biosynthetic reactions, antistress metabolic responses require a supply of energy and anabolic substrates; therefore, they depend on nutrients. Interestingly, the maximum of the stress response is found not with a plentiful supply of nutrients but under substrate limitations [28]; the explanation is presented below (see the section on SCM (Section 4.2)). The opposite is not true; the effect of nutrients can be fruitfully explored separately from the other factors, providing they remain constant.

2.4. SGR and MMCC, the Chicken and the Egg Dilemma

A growing number of studies now deal with the MMCC changes associated with variations of SGR. The results are published under titles like 'Effect of growth rate on expression...' or 'Growth rate dependent expression...' [29–35]. Taking it literally, the time derivative of a cell mass concentration, dx/dt, plays a rather unusual role of a factor like pH, temperature, drug dose, and other independent variables typically characterizing external to cell environmental conditions. Let us see where the problem lies using as representative example the rRNA, one of the best-studied and the most abundant conditionally expressed macromolecule.

The anatomy of experimental approaches. There are two ways of experimental manipulation SGR: (i) a chemostat culture at a series of dilution rates attaining at a steady state the required condition, $D = \mu$, and (ii) a batch cultivation using a series of C-sources supporting different SGR. The first technically more challenging approach has the advantage of keeping the same medium composition and a wider range of the tested SGR (typically, 0.05–0.95 μ_m). The simpler batch method provides just several discrete SGR levels, e.g., in the range 0.4–1.8 h^{-1} for *E. coli* [36]. The essential difference is that microbial batch growth always remains substrate-sufficient ($s \gg K_s$), while the chemostat reproduces various degrees of substrate limitations. Therefore, cells grown batchwise on acetate ($\mu = 0.38$ h^{-1}) are not identical to cells maintained in a glucose-limited chemostat at $D = 0.38$ h^{-1}. The physiological state of the nutrient limitations is preferential for diverse ecological and biomedical studies because in situ (soils and aquatic communities) and human body in vivo (gut microbiome and infection) microbial growths are typically strictly restricted by nutrients [22,28,37,38].

What are the independent and dependent variables in these two types of SGR manipulations? The independent variable is defined as a factor manipulated by the experimenter. The batch method implies manipulation with a medium composition, while the SGR computed from the growth curve is the dependent variable. With the chemostat method, the truly independent factor is also not SGR but the limiting nutrient concentration manipulated by the experimenter by setting up the dilution rate. The desired SGR is established spontaneously in a microbial culture over a transient period that takes hours and days until the steady-state condition $\mu = D$ is achieved. In other words, the experimenter does not have the power to directly manipulate the growth rate. The D management, such as medium flow increase, has an immediate effect on cells through the instant rise of $s(t)$, the residual substrate concentration, the only environmental signal sensed by microorganisms that induces their self-adjustment through differential gene expression. Thus, the SGR is the dependent variable in both methods. Note that condition $\mu = D$ is not absolutely guaranteed even at a steady state because of the easily overlooked effects of wall growth and cell mortality ($\mu < D$) or long-term spontaneous mutations with selection ($\mu > D$) or sustained oscillation of variables x and s making SGR fluctuate around D. Sometimes, at one D, we observe two or more steady states as a memory effect of the prior growth conditions, a phenomenon called bistability [39].

Causation test based on molecular data. Now, we return to the 'SGR-dependent rRNA content'. A typical experimental result is a linear plot of rRNA vs. SGR with a high correlation coefficient, but correlation does not imply causation. The options are:

1. SGR → rRNA—the SGR is the causative factor for the variable rRNA content.
2. rRNA → SGR—reverse causation.
3. s → rRNA and s → SGR—both variables are affected by substrate concentration, s.
4. $s \rightleftharpoons \text{rRNA} \rightarrow \text{SGR}$ —options 2 and 3 are combined.

To select the best option, let us overview the published data on the rRNA expression [40–42]. In brief: each of the seven *rrn* operons in *E. coli* is transcribed by RNAP, the expression is inhibited by the alarmone ppGpp in response to uncharged tRNA (stringent control) and activated by NTP and indirectly by free amino acids (AA) that keep tRNAs charged. Although metabolomic data are limited and rather controversial (Appendices B.2.3 and B.3.3), there is evidence that intracellular pools of NTP and AA (two positive RNAP effectors) positively correlate with s (e.g., glucose), while the negative effector (p)ppGpp correlates with s negatively. It means that two options, 3 and 4, are justified, and s can be considered as the common independent factor affecting the SGR and rRNA expression (the SGR dependence on s was discovered 80 years ago [43] and is beyond any doubt). Finally, all the studied microorganisms, prokaryotic or eucaryotic, uniformly follow the so-called 'ribosomal growth law', which states that the ribosomal content in cells is the principal internal bottleneck [16,17,44,45]; the translation process is recognized as the slowest biosynthetic step; hence, growth acceleration can be achieved

only through an increase of the ribosomal copy numbers. In other words, the SGR depends on the rRNA level, not the opposite! Thus, the rRNA expression data are in a full agreement with the previous section: the limiting substrate concentration in the chemostat is the principal independent factor; it affects the SGR directly (accelerating transport and rising the M-pool) and indirectly through the upregulation of rRNA. Options 2 and 3 are correct but incomplete, while option 1 is unconditionally wrong.

The conclusion derived using rRNA as an example is also applicable to other elements of conditionally expressed MMCC, even if they decline rather than increase with the SGR. The discussion is given below (Sections 4.2 and 6.5 and Appendix B.3.2).

Why is incorrect causation SGR → rRNA so persistent? Probably, it became a part of the commonly accepted lexicon introduced in the middle of 20th century independently by several outstanding scientists [46–48]. The cliché 'growth-dependent gene expression...' conveys a clear message. Afterall, we economize on words speaking about speed-dependent fuel consumption by a car, although it is against the well-known causation sequence: fuel injection → combustion → ... → car speed. Today, the words 'growth-dependent' are often replaced by 'growth-associated', a neutral wording avoiding any causation issue, and it is enough. The only exception presents mathematical models that must follow the causation matters as strictly as possible.

The misrepresenting of SGR as an independent variable in mathematical simulations. Various models, from very simple [32] to advanced GEM [49–51], use functional dependencies like $y = F(\mu)$, where y is a conditionally expressed variable like the rRNA content, and $F(\mu)$ is a simple (linear, hyperbolic, or exponential) function of SGR. In fact, $F(\mu)$ represents mathematically the already discredited option 1 above. What is wrong in using it in GEMs and other mathematical models? The SGR is attributed to the whole cell, integrating the contributions of numerous subcellular parts, including individual conditionally expressed RNA and proteins. The whole is a very awkward factor for predicting its parts because of circular references of the kind: SGR → y_1 → ... → y_n → SGR. Furthermore, SGR is the internal cellular characteristic that cannot convey any specific environmental signal affecting the phenotype. It resembles a hopeless Munchausen attempt to pull himself by the hair out of the swamp:

Newtonian mechanics do not allow such a process, because the rider and horse as a system at rest need an external force to move up the center of the mass. By the same token, a growth model using $F(\mu)$ functions ignoring the external growth conditions misses the processes' driving forces; therefore, it is unable to reproduce microbial biodynamics as a self-regulated and self-evolving process.

Furthermore, implementing the function $y = F(\mu)$ is against the basic principles of the regulated gene expression, aimed to increase the versatility and adaptability of an organism to a changeable environment. It allows the cell to express only those gene products (proteins and RNA) that are required. For example, the lac operon turns on lactose metabolism only in the presence of lactose and absence of glucose. The function $y = F(\mu)$ does not carry any information about the state of the environment. SGR can be below μ_m for many reasons (deficiency of nutrients, toxins, nonoptimal T, pH, etc.), and the model cannot predict based on the $y = F(\mu)$ function the way to optimize the expression profile.

2.5. Drawing the Line between Simple and Complex Biodynamics

At this point, we are ready to formulate a distinction between simple and complex microbial growths. A simple growth must be exponential with a constant SGR and a fixed

MMCC that can occur only under steady environmental conditions (constant physico-chemical and nutritional conditions and lack of self-poisoning). In a continuous culture, all the listed conditions are met when the system attains a steady state. Growth with constant MMCC has been already labeled by the term 'balanced' [52]. Thus, the epithets simple, balanced, and steady state can be applied as interchangeable synonyms. Three conditions (exponential pattern, constancy of SGR and MMCC, and steady environment) are coupled to each other, and if one of them is violated, then the other conditions also become invalidated. The linear growth observed in a fed-batch and perfusion culture seems to be simpler than the exponential, but it is not true, because it implies a progressive decrease of SGR that should be accompanied by respective MMCC changes; therefore, the linear growth is complex. Unsteady environmental conditions inevitably induce SGR and MMCC variations, and, finally, it is impossible to sustain a constant SGR under variable MMCC and vice versa. The complex biodynamics do not have any restrictions; they are observed under changeable conditions and accompanied by parallel changes of the growth rate and the MMCC status, including the expression profile. It is the most natural dynamic state of microorganisms in laboratory, industrial reactors and outdoor ecosystems. The next step of increased complexity is the spontaneous differentiation of cells into two or more subpopulations with the opposite traits, e.g., active–dormant, sessile–motile, attached–planktonic, etc. Below, we are going to demonstrate that a described dynamic complexity does not exclude the possibility of control and monitoring the state of the bioprocess as accurately and reproducibly as it is done with a simple growth.

3. Typical Examples of Simple and Complex Biodynamics

3.1. Batch Culture

The canonical pattern of the batch growth curve as a sequence of the four phases (Figure 2A) becomes an indispensable illustration reproduced in every microbiology text-book. We traced the origin of this sketch back to 1918 [53]. Robert Earle Buchanan (1883–1973) was one of the most influential bacteriologists worldwide, contributing most significantly to the bacterial taxonomy. In 1918, he was 35 and already eight years in charge of bacteriology and the newly appointed Dean at Iowa State College, " . . . hard to think of him working at a bench" [54]. Accordingly, his lucky paper was not just a routine experimental report; rather, it was a generalized summary of the accumulated knowledge on bacterial development analogous to the ontogeny of plants and animals. The paper ignores methodological details and presents growth curves as trends without actual data points. To feel the gap, we addressed the contemporary relevant publications [55–58] and found that bacteria were grown that time on broth (mostly brain infusion) in tubes without stirring or pH control, and cell concentration changes over time were assessed by plating. The real published data differed from the Buchanan profile by two main points: (i) the absence of a stationary phase as an extended period when 'the growth rate is equal to the death rate' [53]; instead, all growth curves passed the sharp peaks with the following decline and (ii) a more extended death phase as compared with the Buchanan pattern. Growth on proteinous broth is accompanied by the release of NH_4^+ and severe alkaline stress (pH curve added to Figure 2A based on our studies), which is another reason for growth halt in addition to the nutrient's depletion. Batch growth under defined conditions (Figure 2B) further confirms two mentioned inadequacies. The exponential phase is immediately followed by the starvation death phase with a progressive decline of the mortality rate over time indicating a self-regulatory stress-response of the starving bacteria.

Batch growth limited by an anabolic (conserved) substrate (Figure 2C) deviates most significantly from the canonical version; it does not stop after depletion of the limiting substrate (phosphate), and cells continue division without nutrient consumption. The entire growth curve is unbalanced and complex. Finally, the fragment of the batch process associated with the biosynthesis of secondary metabolites, such as antibiotics (Figure 2D), is completely beyond the exponential phase of the balanced growth, since the biosynthesis of secondary metabolites is always activated in nearly depleted cultures.
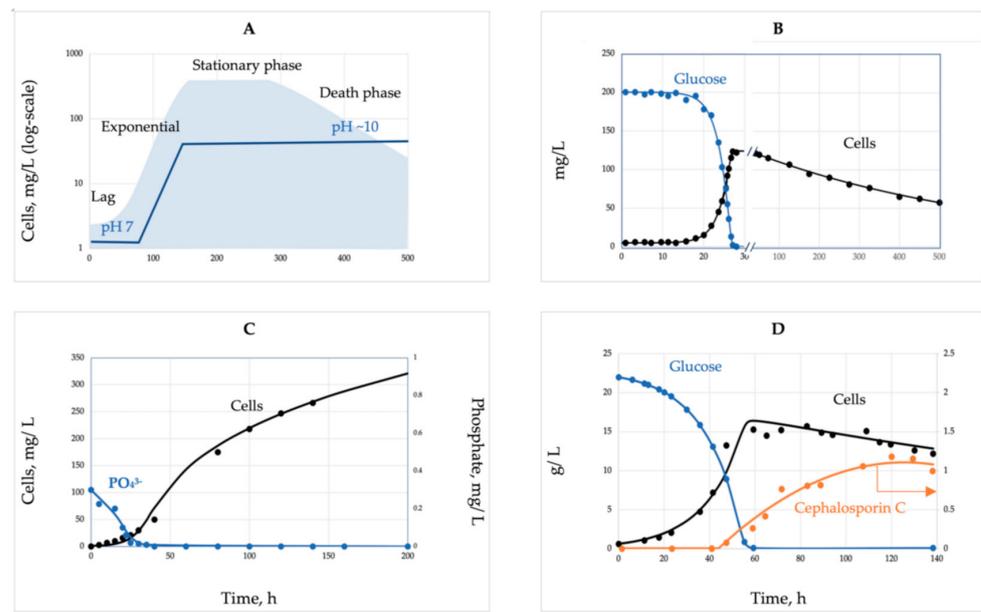
Figure 2. Examples of complex dynamic patterns observed in a batch culture. (**A**) A canonical presentation of batch growth on nutrient broth without a pH control originally deduced by [53]. (**B**) The growth curve of yeasts *Schwanniomyces vanrijiae* on the minimal glucose–mineral medium in a bioreactor [59]. (**C**) Phosphate-limited batch growth of microalga *Selenastrum capricornutum*, under continuous light [59]. (**D**) Growth and antibiotics production in the homogeneous batch culture of filamentous fungus *Cephalosporium acremonium* [60]. The B-D plots fitted to the SCM are reproduced from [59] with permission, Copyright (2019) Elsevier.

To summarize, the simple balanced growth in a batch culture is either absent or can be found only as a relatively short fragment of the entire growth dynamics in the culture limited by C- and energy sources. The stationary phase erroneously shown in every textbook is not confirmed using controlled cultivation; there could not be a balance between the growth and the death rates, i.e., a batch culture can never attain a steady state. However, at the exponential growth phase, the physicochemical conditions remain relatively stable, and the nutrients are not yet depleted, while the toxic products are not yet formed. This relatively short fragment of the entire growth curve is characterized by constant SGR and MMCC and can be qualified as a simple and balanced growth. Formally, the system does not comply with the steady-state conditions, since variables x and s are not constant, but another condition is fulfilled, the constancy of specific rates $(1/x)(dx/dt)$ and $-(1/x)(ds/dt)$. The instrumental pH control provides a longer simple growth with constant SGR and MMCC. The most intriguing and extended dynamic phenomena over time: unusual growth patterns on anabolic substrates, lag phase of a metabolic reconfiguration, starvation survival accompanied by cellular differentiation, and the formation of secondary metabolites all take place beyond the phase of simple balanced biodynamics.

3.2. Continuous Culture

The transient shift-up process after a stepwise D increase (Figure 3A) demonstrates the biological inertia of cells manifested as conspicuous overshoots and undershoots before attaining a new steady state. With successive small-step D-changes (Figure 3B), the transients are much smoother. The third type of hidden transients (Figure 3C) is associated with spontaneous beneficial mutation and selection sweeps, the displacement of a parental line with a mutant acquired a higher affinity to the limiting substrate (lower K_s). The sweep is manifested as a noticeable decrease of s, as well as the interruption in the linear rise of the neutral mutations in phage resistance [61] or drug resistance (Figure 3C, open circles). The interrupted mutant dynamics mark a 'clock resetting', all cells are competitively eliminated, including the WT and resistant variants.
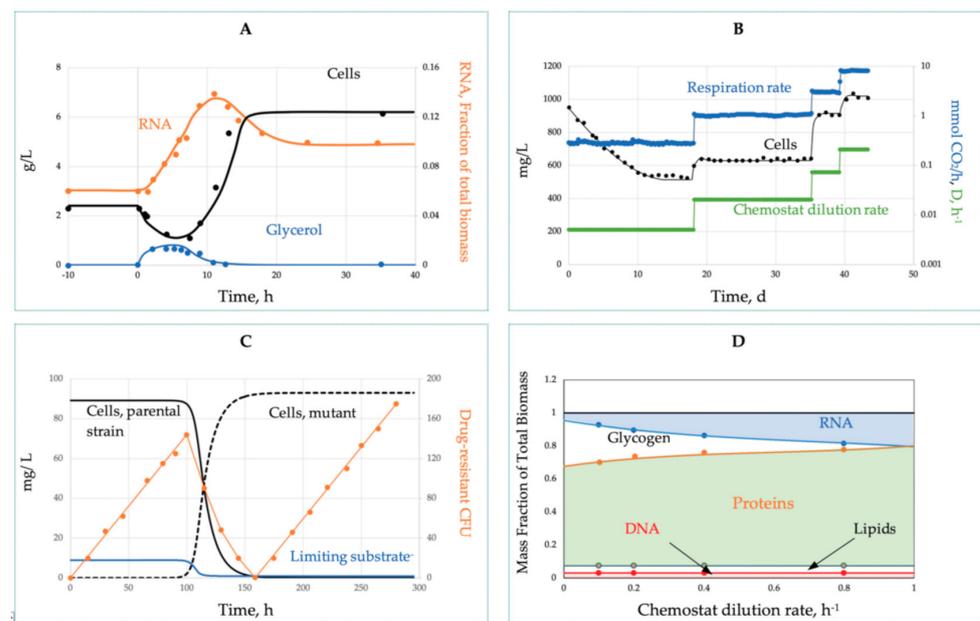
Figure 3. Unbalanced and non-steady-state growth in a continuous culture. (**A**) The shift-up experiment in a chemostat culture of *Aerobacter aerogenes* limited by glycerol; at time zero, the dilution rate was changed from 0.004 to 0.24 h^{-1} [62]. (**B**) Time series of the cell biomass and respiration (CO$_2$ formation rate) in a chemostat culture of *E. coli* K-12 after sequential changes of the dilution rates (Panikov, unpublished). (**C**) Spontaneous mutation and auto-selection in a chemostat. At t = 0, one out of ~10^{11} cells mL^{-1} underwent a beneficial mutation improving the substrate affinity (K_s declined), followed by a complete replacement of the parental strain. A genetic sweep was manifested by a decline of the substrate concentration and resetting of the linear accumulation of the drug-resistant cells (Panikov, unpublished rifampicin-resistant CFU data for *Mycobacterium smegmatis*). (**D**) Changes in the cell composition of *A. aerogenes* grown in a NH4$^+$-limited chemostat culture as dependent on the dilution rate in a chemostat [62]. All curves were calculated using SCM. Plots A and D are reprinted with permission from [59]. Copyright (2019) Elsevier.

As compared with the batch culture, the chemostat allows improved control over microbial growth with an extended duration of the simple balanced and truly steady-state growth under a stable environment. Although we need to watch for a possible selection sweep and biofilm formation (wall growth), nevertheless, the culture stabilized at constant *D* is by far the best representation of simple growth. The transitions between steady states belong to the category of complex biodynamics. The non-steady-state chemostat has another remarkable advantage by fixing over time the starting and terminating states of the transient process, making non-steady-state growth highly reproducible and amenable to vigorous kinetic analyses. Importantly, the pooled chemostat data from several *D* no longer belong to the category of simple growth, even if all the transient data are excluded.

Indeed, at each *D*, the cells acquire unique MMCC and growth characteristics (substrate affinity, metabolic potential, yield, etc.) that are distinct from other *D.* These data can be combined under the umbrella of the complex growth model. Figure 3D illustrates the amplitude of growth-associated changes in the chemical composition of the cells; the proteome expression will be discussed below.

Figure 4 illustrates biodynamics in the continuous culture retaining growing cells (syn: perfusion culture, retentostat, and chemostat with cell recycle) that uses continuous dialysis or tangential filtration to return growing cells back to the cultivation vessel. Contrary to the regular chemostat, the retentostat attains a quasi-steady state: the cell mass *x* increases (as there is no washout), while *ds/dt* ~ 0. Over time, the nutrient supply per unit cell mass dramatically decreases, driving the population to a state of chronic starvation, an extremely deep substrate limitation but without the decay and self-poisoning typical for a depleted batch culture. That is why perfusion is especially beneficial for the cultivation of slowly

growing mammalian cells and fastidious microorganisms. From the 10th day onwards, the division frequency drops down to as low as one division per month, comparable with the reproductive pace of natural microbial populations in the most challenging oligotrophic habitats (lakes, sea pelagic zone, soils, and subsoils) that are otherwise impossible to reproduce in the laboratory [28]. The near-zero growth provokes the formation of nonreplicative VBNC cells and deviates considerably not only from the simple growth discussed above but also from a more familiar complex biodynamic observed in a regular chemostat.
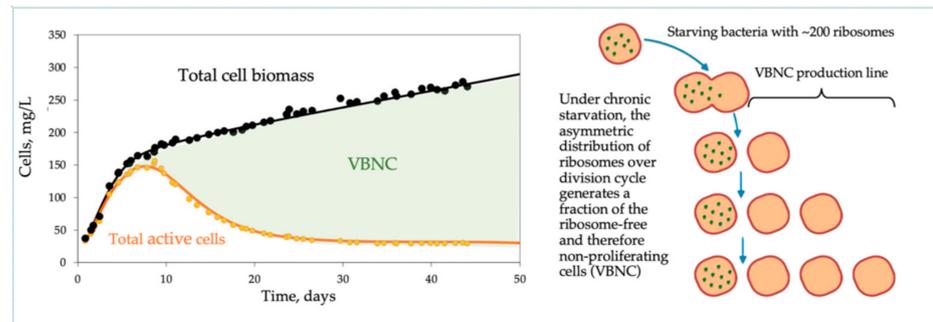

Figure 4. Near-zero growth of Pseudomonas putida F1 in a continuous dialysis culture [28]. **Left**: Growth dynamics and the formation of VBNC (green shaded area). **Right**: The hypothetical mechanism explaining VBNC formation. Similar dynamic patterns were observed in a continuous dialysis culture and in a chemostat with cell recycling. Reprinted with permission from [28]. Copyright (2015) the Society for Applied Microbiology, London, UK.

4. Pre-Genomic Models of Microbial Growth

4.1. From Malthusian Exponent to the Cybernetic Models

Now, the task is to overview the ability of mathematical models to adequately reproduce simple and complex microbial dynamics. We start from the pre-genomic models as predecessors of much more comprehensive genome-scale models. Table 1 presents the selected growth models, the milestones of microbial growth kinetics. The exponential model (4) has a historical significance as the first demographic model, albeit under a naïve assumption of unlimited growth. The next three models, (5)–(7), are more realistic, although still ignore the MMCC changes; they introduce various ways to restrict unlimited growth: the logistic Equation (5) assumes that growth is negatively affected by density-dependent biotic interactions, the Monod model (6) explains the growth restriction by the limited availability of nutrients, and finally, the Monod–Ierusalimsky model (7) considers the combined restrictive effect of substrate limitation and product inhibition. The Monod-type models (6) and (7) were already discussed above; they represent cells as 'proliferating enzymes' ignoring the possibility of self-regulatory changes of the biokinetic qualities of cells. Numerous attempts to improve the Monod equation, such as adding a third parameter or a second variable (e.g., s and x) and using an alternative to hyperbola expressions, were not successful; see the details in [22].

The rest of the models (8)–(11), allow MMCC variations, although to different degrees. The cell quota model (8) was developed for phototrophic microorganisms using light as an unlimited energy source, their growth being controlled by the availability of anabolic substrates, mostly sources of N, P, and vitamins. The cell quota, σ, is the variable content of the deficient element in a cell mass, and the respective growth model was the first attempt to explicitly express the relationship between the SGR and the chemical composition of the cells. The compartmental and structured models (9) are the closest relative to the GEMs, exploiting the idea to combine exchange fluxes (uptakes of exogenous substrates) with internal cellular variables. The simplest models contained just two to three highly aggregated internal variables (e.g., RNA, DNA, and proteins); the more advanced models covered up to 50 real metabolites and the lumped variables such as pools of amino acids, rNTP, dNTP, cell wall precursors, RNAs, DNA, ppGpp, glycogen, and peptidoglycan [63]. They never gained a wide recognition, remaining a caricature of the real metabolic network

for biochemists, cumbersome ODE sets for mathematicians, and an awkward research tool for systems biologists. The obvious disadvantages were (i) inevitable subjectivity in the selection of the most important internal variables, and (ii) an unresolved problem of experimental verification of the incomplete models against the real biochemical data. The cybernetic models (10) were initially introduced as simple structured models, the word 'cybernetic' implying that living cells and electronic gadgets share the common feature of self-regulation. They used special cybernetic variables **v** and **u** as empirical Boolean functions to implement the principle of optimality like FBA models do now. More recent cybernetic models still use the **v** and **u** variables but are scaled up to the genomic level [64–67]. Their value as a special category of GEM is questionable because of the artificial nature of the operational cybernetic variables. The negative feature shared by all three (Equations (8)–(10)) models was that they mostly ignored the growth conditions and did not systematically follow the relationship between the growth and MMCC. The only exception was the Synthetic Chemostat Model (SCM; Equation (11)) that is described below in more detail.

Table 1. Pre-genomic models of microbial growth kinetics.

ODE	Equation	Comments
$\frac{dx}{dt} = \mu x$	(4)	Exponential growth [68]. SGR is constant
$\frac{dx}{dt} = Rx\left(1 - \frac{x}{R}\right),\ R = \mu - a$	(5)	Logistic equation [69]. Growth is restricted by the negative biotic interactions; K is the upper x limit
$\frac{dx}{dt} = \mu x - Dx; \frac{ds}{dt} = D(s_r - s) - \frac{\mu x}{Y} - mx;$ $\mu = \mu_m \frac{(s - s^*)}{K_s + s},\ s^* = \frac{K_s m Y^{max}}{\mu_m},\ Y = -\frac{dx}{ds} = \frac{Y^{max}D}{D + mY^{max}}$	(6)	Monod chemostat model [70] with a maintenance term [14]. SGR is a hyperbolic function of s and is negative below threshold s^*. The yield Y varies because of the maintenance
$\frac{dp}{dt} = Y_p \mu x - Dp; \mu = \mu_m \frac{s}{K_s + s}\frac{K_p}{K_p + p}$	(7)	Monod-Ierusalimsky model [22] accounting self-inhibition by the product, p; x and s are defined by (6)
$\frac{dx}{dt} = \mu x - Dx; \frac{ds}{dt} = D(s_r - s) - \mu x \sigma;$ $\mu = \mu_m \frac{\sigma_m(\sigma - \sigma_0)}{\sigma(\sigma_m - \sigma_0)}$	(8)	Droop model [22,71]. The 'cell quota' σ is the content of deficient nutrient element in cells. The model was specially designed for microbial growth limited by the conserved nutrient substrates
$\frac{dC_i}{dt} = r_1(s, C_1, \ldots, C_n) - r_2(s, C_1, \ldots, C_n) - \mu C_i$ $\sum_{i=1}^{n} C_i = 1$	(9)	Structured models with multiple internal variables C_i [72]. Model includes two types of processes: (i) exchange reactions between cells and surrounding milieu (nutrients uptake, products release, see Equations (3)–(5)) and (ii) the MMCC changes
$\frac{dx}{dt} = x \sum_{i=1}^{2} \mu_i v_i - Dx; v_i = \frac{\mu_i}{\max(\mu_i)}$ $\frac{ds_i}{dt} = D(s_{ri} - s_i) - x\left(\frac{\mu_i v_i}{Y_i} + m_i v_{mi}\right)$ $\frac{de_i}{dt} = q_i u_i - \mu e_i; u_i = \frac{\mu_i}{\mu}, \mu = \sum_{i=1}^{2} \mu_{mi} e_i \frac{s_i}{K_{si} + s_i}$	(10)	Cybernetic models [73] describing the diauxic growth of bacteria on the mixture of glucose and lactose taken up by transporters e_1 and e_2 respectively. The cybernetic variables regulate transcription (u) and uptake rates (v), e.g., $v_1 = \begin{cases} 1, & if\ \mu_2 < \mu_1 \\ 0, & if\ \mu_2 \geq \mu_1 \end{cases}; v_2 = \begin{cases} 0, & if\ \mu_2 < \mu_1 \\ 1, & if\ \mu_2 \geq \mu_1 \end{cases}$
$\frac{ds}{dt} = D(s_r - s) - q_s x,\ q_s = rQ\frac{s}{K_s + s} + (1 - r)\frac{Q's}{K_s' + s}$ $\frac{dx}{dt} = \mu x - Dx,\ \mu = Yq_s - a_0 r$ $\frac{dr}{dt} = \mu\left(\frac{s}{K_r + s} - r\right)$	(11)	Synthetic Chemostat Model [74]. The r-variable integrates the MMCC variation and tunes the kinetic terms earlier assumed to be constants (m, Q, a, and K_s). The ODE dr/dt explains the observed biological inertia. See the Section 4.3 for other details

See the list of common symbols in Appendix B. Red marks the regulatory self-adjustable variables in the cybernetic models and SCM.

4.2. Synthetic Chemostat Model (SCM)

The SCM has been built upon all the chemostat models listed in Table 1 (that explains the epithet synthetic). It pragmatically denies the likelihood of the explicit accounting of thousands of individual MMCC and instead splits them into the P and U clusters with opposite expression trends (Table 2).

Table 2. Conditionally expressed cell constituents.

P-Components	U-Components
Related to the primary metabolism, required for intensive growth	Related to the secondary metabolism, improve stress resistance
Upregulated under optimal growth conditions	Downregulated under optimal conditions
Ribosomes (rRNA and r-proteins), enzymes involved in translation, and other key cellular processes	Protective pigments, C-storage (glycogen and PHB), antioxidants, high-affinity transporters, enzymes of drug resistance, and antibiotics formation

Each component has low ($P_i^{\min}$ and $U_j^{\min}$) and high limits ($P_i^{\max}$ and $U_j^{\max}$). Normalized to the total biomass x vectors, P and U are bound by the conservation condition P + U = 1 (for simplicity, a minor contribution of the small molecules is neglected). Then, any increase in the sum of the P-components with growth acceleration should be compensated by an equivalent decrease of the U-sum and vice versa. The heart of the SCM is the assumption that the sum of the two vectors P + U can be represented as the following linear function of one scalar variable r and three constant vectors:

$$\mathbf{P} + \mathbf{U} = \begin{vmatrix} P_1^{\min} \\ ... \\ P_n^{\min} \end{vmatrix} + \begin{vmatrix} U_1^{\min} \\ ... \\ U_m^{\min} \end{vmatrix} + r \begin{vmatrix} P_1^{\max} - P_1^{\min} \\ ... \\ P_n^{\max} - P_n^{\min} \end{vmatrix} + (1-r) \begin{vmatrix} U_1^{\max} - U_1^{\min} \\ ... \\ U_m^{\max} - U_m^{\min} \end{vmatrix} = |C_0| + r|C_P| + (1-r)|C_U|$$

$$\underbrace{\qquad\qquad}_{\substack{\text{Constitutive part,} \\ |C_0|}} \quad \underbrace{\qquad}_{\substack{\text{Changeable P} \\ \text{part, } |C_P|}} \quad \underbrace{\qquad}_{\substack{\text{Changeable U part,} \\ |C_U|}}$$

(12)

The vectors $|C_0|$, $|C_P|$, and $|C_U|$ remain constant for a given organism under any growth conditions. The r-variable depends on s Equation (11), and other factors (T, pH, and osmolarity) can be added in a multiplicative way. The role of the r-variable is to be a reporter on the status of the entire MMCC. The higher the r, the higher the expression degree of the P-components supporting intensive growth, with a penalty for reduced stress resistance. The r-variable participates in the rate expressions for SUR, SGR, turnover, and the maintenance term, providing a higher flexibility in response to the growth conditions. More technical and interpretive details about the SCM are given in Appendix B.3.3.

4.3. Strength and Limitations of SCM

The main strength comes from introducing the link between the growth kinetics and the state of the MMCC. It explains a long list of the formerly mysterious behaviors repeatedly observed in the cultivation practice (Table 3). Despite its simplicity (just three ODEs in the basic model and a few more in specialized versions), the SCM adequately simulates a wide range of complex microbial behaviors. All the dynamic curves plotted in Figures 1–5 were calculated by using a SCM demonstrating an adequate agreement with the observations. The principal model limitation stems also from its simplicity, giving the disadvantage of low resolution in the interpretation of omics array-style molecular data. This deficiency can be fixed by integrating the SCM with the appropriate GEM.

<p align="center">**Table 3.** Deciphering mechanisms controlling the unbalanced and non-steady-state growth [22].</p>

Phenomena	Traditional Interpretation	SCM-Based Clarification
Lag-phase (τ) in batch culture	Metabolic adjustment (poorly predictable)	Lag-phase is predicted based on r-state of inoculum (r_0); the higher r_0 the shorter is τ.
The stationary phase in batch culture limited by the C-source	The balance between true growth and cell death	There is no steady-state, $dx/dt < 0$, but starving cells differentiate into the active and dormant (VBNC) subpopulations. The cryptic growth is accompanied by upregulation of the U-components improving survival
The death phase	Death rate exceeds growth	
Batch growth limited by the anabolic (conserved) substrates	No explanations to the phenomenon of growth without uptake of deficient element from medium	The total amount of deficient element in the culture remains constant but σ_P progressively declines over time being shared between mother and daughter cells
Shift-up and shift-down in chemostat culture (rise or drop of D)	The observed overshoots and undershoots are vaguely attributed to the biological inertia. Prediction is not available	The SCM reproduces inertia by allowing cells to reconfigure their MMCC (proteome profile, ribosomes number). Time delay is automatically generated by Equation (A24) since the pace of r change depends on SGR
VBNC formation	A hypothetical ontogenetic stage in the natural life cycle of some bacteria	Ribosome-free cells produced under chronic starvation because of asymmetric distribution of rare ribosomes between mother and daughter cells
Cells response to starvation, the impact of maintenance	Cells dye when the energy supply is below the m-level. The maintenance coefficient is constant	Cells survive under deep energy source limitation by adaptive reducing maintenance requirements. Growth becomes slow but never stops
Inverse relationship between SGR and stress resistance	Reasons unknown, interpreted as a descriptive knowledge	Follows immediately from the P- and U-components definition and conservation condition P + U = 1

How reasonable is the assumption expressed by Equation (12)? Essentially, it implies that diverse individual macromolecules vary in response to environmental stimuli in a coordinated synchronous fashion, e.g., if one P-component starts upregulation, then the other should follow the same trend, while U-components should be downregulated to comply with the conservation condition. What is the biological logic behind this synchrony? We list below four direct and indirect supporting arguments.

First, a commonly accepted empirical fact discovered about 100 years ago [75] states that stress resistance is minimal under optimal growth conditions. This empirical generalization has never been rejected. Second, the synchrony between various P-components is expected based on molecular stoichiometry. For example, the ribosomes in *E. coli* contain 55 r-proteins and three rRNAs, which are associated with elongation factor Tu, 21 tRNAs, and the tRNA synthetases charging tRNAs)—all these ~100 giant macromolecules are tightly synchronized by several mechanisms to sustain the uninterrupted synthesis of proteins [41]. Third, transcriptional control of the expression in bacteria is regulated by sigma factors for an extended set of genes rather than for a single gene. The number of sigma factors and their functions vary among taxonomic groups of bacteria; in *E. coli*, the factors σ^S activated under one particular stress, say, starvation, initiate a synchronous expression of up to 100 genes responsible for resistance to various stresses [76]. Similar patterns of stress response were observed in eukaryotic cells, in *S. cerevisiae*, a stress response affects the expression of ~900 genes [77]. Fourth, the most convincing and direct evidence comes from the proteomic data obtained for *E. coli* grown in a glucose-limited

chemostat [21]. After normalizing and replotted vs. SGR, the original data formed two clusters of conditionally expressed proteins behaving in remarkable similarity to the main SCM postulate (Equation (12), Figure 5). Thus, the proteome of *E. coli* is conditionally expressed in coordinated synchrony, indicating that SCM correctly conveys the general trend in gene expression.
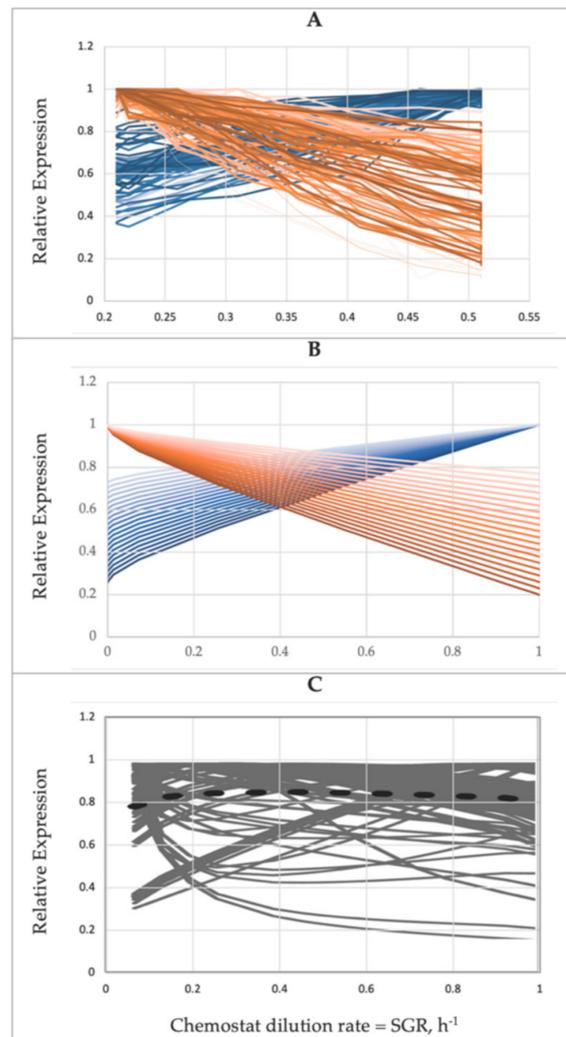

Figure 5. The conditional expression of proteins as dependent on the dilution rate in a glucose-limited chemostat culture of *E coli.* (**A**) Proteomic data [21]. Out of 1525 detected proteins, we selected the most abundant 350 proteins representing 91% of the proteome mass and normalized each protein to its maximum. Blue and orange curves stand, respectively, for the up- and downregulated proteins. (**B**) SCM computing of 40 ad hoc up- and downregulated proteins. (**C**) Profile simulated by the ME model [49].

5. Genome-Scale Models

The first organism to have its entire genome sequenced was *Haemophilus influenzae* in 1995 [78], followed by *E. coli* K-12 and several other species in 1996 to 1997 [79]. The respective genome-scale metabolic reconstructions of these bacteria were completed in few years [80,81]. From that time on, the number of GEM publications has risen exponentially [1,82,83]. This progress would not be possible without the pre-genomic period of 1990–1999 when the constraint-based modeling approach was developed and tested using incomplete genomic data [84–91]. The San Diego team led by Bernhard Palsson made overwhelming contributions to the process with prolific publications, conference talks, developed free public software accompanied with user-friendly tutorials [36], and

easy-to-learn textbooks [12,92,93]; not to be forgotten, numerous alumni have graduated from his SDSU Center and went into academia and the bioindustry across the world to form new hotspots of GEM-based research.

5.1. GEMs Cover Only the Part of the Genome

Some published GEMs represent as low as 2% of the bacterial genomes simulating the core/central metabolism [94]. The most complete metabolic, GEM [95], covers up to 35% of the genome, and it is almost the top. Figure 6 explains why the coverage cannot be higher. From a modeling perspective [92], there are four distinct parts of the genome: (i) junk material and sequencing errors need to be removed (blue background); (ii) the M-matrix (M for metabolism) stands for genes encoding enzymes and catalyzing the entire intermediary metabolism, which is about one-third in prokaryotes and a smaller fraction in eukaryotes [10]; (iii) the E-matrix (E for expression) accommodating genes encoding the enzymes and RNAs responsible for proteins synthesis, transcription, and translation; and (iv) finally, the O-matrix (O for operon) containing the transcriptional regulatory network, a part of the global regulatory machinery orchestrating cellular functions via TFs. The respective GEMs are called M, E, and O models. The reconstruction of metabolic networks (M models) is now a well-established process (see the below FBA and dFBA models). The E models have recently been started by using a similar approach and are produced as combined ME-models. Both models convert genomic data into stoichiometric matrices of intermediates (M-matrix) and biosynthetic reactions (E-matrix) that are made interactive. Both models apply common optimality assumptions and constraints, as well as a quasi-steady state approximation, allowing their easier computational solution. Unfortunately, there is no way to predict in theory or based on sequence similarity the functionality of the O-matrix, 'the brain' of the metabolic network. Therefore, merging the O module into ME models has been claimed [10] to be impossible. However, a fully integrated OEM model was built for the simplest parasitic prokaryote [96], called a whole-cell (WC) model. Recently, WC models were developed for *E. coli* and *S. cerevisiae*, although without addressing the whole genome.
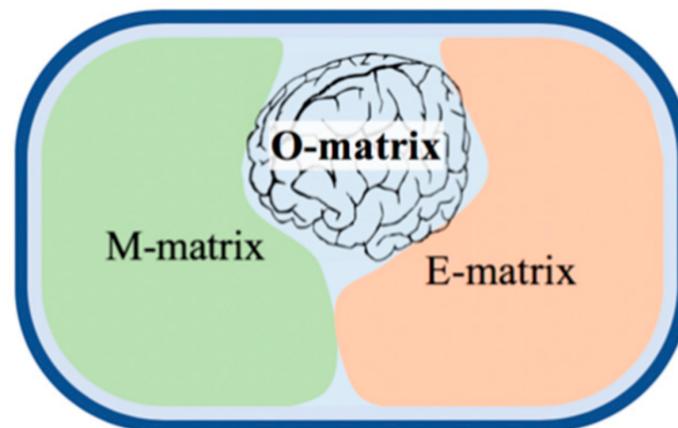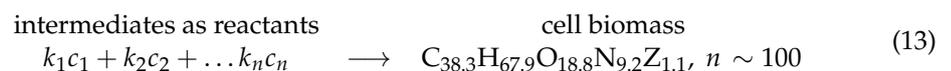

Figure 6. Simplified view on the bacterial genome structure through the prism of GEM modeling.

5.2. Flux Balance Analysis (FBA)

This technique has been introduced as a special case of the simplified genome-scale metabolic reconstruction [12,36,92,93,97–100]: (i) it seeks only steady-state solutions for metabolic fluxes; otherwise, it would be necessary to define the reaction kinetics and the metabolic pool size for hundreds of intermediates, which is unrealistic; (ii) the constraints are used to define a closed solution space for the flux vectors; commonly used constraints include the feasible ranges of each metabolic flux, and additional constraints come from reactions thermodynamics, as well as from the metabolomic, transcriptomic, or proteomic data; (iii) finally, the best solution is found using linear optimization for a certain objective function.

There are both advantages and limitations of the FBA. The main advantage is its computational efficiency: the COBRA toolbox in MATLAB [101] or COBRApy in Python [102] installed on a regular laptop executes full-size M models in seconds. Other advantages include (i) the uniform structure of FBA that is potentially applicable to any sequenced organism after the manual curation of genomic data, including the non-model newly isolated bacteria [103], (ii) the unbiased genome-scale presentation of cellular metabolism making redundant the selection of the 'most important reactions' as in the pre-genomic structured models, and (iii) a good agreement with the available experimental data [36]. The FBA is not designed for dynamic simulations and cannot present growth as a time series of x, s, and p to demonstrate the agreement. Instead, it produces the metabolic map with fluxes for each reaction. It correctly reproduces reconfiguration of the M-matrix after switching from one C-source to another, from aerobic to anoxic growth conditions; it also accurately predicts gene essentiality and the impact of gene knockdown on metabolic flows, including the exometabolic products especially valuable for bioengineering.

The limitations of FBA as admitted by the developers [36] include neglecting the kinetics of the metabolic reactions and their self-regulation. In Appendices B.1 and B.2.1, we provided an explicit solution of the FBA equations that show what specific kinetic data are required for calculation of the steady-state concentrations of metabolic intermediates. By using the introduced above growth terms, the FBA can deal with only the simple, balanced, and steady-state growth under constant optimal environmental conditions. The fixed MMCC assumes a simplified way to present the biomass as the product of a pseudo-reaction with constant stoichiometric coefficients $k_1 \ldots k_n$:

$$\begin{array}{ccc} \text{intermediates as reactants} & & \text{cell biomass} \\ k_1 c_1 + k_2 c_2 + \ldots k_n c_n & \longrightarrow & C_{38.3}H_{67.9}O_{18.8}N_{9.2}Z_{1.1}, \; n \sim 100 \end{array} \tag{13}$$

Here, the 'mole' of biomass is based on the average elemental composition of cells taken from the mid-exponential growth phase. For *E. coli,* the left side of the reaction of Equation (13) contains as reactants 20 amino acids, 4 rNTP, 4 dNTP, 15 inorganic ions, 14 cofactors, and 3 precursors of peptidoglycan and lipids. Using the corrected biomass formulae for two phases of the diauxic growth of *E. coli* instead of a fixed one [88,89] substantially improved the FBA simulation; however, it was an ad hoc solution appropriate only for this model organism.

The FBA has claimed to predict the SGR based on genomic data. However, it is not perfectly true, because we must select the low boundary for the substrate uptake $-q_s$. If it matches exactly the maximum uptake rate, Q_s, then the generated SGR corresponds to the μ_m. The Q_s is to be found in publications or we have to run a simple batch experiment when recording the C-source substrate $s(t)$ and the cell mass $x(t)$ time series. Then, the nonlinear (exponential) regression produces three growth constants: μ_m, Q_s, and Y. In other words, if we enter the experimentally verified $-q_s$ or $-Q_s$, we already know SGR without any genomic simulation! Note also that μ_m is an easier parameter to measure than Q_s. No doubt, the FBA capacity to reproduce μ_m based on the specified Q_s remains to be a nontrivial task, as it involves the computing of the huge M-matrix. If we accidentally or intentionally select $q_s < Q_s$, then FBA generates SGR proportionally decreased vs. μ_m, but metabolic flux maps preserve the same pattern as in the case of unlimited growth. It means that the FBA does not reproduce the substrate-limited growth, as claimed by Reference [36], giving instead just a wrong measure of μ_m.

5.3. The Dynamic FBA (dFBA)

These models apply a quasi-steady-state approximation. Microbial growth can be always viewed as a sum of processes with different characteristic times from extremely slow (microevolution in a chemostat culture, including selection sweep) to slow (cell biomass) and fast (metabolic pool, respiration, SUR, and limiting substrate concentration). The metabolic fluxes can also be differentiated into slow and fast categories, as dependent on the reaction volume [104]. The exchange reactions between the cell and environment

(nutrients uptake and products release) take place in the space from ml (culture tube) to 10-L benchtop bioreactor, the volume range 10^0–10^4 cm^3. The intracellular metabolic reactions occur in a much smaller volume of a single cell, 10^{-8}–10^{-9} cm^3. A volumetric difference by 10^8–10^{13} times implies that intracellular reactions can be safely assumed to be at a steady state, and therefore, the flux distribution can be resolved by the FBA. Few remaining slow exchange variables are to be numerically solved as a set of ODEs.

For the practical implementation of the dFBA [87,105], two approaches were developed: (i) the Dynamic Optimization Approach (DOA) by using NLP and (ii) the Static Optimization Approach (SOA) that discretizes the bioprocess time into small time intervals and uses LP at the beginning of each time interval. The agreement of dynamic simulation with experimental data should be estimated as modest (Figure 7). There were other numerous applications of dFBA to simulate the biodynamics of industrially important organisms: *S. cerevisiae* and *E. coli*, separately or as mixed culture [106–109], *Shewanella oneidensis* [110], *Chlorella* [111], CHO cells [112], *Pichia pastoris* [113], *Aspergillus niger* [114], and *Streptomyces tsukubaensis*, producing an immunosuppressive drug [115]. With rare exceptions, the mismatch between observation and prediction was large. Mostly, the failures occurred in attempts to go beyond a simple growth. It was expected since the exchange reactions were represented by the outdated Monod model.
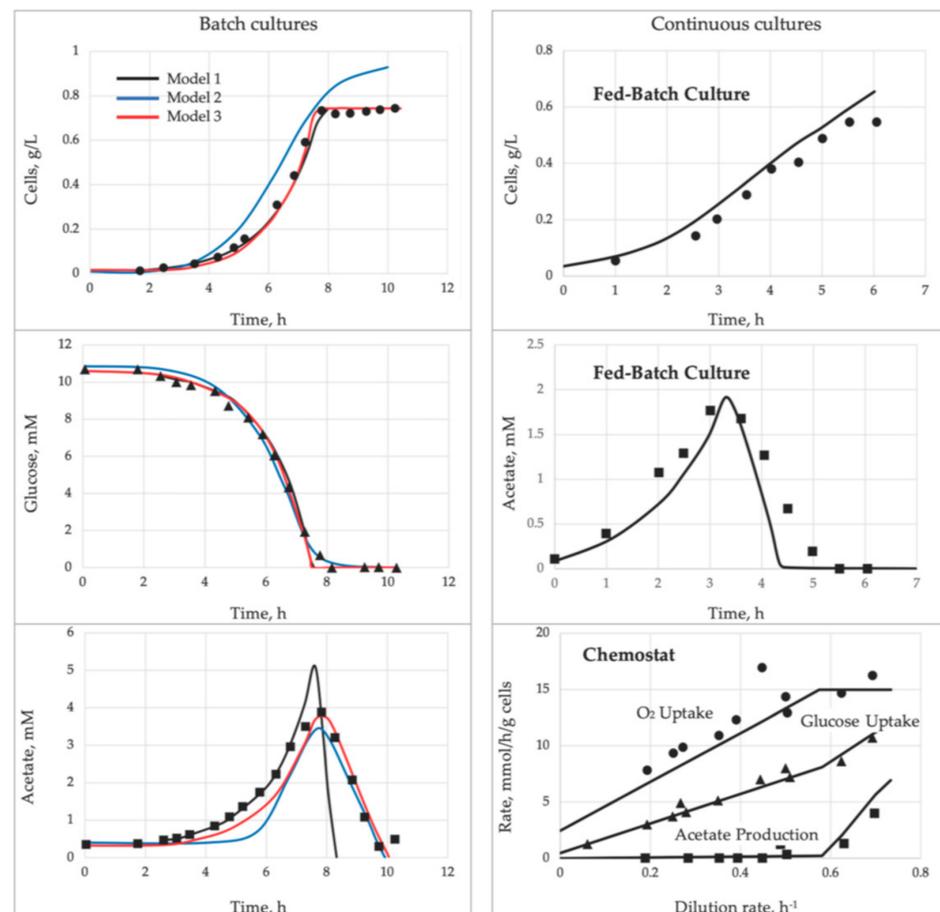

Figure 7. Dynamic FBA simulation of *E. coli* grown in a simple batch culture (**left column**) and in continuous systems (**right column**). The experimental data [86,87] were fitted to the following dynamic FBA models: **Model 1**, the SOA computational approach [87]; **Model 2**, the DOA approach [105]; and **Model 3**, the integrated hybrid FBA/ODE simulation with added transcriptional control [94]. Reprinted with permission from [87], copyright (1994) American Society for Microbiology and with permission from [59], copyright (2019) Elsevier.

5.4. The Modification of dFBA:rFBA and iFBA

The regulatory FBA models [94] dealt with the simulation of the central part of *E. coli* metabolism accounting for 149 genes, encoding 16 regulatory proteins and 73 enzymes, and 113 reactions, 45 of which were under transcriptional control, using Boolean functions. The model can predict the growth curves of *E. coli* on defined media, as well as the substrate uptake, metabolic byproduct secretion, and gene expression under variable growth conditions. The rFBAs were also combined with classic kinetic models based on ODEs [116,117]; the hybrid was called the integrated FBA (iFBA), and it gave a more accurate prediction of phenotypes. However, iFBA lacked generality and could be applied only to the well-studied model organisms. None of these FBA modifications has been applied to the complex biodynamics associated with MMCC variations. It was impossible by default, because the exchange reactions were presented by the Monod model.

5.5. ME-Models

The first ME model coupling metabolic processes with gene expression were developed for *E. coli* with the account of 423 genes [118]; it produced the long-waited result of the successful simulation of growth-associated MMCC self-regulation, demonstrating excellent agreement between the observed and simulated numbers of ribosomes. Thereafter, ME models were applied to the small-genome hyperthermophilic bacteria *Thermotoga maritima*, accounting for 651 genes [119], and to *E. coli* K-12 describing the synthesis and functions of almost 2000 gene products [49,120]. Then, the model was further developed to account for the spatial proteins allocation between the cytosol, periplasm, and inner and outer membranes [121].

The reconstruction was commonly recognized as being both difficult to compute and challenging to understand conceptually. Two improvements were soon offered: (i) the software product solveME provided up to 45% acceleration using a quad-precision NLP solver [122] and (ii) COBRAme [123], which condensed the original ME model around five times without a loss of resolution and functionality, reducing the computing time from 6 h to less than 10 min! COBRAme has been unified and reformulated into a software framework allowing to build and edit the ME models of any sequenced and annotated organism. Soon, it was applied to other organisms, including anaerobic mixotroph *Clostridium ljungdahlii,* a promising biofuel producer [124].

A recent review [1] followed step-by-step the progress in ME model development, including their growing coverage of genome, key findings, and applications. Below, we list their distinctive features as compared with FBA and dFBA based on three reconstructions designed for *E. coli* [49,123,125]. The model covered 1541 unique ORFs and 109 RNA genes, accounting for ~90% of the *E. coli*'s proteome. As compared with prior M models, the ME models face two major challenges. First, the dimension of the E-matrix is about 30 times bigger, and its elements differ by 15 orders of magnitude, presenting a severe ill-conditioned matrix problem. Second, contrary to the M models focusing on fluxome and ignoring reaction kinetics and metabolic pool sizes, the E-models are meaningless without the amounts (contents and concentrations) of conditionally expressed proteins; therefore, the kinetic characterizations and attempts to quantify concentrations are unavoidable. In practice, the ME-models were implemented as follows:

A. The metabolic network was treated similarly to FBA using constraints and objective functions. The maximizing growth rate was combined with minimizing biosynthetic costs of macromolecules. Substrate concentration was never shown as an independent variable; nutrient deficiency was accounted for by using the substrate availability bounds.

B. The mostly unknown k_{cat} values for translational and transcriptional enzymes were set up as a gross average 65 s^{-1}. Considerable improvement in the simulation was achieved by combining the ME model with proteomic data as the model's input [125]. It allowed an estimation of the individual turnover rates k_{eff} for *E. coli*. The range of

k_{eff} varied within eight orders of magnitude, and surprisingly, it was not significantly impacted by variations of the cultivation conditions (four different C-sources).

C. It can be shown (Appendix B.2.5) that the rate constant k_{eff} is dependent on the concentration of the respective intracellular substrate, *M*, e.g., in the form of the Michaelis–Menten equation: $k_{eff} = k_{cat}M/(K_m + M)$. For unlimited growth, $M \gg K_m$, and $k_{eff} = k_{cat}$, while at the nutrient limitation, $k_{eff} < k_{cat}$. However, in vitro kinetic data for the isolated *E. coli* enzymes remain incomplete and most probably not adequately characterize the in vivo metabolic processes (Appendix B.3.5). Instead of using published data, the authors applied a convoluted two-steps binary search procedure involving an auxiliary 'dummy protein' computing (first step) followed by the second binary search to find the minimal ratio k_{eff}/k_{cat} as a condition making dummy protein formation to be zero.

D. The rates of transcription and translation have been assumed to follow the empirical hyperbolic function of SGR, e.g., the translation rate $v = V^m \frac{\mu}{K+\mu}$.

E. The biomass reaction, Equation (13), was completely removed, because most of the components (amino acids, nucleotides, etc.) were already accounted for in the proteins and RNA biosynthesis reactions. Instead, the reduced pseudo-reaction was used for the remaining cell constituents: glycogen; DNA; lipids; peptidoglycans; and several cofactors (NAD, coenzyme A, etc.). The expressed but not incorporated into the model proteins were called the unmodeled protein biomass. Most likely, these proteins are responsible for stress resistance (U cluster in SCM). This fraction has been also added to the modified biomass reaction to comply with the growth–mass balance but excluded from discussion of their biological role.

F. The only sink for macromolecules was assumed to be their dilution caused by growth, so any metabolic rate v_i and translation rates of the respective enzymes were set up as follows:

$$v_{translation,i} = v_{dilution,i} = \mu[E_i]; \; v_i = k_{eff,i}[E_i] \; \therefore \; v_{translation,i} = \frac{\mu}{k_{eff,i}}v_i \qquad (14)$$

The constraining condition providing the v_i flux sustainability was [125]:

$$v_i \; \leq \; \frac{k_{eff,i}}{\mu}v_{translation,i} \qquad (15)$$

Some of the listed assumptions are not justified, specifically, the points B and C (replacing Michaelis–Menten or Hill enzyme kinetics with constant rates found via a binary search) and D (using SGR as an independent variable, discussed above in Section 2.4). Appendix B (Appendices B.1 and B.2) summarizes the published *E. coli* metabolomic data. We conclude that 90–95% of metabolic reactions within the M-matrix are not saturated by the available intracellular substrates ($M \ll K_m$), validating the first-order rather than zero-order approximation. The remaining 5–10% of the metabolic reactions follow the mixed kinetic order, and only a few out of the +5000 reactions approach the state of substrate saturation (zero order assumed by the point B). The inconsistent assumptions made in points B–D could be the reason for the failure of ME models to correctly reproduce a proteomic profile of *E. coli* in the chemostat culture (Figure 5C).

How successful are ME simulations? Unfortunately, none of the published ME models demonstrated an adequate dynamic simulation of a microbial growth and expression profile (transcriptomic or proteomic) under any cultivation scenarios shown above by Figures 1–5. On the other side, M models were able to simulate the perturbation experiment in the batch culture of *E. coli* [125]: bacteria were grown on the minimal glucose–mineral medium and then the second C-source was added (adenine, glycine, tryptophan, or threonine). The expression patterns were 56–100% correctly predicted, but it was a single point of the semi-quantitative simulation, leaving unanswered questions about the dynamic growth pattern. Another successful application was the model OxidizeME [126] simulating an *E. coli* response to oxidative stress, including: (i) ROS-induced auxotrophy for several

unstable amino acids, (ii) a nutrient-dependent variation of stress resistance, and (iii) ROS-induced differential gene expression. The only regret is that the OxidizeME has been customized to accurately reproduce as closely as possible the already known mechanisms and only those affecting iron–sulfur clusters. There are numerous other targets for ROS, and there are many other stresses. We want GEM to assist us in finding the common and specific effects of diverse stress factors, as well as to predict (based on the genome) the dynamic pattern of the microbial stress response, including the progress of recovery.

What is the impact of ME models on modern system biology? Every publication brought some novel results of general biological significance: the role of multiple *rrn* operons [118], synonymous codon usage bias [120], revised interpretation of the growth limitations [49], translational pausing regularities, and synchronization of the cellular synthesis of macromolecules [125].

6. Whole-Cell Genome-Scale Simulations (WC Models)

This type of GEM should be a *summum bonum* of microbial system biology. Ideally, all parts (O, E, and M matrices) of the genome must be accounted for in the simulation. However, it is not realistic now, because the functions of many genes are not yet defined, and we know too little about the global regulation of cellular machinery, how is it orchestrated via the O-matrix or otherwise. Three partially resolved WC models are discussed below.

6.1. Mycoplasma genitalium

The first modeling trial was applied to the free-living parasite bacterium *Mycoplasma genitalium* with the smallest genome accommodating 525 genes, including 382 essential genes [96,127,128]. Even the simplest prokaryotic cell was an exceptionally challenging object for simulation; the problem was solved by using the modularity approach [129]: cellular machinery was split into 28 modules (transcription, translation, metabolism, replication, DNA repair, host interactions, etc.), each of which was solved independently at the integration step 1 s. The modules were represented by ODE and other equations with discrete and continuous variables, mechanistic and empirical (not fully specified in the publication). The most advanced computer technique was implemented (a 128-core Linux cluster), and the simulation took 10 h for a single *M. genitalium* cell to divide once, about the same time the actual cell takes. What have we learned from these remarkable modeling efforts? The authors listed several previously unobserved cellular behaviors, e.g., the in vivo rates of protein–DNA associations and an inverse relationship between the duration of DNA replication initiation and replication. Not too much! However, it is not the final word; presently, the mycoplasma model is available as a research tool for the further exploration of parasitic prokaryotes.

6.2. Escherichia coli K-12

The same research team led by Markus Covert at Stanford University developed a second WC model, this time for *E. coli* MG1655, the best-studied prokaryote with a ~10 times bigger genome [50,130]. The model also had a modular structure and underwent further improvements.

First, the mathematical structure of the model was unified by merging the assorted modules into the unified gigantic ODE set, probably the largest set in the history of science: 1500 ODEs for the M-matrix and 9000 ODEs for the E-matrix, and 4500 + 4500 equations for RNA and proteins, respectively.

Second, the metabolic M-matrix was presented without a relaxing steady-state approximation. Instead, all reactions were supplied with individual kinetic constants (k_{cat}, K_m, V^m, etc.) found from 12,000 publications and presented as a function of the intracellular concentration of the respective metabolites. To avoid erroneous pool depletion at each step of numerical integration, the WC model employed the multi-objective function by adding a homeostatic objective: to minimize the sum $\sum_i |1 - C_i/C_{0,i}|$, where C_i is the ith metabolite concentration, and $C_{0,i}$ is the respective set point.

Third, the metabolic regulation in the M- and E-matrices was introduced through 22 individual transcription factors (TF) controlling 355 genes in one of three regulatory classes: (i) zero-component (TF becomes instantly active when expressed), (ii) one-component (TF are directly activated or inhibited by a small molecule ligand), and (iii) two-component (TF paired with a membrane protein sensing environment, involving condition-dependent phosphorylation).

Fourth, by improving file I/O and writing inner loops, the simulation runtime of the model was decreased from 10 h to 15 min!

The developed model has been introduced as a tool to integrate and cross-evaluate the extensive molecular data obtained for *E. coli* by various research groups over decades; the found inconsistencies (mostly mismatching published kinetic parameters) were suggested to correct through the sensitivity analyses of the WC model by correcting the suspicious coefficient to a value minimizing simulation error. Three specific inconsistencies were discussed: (i) the underestimated transcription probability of RNA polymerase and ribosomal subunits that failed to reproduce the observed SGR of bacteria, (ii) wrong turnover rates (half-life time) of ~15% of cellular proteins estimated in the published sources by the approximate "N-end" rule, and (iii) the most extensive inconsistencies found in the upgraded M-matrix. The reason for the last inconsistency was explained by the fact that enzyme kinetics in vitro (published data on parameters of the isolated enzymes) were different from that in vivo in the overcrowded intracellular space, especially for membrane-bound enzymes. The problem had to be fixed by adjusting the Michaelis–Menten parameters and use of the multiple objection functions mentioned above. The final intriguing inconsistency was a subgenerational transcription: >50% of the transcriptome, including 72 essential genes of *E. coli*, were found to be transcribed less than once per cell cycle. The survival of cells without essential genes remained a mystery for the authors despite being observed experimentally and reproduced by the model.

In the Appendix B, we discuss the last two extreme cases: (i) a too-low enzyme content in cells with an enigmatic phenomenon of less than one copy of enzyme per cell (Appendix B.3.4), and (ii) a too-high protein content, causing macromolecular overcrowding (Appendix B.3.5). In the first case, we concluded that rare enzymes are likely to be an artefact of using proteomic surveys of heterogeneous populations of cells. These enzymatic proteins should not necessarily represent the low number of copies per cell; rather, their source in proteomic analyses could be the rare cells of spontaneous mutants coexisting with the WT. The overcrowding has been reproduced by simple kinetic models that can potentially be applied in GEM to mimic a specific overcrowded environment.

6.3. Saccharomyces cerevisiae BY4741

The third WC model [51] was constructed by using the same modular approach as two other WC models. Moreover, the acknowledgment informs us that this reconstruction was done using the WC software developed by the Markus Covert lab for mycoplasma and *E. coli*. Sadly, the WC model was not the only part of the publication, being mixed up with a bulky experimental component on metabolomics and genetic manipulation studies. Consequently, the overloaded publication left many modeling specifics without a clear explanation. Hopefully, there will be other reports using this model and disclosing its features. Meanwhile, we assume that WC models for *E. coli* and *S. cerevisiae* are closely related.

6.4. Limitations of the WC Models

Functionally, the largest *E. coli* WC model covers 1214 genes, 28% from the total genome and 45% from the annotated genome. It was lower as compared with the published ME models [1,49,131], and the only apparent progress of the WC models was the upgraded dynamic state for the M-matrix; the positive impact of this upgrade on the model's performance remains to be proven in future. The skipped genes/processes include: (i) proteases eliminating redundant proteins, (ii) chaperones refolding damaged proteins,

(iii) the sigma factors, and a few other factors controlling the transcription. All three groups of macromolecules are crucially important for the self-regulatory behavior of bacteria under nonoptimal growth conditions, especially under stresses. Without them, the model reproduces a defective phenotype of cells able to grow under optimal conditions but expected to be super-sensitive to the smallest deviations from the growth optimum. Thus, the first limitation of the model is the overlooked stress resistance in the reconstructed phenotype.

The indicated deficiency hardly could be revealed in the undertaken study because of the second limitation, neglecting the environmental factors. Even though the growth of bacteria was simulated on three media (minimal and complex aerobic and minimal anoxic), it was not enough to claim that the study 'explored the effect of growth conditions'. The model never addressed the issue of nutrient limitations or starvation stress or any other stresses. Moreover, the SGR (or corresponding doubling time) was a fixed value throughout the entire simulation.

The third deficiency of the study was the minimal, if any, attempt to verify the model against experimental data by testing the model's ability to reproduce distinctive behavioral patterns, e.g., the shapes of the growth curves, respiration and fermentation dynamics, proteome profile vs. SGR, etc. Without rigorous experimental verification, the WC model cannot be trusted for its claimed goal to be a tool for the quality control of molecular data.

The fourth limitation is the overstated attention to the cell cycle events, ignoring the other important behaviors of cells (see below).

6.5. Cell Cycle as a Part of Growth Kinetics

Cell division cycle is an important part of bacteriology using specialized techniques (time-lapse live-cell microscopy, microfluidics, cells sizing, fluorescent probes, etc.) and dedicated mathematical tools operating with discrete and continuous variables, partial differential equations (PDE), accounting for changes across the time and cell ages [132] and statistics of the frequency distributions of the cells aimed at the discrimination of alternative division mechanisms [133]. It would be wrong to think that a single-cell study is literally limited to one cell; more appropriate is to define it as a bottom-up approach that starts from single-cell observations followed by their assembling into a holistic picture of a heterogeneous population. The source of heterogeneity is the combined effect of numerous factors, including the passing through division cycles and intrinsic stochastic noise.

The opposite top-down approach operates with bulk population measurements over many cells, followed by inferring the average cell status by using mostly deterministic mathematical models. It is faster, easier, allowing more comprehensive and accurate molecular assays (for comparison, the single-cell proteomics reveals only a fraction of the global proteome assessed by the bulk analyses). The top-down mainstream in microbiology has always been accompanied by bottom-up single-cell studies, which recently have attracted more attention due to remarkable opportunities provided by microfluidics [134–138]. We need both types of research and realize the inherent limitations of each one: the bottom-up single-cell approach focuses on cell division, missing other cellular processes, while the top-down approach neglects cellular heterogeneity and can overlook important mechanisms involving the 'division of labor' between individual cells. Most GEMs are deterministic models using ODE and continuous variables (rather than PDE and discrete variables) and, therefore, are better aligned with the top-down data format. The standard output of such GEM includes dynamic data on the asynchronous population of cells evenly distributed across the division cycle, but it can easily be reconfigured into a synchronized format [22] to simulate cell cycle events controlled by the limited number of genes (DnaA—DnaC, FtsZ, septation proteins, etc.). The opposite is not true, and reconfiguration of the specialized cell cycle GEM into the broad-spectrum style is likely to fail because of its narrow scope. Based on this explanation, the only focus of the WC model on the cell cycle seems to be self-restrictive: the whole-cell model should be more inclusive than the single-cell cycle model.

7. Conclusions: Strengths, Weaknesses, and Prospects of Specific GEMs

Let us summarize. The undertaken critical review was aimed to reveal the pros and cons of the existing GEMs to identify the most promising directions to move in. We started by defining the terms of simple and complex biodynamics and introduced several basic principles of microbial biokinetics that have not yet been commonly accepted. These principles were summarized as three required conditions to be met by any GEM intending to reproduce the complex microbial biodynamics:

(1)　to account for the effects of environmental factors on cell dynamics, especially related to the nutrient supply (limiting substrate, its concentrations, and regime of delivery).

(2)　to simulate the adaptive MMCC changes as conditional gene/protein expressions in response to a changeable environment.

(3)　to predict the SGR as a complex function of the MMCC status and environmental conditions rather than a fixed input parameter.

The basic FBA model meets none of the three conditions, and even a simple growth is presented ambiguously. However, the FBA is still the most popular GEM that gained enviable recognition among diverse users. The FBA is really very good in static applications (e.g., gene knockouts and essentiality prediction), and it is attractive because of the elegant computational resolution of the entire M-matrix. This quality should be preserved but how do we improve the dynamic strength of the FBA without making it too cumbersome?

A feasible solution could be to couple the basic FBA with the SCM. The SCM itself meets all three conditions above, including a coarse-grained nongenomic reproduction of the MMCC profiles (Figures 3D and 5B). Easy computational steps are required to transform the MMCC profile into the self-adjustable biomass reaction, replacing the fixed reaction (Equation (13)). Then, the stoichiometric coefficients $k_i s_i$ on the left side and the biomass formulae on the right side can be presented as functions of the r-variable (Appendix B.3.3).

The adjustable biomass reaction links the SCM to the steady-state FBA, making two models interactive. The expected outcome is that every FBA-generated fluxome would become associated with a specific growth pattern. On the other side, any step in the complex biodynamics can be supplemented by a corresponding fluxome screenshot. Each screenshot is static, but their time series produces something like a series of frames in a Walt Disney animation. The range of accessible conditions is practically unlimited due to the flexibility of SCM: any cultivation systems (chemostat, perfusion, fed-batch, and the entire batch process from the lag to death phases); variable media composition; single or multiple limitations; optimal and suboptimal pH; temperature; and other physicochemical conditions, including physiological stresses. Probably, it would be possible to predict the fluxome of mutants vs. the WT under optimal and any challenging growth conditions important for industrial setups. Especially beneficial it would be for the bioprocessing of antibiotics and other secondary metabolites produced only under suboptimal growth conditions.

The dynamic FBA models containing the Monod-style exchange module and the fixed biomass reaction do not meet requirements (ii) and (iii). Therefore, they adequately reproduce only simple growth that is of limited use in bioindustry and biomedical applications. A feasible solution could be replacing the outdated Monod model with SCM and upgrading the biomass reaction, as discussed above. The only difference from the FBA–SCM hybrid is that the SCM must be fully integrated into the dFBA, and the biomass reactions need to be updated at each integration step using the SCM r-variable as the input. Then, the dFBA–SCM hybrid will be able to simulate the steady-state and transient bioprocesses under diverse conditions with partial characterization of the expression profiles. The expected advantage of the dFBA–SCM hybrid over the state-of-the-art ME and WC models is a relative simplicity and robustness, making it accessible to unexperienced users.

The published ME models comply with all three requirements but incompletely: (i) the effect of environmental factors has been incorporated in a 'relaxed' way, e.g., by setting constraints on the s-variation rather than explicitly by using kinetic equations (Table A1); (ii) the adaptive MMCC changes were simulated only for rRNA and proteins, neglecting other variable cell constituents, e.g., glycogen, polyphosphate, etc.; (iii) the protein expression was

simulated using LP and multiple objective functions (maximizing the SGR and protein yield and minimizing the biosynthetic cost) without a clear relationship with the environmental factors; and (iv) the SGR has been both the final output of reconstruction and the factor affecting numerous intermediary steps (e.g., the transcription and translation reactions). Due to the listed omissions, the ME models were not fully successful in a simulation of the complex biodynamics. We showed (Appendix B.3.2 on the conditional expression of transporters) that introducing a limiting substrate concentration as the independent environmental factor would fix the problem. Other prospects are discussed below, jointly with the WC models.

WC models. None of the three required conditions are met in these models [50,130] because of the fixed status of the SGR, MMCC, and growth conditions. Most likely, it was done on purpose to avoid nuisances of a fluctuating environment and stay focused on the molecular complexity. However, the disadvantage of the narrow model's scope is that its experimental verification cannot be performed completely; the explanation is given below.

GEM verification against a simple growth can never be complete. Many modelers prefer simple growth data, assuming that only these data are accurate and reproducible. This is not true; looking back (Figures 1–5), we can testify roughly the same experimental scatter for simple growths and complex biodynamics. All depend on the selected experimental techniques, for instance, a fully controlled bioreactor allows recordings of the most challenging unbalanced transient growths with higher accuracy and precision than the simplest exponential growth in shaking flasks. Another misconception comes from a pearl of wisdom: learn to walk before you run. It assumes a step-by-step evolution from a basic model handling simple growth to more advanced versions by adding new modules able to cope with complex biodynamics. However, such a strategy usually badly fails, because the complex biodynamics cannot be decomposed into the series of simple growth steps. It is like trying to learn riding a bike starting from a 'simplified' one wheel machine and then step-by-step adding the second wheel, frame, handlebar, and other parts.

WC models are the largest reconstruction in modern microbiology, dealing with a gigantic set of 10,500 mostly nonlinear ODEs without simplifying the steady-state assumptions. This goliath has been applied to the exponential growth between two consecutive divisions, the simplest possible dynamic process. The most advanced part of the model is the combined M- and E-matrices representing TF-regulated metabolic reactions. Some of them are involved in cell division; therefore, the omics data obtained in the synchronized batch culture can be used for experimental verification. However, the rest of the self-regulatory machinery responds to a changeable environment and, therefore, will be not observed in a simple growth experiment! Only the complex biodynamic data carry sufficient information to compare the WC model prediction with observation. It has already been shown in enzymology: non-steady-state (complex) kinetics is much more informative than steady-state kinetics (simple) for the elucidation of catalytic mechanisms [24]. Finally, complex biodynamics is worth studying because of its practical significance; none of the known industrial, in situ, and in vivo processes follow a simple growth pattern.

Important macromolecules disregarded by ME and WC models. The SCM splits conditionally expressed macromolecules into P and U clusters (Table 2), supporting the primary and secondary metabolisms responsible for the growth and stress resistance of cells, respectively [22]. These two clusters are revealed in a chemostat at a series of *D* (Figure 5A) or in specially designed titrated batch cultures [139]. The downregulated at high SGR proteins were erroneously called 'unnecessary' and even self-inhibitory [140] and then correctly identified as stress response constituents [141,142]. WC and ME models categorized them as unmodeled proteins to be an inert part of the cell mass [123]. It was an unfortunate omission, because the conditionally expressed 'in inverse order' (upregulated at a low SGR) macromolecules are crucially important from the system biology and evolutionary perspectives. The microbial growth rate alone is meaningless measure of the fitness without parallel characterization of stress resistance and self-defense against numerous hazardous factors.

Revisiting the optimality principle. The philosophy of life is to keep the balance between two general trends: (i) to reproduce themselves as fast as possible under favorable conditions and (ii) to minimize mortality under stress conditions. Cellular resources are limited; therefore, intensive growth and high stress resistance cannot be expressed simultaneously. GEMs aimed at the simulation of a wide spectrum of microbial phenotypes should radically revise the application of the optimality principle. LP in combination with a single objective function such as SGR maximization is appropriate only for nutrient-sufficient optimal conditions. Under restrictive conditions (substrate limitation, nonoptimal temperature, a gradient of inhibitory compounds, etc.), both growth-promoting and stress-resistance subsets of macromolecules should be expressed in proportions that depend on the restriction degree. The older strategy-seeking SGR maximizing under nonoptimal conditions is at risk to produce a degenerative phenotype, a microbial equivalent of immuno-compromised mammals.

Appendix A

Appendix A.1. Common Abbreviations

GEM	Genome-scale models
ODE	Ordinary differential equation
PDE	Partial differential equation
MMCC	Macro-Molecular Cell Composition
FBA	Flux Balance Analysis
COBRA	COnstraint-Based Reconstruction and Analysis
LP	linear programming
NLP	non-linear programming
dFBA	Dynamic Flux Balance Analysis
SOA	static optimization approaches for solving dFBA
DOA	dynamic optimization alternative
rFBA	Regulated Flux Balance Analysis
ME-models	GEM simulating metabolism and gene expression
WC models	whole cell GEM
VBNC	Viable but nonculturable cells
NZG	near-zero growth under chronic starvation
SCM	Synthetic Chemostat Model
SOD	superoxide dismutase
TF	transcription factor
PHB	polyhydroxybutyrate, cellular C-storage
ROS	reactive oxygen species
WT	wild type distinct from mutants

t	time
V	the volume of cell culture, kept constant in chemostat
F	the medium feeding rate
$D = F/V$	chemostat dilution rate
x	cell mass concentration
s	limiting nutrient substrate concentration
s_r	input substrate concentration in chemostat culture
p	product concentration
M	concentration of intracellular metabolites
σ	mass of element (N, P, S, etc.) per g of total cell mass
σ_0 and σ_m	resp. low and upper limits of σ-variation
$\mu = (1/x)\cdot dx/dt$	specific growth rate (SGR)
μ_m	the maximum SGR, $\mu \to \mu_m$ as $s \to \infty$
$Y = -dx/ds$	cell yield per unit of consumed substrate
$q_s = (1/x)\cdot ds/dt$	specific substrate uptake rate (SUR)
m	maintenance term, the catabolic rate $q_s \to m$, as $\mu \to 0$
Q_s	maximum SUR
$q_p = (1/x)\cdot dp/dt$	specific product formation rate (SPR)
Q_p	maximum SPR
K_s	saturation constant, at $s = K_s$, $q_s = 0.5\,Q_s$ or $\mu = 0.5\,\mu_m$
K_p	product inhibition constant, at $p = K_p$, $q_s = 0.5\,Q_s$
K_r	saturation constant for r-variable
r	SCM variable [0,1] reporting the status of MMCC

Appendix A.2. Glossary

Affinity to substrate	**Auto-selection, selection sweep**
Substrate binding strength of transmembrane transporters characterized by the parameter K_s. The lower K_s the higher affinity	A process observed in long-term continuous culture, a displacement of the parental population by a spontaneous mutant having higher SGR
Balanced growth	**Steady-state continuous growth**
A proportional increase in the content of all cell constituents (constancy of MMCC) observed in a steady state chemostat or exponential phase of batch culture.	A situation in which all state variables (cell mass, nutrients, and products concentration, MMCC) remain constant as the opposite processes (input–output and growth–washout) match each other
Quasi-steady state growth	**Transient processes**
Only fast variables are stabilized, other not	A movement from one steady state to another
Catabolic nutrient substrates	**Conserved or anabolic nutrient substrates**
Sources of energy; undergo irreversible oxidation to form nonutilizable waste products (H_2O and CO_2) coupled to ATP generation	Sources of biogenic elements (N, P, K, Fe, etc.) that are consumed reversibly (can leak out and re-consumed) and incorporated into cellular constituents
Kinetics	**Metabolic stoichiometry**
Scientific discipline studying the development of biological or physicochemical processes over time; combines experiments with mechanistic mathematical models.	A quantitative relationship between reactants and products in bioprocesses complying with the mass- and energy conservation conditions

Maintenance energy	Primary and secondary metabolism
Some fraction of available energy used for other than growth functions: osmoregulation; motility; and turnover of macromolecules (proteins, nucleic acids, and cell wall)	The primary metabolism produces new cell mass (growth) while the secondary metabolism supports enhanced stress resistance and other functions
Starvation	**Chronic starvation**
Self-degradation of cells (net SGR ≤ 0) caused by a complete absence of one or more essential nutrients	Deep substrate limitation in continuous culture causing a significant decrease of growth rate, SGR $\ll \mu_m$
Growth limitation and inhibition	**Single or multiple substrate limitation**
Growth condition decreasing SGR ($<\mu_m$) by low nutrients concentration (substrate limitation) or by toxic compounds (growth inhibition)	Respectively, one or several nutrients concurrently affect growth, including SGR, SUR, yield, and MMCC

Appendix B. Derivation and Explanation of Biokinetic Equations

Appendix B.1. Structured Models, the Internal and External State Variables

All GEMs belong to the category of structured growth models containing the internal variables, the intracellular quantities of metabolites, or cell constituents expressed in grams or moles per g DW (dry weight of total cell mass). There are also the external extracellular variables, such as concentrations of nutrients in growth media and exometabolites expressed as grams or moles per unit volume of bioreactor. Figure A1 presents the toy metabolic network with four internal (M_1–M_4) and three external variables (S, P_1, and P_2).
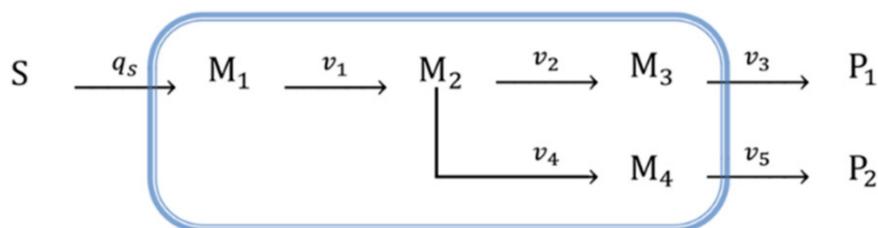

Figure A1. Simplified metabolic network of growing microbial cells. The limiting nutrient substrate S is transformed via intracellular intermediates M1–M4 into the extracellular products P1 and P2.

According to the FBA modeling style, the biomass formation is not shown explicitly on the flowchart; instead, it is specified as the following pseudo-reaction:

$$\theta_1 M_1 + \theta_2 M_2 + \theta_3 M_3 + \theta_4 M_4 \overset{growth}{\rightarrow} CH_h O_o N_n \tag{A1}$$
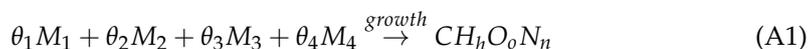
where stoichiometric coefficients θ_1–θ_4 indicate the contribution of each precursor to the biomass $CH_h O_o N_n$.

The extracellular variables only have the option to be expressed as mM or g/L of a culture volume, while the internal variables have two expression options: (i) g or mmol per liter of culture ($m_1, \dots m_4$) and (ii) g or mmol per g DW ($M_1, \dots M_4$). The variables m_i and M_i are interconverted via cell mass concentration (x, g DW/L): $m_i = x M_i$.

The starting point in developing the metabolic dynamic model is to write down the set of mass–balance ODEs for both internal and external variables expressed uniformly as mass per culture volume (e.g., g/L), the right side of each ODE containing the sum of the sources (positive terms) and sinks (negative terms):

Extracellular substrate:

$$\frac{ds}{dt} = -q_s \tag{A2}$$

Cell mass:

$$\frac{dx}{dt} = Yq_s x = \mu x \tag{A3}$$

Intracellular metabolite M1:

$$\frac{dm_1}{dt} = q_s x - v_1 x \tag{A4}$$

Intracellular metabolite M2:

$$\frac{dm_2}{dt} = v_1 x - v_2 x - v_4 x \tag{A5}$$

Intracellular metabolite M3:

$$\frac{dm_3}{dt} = v_2 x - v_3 x \tag{A6}$$

Intracellular metabolite M4:

$$\frac{dm_4}{dt} = v_4 x - v_5 x \tag{A7}$$

Secreted product P1:

$$\frac{dp_1}{dt} = v_3 x \tag{A8}$$

Secreted product P2:

$$\frac{dp_2}{dt} = v_5 x \tag{A9}$$

With the first intermediate M1 as an example, we demonstrate the conversion from derivatives dm_i/dt to dM_i/dt. First, we make the substitution $m_1 = xM_1$ and apply the product rule of differentiation:

$$\frac{dm_1}{dt} = \frac{d(xM_1)}{dt} = x\frac{dM_1}{dt} + M_1\frac{dx}{dt} = q_s x - v_1 x$$

Then, we divide both parts of the equation by x and rearrange it:

$$\frac{dM_1}{dt} = q_s - v_1 - M_1\frac{1}{x}\frac{dx}{dt} = q_s - v_1 - \mu M_1 \tag{A4a}$$

The rest of the internal variables were derived in the same way:
Intracellular metabolite M2:

$$\frac{dM_2}{dt} = v_1 - v_2 - v_4 - \mu M_2 \tag{A5a}$$

Intracellular metabolite M3:

$$\frac{dM_3}{dt} = v_2 - v_3 - \mu M_3 \tag{A6a}$$

Intracellular metabolite M4:

$$\frac{dM_4}{dt} = v_4 - v_5 - \mu M_4 \tag{A7a}$$

The negative term $-\mu M_i$ stands for dilution of the *i*th component due to cell growth and should not be confused with the washout term in the chemostat model of Equation (6) of the main text. The Equations (A4a)–(A7a) containing the normalized per g DW variables remain the same for any cultivation system.

Appendix B.2. The Effects of Substrate Concentration

Appendix B.2.1. Steady-State Concentrations of Metabolic Intermediates

The classic FBA has several limitations, one of them being the inability to predict the concentrations of metabolites but able to resolve the steady-state metabolic fluxes [36]. As applied to our toy example, the concentrations of four metabolites, $M_1, \dots M_4$, approach their respective steady-state values, $\overline{M_1}, \dots, \overline{M_4}$. At a steady state, the derivatives are set to zero, and the concentrations of the metabolites can be expressed through the known quantities of q_s, v_1–v_5, and μ:

$$\frac{dM_1}{dt} = 0, \therefore \overline{M_1} = \frac{q_s - v_1}{\mu} \tag{A4b}$$

$$\frac{dM_2}{dt} = 0, \therefore \overline{M_2} = \frac{v_1 - v_2 - v_4}{\mu} \tag{A5b}$$

$$\frac{dM_3}{dt} = 0, \therefore \overline{M_3} = \frac{v_2 - v_3}{\mu} \tag{A6b}$$

$$\frac{dM_4}{dt} = 0, \therefore \overline{M_4} = \frac{v_4 - v_5}{\mu} \tag{A7b}$$

However, these expressions cannot be used for finding the metabolic pool sizes. The reason is that all the fluxes on the right sides of Equations (A4b)–(A7b) contain $\overline{M_1}, \dots, \overline{M_4}$ in a hidden form, being dependent on the metabolite concentrations. Using two Equations (A4a) and (A5a) as an example:

$$\frac{dM_1}{dt} = q_s - v_1 - \mu M_1 = Q_s - \frac{k_1[E_1]M_1}{(K_{m1} + M_1)} - \mu M_1 = 0 \tag{A4c}$$

$$\frac{dM_2}{dt} = v_1 - v_2 - v_4 - \mu M_2 = A - \frac{k_2[E_2]M_2}{(K_{m2} + M_2)} - \frac{k_4[E_4]M_2}{(K_{m4} + M_2)} - \mu M_2 = 0;$$
$$A = \frac{k_1[E_1]\overline{M_1}}{(K_{m1} + \overline{M_1})} \tag{A5c}$$

Here, we presented the fluxes v_1, v_2, and v_4 in the Michaelis–Menten format, assuming the uptake flux $q_s = Q_s$ (unlimited batch growth). The steady-state flux v_1 depends on M_1 and, therefore, in the second Equation (A5c) (that does not contain M_1), remains constant at $v_1 = A$. The solutions were found manually or by symbolic computing as the real non-negative roots of the quadratic ($\overline{M_1}$) and cubic ($\overline{M_2}$) equations:

$$\overline{M_1} = \frac{Q_s - k_1[E_1] - \mu K_{m1} + \sqrt{(Q_s - k_1[E_1] - \mu K_{m1})^2 - 4Q_s\mu K_{m1}}}{2\mu} \tag{A10}$$

$$\overline{M_2} = \frac{\sqrt[3]{-2a^3 + \sqrt{(-2a^3 + 9ab\mu + 27c\mu^2)^2 + 4(3b\mu - a^2)^3} + 9ab\mu + 27c\mu^2}}{3\sqrt[3]{2}\mu} -$$
$$\frac{-\sqrt[3]{2}(3b\mu - a^2)}{3\mu\sqrt[3]{-2a^3 + \sqrt{(-2a^3 + 9ab\mu + 27c\mu^2)^2 + 4(3b\mu - a^2)^3} + 9ab\mu + 27c\mu^2}} - \frac{a}{3\mu} \tag{A11}$$

$$a = k_2[E_2] + k_4[E_4] + \mu(K_{m2} + K_{m4}) - A; \quad b = k_2[E_2]K_{m4} + \mu K_{m2}K_{m4} - A(K_{m2} + K_{m4}); \quad c = AK_{m2}K_{m4}$$

At very low or very high metabolite concentrations, the Michaelis–Menten equation can be reduced, respectively, to the first- and zero-order (see the Appendix B.2), and the steady-state solutions become simpler:

First-order approximation: $M_1 \ll K_{m1}, M_2 \ll K_{m2} \; K_{m4}$

$$\overline{M_1} = \frac{Q_s}{\mu + \frac{k_1[E_1]}{K_{m1}}}; \; \overline{M_2} = \frac{A}{\mu + \frac{k_2[E_2]}{K_{m2}} + \frac{k_4[E_4]}{K_{m4}}} \tag{A12}$$

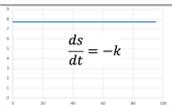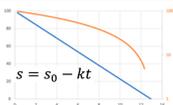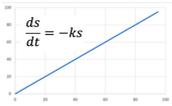Zero-order approximation: $M_1 \gg K_{m1}, M_2 \gg K_{m2} \; K_{m4}$
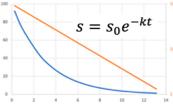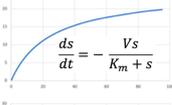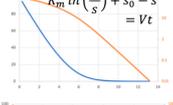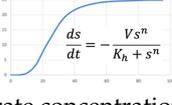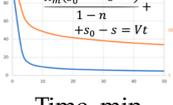
$$\overline{M_1} = \frac{Q_s - k_1[E_1]}{\mu}; \overline{M_2} = \frac{A - k_2[E_2] - k_4[E_4]}{\mu} \tag{A13}$$

Any solution, complete or simplified, contains kinetic parameters of the enzymes involved in the transformation process; thus, we confirmed the verdict [36] stating that the FBA is unable to predict the metabolite concentrations without independently obtained data on enzyme concentrations, their catalytic constants, and K_m. There are also two additional precautionary notes:

Currently, there is no accurate method for measuring the in vivo kinetic constants. The published in vitro data are available for only the well-studied model organisms like *E. coli.* Even for them, the in vitro enzymological data could be not a perfect representation of the in vivo kinetics. One complication comes from the possible reversibility of metabolic reactions. The second interfering factor is the in vivo/in vitro differences in the physico-chemical conditions, e.g., the molecular crowding effects, high viscosity, pH shift, presence of activators and inhibitors, etc. [143,144].

Another systematic error stems from the basic nature of FBA that excludes the biomass formation from metabolic stoichiometry using instead the standalone biomass pseudo-reaction A1. A hidden withdrawal of metabolites for biosynthesis underestimates their sink, resulting in an overestimation of the metabolic pools if using the Equations (A4)–(A7). The ME models resolve the problem but only partially, because the computed E-matrix covers only the proteins and the RNA (about half of the global cell mass); other constituents (glycogen, PHB, other storage components, cell wall, etc.) are not included.

Table A1. Basic equations used in chemical and enzyme kinetics.

Kinetic Order	Rate vs. Substrate Concentration	Reaction Progress over Time
Zero-order	$\frac{ds}{dt} = -k$	$s = s_0 - kt$
First-order	$\frac{ds}{dt} = -ks$	$s = s_0 e^{-kt}$
Second order	$\frac{ds}{dt} = -ks^2$	$s = \frac{s_0}{1 + ks_0 t}$
Michaelis-Menten equation	$\frac{ds}{dt} = -\frac{Vs}{K_m + s}$	$K_m \ln\left(\frac{s_0}{s}\right) + s_0 - s = Vt$
Hill-Langmuir equation	$\frac{ds}{dt} = -\frac{Vs^n}{K_h + s^n}$	$\frac{K_m(s_0^{1-n} - s^{1-n})}{1 - n} + s_0 - s = Vt$
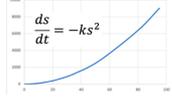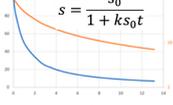
Rate, mmol per min — Substrate concentration, mM

Residual substrate, linear scale — Residual substrate, log scale — Time, min

Appendix B.2.2. Kinetic Order of Metabolic Reactions

The topic is covered in detail elsewhere [25]. Table A1 presents the five typical cases of chemical reaction kinetics (the zero, first, and second orders) and enzyme kinetics (the Michaelis–Menten and Hill–Langmuir equations).

All enzymatic reactions follow the mixed kinetic order. Specifically, the Michalis–Menten equation is reduced to the first order at low *s* and to the zero order at high *s*. The Hill–Langmuir equation is applied to enzymes containing several interacting subunits: when a substrate binds to one subunit, other subunits either increase the substrate affinity (positive cooperativity, $n > 1$) or decrease it (negative cooperativity, $n < 1$). Such a regulation is especially important for transcription and translation processes (an unfortunately overlooked feature in most published ME and WC models). The mixed kinetic order for the Hill–Langmuir equation is implemented as follows: at $s \ll K_h$, the reaction rate is $v = (V/K_h)s^n$; therefore, if $n \sim 2$, then it is close to the second-order kinetics; at $s \gg K_h$, reaction rate approaches the constant maximum level V faster than predicted by the Michaelis-Menten kinetics at $n > 1$ and slower at $n < 1$.

Before further discussion of the biokinetic issues related to the intracellular reactions, we should clarify what is the real concentration range of metabolic intermediates in microbial cytosol.

Appendix B.2.3. In Vivo Concentration of Metabolic Intermediates

Experimental metabolomic data are very limited even for model microorganisms. Below, we present a summary of a recently published *E. coli* study by using the state-of-the-art technique (LC-MS/MS) combined with fast metabolism quenching and applying an extensive set of 13C internal standards [145,146]. Three minimal media were used with glucose, glycerol, or acetate as the only C-source. The number of detected intracellular metabolites varied from 68 (acetate) to 103 (glucose) of the overall cytosolic concentration, 129–264 mM, that we split into three clusters (Table A2).

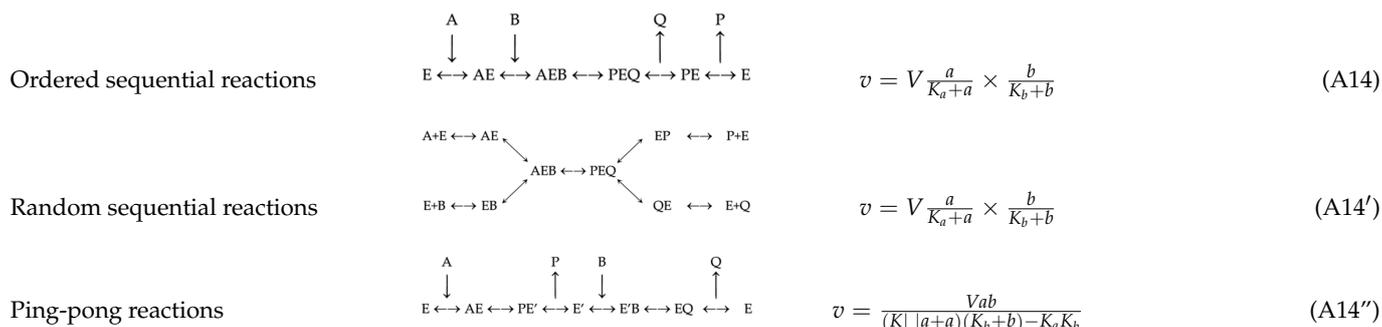**Table A2.** Summary of the metabolomic analysis of *E. coli* grown on glucose, based on Reference [145].

Cluster	Pool Size Range	Contribution to Metabolome		Representatives, Values in Brackets Are Concentrations (mM) for the Most Abundant Compounds
		Mass %	Compounds Number	
1	10–100 mM	59	4	Glutamate (96), glutathione (17), FDP (15), ATP (10)
2	0.1–10 mM	40	57	NTP, NDP, NAD⁺, FAD, 13 AA, glycolysis, TCA, and pentose pathway intermediates
3	0.1–100 μM	<1	42	5 AA, dNTP, FMN, NADP, and other metabolites

The full genomic inventory provides a much longer list of metabolic intermediates, up to 1210 entries for *E. coli* K-12 MG1655 in the Metabolome Database and up to 1192 intermediates in the most recent metabolic reconstructions [95]. Therefore, the cited experimental study deals with no more than a core metabolome. Based on the molar mass and the cytosol-to-cell mass conversion factor used in the original publication [146], we estimated the cumulative contribution of the detected metabolites to the total cell mass as 0.09–0.15 g/g DW. It is already either close to the top or even exceeds the 10% quota allocated for the sum of monomeric compounds in *E. coli* cells [147,148]. In other words, the missed genome-predicted metabolites should form the fourth cluster containing more than 1000 compounds but at extremely low concentrations. Using the lowest detected quantity (adenosine, ~100 nM) as the upper boundary for the missed metabolites, the expected overall contribution of the fourth cluster to the global metabolome mass is negligible, 0.04–0.08%. To summarize, the global *E. coli* metabolome consists of 50–60 major metabolites (clusters 1 and 2) that account for more than 99% of the metabolome mass. They are mostly common cofactors (NAD, NTP, and FAD); amino acids; and intermediates of the central carbon metabolism. The most abundant intermediates also serve other physiological functions, e.g., as a counterion to cytosolic K^+ and the nitrogen donor (glutamate) and the antioxidants (glutathione).

Appendix B.2.4. Substrate Saturation (SS) of Enzymes Catalyzing Polysubstrate Reactions

The experimentally established concentrations of metabolites were compared with the published K_m data for enzymes responsible for their transformations [145]. The degree of substrate saturation was calculated as the ratio SS = $M/(K_m + M)$. Figure A2 (blue curve, panel A) indicates near-complete saturation (SS range 0.9–1.0) for 54% of the tested enzyme–metabolite pairs. About 80% of all the enzymes are above the state of half-saturation ($M > K_m$). The low SS was found mostly for the third cluster of metabolites involved in the degradation of free nucleotides, nucleosides, and amino acids. Based on the undertaken survey, the authors conclude that *E. coli* enzymes 'are reliably saturated' by intracellular metabolites, implying that most metabolic reactions in the bacterial cell are not limited by the availability of metabolites present in the cytosol.

We need to revise this conclusion, because most intracellular reactions involve several reactants and, therefore, do not obey the simple Michaelis–Menten equation. Even without the abundant background reactants, such as the water in hydrolytic reactions or CO_2/HCO_3^- in carboxylation reactions, as many as 82% of the listed 395 reactions [145] involve two and more nontrivial substrates. Often, one out of two or two out of three reactants are present at saturating concentrations, while one deficient substrate controls the overall reaction rate playing the bottleneck role. There are three types of polysubstrate reactions illustrated below for the bisubstrate reactions [24]:

Ordered sequential reactions

$$\text{E} \longleftrightarrow \text{AE} \longleftrightarrow \text{AEB} \longleftrightarrow \text{PEQ} \longleftrightarrow \text{PE} \longleftrightarrow \text{E}$$

$$v = V \frac{a}{K_a + a} \times \frac{b}{K_b + b} \tag{A14}$$

Random sequential reactions

$$v = V \frac{a}{K_a + a} \times \frac{b}{K_b + b} \tag{A14$'$}$$

Ping-pong reactions

$$\text{E} \longleftrightarrow \text{AE} \longleftrightarrow \text{PE}' \longleftrightarrow \text{E}' \longleftrightarrow \text{E}'\text{B} \longleftrightarrow \text{EQ} \longleftrightarrow \text{E}$$

$$v = \frac{Vab}{(K| |a+a)(K_b + b) - K_a K_b} \tag{A14$''$}$$

Here, a and b are the molar concentrations of two reactants, K_a and K_b are the respective Michaelis constants, and V is the maximum reaction rate. The first two mechanisms imply the identical rate expressions (A14) and (A14$'$). Equation (A14$''$) differs by the small negative term $-K_a K_b$ in the denominator, predicting a slightly higher rate for the ping-pong mechanism, the difference being negligible. Therefore, Equation (A14) can be applied to all the polysubstrate reactions as follows:

$$v = V \frac{M_1}{K_1 + M_1} \times \frac{M_2}{K_2 + M_2} \cdots \times \frac{M_n}{K_n + M_n}, 2 \leq n \leq 6 \tag{A15}$$

As many as 21 out of 395 reactions were found to use metabolites missed from the metabolome analysis. For example, the acetaldehyde dehydrogenase EC 1.2.1.10 recognized [145] as a 'reliably saturated' enzyme catalyzes the following trimolecular reaction:

acetaldehyde + CoA + NAD+ → acetyl-CoA + NADH + H+
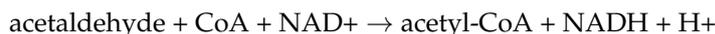
However, acetaldehyde was not detected. Assuming its maximum possible concentration of 100 nM with the published K_m 10 mM (Brenda), we find from Equation (A15) that, even at saturating concentrations of CoA and NAD$^+$, the rate of this reaction is extremely low and cannot exceed $v = 10^{-8} \times V$.
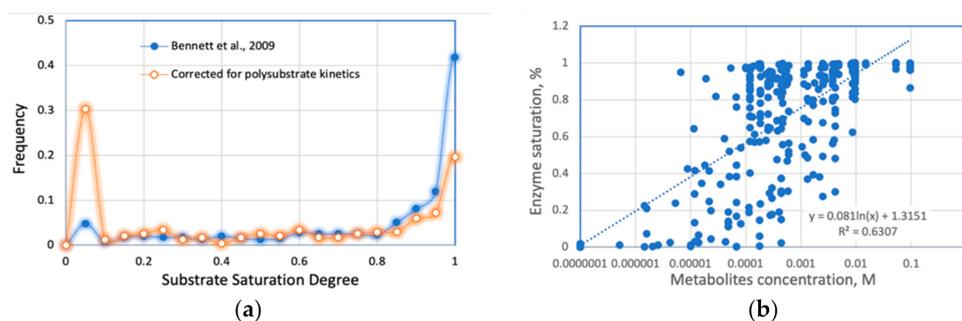
Figure A2. Experimental data on metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli.* Replotted from the supplemental data of Reference [145]. (**a**) Frequency distribution of enzyme substrate saturation before (blue) and after correction (orange) according to Equation (A15). (**b**) The plot of substrate saturation vs. log-scale metabolite concentrations.

The described polysubstrate correction dramatically affected the SS frequency distribution (Figure A2, panel a, the orange curve): the distribution peak moved to the lowest SS range, 0–0.05, the fraction of fully saturated enzymes with SS 0.9–1.0 dropped from 54 to 27%, and the fraction of SS 0.5–1.0 declined from 81 to 51%.

Appendix B.2.5. Simplified Kinetic Order of Metabolic Reactions

The Michaelis–Menten equation can be simplified to avoid cumbersome analytical solutions (compare (A10)–(A11) with (A12)–(A13)). The price paid is the restricted application scope of the simplified models. By switching from the mixed to the zero or the first kinetic order, we assume, respectively, the condition of either $M_i \gg K_m$ or $M_i \ll K_m$ to be applied to a selected reactions subset. If we want to keep the simulation error below 10%, then the first-order rate equations are applicable to SS < 0.1 and the zero-order equations to SS > 0.9. Then, as applied to the core metabolome, the first-order approximation can be justified for 75 out of 234 reactions (32%) and the zero-order approximation for 64 reactions (27%). The remaining 41% of the reactions are nearly half-saturated (0.1 < SS < 0.9), implying that neither approximation is appropriate: actual rates are to be overestimated by the first-order and underestimated by the zero-order equations.

What about the whole M-matrix reconstruction with its +1000 metabolites? Most genome-predicted metabolic intermediates are present in the cytosol at undetectable concentrations below 10^{-7} M. There is a clear semi-log trend of SS vs. M (Figure 2B), while K_m values do not correlate with M ($R^2 \sim 0$, data not shown). Thus, it would be safe to assume that undetectable metabolites provide in vivo a very low SS degree compatible with the first-order kinetics. Then, out of 2719 metabolic reactions accounted for by the iML1515 [95], high SS > 0.9 can be attributed to no more than 2% of the reactions. The half-saturated reactions contribute about 3%, and the other 95% of the cellular metabolic reactions should follow the first-order kinetics. This estimate was done for the nutrient-sufficient growth conditions, the nutrient-limitation is expected to decrease the concentration of all metabolites (see, as an example, the ATP data [40]). Under deep and even moderate substrate limitations, nearly 100% of metabolic reactions should obey the first kinetic order. The only rare exception could be a few reactions involving only the first metabolic cluster (Table A2) without the participation of other clusters as co-substrates. It would be reasonable to present these relatively rare reactions in GEMs in the mixed kinetic format without simplification.

Note 1. The predominance of the first-order kinetics is referred to as the number of reactions, not the fluxes. The rare mixed-order reactions engage the most abundant metabolite and, hence, are expected to make a disproportionally higher contribution to the mass and energy flow.

Note 2. Even under deep substrate limitations, the first-order simplification is justified for only the steady-state regimes. Transiently, a perturbed microbial population can considerably depart from the first-order kinetics.

Note 3. Ironically, the published ME models [49,122–124,149] do not use any kinetic expressions (Table A1) containing the concentration terms *s* or *M*. Instead, they rely on a binary search and constraints, such as defining the top and low limits of *s* while applying objective functions. Three disadvantages are obvious: (i) the subjectivity of constraints, (ii) a longer computation time, and (iii) missing the opportunity to reproduce a fine metabolic control in growing cells, an instant adjustment of enzymatic activity to the available reactant concentration in the cytosol. Two important system qualities of the mixed kinetic order are well-known: (i) stabilizing negative feedback to perturbation (an increase of *M* boosts its uptake and vice versa; both mechanisms accelerate the restoration of the steady state) and (ii) the asymptotic saturation kinetic that never exceeds a maximum. The first-order kinetics still comply with the first quality, but the zero-order processes are completely unregulated. By the way, the WC models roughly of the same level of complexity as ME models employ the mixed-order Michaelis kinetics for all metabolic reactions, and it did not take a heavy toll on the computing time [50].

Appendix B.3. The Effect of Enzymatic Proteins Concentrations

Extensive proteomic studies made it possible to quantify individual enzymatic proteins and then use these data for the verification of the ME and WC models and testing alternative hypotheses of system biology. This section covers several technical issues of incorporating protein concentrations into kinetic equations.

Appendix B.3.1. The Simplest Case of Pure Enzyme

The isolated and purified enzymes in vitro display proportionality between the reaction rate, *v*, and the molar concentration of the enzyme active sites, [*E*]:

$$v = k_{cat}[E]F(s, T, pH, \ldots) \tag{A16}$$

where $F(s, T, pH,$ etc.) is the functional relationship between the enzyme activity and the variable environmental factors.

The situation is more complicated in vivo. Below, we explore the following effects: (i) variable status of conditionally expressed proteome, (ii) deficiency of rare proteins presents an amount less than one copy per cell, and (iii) the opposite extreme of a very high protein concentration leading to macromolecular crowding.

Appendix B.3.2. Conditional Expression of Proteins with Glucose Transporters as Example

A typical kinetic model in classic enzymology is designed to predict the effect of various factors on the activity of selected enzymes taken at specified concentrations. The genome-scale reconstructions are charged with the same task but under much more challenging conditions; we do not know what enzymes are engaged and what are their concentrations. GEMs should predict both unknowns, using as the input (i) the specified environmental conditions and (ii) the annotated genome of a selected organism with a long list of potentially expressible gene products. Below, we show several approaches to solve the problem using, as an example, the process of nutrient uptake. There are many genes/proteins associated with nutrient consumption (porins, transporters, and binding proteins), accounting for up to 13–15% of the global genome and proteome of microorganisms [150]. For simplicity, we restrict our scope to the model organisms, *E. coli* K-12 and *Saccharomyces cerevisiae*, and select only one nutrient substrate, glucose.

E. coli. Bacteria express low-affinity uptake system under nutrient excess in batch culture and a high-affinity system under C-limitation in chemostat or depleted batch culture [151]; the switch between the two systems is controlled by cAMP/Crp at the transcriptional level [68]. The low-affinity system contains two components: (i) the porins OmpF and OmpC providing passive substrate diffusion through the outer membrane and (ii) the PTS (PEP-carbohydrate phosphotransferase system containing four enzymes: EI, EIIA, EIIB, and EIIC, and phospho-carrier protein HPr) transferring glucose through the

inner membrane with simultaneous phosphorylation. The affinity varies (K_m 10–1500 µM) because of the loose specificity of the EII component for structurally related sugars. The high-affinity system also contains two components: the LamB porin that, contrary to OmpF/C, has a sugar-binding site and follows Michaelis–Menten kinetics, K_m 64 µM, [152]), and the ABC (ATP-Binding Cassette) transporter Mgl containing the periplasmic binding protein with a dissociation constant of K_d 0.2 µM [151]. The process of glucose uptake by *E. coli* is completed in two steps: (i) glucose passing through the outer membrane to the periplasm and (ii) crossing the inner membrane. Respectively, there are two transfer processes, each one including low- and high-affinity systems:

Step 1:

$$v_1 = k_1[E_1]s + k_2[E_2]\frac{s}{K_{m2} + s} \tag{A17}$$

Step 2:

$$v_2 = k_3[E_3]\frac{S}{K_{m3} + S} + k_4[E_4]\frac{S}{K_{m4} + S} \tag{A18}$$

The variable *S* (the upper-case symbol in Equation (A18)) is the concentration of glucose in the periplasm distinct from *s* (the low-case symbol), the glucose extracellular concentration. Other symbols are common to previous equations. The Step 1 low-affinity process follows the first-order diffusion kinetics with the rate constant k_1, mM^{-1} h^{-1}; the other rates constants have dimension (time)$^{-1}$.

Saccharomyces cerevisiae. Yeasts have seven conditionally expressed hexose transporters, Hxt1–Hxt7, that cover the K_m range from 1 to 110 mM (Figure A3). All seven transporters follow the same mechanism of facilitated diffusion. Other members of the Hxt family participate in the regulation of the process as glucose sensors and transcriptional repressors [153]. The rate of glucose uptake depends on the concentrations $[E_i]$ of all seven conditionally expressed transporters and their kinetic characteristics:

$$v = \sum_{i=1}^{n} k_i[E_i]\frac{s}{K_{mi} + s}_1, n = 7 \tag{A19}$$
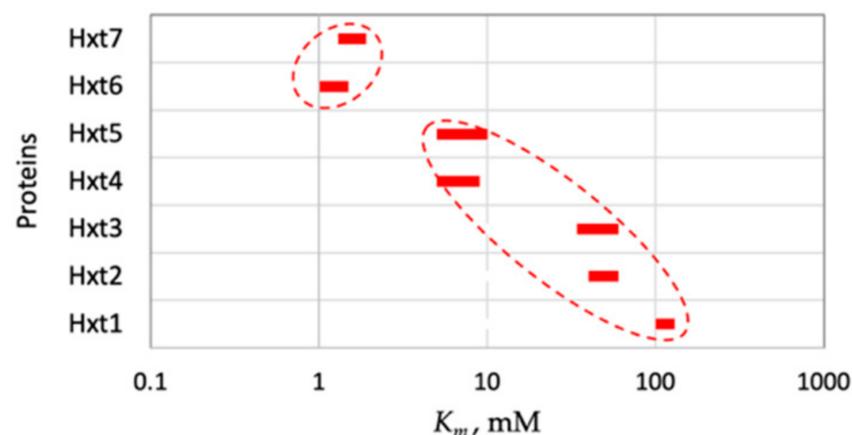

Figure A3. Conditionally expressed hexose transporters of *S. cerevisiae*, based on Reference [153]. Encircled clusters are assumed to be the high- (**left**) and low-affinity transporters (**right**).

There are several ways to simulate the process of differential gene expressions that determine the concentrations of the enzymatic proteins E$_1$–E$_4$ (*E. coli*) and E$_1$–E$_7$ (*S. cerevisiae*).

Approach 1 uses the optimality principle developed for FBA and more elaborated GEMs [36,131,154]. As applied to the small part of the entire M-matrix, which is just one exchange reaction of glucose uptake, the most common global objective function of the SGR maximization is equivalent to the maximization of the glucose uptake as dependent on its extracellular concentration. The high-affinity system (low K_m and low k_{cat}) is beneficial under substrate deficiency but becomes a burden at a high glucose concentration; therefore,

at each *s*-level, there should be a unique combination of the expressed transporters that provide the global maximum of the glucose uptake. The penalty for the energy (NTP) expenses, as well as the low and upper bounds, can be added to further refine the optimal solution, but we skip these options.

Approach 2 has been applied to several rFBA and iFBA models [94,116,117] by using Boolean algebra in which gene products are either available (ON) or unavailable (OFF) to the cell. As applied to our reduced uptake model (there are no variables responsible for regulatory metabolites like ppGpp or cAMP), the conditions can be set up for turning ON or OFF specific transporters as dependent directly on the extracellular glucose concentration; then, the Equations (A17) and (A18) should be appended with the following conditions:

$$[E_1] = \begin{cases} [E_1^{max}] & if\, s \geq s^* \\ [E_1^{min}] & if\, s < s^* \end{cases}, \ldots, [E_4] = \begin{cases} [E_4^{min}] & if\, S \geq S^* \\ [E_4^{max}] & if\, S < S^* \end{cases} \tag{A20}$$

where s^* and S^* are the threshold glucose concentrations outside the cells and in the periplasm, respectively, triggering the expression switch from one transporter to another. The minimal expression $[E_i^{min}]$ can be set to zero for simplicity.

Approach 3, the SCM-FBA hybrid model (see the main text and Section 4.2 'SCM' and Section 7 'Conclusions'). Firstly, the conditionally expressed proteins are attributed either to the P- or U clusters. In the glucose uptake sub-model, the P-cluster is unequivocally identified as the low-affinity/high-intensity transporters and the U cluster as the high-affinity transporters. The whole proteome differentiation into the P and U slices was more challenging. The complete empirical solution requires the global proteomic analyses of cells grown in a chemostat at a series of *D* (Figure 5); the proteins displaying the positive correlation with SGR ($R^2 > 0.5$) form the P cluster, the negatively correlating proteins ($R^2 < -0.5$) are assigned to the U cluster, and the rest can be categorized as either stochastic noise (low abundance proteins) or the constitutively expressed proteins. Such data are already available for the model organisms. Potentially, the COG classification system (cluster of orthologues groups [155]) can be updated to perform P/U differentiation directly from annotated genomes without the obligatory requirement of the chemostat data. Secondly, the upper and low boundaries for transporters are to be found also from proteomic data (Figure 5) or other published experimental results obtained at a series of SGR. For the glucose uptake sub-model, the most essential information is K_m, and the maximum glucose uptake rate, $Q = k_{cat}[E]$, is proportional to the enzyme concentration (Figure 1). The upper limits of the individual P and U proteins are to be established by extrapolation to the μ_m and $\mu = 0$, respectively. The low boundary for the fully inducible proteins is zero or found by extrapolation to $\mu = 0$ (P-cluster) or μ_m (U cluster). Thirdly, the rate Equations (A17) and (A18) are presented in the condensed SCM format as follows:

E. coli, step 1:

$$v = rQ_1 s + (1 - r)Q_2 \frac{s}{K_{m2} + s} \tag{A21}$$

E. coli, step 2:

$$v = rQ_3 \frac{S}{K_{m3} + S} + (1 - r)Q_4 \frac{S}{K_{m4} + S} \tag{A22}$$

S. cerevisiae:

$$v = rQ_1 \frac{s}{K_{m1} + s} + (1 - r)Q_2 \frac{s}{K_{m2} + s} \tag{A23}$$

Global *r*-variable:

$$\frac{dr}{dt} = \mu \left(\frac{s}{K_r + s} - r \right) \tag{A24}$$

Here, Q_1–Q_4 are the maximum glucose uptake rates via respective transporters. Every $Q_i = k_i [E_i^{max}]$, where E_i^{max} is the maximum attainable content of the *i*th transporter, then $rQ_i = k_i[E_i]$, and $r = \frac{[E_i]}{[E_i^{max}]}$.

The *S. cerevisiae* transporters were assigned as follows: the U cluster was formed by transporters Hxt6 and Hxt7 as having the lowest K_m and being glucose-suppressed [153]; other transporters form the P cluster. Therefore, $rQ_1 = k_1[E_1] + k_2[E_2] + k_3[E_3] + k_4[E_4] + k_5[E_5]$ and $(1-r)Q_2 = k_6[E_6] + k_7[E_7]$.

The last Equation (A24) is applied not only to the nutrient's consumption sub-model but to the entire metabolic reconstruction (not shown). The *r*-variable is a reporter of the proteome status; it is defined as a common to all P protein (including the low-affinity transporters) degrees of expression. The high-affinity transporters and other U proteins are set to be proportional to the difference $1 - r$.

Comparison of the three approaches (Figure A4). For clarity of illustration, we simulated glucose uptake by *E. coli* as a single step of a substrate transfer through the internal membrane: $v = Q_1 \frac{s}{K_1+s} + Q_2 \frac{s}{K_2+s}$. Bacterial growth in the glucose-limited chemostat was reconstructed in the range of s from 10^{-3} to 10^3 mg/L, corresponding to the *D* range from the near-zero to the washout point. The first approach (uptake maximization) produced a square profile of the abrupt transition from high- to low-affinity transporters at a residual glucose concentration $s \geq s^* = (Q_1 K_2 - Q_2 K_1)/(Q_1 - Q_2)$. The second approach brings the identical result if we select the same threshold, and this is not surprising for such a narrow solution space. The SCM–FBA hybrid turned out to generate a more realistic expression profile of 11 individual proteins representing low- and high-affinity systems of *E. coli* [156] (Figure A4, panel C).
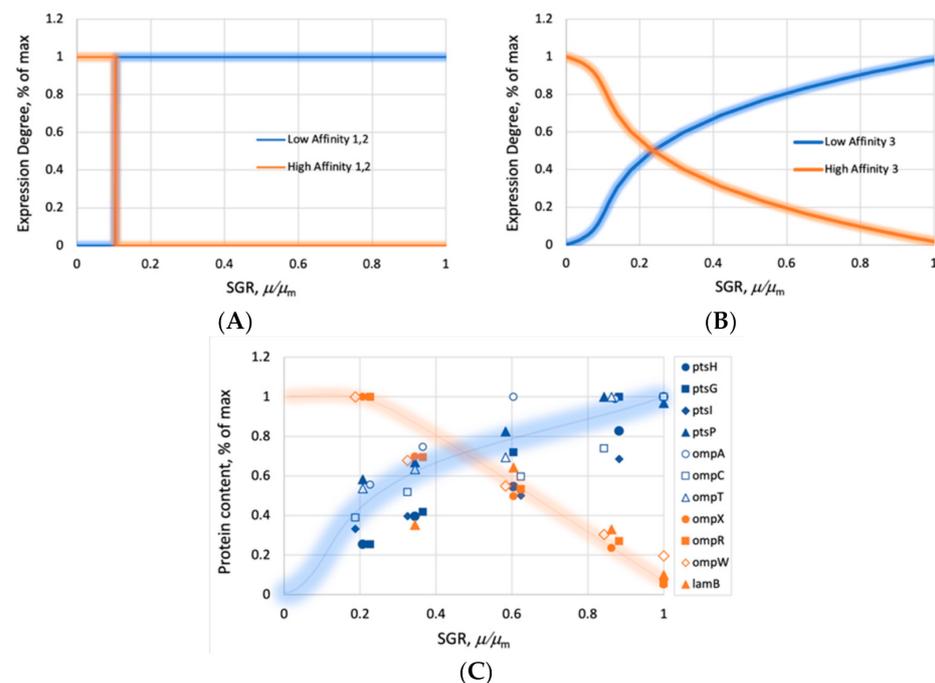

Figure A4. Simulation of conditionally expressed glucose transporters using three methods. (**A**) The optimality and Boolean approaches (results turned out identical). (**B**) Expression predicted by the SCM/FBA hybrid model. (**C**) *E. coli* proteomic data [156]. Blue and orange curves and symbols are applied, respectively, to the low and high affinity transporters.

Appendix B.3.3. Technical Details of Using the SCM

Solving direct and inverse problems. In the proposed SCM–FBA hybrid model, the SCM is solved independently of the FBA. The inverse problem stands for the task of finding the model's parameters that reproduce a given set of experimental data, such as a time series of *x*, *s*, and *r* or the steady-state values of these variables. There are diverse computational approaches to minimize the simulation errors; the topic has been covered specifically for the SCM in Reference [22]. The direct (forward) problem has the opposite purpose, to compute the state variables *x*, *s*, and *r* for a given set of ODEs with specified

model coefficients and initial conditions. The transient SCM dynamics require numerical integration, preferentially by using stiff-resistant algorithms (e.g., the ode15s solver in MATLAB), because the model includes fast (s) and slow variables (x and r), and there is a potential danger of too-small integration steps over extended time intervals [22]. The steady-state SCM solution ($\bar{x}, \bar{s}, \bar{p}$, and $\bar{r}$) for the chemostat cannot be resolved explicitly for D because of multiple nonlinearities. However, there is a simple implicit solution:

$$\text{Input } \bar{s}, \implies \bar{r} = \frac{\bar{s}}{K_r + \bar{s}} \implies q_s = \bar{r}Q\frac{\bar{s}}{K_s + \bar{s}} \implies \mu = D = Yq_s - a_o\bar{r} \implies \bar{x} = D\frac{s_r - \bar{s}}{q_s} \tag{A25}$$

Selection of experimental data. Experimental verification of the SCM requires chemostat experiments combined with proteomic or transcriptomic analyses at several D; if these data are not available, then the growth-associated changes in the RNA content (total or rRNA) would be the best proxy for the r-variable. Another key variable is the limiting substrate concentration. In theory, it plays the most essential regulatory role, but unfortunately, it is notoriously known as a very problematic category of chemostat data [14,22,157,158]: below detection limit at low and intermediate D and too-high turnover rate at a standard cell density of ~1 g/L (OD_{600}~1.0). The accurate recording of the s-variable for microorganisms with high-affinity transporters requires specialized sampling devices (tangential flow filtration or liquid nitrogen trap) applied to a continuous culture with as low as possible the input concentration of limiting nutrient, e.g., 50–100-mg glucose/L for *E. coli* [158]. If these conditions are not met, then direct analytical data are of limited value and a better solution could be an indirect estimation of s based on the other accurately recorded variables functionally dependent on s, e.g., the instrumentally recorded DO uptake rate, CO_2 production, pH titration rate, etc. Figure A5 illustrates such an indirect solution based on the measurement of q_s (glycerol uptake rate) and RNA, the missed s data are treated as unknowns and found together with the model's parameters by the standard inverse problem procedure.

Customizable biomass formula and reaction. An adjustable biomass reaction has been used once [88] for one specific case of *E. coli* batch growth on glucose (generation time 70 min) or acetate (140 min). Two respective cell compositions were reproduced using the linear regression of published chemostat data for the content of the RNA, DNA, glycogen, and total proteins. The SCM offers an easier and instant customization procedure implementing the r-variable (Figure A6). Two of the most abundant conditionally expressed constituents of *E. coli* are stable RNA (rRNA and tRNA) and glycogen, the C-storage polymers. According to the SCM, the RNA content is proportional to r, and the content of the C-storage is proportional to the difference $(1-r)$. The elemental composition of the RNA (AGCU) and glycogen are, respectively, $(C_{38}H_{45}O_{26}N_{15}P_3)_n$ and $(C_6H_{10}O_5)_n$ or $CH_{1.18}O_{0.68}N_{0.39}P_{0.08}$ and $CH_{1.67}O_{0.83}$ if expressed per carbon atom. The empirical biomass formula of *E. coli* cells at the maximum SGR ($r = 1.0$, the RNA content 0.2) is $CH_{1.77}O_{0.49}N_{0.24}P_{0.03}$. At another extreme, under chronic starvation when $s \to 0$ and $r \to 0$, the conditionally expressed RNA is completely substituted with glycerol that attains its maximum content of 0.2. Based on difference in elemental composition of RNA and glycogen, the empirical formula of starving cells becomes $CH_{1.866}O_{0.52}N_{0.16}P_{0.008}$. Between these two extremes, the elemental composition of cells is linearly related to the r-variable. At each steady state in a chemostat culture or at every integration step simulating transient growth, the SCM generates unique r-value that can be instantly recalculated into the biomass formula to be used by the FBA model for computing the fluxome. The coefficients of the biomass reaction (Equation (13)) are corrected in a similar way as they display a linear relationship with r-variable.
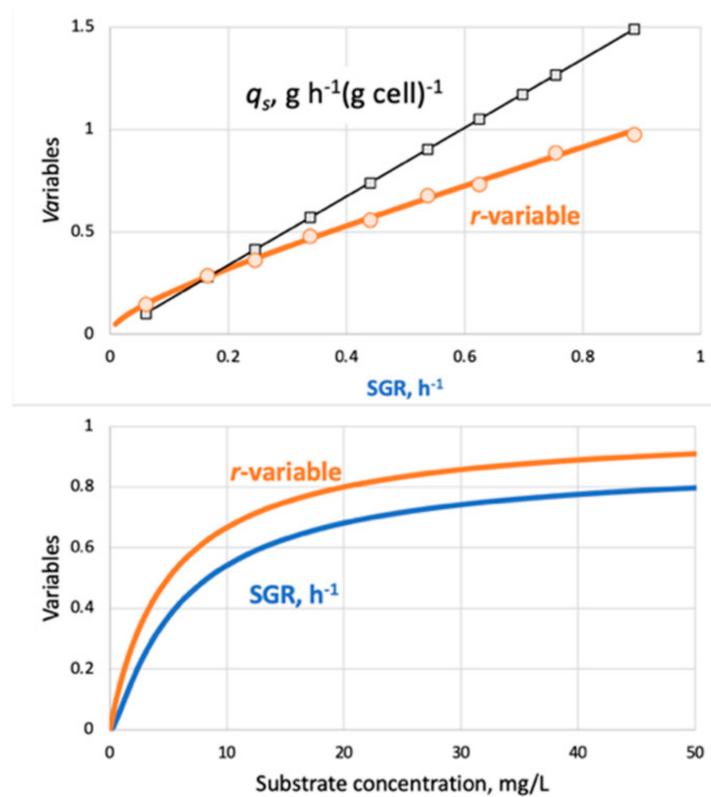
Figure A5. Reconstruction of the missed data on limiting the substrate concentration. **Top**: Glycerol-limited chemostat, SUR (q_s), and cell RNA (proxy for *r*-variable) are plotted vs. $\mu = D$. **Bottom**: Best-fit reconstruction of the SGR and RNA plots vs. s. The curves were calculated from Equation (A25), minimizing the residual squares with MS Solver.
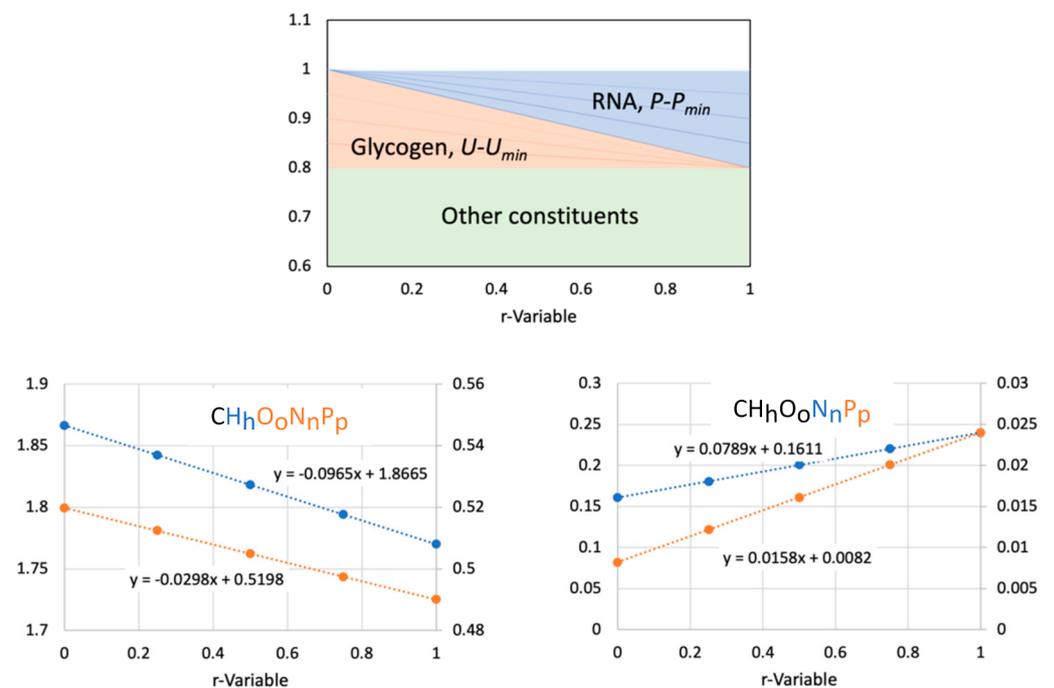

Figure A6. Adjustment of the cell biomass formula. (**Top**) RNA and glycogen, two major changeable cell components. (**Bottom**) Empirical biomass formula with H and O subscripts linearly decreasing with *r* (**left**) and the N and P contents increasing with *r* (**right**).

Appendix B.3.4. Extremely Low Proteins Concentration: Down to One Enzyme Molecule Per Cell

The WC modeling of *E. coli* [50] predicted that as many as 72 essential genes of *E. coli* were transcribed less than once per cell cycle. This result is compatible with proteomic analyses of these bacteria [156] consistently indicating the presence of extremely rare proteins, e.g., one or less copy per cell for 294 out of 2359 identified proteins (12.5%) at least at one tested cultivation condition (complex and minimal media, a series of *D* in chemostat culture, and starvation). These rare proteins might be downregulated under the given growth conditions; however, some of the proteins (2.5%) remained at a low level under any growth conditions.

How might cells function and grow when some of their essential genetic content is not transcribed during a typical division cycle? How might a single enzyme molecule perform their destined function inside a regular bacterial cell? Let us picture a rare enzymatic protein (assuming its size 50 kDa, the middle distribution for all bacterial proteins) inside the average *E. coli* cell of a volume of 0.7 μm^3. If we zoom in on both objects proportionally, then our rare enzyme would be like one human in a 10-million giant city like London or Paris. The probability of a random colliding into a respective substrate in the cytosol is not zero, but it is certainly an extremely rare stochastic event with a minor impact on the whole-cell metabolism.

What is the origin of rare proteins? First, it can be just an analytical error, resulting in the underestimation of the actual number of enzyme copies. Second, rare proteins could be expressed in the past under the prior environmental conditions; it takes some time to degrade and dilute them with the newly synthesized molecules. Third, what we found as 'rare proteins' could be expressed only in a small subpopulation of heterogeneous microbial culture. In other words, rare proteins might belong to some spontaneous mutants having a distinct proteome profile as compared with the WT and coexisting with the WT.

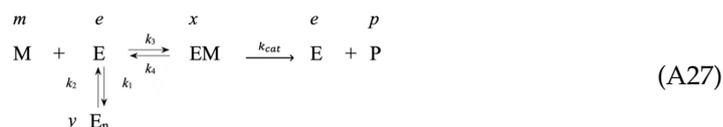Appendix B.3.5. Extremely High Proteins Concentration: Macromolecular Crowding

The cytosol is called to be 'crowded' rather than 'concentrated', because no single macromolecular species occurs at a high concentration, but taken together, it makes up 300–400 g/L, and the macromolecules occupy, by different estimates, from 20–30% [143,144] to 34–44% of the total cellular volume [159]. The portion of excluded volume depends on the growth conditions and steeply increases with the SGR. The total number of protein molecules is more or less constant over the whole range of the SGR; some reports indicate a minor decline [21] or increase [156] of the total proteins content with the SGR. However, the sizes of the conditionally expressed macromolecules are dramatically different in slowly and rapidly growing cells. Growth acceleration requires a higher number of ribosomes and enzymes associated with translation that take up much more intracellular space. A single ribosome of *E. coli* is about 20 nm in diameter and a volume of 3400 nm^3 [160]; cells (average volume 0.7 μm^3) contain 30,000–50,000 ribosomes that alone occupy 15–25% of the cellular space. In addition, there are other large and abundant proteins supporting translations, such as elongation factor Tu (EF-Tu) forming a ternary complex with tRNA and tRNA synthetases; all these tRNA-affiliated protein masses are about two-thirds that of the ribosomal protein mass in moderate-to-fast growth [161]. The effect of molecular crowding is the slowing down of all metabolic rates because of increased viscosity and restricted diffusion, as well as changes in the biochemical equilibria to favor the association of macromolecules into aggregates [159,161–164]. The effect of limited diffusion has been accounted for by using the empirical equations incorporated into the FBA and other GEMs [159]. The impact of aggregation remains unexplored; to fill in the gap, we outline below one simple kinetic mechanism.

Let us view the aggregation as a reversible process involving the binding of *n* individual enzyme molecules E to each other with the formation of aggregate E_n having no

enzymatic activity. The dissociation of E_n into monomeric units restores the activity, e.g., via chaperone-assisted refolding:

$$\underbrace{E + E + \ldots + E}_{n} \quad \overset{k_1}{\underset{k_2}{\rightleftharpoons}} \quad E_n \tag{A26}$$
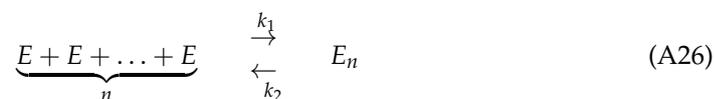
Combining Aggregation (26) with the Michaelis–Menten process gives the following reaction scheme:

$$\begin{array}{ccccccc} m & & e & & x & & e & & p \\ M & + & E & \overset{k_3}{\underset{k_4}{\rightleftharpoons}} & EM & \overset{k_{cat}}{\longrightarrow} & E & + & P \\ & & {\scriptstyle k_2}\updownarrow{\scriptstyle k_1} & & & & & & \\ & & y\;E_n & & & & & & \end{array} \tag{A27}$$

The process is described by four ODEs and one conservation condition:
The initial reaction rate:

$$\frac{dp}{dt} = \frac{-dm}{dt} = k_{cat}x \tag{A28}$$

Enzyme–substrate complex:

$$\frac{dx}{dt} = k_3 me - k_4 x - k_{cat}x \tag{A29}$$

Inactive aggregates:

$$\frac{dy}{dt} = k_1 e - k_2 y^n \tag{A30}$$

Free active enzymes:

$$\frac{de}{dt} = k_4 x - k_3 me + k_2 y^n - k_1 e + k_{cat}x \tag{A31}$$

E balance:

$$e_0 = e + x + ny \tag{A32}$$

Assuming a steady state for the intermediary complexes, $\bar{x} = \frac{k_3 me}{k_4 + k_{cat}}$ and $\bar{y} = \sqrt[n]{\frac{k_1 e}{k_2}}$, we exclude the third variable e using the mass–balance (Equation (A32)) and arrive at the final expression for the steady-state reaction rate. It remains compact for the simplest case of $n = 1$:

$$\frac{-dm}{dt} = k_{cat}x = k_{cat}\frac{e_0 m}{K_m^{app} + m}; K_m^{app} = K_m(1 + k_1/k_2); K_m = \frac{k_4 + k_{cat}}{k_3} \tag{A33}$$

As we can see, the molecular crowding decreases the substrate affinity by making the apparent K_m be enlarged by the quotient $(1 + k_1/k_2)$. Like competitive inhibition, the negative impact is inversely related to the substrate concentration and is minimal at a high m. The complete solution for arbitrary n (multiple aggregates) is too cumbersome and should be explored by using the numeric integration of the entire set (Equations (A28)–(A32)). However, we can present the approximate solution for any n that allows the evaluation of the relationship between the total enzymatic proteins in cytosol, e_0, and the content of the nonaggregated enzyme-preserving activity, e:

$$e_0 = e\left(nK_e e^{n-1} + 1\right), K_e = \frac{k_1}{k_2} \tag{A34}$$

This equation is perfectly valid at a zero concentration of the internal substrate $m = 0$ and remains a reasonably good approximation at a low m, when $x \ll e$, and we can neglect the defensive effect of the substrate binding that prevents aggregation. The inhibitory

effect of molecular crowding (Figure A7) increases with the n. The stronger the aggregation (higher n and K_e), the higher the loss of enzymatic activity, with an asymptotic trend to a global maximum of the remaining metabolic activity, as $e_0 \rightarrow \infty$. The complete numeric solution of the aggregation mechanism will make sense when more experimental data becomes available.
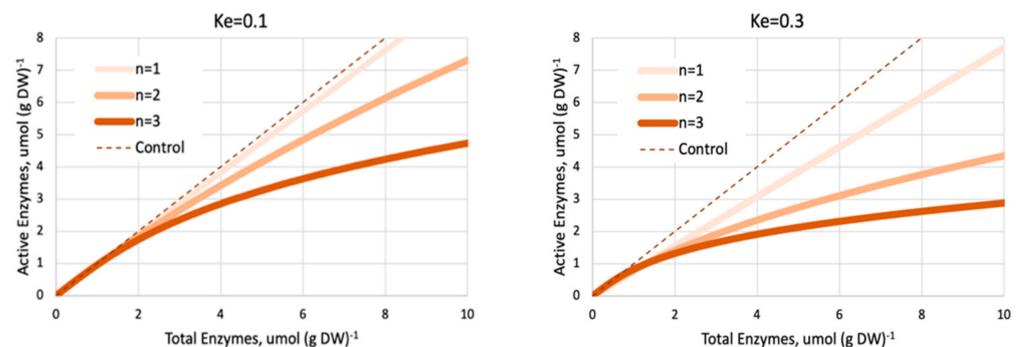

Figure A7. The modeling of the molecular crowding effect. The plots show progressive disparity between the active and total enzymes as the proteins' concentration increases at two levels of $K_e = k_1/k_2$ and three levels of n, the kinetic order of the aggregation reaction. The dotted line stands for negative control, no crowding.

References

1. Fang, X.; Lloyd, C.J.; Palsson, B.O. Reconstructing organisms in silico: Genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* **2020**, *18*, 731–743. [CrossRef] [PubMed]
2. Ahmed, N. A flood of microbial genomes-do we need more? *PLoS ONE* **2009**, *4*, e5831. [CrossRef] [PubMed]
3. Tavassoly, I.; Goldfarb, J.; Iyengar, R. Systems biology primer: The basic methods and approaches. *Essays Biochem.* **2018**, *62*, 487–500. [CrossRef] [PubMed]
4. Mendoza, S.N.; Olivier, B.G.; Molenaar, D.; Teusink, B. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* **2019**, *20*, 1–20. [CrossRef]
5. Kumar, M.; Ji, B.; Babaei, P.; Das, P.; Lappa, D.; Ramakrishnan, G.; Fox, T.E.; Haque, R.; Petri, W.A.; Bäckhed, F.; et al. Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling. *Metab. Eng.* **2018**, *49*, 128–142. [CrossRef]
6. Liu, Y.; Rousseaux, S.; Tourdot-Maréchal, R.; Sadoudi, M.; Gougeon, R.; Schmitt-Kopplin, P.; Alexandre, H. Wine microbiome: A dynamic world of microbial interactions. *Crit. Rev. Food Sci. Nutr.* **2017**, *57*, 856–873. [CrossRef]
7. Levering, J.; Dupont, C.L.; Allen, A.E.; Palsson, B.O.; Zengler, K. Integrated regulatory and metabolic networks of the marine diatom *Phaeodactylum tricornutum* predict the response to rising CO2 levels. *mSystems* **2017**, *2*, e00142-16. [CrossRef]
8. De Jong, H.; Casagranda, S.; Giordano, N.; Cinquemani, E.; Ropers, D.; Geiselmann, J.; Gouzé, J.-L. Mathematical modelling of microbes: Metabolism, gene expression and growth. *J. R. Soc. Interface* **2017**, *14*, 20170502. [CrossRef]
9. Alper, H.S.; Wittmann, C. Editorial: How multiplexed tools and approaches speed up the progress of metabolic engineering. *Biotechnol. J.* **2013**, *8*, 581–594. [CrossRef]
10. Hefzi, H.; Palsson, B.O.; Lewis, N.E. Reconstruction of genome-scale metabolic networks. In *Handbook of Systems Biology*; Vidal, M., Dekker, J., Eds.; Academic Press: San Diego, CA, USA, 2013; pp. 229–250.
11. Simeonidis, E.; Price, N.D. Genome-scale modeling for metabolic engineering. *J. Ind. Microbiol. Biotechnol.* **2015**, *42*, 327–338. [CrossRef]
12. Palsson, B.Ø. *Systems Biology: Constraint-Based Reconstruction and Analysis*; Cambridge University Press: Cambridge, UK, 2015.
13. Ruderman, D. The emergence of dynamic phenotyping. *Cell Biol. Toxicol.* **2017**, *33*, 507–509. [CrossRef]
14. Pirt, S.J. *Principles of Microbe and Cell Cultivation*; Blackwell Sci.: Oxford, UK; London, UK; Edinburgh, Scotland; Melbourne, FL, USA, 1975.
15. Panikov, N.S. Kinetics of Microbial Processes. In *Reference Module in Earth Systems and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2016; pp. 1–19.
16. Herbert, D. The chemical composition of micro-organisms as a function of their environment. *Symp. Soc. Gen. Microbiol. Cambridge Univ. Press* **1961**, *11*, 7.
17. Neidhardt, F.C.; Magasanik, B. Studies on the role of ribonucleic acid in the growth of bacteria. *Biochim. Biophys. Acta* **1960**, *42*, 99–116. [CrossRef]
18. Holme, T. Continuous culture studies on glycogen synthesis in *Escherichia coli* B. *Acta Chem. Scand.* **1957**, *11*, 763–775. [CrossRef]
19. Vemuri, G.N.; Altman, E.; Sangurdekar, D.P.; Khodursky, A.B.; Eiteman, M.A. Overflow metabolism in *Escherichia coli* during steady-state growth: Transcriptional regulation and effect of the redox ratio. *Appl. Environ. Microbiol.* **2006**, *72*, 3653–3661. [CrossRef]

20. Schmidt, B.J.; Ebrahim, A.; Metz, T.; Adkins, J.N.; Palsson, B.Ø.; Hyduke, D.R. GIM3E: Condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics* **2013**, *29*, 2900–2908. [CrossRef]

21. Peebo, K.; Valgepea, K.; Maser, A.; Nahku, R.; Adamberg, K.; Vilu, R. Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Mol. Biosyst.* **2015**, *11*, 1184–1193. [CrossRef]

22. Panikov, N.S. *Microbial Growth Kinetics*; Chapman & Hall: London, UK; Springer: Amsterdam, The Netherlands, 1995; 378p.

23. Doran, P.M. *Bioprocess. Engineering Principles*, 2nd ed.; Elsevier Ltd.: Amsterdam, The Netherlands; Boston, MA, USA, 2013; 439p.

24. Cornish-Bowden, E. *Fundamentals of Enzyme Kinetics*, 3rd ed.; Portland Press: London, UK, 2004.

25. Panikov, N.S. Kinetics, Microbial Growth. In *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*; Flickinger, M.C., Ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009; pp. 1–34.

26. Stewart, P.A. *How to Understand Acid-Base: A Quantitative Acid-Base Primer for Biology and Medicine*; Elsevier: New York, NY, USA, 1981; 186p.

27. Dash, M.C. *Fundamentals of Ecology*; Tata McGraw-Hill Education: New York, NY, USA, 2001.

28. Panikov, N.S.; Mandalakis, M.; Dai, S.; Mulcahy, L.R.; Fowle, W.; Garrett, W.S.; Karger, B.L. Near-zero growth kinetics of *Pseudomonas putida* deduced from proteomic analysis. *Environ. Microbiol.* **2015**, *17*, 215–228. [CrossRef]

29. Chiaramello, A.E.; Zyskind, J.W. Expression of Escherichia coli dnaA and mioC genes as a function of growth rate. *J. Bacteriol.* **1989**, *171*, 4272–4280. [CrossRef]

30. Dror, T.W.; Morag, E.; Rolider, A.; Bayer, E.A.; Lamed, R.; Shoham, Y. Regulation of the cellulosomal celS (cel48A) gene of Clostridium thermocellum is growth rate dependent. *J. Bacteriol.* **2003**, *185*, 3042–3048. [CrossRef]

31. Hellmuth, K.; Korz, D.; Sanders, E.; Deckwer, W.-D. Deckwer, Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high cell density cultivation of Escherichia coli TG1. *J. Biotechnol.* **1994**, *32*, 289–298. [CrossRef]

32. Hintsche, M.; Klumpp, S. Dilution and the theoretical description of growth-rate dependent gene expression. *J. Biol. Eng.* **2013**, *7*, 1–3. [CrossRef] [PubMed]

33. Klumpp, S.; Zhang, Z.; Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **2009**, *139*, 1366–1375. [CrossRef] [PubMed]

34. Lin-Chao, S.; Bremer, H. Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*. *Mol. Gen. Genet. MGG* **1986**, *203*, 143–149. [CrossRef]

35. Nahku, R.; Valgepea, K.; Lahtvee, P.-J.; Erm, S.; Abner, K.; Adamberg, K.; Vilu, R. Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures. *J. Biotechnol.* **2010**, *145*, 60–65. [CrossRef]

36. Orth, J.D.; Thiele, I.; Palsson, B.O. What is flux balance analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248. [CrossRef]

37. Brock, T.D. Microbial growth rates in nature. *Bacteriol. Rev.* **1971**, *35*, 39–58. [CrossRef]

38. Chowdhury, S.; Fong, S.S. Computational modeling of the human microbiome. *Microorganisms* **2020**, *8*, 197. [CrossRef]

39. Vet, S.; Gelens, L.; Gonze, D. Mutualistic cross-feeding in microbial systems generates bistability via an Allee effect. *Sci. Rep.* **2020**, *10*, 7763. [CrossRef]

40. Gaal, T.; Bartlett, M.S.; Ross, W.; Turnbough, C.L.; Gourse, R.L. Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **1997**, *278*, 2092–2097. [CrossRef]

41. Burgos, H.L.; O'Connor, K.; Sanchez-Vazquez, P.; Gourse, R.L. Roles of transcriptional and translational control mechanisms in regulation of ribosomal protein synthesis in *Escherichia coli*. *J. Bacteriol.* **2017**, *199*, 199. [CrossRef]

42. Murray, H.; Schneider, D.A.; Gourse, R.L. Gourse, Control of rRNA expression by small molecules Is dynamic and nonredundant. *Mol. Cell* **2003**, *12*, 125–134. [CrossRef]

43. Monod, J. *Recherches sur la Croissancu des Cultures Bacteriennes*; Hermann & C: Paris, France, 1942; pp. 200–210.

44. Kleijn, I.T.; Martínez-Segura, A.; Bertaux, F.; Saint, M.; Kramer, A.; Shahrezaei, V.; Marguerat, S. Growth-Rate Dependent and Nutrient-Specific Gene Expression Resource Allocation In Fission Yeast. *bioRxiv* **2021**. [CrossRef]

45. Schaechter, M.; Maaløe, O.; Kjeldgaard, N.O. Dependency on medium and temperature of cell size and chemical composition during balanced growth of Salmonella typhimurium. *Microbiology* **1958**, *19*, 592–606. [CrossRef]

46. Schaechter, M. A brief history of bacterial growth physiology. *Front. Microbiol.* **2015**, *6*, 289. [CrossRef]

47. Herbert, D. Some principles of continuous culture. In *Recent Progress in Microbiology*; Tunevail, G., Ed.; Blackwell Scientific: Oxford, UK, 1959; pp. 381–396.

48. Malek, I. Presidential Address. Environmental Control of Cell Synthesis and Function. *J. Appl. Chem. Biotechnol.* **1972**, *22*, 65–70. [CrossRef]

49. O'Brien, E.J.; Lerman, J.; Chang, R.; Hyduke, D.R.; Palsson, B. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **2013**, *9*, 1–13. [CrossRef]

50. 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.; et al. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Science* **2020**, *369*, eaav3751. [CrossRef]

51. Ye, C.; Xu, N.; Gao, C.; Liu, G.-Q.; Xu, J.; Zhang, W.; Chen, X.; Nielsen, J.; Liu, L. Comprehensive understanding of *Saccharomyces cerevisiae* phenotypes with whole-cell model WM_S288C. *Biotechnol. Bioeng.* **2020**, *117*, 1562–1574. [CrossRef]

52. Campbell, A. Synchronization of cell division. *Bacteriol. Rev.* **1957**, *21*, 263–272. [CrossRef]

53. Buchanan, R.E. Life phases in a bacterial culture. *J. Infect. Dis.* **1918**, *23*, 109–125. [CrossRef]

54. Singleton, R., Jr. Robert Earle Buchanan: An unappreciated scientist. *Yale J. Biol. Med.* **1999**, *72*, 329–339.

55.　Lane-Claypon, J.E. Multiplication of bacteria and the influence of temperature and some other conditions thereon. *J. Hyg.* **1909**, *9*, 239–248. [CrossRef]

56.　Penfold, W.; Norris, D. The relation of concentration of food supply to the generation-time of bacteria. *Epidemiol. Infect.* **1912**, *12*, 527–531. [CrossRef]

57.　Clark, P.F.; Ruehl, W. Morphological changes during the growth of bacteria. *J. Bacteriol.* **1919**, *4*, 615. [CrossRef]

58.　Graham-Smith, G. The behaviour of bacteria in fluid cultures as indicated by daily estimates of the numbers of living organisms. *Epidemiol. Infect.* **1920**, *19*, 133–204. [CrossRef]

59.　Panikov, N.S. 1.18—Microbial Growth Dynamics. In *Comprehensive Biotechnology*, 3rd ed.; Moo-Young, M., Ed.; Pergamon: Oxford, UK, 2019; Volume 1, pp. 231–273.

60.　Silva, A.; Cruz, A.J.; Araujo, M.L.G.; Hokka, C.O. The effect of the addition of invert sugar on the production of cephalosporin C in a fed-batch bioreactor. *Braz. J. Chem. Eng.* **1998**, *15*, 320–325. [CrossRef]

61.　Novick, A. Experimentation with chemostat. In *Recent Progress in Microbiology*; Blachwell Science Publication: Oxford, UK, 1959; pp. 403–415.

62.　Tempest, D.W.; Herbert, D.; Phipps, P.J. Studies on the growth of *Aerobacter aerogenes* at low dilution rates in a chemostat. In *Continuous Cultivation of Microorganisms*; Stat. Office: Salisbury, UK, 1967; pp. 240–253.

63.　Domach, M.M.; Leung, S.K.; Cahn, R.E.; Cocks, G.G.; Shuler, M.L. Computer model for glucose-limited growth of a single cell of *Escherichia coli B/r-A. Biotechnol. Bioeng.* **1984**, *26*, 203–216. [CrossRef]

64.　Young, J.D.; Ramkrishna, D. On the matching and proportional laws of cybernetic models. *Biotechnol. Prog.* **2007**, *23*, 83–99. [CrossRef]

65.　Kim, J.I.; Varner, J.D.; Ramkrishna, D. A hybrid model of anaerobic *E. coli* GJT001: Combination of elementary flux modes and cybernetic variables. *Biotechnol. Prog.* **2008**, *24*, 993–1006. [CrossRef]

66.　Young, J.D.; Henne, K.L.; Morgan, J.A.; Konopka, A.E.; Ramkrishna, D. Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. *Biotechnol. Bioeng.* **2008**, *100*, 542–559. [CrossRef] [PubMed]

67.　Song, H.-S.; Ramkrishna, D. Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. *Biotechnol. Bioeng.* **2011**, *108*, 127–140. [CrossRef] [PubMed]

68.　Malthus, T.R. *An Essay on the Principle of Population*; J. Johnson: London, UK, 1798; 360p.

69.　Verhulst, P.F. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.* **1838**, *10*, 113–121.

70.　Monod, J. La technique de culture continue: Theorie et applications. *Ann. Instit. Pasteur* **1950**, *79*, 390–410.

71.　Droop, M. Some thoughts on nutrient limitation in algae. *J. Phycol.* **1973**, *9*, 264–272. [CrossRef]

72.　Fredrickson, A.G. Formulation of structured growth models. *Biotechnol. Bioeng.* **1976**, *18*, 1481–1486. [CrossRef]

73.　Kompala, D.S.; Ramkrishna, D.; Jansen, N.B.; Tsao, G.T. Investigation of bacterial growth on mixed substrates: Experimental evaluation of cybernetic models. *Biotechnol. Bioeng.* **1986**, *28*, 1044–1055. [CrossRef]

74.　Panikov, N.S. Mechanistic mathematical models of microbial growth in bioreactors and in natural soils: Explanation of complex phenomena. *Math. Comput. Simul.* **1996**, *42*, 179–186. [CrossRef]

75.　Sherman, J.M.; Albus, W.R. Physiological youth in bacteria. *J. Bacteriol.* **1923**, *8*, 127. [CrossRef]

76.　Hengge-Aronis, R. Signal transduction and regulatory mechanisms involved in control of the sigma^s (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 373–395. [CrossRef]

77.　Gasch, A.P.; Spellman, P.T.; Kao, C.M.; Carmel-Harel, O.; Eisen, M.B.; Storz, G.; Botstein, D.; Brown, P.O. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **2000**, *11*, 4241–4257. [CrossRef]

78.　Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **1995**, *269*, 496–512. [CrossRef]

79.　Blattner, F.R.; Plunkett, G., III; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1462. [CrossRef]

80.　Edwards, J.S.; Palsson, B.O. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **1999**, *274*, 17410–17416. [CrossRef]

81.　Edwards, J.S.; Palsson, B.O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5528–5533. [CrossRef]

82.　Yan, Q.; Fong, S.S. Challenges and advances for genetic engineering of non-model bacteria and uses in consolidated bioprocessing. *Front. Microbiol.* **2017**, *8*, 2060. [CrossRef]

83.　Reed, J.L.; Palsson, B.Ø. Thirteen Years of building constraint-based In silico models of *Escherichia coli*. *J. Bacteriol.* **2003**, *185*, 2692–2699. [CrossRef]

84.　Majewski, R.A.; Domach, M.M. Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol. Bioeng.* **1990**, *35*, 732–738. [CrossRef]

85.　Savinell, J.M.; Palsson, B.O. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J. Theor. Biol.* **1992**, *154*, 421–454. [CrossRef]

86.　Varma, A.; Boesch, B.W.; Palsson, B.O. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* **1993**, *59*, 2465–2473. [CrossRef]

87.　Varma, A.; Palsson, B.O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **1994**, *60*, 3724–3731. [CrossRef]

88. Pramanik, J.; Keasling, J.D. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* **1997**, *56*, 398–421. [CrossRef]

89. Pramanik, J.; Keasling, J.D. Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnol. Bioeng.* **1998**, *60*, 230–238. [CrossRef]

90. Schilling, C.H.; Edwards, J.S.; Palsson, B.O. Toward metabolic phenomics: Analysis of genomic data using flux balances. *Biotechnol. Prog.* **1999**, *15*, 288–295. [CrossRef]

91. Schilling, C.H.; Schuster, S.; Palsson, B.O.; Heinrich, R. Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **1999**, *15*, 296–303. [CrossRef]

92. Palsson, B.Ø. *Systems Biology: Properties of Reconstructed Networks*; Cambridge University Press: Cambridge, UK, 2006.

93. Palsson, B.Ø. *Systems Biology: Simulation of Dynamic Network States*; Cambridge University Press: Cambridge, UK, 2011.

94. Covert, M.W.; Palsson, B.O. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **2002**, *277*, 28058–28064. [CrossRef]

95. Monk, J.M.; Lloyd, C.J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. iML1515, a knowledgebase that computes Escherichia coli traits. *Nat. Biotechnol.* **2017**, *35*, 904–908. [CrossRef]

96. Karr, J.R.; Sanghvi, J.C.; Macklin, D.N.; Gutschow, M.V.; Jacobs, J.M.; Bolival, B., Jr.; Assad-Garcia, N.; Glass, J.I.; Covert, M.W. A whole-cell computational model predicts phenotype from genotype. *Cell* **2012**, *150*, 389–401. [CrossRef]

97. Edwards, J.S.; Ibarra, R.U.; Palsson, B.O. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* **2001**, *19*, 125–130. [CrossRef]

98. Kauffman, K.J.; Prakash, P.; Edwards, J.S. Edwards, Advances in flux balance analysis. *Curr. Opin. Biotechnol.* **2003**, *14*, 491–496. [CrossRef]

99. Becker, S.A.; Feist, A.M.; Mo, M.L.; Hannum, G.; Palsson, B.Ø.; Herrgard, M.J. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. *Nat. Protoc.* **2007**, *2*, 727–738. [CrossRef] [PubMed]

100. Segré, D.; Zucker, J.; Katz, J.; Lin, X.; D'Haeseleer, P.; Rindone, W.P.; Kharchenko, P.; Nguyen, D.H.; Wright, M.A.; Church, G.M. From annotated genomes to metabolic flux models and kinetic parameter fitting. *OMICS* **2003**, *7*, 301–316. [CrossRef] [PubMed]

101. Heirendt, L.; Arreckx, S.; Pfau, T.; Mendoza, S.N.; Richelle, A.; Heinken, A.; Haraldsdóttir, H.S.; Wachowiak, J.; Keating, S.M.; Vlasov, V.; et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v3.0. *Nat. Protoc.* **2019**, *14*, 639–702. [CrossRef] [PubMed]

102. Ebrahim, A.; Lerman, J.A.; Palsson, B.O.; Hyduke, D.R. Hyduke, COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **2013**, *7*, 74. [CrossRef]

103. Kulyashov, M.; Peltek, S.E.; Akberdin, I.R. A genome-scale metabolic MODEL of 2, 3-butanediol production by thermophilic Bacteria Geobacillus icigianus. *Microorganisms* **2020**, *8*, 1002. [CrossRef]

104. Stephanopoulos, G.; Aristidou, A.A.; Nielsen, J. *Metabolic Engineering: Principles and Methodologies*; Elsevier: Amsterdam, The Netherlands, 1998.

105. Mahadevan, R.; Edwards, J.S.; Doyle, F.J., III. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* **2002**, *83*, 1331–1340. [CrossRef]

106. Hanly, T.J.; Henson, M.A. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol. Bioeng.* **2011**, *108*, 376–385. [CrossRef]

107. Lisha, K.P.; Sarkar, D. Dynamic flux balance analysis of batch fermentation: Effect of genetic manipulations on ethanol production. *Bioprocess. Biosyst. Eng.* **2014**, *37*, 617–627. [CrossRef]

108. von Wulffen, J.; RecogNice, T.; Sawodny, O.; Feuer, R. Transition of an anaerobic *Escherichia coli* culture to aerobiosis: Balancing mRNA and protein levels in a demand-directed dynamic flux balance analysis. *PLoS ONE* **2016**, *11*, e0158711. [CrossRef]

109. Meadows, A.L.; Karnik, R.; Lam, H.; Forestell, S.; Snedecor, B. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. *Metab. Eng.* **2010**, *12*, 150–160. [CrossRef]

110. Feng, X.; Xu, Y.; Chen, Y.; Tang, Y.J. Integrating flux balance analysis into kinetic models to decipher the dynamic metabolism of *Shewanella oneidensis* MR-1. *PLoS Comput. Biol.* **2012**, *8*, e1002376. [CrossRef]

111. Muthuraj, M.; Palabhanvi, B.; Misra, S.; Kumar, V.; Sivalingavasu, K.; Das, D. Flux balance analysis of *Chlorella sp.* FC2 IITG under photoautotrophic and heterotrophic growth conditions. *Photosynth. Res.* **2013**, *118*, 167–179. [CrossRef]

112. Villaverde, A.; Bongard, S.; Mauch, K.; Balsa-Canto, E.; Banga, J.R. Metabolic engineering with multi-objective optimization of kinetic models. *J. Biotechnol.* **2016**, *222*, 1–8. [CrossRef]

113. Saitua, F.; Torres, P.; Pérez-Correa, J.R.; Agosin, E. Dynamic genome-scale metabolic modeling of the yeast *Pichia pastoris*. *BMC Syst. Biol.* **2017**, *11*, 27. [CrossRef]

114. Upton, D.J.; McQueen-Mason, S.J.; Wood, A.J. An accurate description of *Aspergillus niger* organic acid batch fermentation through dynamic metabolic modelling. *Biotechnol. Biofuels* **2017**, *10*, 258. [CrossRef]

115. Wang, C.; Liu, J.; Liu, H.; Wang, J.; Wen, J. A genome-scale dynamic flux balance analysis model of *Streptomyces tsukubaensis* NRRL18488 to predict the targets for increasing FK506 production. *Biochem. Eng. J.* **2017**, *123*, 45–56. [CrossRef]

116. Covert, M.W.; Xiao, N.; Chen, T.J.; Karr, J. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics* **2008**, *24*, 2044–2050. [CrossRef]

117. Jensen, P.A.; Lutz, K.A.; Papin, J.A. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Syst. Biol.* **2011**, *5*, 147. [CrossRef]

118. Thiele, I.; Jamshidi, N.; Fleming, R.M.; Palsson, B.O. Genome-scale reconstruction of *Escherichia coli's* transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput. Biol.* **2009**, *5*, e1000312. [CrossRef]

119. Lerman, J.; Hyduke, D.R.; Latif, H.; Portnoy, V.A.; Lewis, N.; Orth, J.D.; Rutledge, A.; Smith, R.; Adkins, J.; Zengler, K.; et al. *In silico* method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* **2012**, *3*, 929. [CrossRef]

120. Thiele, I.; Fleming, R.M.T.; Que, R.; Bordbar, A.; Diep, D.; Palsson, B.O. Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PLoS ONE* **2012**, *7*, e45635. [CrossRef]

121. Liu, J.K.; O'Brien, E.J.; Lerman, J.A.; Zengler, K.; Palsson, B.O.; Feist, A.M. Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. *BMC Syst. Biol.* **2014**, *8*, 110. [CrossRef]

122. Yang, L.; Ma, D.; Ebrahim, A.; Lloyd, C.J.; Saunders, M.A.; Palsson, B.O. solveME: Fast and reliable solution of nonlinear ME models. *BMC Bioinform.* **2016**, *17*, 391. [CrossRef]

123. Lloyd, C.J.; Ebrahim, A.; Yang, L.; King, Z.; Catoiu, E.; O'Brien, E.J.; Liu, J.K.; Palsson, B.O. COBRAme: A computational framework for genome-scale models of metabolism and gene expression. *PLoS Comput. Biol.* **2018**, *14*, e1006302. [CrossRef]

124. Liu, J.K.; Lloyd, C.; Al-Bassam, M.M.; Ebrahim, A.; Kim, J.-N.; Olson, C.; Aksenov, A.; Dorrestein, P.; Zengler, K. Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen. *PLoS Comput. Biol.* **2019**, *15*, e1006848. [CrossRef]

125. Ebrahim, A.; Brunk, E.; Tan, J.; O'Brien, E.J.; Kim, D.; Szubin, R.; Lerman, J.A.; Lechner, A.; Sastry, A.; Bordbar, A.; et al. Multi-omic data integration enables discovery of hidden biological regularities. *Nat. Commun.* **2016**, *7*, 13091. [CrossRef]

126. Yang, L.; Mih, N.; Anand, A.; Park, J.H.; Tan, J.; Yurkovich, J.T.; Monk, J.M.; Lloyd, C.J.; Sandberg, T.E.; Seo, S.W.; et al. Cellular responses to reactive oxygen species are predicted from molecular mechanisms. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 14368–14373. [CrossRef]

127. Gibson, D.G.; Benders, G.A.; Andrews-Pfannkoch, C.; Denisova, E.A.; Baden-Tillson, H.; Zaveri, J.; Stockwell, T.B.; Brownley, A.; Thomas, D.W.; Algire, M.A.; et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **2008**, *319*, 1215–1220. [CrossRef]

128. Karr, J.R.; Takahashi, K.; Funahashi, A. The principles of whole-cell modeling. *Curr. Opin. Microbiol.* **2015**, *27*, 18–24. [CrossRef]

129. Hartwell, L.H.; Hopfield, J.J.; Leibler, S.; Murray, A.W. From molecular to modular cell biology. *Nature* **1999**, *402*, C47. [CrossRef]

130. Covert, M. Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Biophys. J.* **2019**, *116*, 451a. [CrossRef]

131. O'Brien, E.J.; Palsson, B.O. Computing the functional proteome: Recent progress and future prospects for genome-scale models. *Curr. Opin. Biotechnol.* **2015**, *34*, 125–134. [CrossRef] [PubMed]

132. Keyfitz, B.L.; Keyfitz, N. The McKendrick partial differential equation and its uses in epidemiology and population study. In *Mathematical and Computer Modelling*; Elsevier Science Ltd.: Amsterdam, The Netherlands, 1997; pp. 1–9.

133. Collins, J.F.; Richmond, M.H. Rate of growth of Bacillus *cereus* between divisions. *J. Gen. Microbiol.* **1962**, *28*, 15–33. [CrossRef] [PubMed]

134. Wang, P.; Robert, L.; Pelletier, J.; Dang, W.L.; Taddei, F.; Wright, A.; Jun, S. Robust growth of *Escherichia coli*. *Curr. Biol.* **2010**, *20*, 1099–1103. [CrossRef] [PubMed]

135. Wallden, M.; Fange, D.; Lundius, E.G.; Baltekin, O.; Elf, J. The synchronization of replication and division cycles in individual *E. coli* cells. *Cell* **2016**, *166*, 729–739. [CrossRef]

136. Osella, M.; Tans, S.J.; Cosentino Lagomarsino, M. Step by step, cell by cell: Quantification of the bacterial cell cycle. *Trends Microbiol.* **2017**, *25*, 250–256. [CrossRef]

137. Kleckner, N.E.; Chatzi, K.; White, M.A.; Fisher, J.K.; Stouf, M. Coordination of growth, chromosome replication/segregation, and cell division in *E. coli*. *Front. Microbiol.* **2018**, *9*, 1469. [CrossRef]

138. Lee, S.; Wu, L.J.; Errington, J. Microfluidic time-lapse analysis and reevaluation of the *Bacillus subtilis* cell cycle. *Microbiologyopen* **2019**, *8*, e876. [CrossRef]

139. Hui, S.; Silverman, J.M.; Chen, S.S.; Erickson, D.W.; Basan, M.; Wang, J.; Hwa, T.; Williamson, J.R. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **2015**, *11*, 784. [CrossRef]

140. Scott, M.; Gunderson, C.W.; Mateescu, E.M.; Zhang, Z.; Hwa, T. Interdependence of cell growth and gene expression: Origins and consequences. *Science* **2010**, *330*, 1099–1102. [CrossRef]

141. Battesti, A.; Majdalani, N.; Gottesman, S. The RpoS-mediated general stress response in Escherichia coli. *Annu. Rev. Microbiol.* **2011**, *65*, 189–213. [CrossRef]

142. Basan, M. Resource allocation and metabolism: The search for governing principles. *Curr. Opin. Microbiol.* **2018**, *45*, 77–83. [CrossRef]

143. Davidi, D.; Milo, R. Lessons on enzyme kinetics from quantitative proteomics. *Curr. Opin. Biotechnol.* **2017**, *46*, 81–89. [CrossRef]

144. Davidi, D.; Noor, E.; Liebermeister, W.; Bar-Even, A.; Flamholz, A.; Tummler, K.; Barenholz, U.; Goldenfeld, M.; Shlomi, T.; Milo, R. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3401–3406. [CrossRef]

145. Bennett, B.D.; Kimball, E.H.; Gao, M.; Osterhout, R.; van Dien, S.J.; Rabinowitz, J.D. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **2009**, *5*, 593–599. [CrossRef]

146. Bennett, B.D.; Yuan, J.; Kimball, E.H.; Rabinowitz, J.D. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat. Protoc.* **2008**, *3*, 1299–1311. [CrossRef]

147. Neidhardt, F.C. *Escherichia coli and Salmonella: Cellular and Molecular Biology*; ASM Press: Washington, DC, USA, 1996; Volume 1.

148. Watson, J.D. *Molecular Biology of the Gene*; Pearson Education India: Noida, India, 2004.

149. Heckmann, D.; Campeau, A.; Lloyd, C.J.; Phaneuf, P.V.; Hefner, Y.; Carrillo-Terrazas, M.; Feist, A.M.; Gonzalez, D.J.; Palsson, B.O. Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 23182–23190. [CrossRef]

150. Niewerth, H.; Schuldes, J.; Parschat, K.; Kiefer, P.; Vorholt, J.A.; Daniel, R.; Fetzner, S. Complete genome sequence and metabolic potential of the quinaldine-degrading bacterium *Arthrobacter* sp. Rue61a. *BMC Genom.* **2012**, *13*, 534. [CrossRef]

151. Ferenci, T. Adaptation to life at micromolar nutrient levels: The regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. *FEMS Microbiol. Rev.* **1996**, *18*, 301–317. [CrossRef]

152. Andersen, C.; Jordy, M.; Benz, R. Evaluation of the rate constants of sugar transport through maltoporin (LamB) of *Escherichia coli* from the sugar-induced current noise. *J. Gen. Physiol.* **1995**, *105*, 385–401. [CrossRef]

153. Boles, E.; Hollenberg, C.P. The molecular genetics of hexose transport in yeasts. *FEMS Microbiol. Rev.* **1997**, *21*, 85–111. [CrossRef]

154. Monk, J.M.; Charusanti, P.; Aziz, R.K.; Lerman, J.A.; Premyodhin, N.; Orth, J.D.; Feist, A.M.; Palsson, B.O. Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20338–20343. [CrossRef]

155. Koonin, E.V. The Clusters of Orthologous Groups (COGs) Database: Phylogenetic classification of proteins from complete genomes. In *The NCBI Handbook*; National Center for Biotechnology Information (US): Bethesda, MD, USA, 2002.

156. Schmidt, A.; Kochanowski, K.; Vedelaar, S.; Ahrne, E.; Volkmer, B.; Callipo, L.; Knoops, K.; Bauer, M.; Aebersold, R.; Heinemann, M. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **2016**, *34*, 104–110. [CrossRef]

157. Button, D. Kinetics of nutrient-limited transport and microbial growth. *Microbiol. Rev.* **1985**, *49*, 270–297. [CrossRef]

158. Wick, L.M.; Quadroni, M.; Egli, T. Short-and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. *Environ. Microbiol.* **2001**, *3*, 588–599. [CrossRef]

159. Vazquez, A. Optimal cytoplasmatic density and flux balance model under macromolecular crowding effects. *J. Theor. Biol.* **2010**, *264*, 356–359. [CrossRef]

160. Gabashvili, I.S.; Agrawal, R.K.; Spahn, C.M.; Grassucci, R.A.; Svergun, D.I.; Frank, J.; Penczek, P. Solution structure of the *E. coli* 70S ribosome at 11.5 A resolution. *Cell* **2000**, *100*, 537–549. [CrossRef]

161. Klumpp, S.; Scott, M.; Pedersen, S.; Hwa, T. Molecular crowding limits translation and cell growth. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16754–16759. [CrossRef]

162. Ellis, R.J. Macromolecular crowding: Obvious but underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597–604. [CrossRef]

163. Vazquez, A.; Beg, Q.K.; de Menezes, M.A.; Ernst, J.; Bar-Joseph, Z.; Barabási, A.-L.; Boros, L.G.; Oltvai, Z.N. Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Syst. Biol.* **2008**, *2*, 7. [CrossRef]

164. Weiss, M. Crowding, diffusion, and biochemical reactions. *Int. Rev. Cell Mol. Biol.* **2014**, *307*, 383–417. [PubMed]