HHS Public Access

Author manuscript

Isr J Chem. Author manuscript; available in PMC 2016 March 15.

Published in final edited form as:

Isr J Chem. 2014 August ; 54(8-9): 1219–1229. doi:10.1002/ijch.201300147.

Whole Cell Modeling: From Single Cells to Colonies

John A. Cole[a] and **Zaida Luthey-Schulten**[b]

Zaida Luthey-Schulten: zan@illinois.edu

[a]Department of Physics, University of Illinois, Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801 (USA)

[b]Department of Chemistry, University of Illinois, Urbana-Champaign, 505 South Mathews Avenue, Urbana, IL 61801 (USA)

Abstract

A great deal of research over the last several years has focused on how the inherent randomness in movements and reactivity of biomolecules can give rise to unexpected large-scale differences in the behavior of otherwise identical cells. Our own research has approached this problem from two vantage points – a microscopic kinetic view of the individual molecules (nucleic acids, proteins, etc.) diffusing and interacting in a crowded cellular environment; and a broader systems-level view of how enzyme variability can give rise to well-defined metabolic phenotypes. The former led to the development of the Lattice Microbes software – a GPU-accelerated stochastic simulator for reaction-diffusion processes in models of whole cells; the latter to the development of a method we call population flux balance analysis (FBA). The first part of this article reviews the Lattice Microbes methodology, and two recent technical advances that extend the capabilities of Lattice Microbes to enable simulations of larger organisms and colonies. The second part of this article focuses on our recent population FBA study of *Escherichia coli*, which predicted variability in the usage of different metabolic pathways resulting from heterogeneity in protein expression. Finally, we discuss exciting early work using a new hybrid methodology that integrates FBA with spatially resolved kinetic simulations to study how cells compete and cooperate within dense colonies and consortia.

Keywords

colony dynamics; flux balance analysis; kinetics; metabolism; stochastic modeling

1 Introduction

The underlying randomness inherent in chemical processes can be difficult to discern in bulk experiments conducted under well-stirred conditions, where variations from mean behavior are swamped by the large concentrations studied. The biochemistry inside the living cell, however, takes place in a compact and densely crowded environment,[1–3] and often involves only a few copies of a given reactant. Under these conditions, the behavior of systems of chemical reactions – such as those involved in gene expression and regulation –

Correspondence to: Zaida Luthey-Schulten, zan@illinois.edu.

have been shown to be highly variable from one cell to another, giving rise to remarkably divergent fates among otherwise identical cells.[4–6]

Sources of cellular noise can be categorized into two main types: “intrinsic” noise that independently affects each reaction within a cell, and “extrinsic” noise that affects many reactions within a cell in the same way, but differs from cell to cell.[7–9] Differences in polymerase or ribosome copy numbers, perhaps stemming from unequal partitioning during division, are examples of extrinsic variability.[10–12] Because extrinsic sources of noise tend to impact the expression of many genes within a cell simultaneously, counts of proteins in the extrinsic noise-dominated regime – in the case of slow-growing *Escherichia coli*, those whose mean count is larger than about 10 – exhibit a significant amount of correlation.[13]

Several theoretical studies of stochastic gene expression have offered insights into both its causes and implications. Among the most important of these has been the prediction by several convergent analyses that bacterial protein counts are approximately gamma distributed, and that each distribution is parameterized by only two values: the gene's average number of mRNA bursts per cell cycle and the average number of proteins transcribed per mRNA.[14–16] Simplifying assumptions, such as the form of the response of regulatory machinery to perturbation, have made more sophisticated analyses possible. The effects of extrinsic sources of noise on genetic circuits with simplified (e.g., Hill-type) regulation have also been studied analytically,[17] and similar assumptions have been employed to study the behavior of more complex models of gene expression, such as those that arise when considering multiple transcriptional and DNA looping states.[18] These studies have focused largely on analytical and computational solutions to the chemical master equation (CME), which ignores spatial heterogeneity within the cell. Despite their successes, these analyses cannot shed light on the roles crowding and spatial localization of cellular components play within the cell.

A mounting body of evidence suggests that cells are not the disorganized bags of proteins and nucleic acids they sometimes appear to be; rather, increasingly sophisticated microscopy experiments depict the cytoplasm as a dynamic, yet highly organized place. Even bacterial cells, which lack the conspicuous forms of compartmentalization seen in eukaryotes, exhibit highly nonuniform distributions of key cellular components. Examples range from the bacterial genome, which is known to condense and reorganize through interaction with H-NS and other nucleoid-associated proteins;[19–21] to the tendency of ribosomes to congregate in the polar regions of cells;[22–25] to emerging evidence that functionally-related enzymes may form clusters *in vivo*.[26]

Explicitly accounting for spatial effects in models of biochemical reaction networks can be exceedingly difficult, and analytical solutions are rarely possible. One successful analytical treatment of a reaction-diffusion system of biological interest is the recent study by Frey *et al*. of the dynamics of MinD and MinE – key proteins in the *E. coli* divisome thought to be responsible for nucleoid segregation, and the localization of division machinery to the center of the cell.[27–30] Although this study recovered the correct spatio-temporal dynamics of the system, both reactions and diffusion were treated in a deterministic way. To model

stochasticity in reaction-diffusion systems, as has been shown to be important even in the Min system,[31] a computational approach is generally required.

There are a number of programs designed for the stochastic simulation of large numbers of diffusing and reacting species.[32–34] These programs typically fall into one of two categories: particle-based methods, such as Smoldyne,[35] which track the locations of individual molecules in a three-dimensional space; and lattice-based methods, such as MesoRD[36] and our own highly-efficient GPU-accelerated Lattice Microbes,[37] which track the occupancy of each site in a three-dimensional lattice. Previous studies have used the Lattice Microbes software to model the effects of cytoplasmic crowding and cellular architecture on the inducible *lac* genetic switch,[23] as well as how the presence of membrane-embedded obstacles affects the oscillation frequency of the MinDE system.[60] All kinetic models benefit greatly from prior experimental studies, and our own work is no exception. In particular, both *in vivo* and *in vitro* rate constants, proteomics studies of cytoplasmic crowding, and cryoelectron tomography data have all been used to build our models.

Reaction-diffusion master equation (RDME) sampling codes, like Lattice Microbes, offer unprecedented views into the mechanisms by which reaction stochasticity and spatial heterogeneity give rise to large-scale differences between living cells, but the reaction parameters on which they rely are often not available in the literature, and at best, must be fitted or approximated.[39] To understand how stochastic gene expression affects reaction networks with sizes of the order of thousands of reactions, a different modeling paradigm is necessary. Trading detailed spatio-temporal dynamics for a steady-state approximation, flux balance analysis (FBA) offers a method for predicting the reaction flux through networks as large as bacterial metabolism, without requiring rate parameters. FBA poses the biochemical reaction network as a linear programming problem, and seeks to maximize the production of some component, which, in the case of metabolism, is usually biomass (an element in the network that accounts for the lipids, nucleotides, amino acids, and other “building blocks” necessary to produce more cells).[40] The optimization is subject to sets of constraints, which include, for example, the uptake of nutrients from the modeled environment, and maximum reaction fluxes based on enzyme availability.

FBA has been used in the past to study how variability in the counts, or efficiency of metabolic enzymes, affect overall metabolic behavior. Two recent studies of particular significance are outlined here. It had been predicted previously that a cell's proteome is maximally efficient in a near critical temperature regime beyond which a “catastrophic” proportion of proteins denature.[41] Building on this, a study of the thermostability of metabolic proteins by the Palsson laboratory used FBA to show that a relatively small number of enzymes exert a disproportionate amount of influence over a cell's ability to grow.[42] Recent work by the Covert laboratory also used a flux balance metabolic model; in this case, as a lynchpin in integrating a number of disparate modeling techniques into a genome-complete, temporally-resolved model of *Mycoplasma genitalium*. The model consists of 28 submodels which govern processes such as DNA replication, translation, ribosome formation, and ftsZ polymerization. Each uses the modeling technique most suited to it: from ODE (ordinary differential equation) integration; to various Monte Carlo

techniques; to FBA, in the case of the metabolism submodel. Building blocks like nucleic and amino acids generated at a given time step by the FBA model are used by the other submodels to produce cellular components like ribosomes, DNA, and metabolic enzymes. Future time steps then reflect these changes by updating, for example, the rates of translation due to the additional ribosomes, or the upper bounds on the flux through metabolic reactions catalyzed by the newly formed enzymes.[43]

While both of these studies represent important advances for systems biology modeling, neither is fully capable of predicting the population-level effects of variability in gene expression or enzyme function. The thermostability study deals in averages – average enzyme efficiency at a given temperature, average cell growth rate, etc. – and the model of *M. genitalium*, while capable of accounting for randomness in protein expression, is intended to model the detailed time evolution of only a handful of independent cells. To capture the behavior of cells in rare states of enzyme expression, large numbers of replicates need to be modeled. Our own population FBA method used experimentally determined protein count distributions to predict the instantaneous metabolic states of a population of cells numbering as large as one million.[44] This study revealed cells growing at a broad range of growth rates, and pursuing a number of very different metabolic behaviors, despite the cells living in identical environments and sharing a common genome.

Localization within colonies and biofilms, and the resulting competition among neighbors for diffusing resources, must impact each cell's behavior as much as its own unique enzyme expression state. Moreover, these effects appear to be nonlinear; the behavior of a given cell depends on its local microenvironment, but the local microenvironment is itself a product of the resources depleted, and byproducts generated, by the cell and its neighbors. To model not only the effects of heterogeneity in components inside the cell, but also components outside in the environment, we are developing new types of mixed-methodology simulations. By exploiting a natural time-scale separation that exists between reactive and diffusive events, and the timescales at which the assumptions of FBA are valid, we are beginning to mix FBA models of the individual metabolisms of large numbers of cells with RDME models of the shared extracellular environment, to unravel features of the collective metabolism of cellular consortia, ranging from simple bacterial colonies all the way up to tissues.

This mini-review initially focuses on the technical details of the Lattice Microbes software, as well as two recent advances that enable larger simulations and rapid integration of Lattice Microbes with other modeling packages. The focus of the article then shifts to review recent work using population FBA that predicted the emergence of a number of metabolic phenotypes arising from stochasticity in enzyme expression. Finally, we describe some early work in creating a hybrid RDME/FBA method[60] designed to resolve spatially the metabolisms of individual cells in a shared environment.

2 Lattice Microbes

Lattice Microbes[37,38] is a set of highly optimized algorithms for simulating realizations of stochastic biochemical systems in realistic models of whole cells. It was built from the

ground up to take advantage of the massive parallelism afforded by GPU computing, which enables users to model millions of molecules simultaneously diffusing and reacting on a timescale of an hour or more. Here we briefly describe the Lattice Microbes software and performance, and then review several recent developments of the code.

Lattice Microbes contains algorithms for sampling both the CME and RDME. The CME governs the time evolution of the probability of a system being in a given chemical state, **x**, defined as the instantaneous numbers of each species (protein, RNA, etc.) in the reaction network, and is given as:

$$\partial_t P(\mathbf{x},t) = \sum_{reactions\ r} -a_r(\mathbf{x})P(\mathbf{x},t) + a_r(\mathbf{x}-\mathbf{S_r})P(\mathbf{x}-\mathbf{S_r},t)$$

where $a_r(\mathbf{x})$ is the probability per unit time of the r^{th} reaction taking place, given that the current state is **x**, and $\mathbf{S}_r$ is the stoichiometry vector for the r^{th} reaction. The first term on the right represents transitions out of state **x**, while the second term represents transitions into state **x** from all states $\mathbf{x}–\mathbf{S_r}$. Lattice Microbes uses Gillespie's stochastic simulation algorithm (SSA)[45,46] to sample trajectories out of the solution space of the CME when the reaction volume is assumed to be well-stirred.

The RDME extends the CME in order to describe the time evolution of the spatially-resolved chemical state, **X**, of a reaction system:

$$\begin{aligned}\partial_t P(\mathbf{X},t) = &\sum_{subvolumes\ i} \sum_{reactions\ r} \\ &- a_{i,r}(\mathbf{X})P(\mathbf{X},t) \\ &+ a_{i,r}(\mathbf{X} \\ &- \mathbf{S_{i,r}})P(\mathbf{X} \\ &- \mathbf{S_{i,r}},t) \\ &+ \sum_{neighbor\ subvolumes\ i\&j} \sum_{species\ n} \\ &- d^n_{i,j}(\mathbf{X})P(\mathbf{X},t) \\ &+ d^n_{j,i}(\mathbf{X} \\ &- \Delta^n_{i,j},t)\end{aligned}$$

The RDME can be understood as consisting of a reaction part and a diffusion part. The reaction part, the first term on the right hand side, is essentially the CME, but it is applied independently to every subvolume in the system. The probability per unit time of a given reaction taking place, $a_{i,r}(\mathbf{X})$ is dependent not only on the state of the system, but also on the location within the reaction volume. This allows, for example, for different reactions to take place in a cell's cytoplasm and on its membrane. The diffusion part, the second term on the right hand side, accounts for the diffusive motions of particles between sites. Here, $d^n_{i,j}(\mathbf{X})$ is the probability per unit time that a particle of type *n* will diffuse from site *i* to

neighboring site *j* given that the system is in the spatially-resolved chemical state **X**, and $\Delta_{i,j}^{n}$ represents the difference between the state of a system before and after a diffusive event.

Lattice Microbes includes two algorithms for sampling trajectories from the solution space of the RDME; the first is a CPU-based, GPU-accelerated implementation of the next-subvolume method,[36,47] and the second, the multi-particle diffusion RDME (MPD-RDME), is a GPU-based algorithm combining the multi-particle (MP) diffusion method of Chopard *et al*.[48] with Gillespie's SSA, in a manner similar to the Gillespie multi-particle (GMP) method of Rodríguez *et al*.[49] Short time steps are chosen such that during any iteration, the probability that a given particle will react or diffuse is small, rendering the subvolumes independent, and allowing each to be computed separately, in parallel, on the GPU.[50]

During a simulation, for each of the *x*-, *y*-, and *z*-dimensions, Lattice Microbes draws from a standard uniform distribution to determine whether each particle diffuses in either the positive or the negative direction, or stays in its current subvolume. The probability of a particle diffusing either forward or backward in, for example, the *x*-direction, is approximately $\frac{2D\tau}{\lambda^2}$, where τ is the time step, λ is the lattice spacing, and *D* is the diffusion constant for the particle of interest. This sets an absolute upper bound on the simulation time step of $\tau = \frac{\lambda^2}{2D_{\max}}$ Where $\boldsymbol{D}_{\max}$ is the diffusion coefficient of the fastest-diffusing particle. In this case, these fastest-diffusing particles would be making diffusive transitions every time step, which can lead to unintended scenarios wherein particles may never interact. To avoid this, the time step should be set such that $\tau \leq \frac{\lambda^2}{4D_{\max}}$.[37]

After the positions of the diffusing particles have been updated, Lattice Microbes computes the probability of a reaction occurring within each subvolume during the time interval, τ, as $\int_0 a_{tot} e^{-a_{tot}t} dt$. Here, a_{tot} represents the sum of reaction propensities for all possible reactions, given the species currently within the subvolume of interest. This value is then compared to a random number drawn from a uniform distribution to determine if a reaction will take place. If one does, a second uniform random number is drawn to determine which reaction takes place, and the local species counts are updated to reflect the change.[37]

The MPD-RDME implementation in Lattice Microbes is of sufficient performance to simulate reaction-diffusion dynamics within spatial models of the cell that include realistic *in vivo* crowding. The spatial model can be shaped to mimic the cell being simulated, and its cytoplasm can be based on a number of different types of experimental data, including proteomics studies such as in Refs. [1,2], and imaging techniques such as cryoelectron tomography.[23] A model of densely packed crowders – hard spheres of varying sizes, corresponding to ribosomes, polymerases, and all the way down to small proteins – is generated and discretized to a lattice. Subvolumes within the cell are then either defined as "free" or "occupied", based on the volume fraction of crowders they contain. During the simulation, occupied sites represent reflective boundaries to the actively diffusing particles,

and no transitions into them are allowed.[50] A schematic of the MPD-RDME methodology is shown in Figure 1.

Studies of diffusion with and without crowding using the MPD-RDME methodology show excellent agreement with both experimental results and other theoretical treatments.[50] Of particular interest with regards to cellular function, is the anomalous subdiffusive behavior observed among simulated particles in crowded environments, and the impact of crowding on repressor-operator rebinding rates.[37,50] A comparison of 100,000 trajectories, sampled using the MPD-RDME methodology, showed good agreement with the next-subvolume method for exact sampling of the RDME. Relative errors of the order of 10^{-3}, or lower, across a wide range of time steps were reported.[37]

The MPD-RDME has been used to simulate the *lac* genetic switch in *E. coli*.[23] These simulations used a capsule-shaped cellular geometry, approximately 2–3 microns long and 0.5–1 micron in diameter, and filled with crowders to a total volume fraction of 50 %.

In terms of performance, Lattice Microbes compares favorably with other RDME-sampling software packages, especially in the large particle number limit, where 400-fold speed-up can be achieved, relative to competitors.[50]

2.1 Ongoing Development

Development of the Lattice Microbes code is ongoing, with two recent advances expanding the scale and performance of the software, as well as its user interface. The first major advance was the extension of the code to take advantage of multiple GPUs to perform larger and/or more finely resolved simulations.[38] The second major advance was the development of the Lattice Microbes problem-solving environment (PSE) – a set of libraries designed to facilitate simulation setup, execution, and analysis, using the popular and intuitive Python scripting language.[51]

2.1.1 Multi-GPU Lattice Microbes—The Lattice Microbes software has been used to study several reaction systems in models of single bacterial cells.[23,38,51] In the past, the time to complete a cell-cycle-long simulation of a large system like yeast has proven impractical, while the limited memory available on a single GPU has made some very large or very finely-resolved simulations, like those of a cell colony, impossible. While newer GPUs have enabled larger simulations, onboard GPU memory constrains the possible number of lattice sites and particles that can be stored. Because each lattice site requires 9 bytes of memory,[50] a modern GPU with 6 GB memory can hold a maximum of ∼358 million sites. This upper limit on the number of sites, in turn, implies an upper limit of 2.8 billion on the number of actively diffusing particles. A new version of Lattice Microbes, capable of spatially decomposing a simulation volume and distributing it across multiple GPUs connected to a common host CPU, has been developed to aggregate on-board GPU memory and enable simulations of larger volumes with greater spatial resolution. This “intra-node” version is targeted at computing resources like Georgia Tech's Keeneland machine,[52] which boasts three NVIDIA M2090GPUs per compute node. An MPI-based “inter-node” version is under development, that will be capable of distributing simulations across hundreds of GPUs.

The approach is straightforward, but highly optimized; the simulation volume is split along its longest dimension, and the sublattices are distributed among the available GPUs (see Figure 2A). A specialized load-balancer was developed to ensure that each GPU is utilized efficiently. This not only optimizes performance in cases where there may be an uneven distribution of particles in the simulation volume, but also in cases where the available GPUs may not be evenly matched, such as in a machine containing two GPUs of different generations. Exchanging information between GPUs is unavoidable; the diffusion calculations in the first and/or last planes on each GPU depend on the state of neighboring planes that are computed on another GPU, and so these planes must be exchanged between devices. Lattice Microbes takes advantage of the ability of GPUs to simultaneously execute kernel operations and perform memory transfers, so as to hide the cost of these data exchanges. This is done by launching two streams simultaneously – one begins performing diffusion computations on the interior of a given GPU's sublattice (those subvolumes that do not depend on neighboring planes), while the other stream copies the data necessary to perform diffusion calculations on the edge subvolumes, and then performs them. The multi-GPU Lattice Microbes code achieves outstanding parallel efficiency, especially for very large simulation volumes on the order of the size of a yeast cell (approximately one order of magnitude larger than E. coli). Details on the methodology and performance characteristics can be found in Hallock et al.[38]

2.1.2 Lattice Microbes Problem Solving Environment—Concurrent with the extension of the Lattice Microbes' technical capabilities, a significant amount of work was also devoted to enhancing its user interface. The Lattice Microbes PSE consists of pyLM and pySTDLM – Python libraries designed to enable users to create clear, concise, and unified simulation setup, execution, and analysis scripts, as well as a set of example scripts featuring case studies involving simulations of the lac genetic switch, and MinD and MinE dynamics in E. coli. The PSE enables users to interact directly with the main lattice data structures, either during initial setup or during timed interruptions of a simulation. This allows the state of the simulation (such as species counts or physical properties of the lattice) to be altered and analysis to be performed on-the-fly as the simulation runs.

In addition to simulation setup and analysis, the PSE offers users a natural way to combine different modeling techniques. A number of Python-based computational biology tools have emerged over the past several years including Biopython, ChemoPy, and COBRApy, to name a few.[53–55] These can be used to perform different types of computations on a Lattice Microbes trajectory, enabling either powerful analysis capabilities, or, as will be discussed in Section 4, new types of simulation schemes involving the mutual interaction of an RDME simulation with, for example, an FBA model of metabolism (see Figure 2B). Details on the PSE and its applications can be found in Peterson et al.[51]

3 Protein Variability in Metabolism

The ongoing development over recent years of single-molecule fluorescence techniques has made single-molecule “-omics” studies possible. A ground-breaking study of the *E. coli* proteome by Taniguchi *et al*. revealed strong agreement between the distributions of protein copy numbers observed *in vivo* and the gamma distribution that had been theoretically

predicted.[13] Studies such as this offer profound insights into the make-up of a cell, but not necessarily how the cell is functioning. To that end, we developed the population FBA technique which generates thousands or millions of unique realizations of bacterial cells by sampling the copy number distributions for hundreds of metabolic enzymes. Each modeled cell is in a unique state of protein expression, and as a result, pursues a unique metabolic strategy. Here we will review some details of the methodology, as well as several results and theoretical predictions from our study of *E. coli*.[44]

3.1 Population FBA Methodology

The fundamental assumption of population FBA is that the copy number of each metabolic enzyme in a given cell sets V_{max}, the upper bound for the flux through that enzyme's associated reactions, in accordance with Michaelis-Menten enzyme kinetics, i.e., $V_{max}=[E]k_{cat}$, where $[E]$ is the dry-weight normalized enzyme concentration, and k_{cat} is the enzyme turnover rate.[44] By using these upper bounds to constrain reactions within the *E. coli* metabolic network – in our case *iJO1366*,[56] the most recent and comprehensive constraint-based model of its type – we were able to predict how individual cells could best utilize their metabolic machinery to grow. In all, 352 metabolic enzymes from the data of Taniguchi *et al*.[13] were used. For each realization of a modeled cell, the copy number distributions of each of these enzymes were sampled randomly. The results were then paired with k_{cat} values from the BRENDA database[57] to set constraints within the metabolic network. FBA and parsimonious FBA (pFBA) – wherein the flux distribution that simultaneously maximizes cellular growth and minimizes total enzyme-mediated reaction flux is sought – were then used to predict the maximum growth rate and network reaction fluxes for each cell. This process was repeated until sizable populations, of the order of one million cells, were modeled. Because extrinsic noise has been shown to give rise to strong correlations between protein copy numbers,[13] we have also investigated the effects of correlated protein sampling. In one case, proteins in the extrinsic noise-dominated regime (those with mean copy number greater than 10) were correlated with coefficient of 0.66, in keeping with experimental results.[13] In another case, we investigated the effects of correlating (with a coefficient of 1) proteins encoded by genes transcribed within the same operon. Although the distributions we sample do account for noise intrinsic to protein expression, it is worth noting that FBA remains a deterministic treatment of a biochemical process, and as such, it cannot capture stochasticity in the reactions that the sampled proteins catalyze.

Our work has so far relied on sampling experimentally determined distributions, but in principle, there is no reason that the distributions cannot be determined theoretically, for example, by simulation using Lattice Microbes, or by some other means.

3.2 Enzyme Noise Gives Rise to Metabolic Variability

3.2.1 Growth Rate Variability—Our analysis of *E. coli* metabolism yielded a number of interesting results and predictions. We found that the specific growth rates of our modeled cells spanned a range from nearly zero, up to approximately 0.55 hr^{-1}, with a mean of 0.37 hr^{-1}, near the value measured in bulk (see Figure 3). The growth rate distribution was found to have a distinctive peak above a value of approximately 0.4 hr^{-1}; our analysis showed that

this was due to these cells sharing a common limitation to growth, namely their glucose uptake rate (see Labhsetwar *et al*.[44]). Increasing the upper bound on the glucose uptake rate led to shifts of this peak to the right, highlighting the need for single-cell nutrient uptake rate measurements. A recent study of the growth and proteome of *Saccharomyces cerevisiae* microcolonies yielded a growth rate distribution of remarkably similar form.[58] We also found that sampling proteins in a correlated manner generally tended to shift the distribution toward higher growth rates, but that the correlations imposed among cotranscribed genes were too few to significantly impact the distribution. Interestingly, we found that most of the metabolic variability observed in our study could be recovered by sampling a few (roughly 15) key proteins; in their study of protein thermostability, Chang *et al*. also found that just a handful of proteins controlled the metabolic behavior of the modeled cells across a wide range of temperatures.[42] It is possible that the structure of the metabolic network itself gives rise to relatively small sets of bottleneck enzymes that exert farreaching control.

3.2.2 Acetogenic Phenotype

PCA analysis of the flux distribution of thousands of modeled cells revealed the existence of several metabolic phenotypes. The starkest difference in metabolic behavior occurred between cells growing at rates above and below approximately 0.38 hr^{-1} – the rate at which the glucose uptake rate in most cells tends to reach its upper bound. Those cells growing above this rate were found to produce significant amounts of CO_2, but very little or no acetate, whereas those growing below this rate tended to make heavy use of the acetate overflow pathway. This phenotypic differentiation was determined to be the result of a trade-off between ATP-production efficiency and enzyme usage efficiency. The slow-growing cells, having not reached their glucose uptake upper bound, experience a glucose-rich environment. These cells do not require use of the energetically efficient tricarboxylic acid cycle, and instead use the acetate overflow pathway for its low enzyme-mediated flux. Conversely, the fast growing cells experience relative glucose-deprivation, and require the highly efficient ATP generating machinery of the TCA cycle and downstream electron transport chain (see Figures 3A, 3C, 4A, 4B). These predictions were partially validated through experimental determination of acetogenesis in batch cultures (see Section 3.2.4 and Figure 5). These experiments used strains and culturing procedures identical to those of Taniguchi et al.[13,44]

3.2.3 Entner-Doudorof Pathway Phenotype—Notable among the other phenotypes that emerged from our simulations was the existence of a large proportion of cells at intermediate growth rates that rely on use of the Entner-Doudorof pathway, while the slower- and faster-growing cells that flank them rely almost exclusively on the main Embden-Meyerhof-Parnas glycolytic pathway. This two-fold switching in pathway usage was found to hinge largely on a cell's sampled enolase copy number, which represents a key glycolytic bottleneck for most of the population. The slowest-growing cells, having a low glucose requirement, were able to metabolize all of their glucose products through the EMP pathway without being constrained by their enolase copy numbers. Cells at intermediate growth rates, on the other hand, tended to have larger glycolytic requirements that could not be met by the EMP pathway and their sampled enolase copy numbers alone. For these cells, the ED pathway, which has a significantly smaller enolase flux requirement than does the

EMP pathway, offers a way of processing greater amounts of glucose. This increase in glycolytic flux comes at a cost, however; the ED pathway forgoes some substrate-level ATP generation. The fastest-growing cells, which require the greatest metabolic efficiency, cannot afford the ATP opportunity cost associated with the ED pathway, and instead tend to be those cells whose enolase copy numbers were sampled high, allowing them to make use of the EMP pathway alone (see Figures 3B, 3D, 4C). It has been suggested that the ED pathway may be more favorable in terms of protein requirement than the EMP pathway.[59] This could be a reason why the enolase copy numbers measured in vivo appear to require simultaneous use of both pathways.

3.2.4 Growth Rate Heritability—Experimental verification of many of the predictions of our work remain technically challenging at the single-cell level, but by modeling the growth of a colony of cells over several generations and monitoring its total acetate production, we were able to draw a comparison between our results and a measurement of acetogenesis in an experimental batch culture. Our modeling approach involved removing mother cells at their time of division, and adding two daughter cells that were sampled from a population generated by our population FBA technique. Because we expect proteins – and therefore growth rate – to be heritable from generation to generation, we introduced some correlation between the growth rates of each mother cell and its daughters. Overall, we found good agreement between the acetogenesis we observed experimentally and that of several colonies modeled with generational growth rate correlations spanning a broad range of values (see Figure 5A). It was expected that due to the slow growth of the cells studied (a bulk doubling time of approximately 109 minutes was measured), very little acetate would be formed, as overflow metabolism of this kind is generally associated with fast growing cultures, but surprisingly, our predicted values actually underestimated the acetate production observed.

Naively, one might expect that as a colony grows, the faster growing cells would tend to swamp the slow growing subpopulation. In practice, our colony simulations showed that this tendency is highly dependent on the degree to which the growth rates of mother and daughter cells are correlated, and surprisingly, a low correlation co-efficient gives rise to precisely the opposite effect. We found that correlation coefficients above approximately 0.5 led to comparative over-representation of the fast-growing cells relative to the initial population, whereas coefficients less than 0.5 tended to lead to comparative over representation of the slow-growers (see Figure 5B). The reason for this is fairly simple. Without correlation, the likelihood that a fast-growing mother will give rise to a fast-growing daughter is comparable to the likelihood that it will give rise to a slow-grower. The same is true for slow-growing mother cells; but because the slow-growing cells divide so much less frequently than do the fast-growing cells, there is a lower flow of cells from the slow-growing subpopulations to the fast-growing subpopulations than there is from fast to slow. In essence, the fast-growing cells randomly produce slow-growing offspring, and the slow-growing cells simply accumulate as they wait to divide.

4 Spatially-resolved FBA: A mixed-methodological approach

From isogenic colonies to complex biofilms and tissues, cells often live in dense consortia, each vying for the same limited resources diffusing through their shared microenvironments. While our population FBA technique offers insight into how heterogeneity in the internal constituents of cells can affect their behavior, it says little about how nonuniform external conditions can affect behavior. To that end, we are working to leverage our expertise in multi-GPU reaction-diffusion modeling with our experience studying constraint-based models of the metabolism of large populations of cells to construct an exciting new type of mixed-methodological simulation technique. This approach is unique in its capacity to shed light on how spatial organization can influence the collective metabolic behaviors of groups of cells. These simulations use Lattice Microbes to treat the diffusion and transport reactions of metabolites (such as glucose and oxygen) in large extracellular spaces containing many cells, and FBA to treat the metabolism inside each cell based on its individual substrate uptake rates. This approach benefits greatly from both the multi-GPU version of Lattice Microbes and the Python based problem solving environment. Without the former, the size of the colonies and the concentrations of diffusing particles would be constrained to values much less than those generally studied experimentally, and without the latter, the marriage of Lattice Microbes with the Python-based FBA package COBRApy[53] would be cumbersome.

Details on our hybrid RDME/FBA methodology can be found in Cole *et al*.[60] and Peterson *et al*.[51] The basic approach revolves around the iterative use of short spatio-temporal simulations of the diffusion and uptake of substrates by modeled cells, and FBA to determine how the cells respond metabolically. Substrates that are predicted to be used by a cell are removed, and byproducts predicted to be generated are added to the simulation volume prior to the next round of simulation (see Figure 6). Other attempts have been made in the past to integrate dynamics into genome-scale constraint-based models in similar ways, but none have accounted for the type of detailed spatial resolution possible with our method.[43,61] Our simulations – although still at an early stage – have already begun to hint at the natural emergence of spatially-dependent metabolic phenotypes within isogenic colonies. Aerobicity appears to be of primary importance in this regard; we find that a rapid drop-off in oxygen concentration occurs within approximately ten microns from the surface of the colony, and that this differentiates aerobic surface cells from anaerobic cells in the colony interior (see Figure 7). This technique has obvious applications in models of quorum sensing, microbial cooperativity or competition,[62,63] and of intercellular communication in tissues and tumors.

Acknowledgments

The authors wish to thank the Weizmann Institute of Science, and “Computational Biology: Then and Now” organizers Koby Levy, Tamar Schlick, and Michael Levitt. We also thank Michael Hallock, Joseph Peterson and Piyush Labhsetwar for their help in preparing this manuscript. Development and applications of Lattice Microbes software and population FBA methodology are supported by the Department of Energy Office of Science (BER) under grant DE-FG02–10ER6510, the National Science Foundation under grants MCB-12-44570 and PHY-0822613. Additional software support by the National Institutes of Health grant NIH 9 P41 GM 104601-23.

References

1. Ridgway D, Broderick G, Lopez-Campistrous A, Ru'aini M, Winter P, Hamilton M, Boulanger P, Kovalenko A, Ellison MJ. Biophys J. 2008; 94:3748–3759. [PubMed: 18234819]
2. Sundararaj S, Guo A, Habibi-Nazhad B, Rouani M, Stothard P, Ellison M, Wishart DS. Nucleic Acids Res. 2004; 32:D293–295. [PubMed: 14681416]
3. McGuffee SR, Elcock AH. PLoS Comput Biol. 2010; 6:e1000694. [PubMed: 20221255]
4. Munsky B, Neuert G, van Oudenaarden A. Science. 2012; 336:183–187. [PubMed: 22499939]
5. Perkins, T.; Weise, A.; Swain, P. Quantitative Biology: From Molecular to Cellular Systems. Wall, ME., editor. CRC Press; Boca Raton: 2013. p. 51-72.
6. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Science. 2002; 297:1183–1186. [PubMed: 12183631]
7. Komorowski M, Miekisz J, Stumpf MP. Biophys J. 2013; 104:1783–1793. [PubMed: 23601325]
8. Johnston IG. Significance. 2012; 9:17–21.
9. Tkacik G, Gregor T, Bialek W. PLoS One. 2008; 3:e2774. [PubMed: 18648612]
10. Huh D, Paulsson J. Proc Natl Acad Sci U S A. 2011; 108:15004–15009. [PubMed: 21873252]
11. Pedraza JM, van Oudenaarden A. Science. 2005; 307:1965–1969. [PubMed: 15790857]
12. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N. Nat Genet. 2006; 38:636–643. [PubMed: 16715097]
13. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. Science. 2010; 329:533–538. [PubMed: 20671182]
14. Shahrezaei V, Swain PS. Proc Natl Acad Sci U S A. 2008; 105:17256–17261. [PubMed: 18988743]
15. Friedman N, Cai L, Xie XS. Phys Rev Lett. 2006; 97:168302. [PubMed: 17155441]
16. Walczak AM, Sasai M, Wolynes PG. Biophys J. 2005; 88:828–850. [PubMed: 15542546]
17. Assaf M, Roberts E, Luthey-Schulten Z, Goldenfeld N. Phys Rev Lett. 2013; 111:058102. [PubMed: 23952448]
18. Earnest TM, Roberts E, Assaf M, Dahmen K, Luthey-Schulten Z. Phys Biol. 2013; 10:026002. [PubMed: 23406725]
19. Wang W, Li GW, Chen C, Xie XS, Zhuang X. Science. 2011; 333:1445–1449. [PubMed: 21903814]
20. Fisher JK, Bourniquel A, Witz G, Weiner B, Prentiss M, Kleckner N. Cell. 2013; 153:882–895. [PubMed: 23623305]
21. Kuhlman TE, Cox EC. Mol Syst Biol. 2012; 8:610. [PubMed: 22968444]
22. Mascarenhas J, Weber MH, Graumann PL. EMBO Rep. 2001; 2:685–689. [PubMed: 11463749]
23. Roberts E, Magis A, Ortiz JO, Baumeister W, Luthey-Schulten Z. PLoS Comput Biol. 2011; 7:e1002010. [PubMed: 21423716]
24. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Mol Microbiol. 2012; 85:21–38. [PubMed: 22624875]
25. Llopis PM, Jackson AF, Sliusarenko O, Surovtsev I, Heinritz J, Emonet T, Jacobs-Wagner C. Nature. 2010; 466:77–81. [PubMed: 20562858]
26. An S, Kumar R, Sheets ED, Benkovic SJ. Science. 2008; 320:103–106. [PubMed: 18388293]
27. Halatek J, Frey E. Cell Rep. 2012; 1:741–752. [PubMed: 22813748]
28. Loose M, Kruse K, Schwille P. Ann Rev Biophys. 2011; 40:315–336. [PubMed: 21545286]
29. Di Ventura B, Knecht B, Andreas H, Godinez WJ, Fritsche M, Rohr K, Nickel W, Heermann DW, Sourjik V. Mol Syst Biol. 2013; 9:686. [PubMed: 24022004]
30. Akerlund T, Gullbrand B, Nordstrom K. Microbiology. 2002; 148:3213–3222. [PubMed: 12368455]
31. Fange D, Elf J. PLoS Comput Biol. 2006; 2:e80. [PubMed: 16846247]
32. Kerr RA, Bartol TM, Kaminsky B, Dittrich M, Chang JJ, Baden SB, Sejnowski TJ, Stiles JR. SIAM J Sci Comput. 2008; 30:3126. [PubMed: 20151023]

33. van Zon JS, ten Wolde PR. Phys Rev Lett. 2005; 94:128103. [PubMed: 15903966]
34. Plimpton, S.; Slepoy, A. DOE SciDAC 2005 Meeting, Journal of Physics: Conference Series 16. Institute of Physics Publishing; Bristol: 2005. p. 305-309.
35. Andrews SS, Addy NJ, Brent R, Arkin AP. PLoS Comput Biol. 2010; 6:e1000705. [PubMed: 20300644]
36. Hattne J, Fange D, Elf J. Bioinformatics. 2005; 21:2923–2924. [PubMed: 15817692]
37. Roberts E, Stone JE, Luthey-Schulten Z. J Comput Chem. 2013; 34:245–255. [PubMed: 23007888]
38. Hallock MJ, Stone JE, Roberts E, Fry C, Luthey-Schulten Z. Parallel Comput. 2014; 40:86–99. [PubMed: 24882911]
39. Stanford NJ, Lubitz T, Smallbone K, Klipp E, Mendes P, Liebermeister W. PLoS One. 2013; 8:e79195. [PubMed: 24324546]
40. Lewis NE, Nagarajan H, Palsson BO. Nature Rev Microbiol. 2012; 10:291–305. [PubMed: 22367118]
41. Dill KA, Ghosh K, Schmit JD. Proc Natl Acad Sci U S A. 2011; 108:17876–17882. [PubMed: 22006304]
42. Chang RL, Andrews K, Kim D, Li Z, Godzik A, Palsson BØ. Science. 2013; 340:1220–1223. [PubMed: 23744946]
43. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, Assad-Garcia N, Glass JI, Covert MW. Cell. 2012; 150:389–401. [PubMed: 22817898]
44. Labhsetwar P, Cole JA, Roberts E, Price ND, Luthey-Schulten Z. Proc Natl Acad Sci U S A. 2013; 110:14006–14011. [PubMed: 23908403]
45. Gillespie DT. J Comput Phys. 1976; 22:403–434.
46. Gillespie DT. J Phys Chem. 1977; 81:2340–2361.
47. Elf J, Ehrenberg M. Syst Biol. 2004; 1:230–236.
48. Chopard, B.; Droz, M. Modeling of Physical Systems. Cambridge University Press; Cambridge: 1998.
49. Rodríguez JV, Kaandorp JA, Dobrzyński M, Blom JG. Bioinformatics. 2006; 22:1895–1901. [PubMed: 16731694]
50. Roberts, E.; Stone, JE.; Sepúlveda, L.; Hwu, WM.; Luthey-Schulten, Z. Parallel & Distributed Processing. IPDPS 2009, IEEE International Symposium; 2009.
51. Peterson JR, Hallock MJ, Cole JA, Luthey-Schulten Z. Workshop for Python for High Performance and Scientific Computing (PyHPC 2013), Supercomputing 2013. 2013
52. Vetter JS, Glassbrook R, Dongarra J, Schwan K, Loftis B, McNally S, Meredith J, Rogers J, Roth P, Spafford K, Yalamanchili S. Comput Sci Eng. 2011; 13:90–95.
53. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. BMC Syst Biol. 2013; 7:74. [PubMed: 23927696]
54. Cao DS, Xu QS, Hu QN, Liang YZ. Bioinformatics. 2013; 29:1092–1094. [PubMed: 23493324]
55. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. Bioinformatics. 2009; 25:1422–1423. [PubMed: 19304878]
56. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ. Mol Syst Biol. 2011; 7:535. [PubMed: 21988831]
57. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Sçhngen C, Stelzer M, Thiele J, Schomburg D. Nucleic Acids Res. 2011; 39:D670–676. [PubMed: 21062828]
58. Levy SF, Ziv N, Siegal ML. PLoS Biol. 2012; 10:e1001325. [PubMed: 22589700]
59. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Proc Natl Acad Sci U S A. 2013; 110:10039–10044. [PubMed: 23630264]
60. Cole, JA.; Hallock, MJ.; Labhsetwar, P.; Peterson, JR.; Stone, JE.; Luthey-Schulten, Z. Computational Systems Biology Second Edition: From Molecular Mechanisms to Disease. Kriete, A.; Eils, R., editors. Academic Press; Amsterdam: 2014. p. 278-295.
61. Covert MW, Xiao N, Chen TJ, Karr JR. Bioinformatics. 2008; 24:2044–2050. [PubMed: 18621757]

62. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, Stahl DA. Mol Syst Biol. 2007; 3:92. [PubMed: 17353934]
63. Zhuang K, Izallalen M, Mouser P, Richter H, Risso C, Mahadevan R, Lovley DR. ISME J. 2011; 5:305–316. [PubMed: 20668487]
64. Humphrey W, Dalke A, Schulten K. J Mol Graph. 1996; 14:33–38. [PubMed: 8744570]

Biographies

John A. Cole is a graduate student in the laboratory of Professor Luthey-Schulten. Mr. Cole received a B. S. in Physics from Rutgers University in 2005, and an M. S. in Computational Biology from the New Jersey Institute of Technology in 2008.

Professor Schulten received a B. S. in Chemistry from the University of Southern California in 1969, an M. S. in Chemistry from Harvard University in 1972, and a Ph.D. in Applied Mathematics from Harvard University in 1975. From 1975 to 1980, she was a Research Fellow at the Max-Planck Institute for Biophysical Chemistry in Göttingen, and from 1980 to 1985, she was a Research Fellow in the Department of Theoretical Physics at the Technical University of Munich. Professor Schulten has been at the University of Illinois since 1987, where she is currently the William and Janet Lycan Professor of Chemistry, and maintains affiliations in the Department of Physics, the Beckman Institute, and the Institute for Genomic Biology.

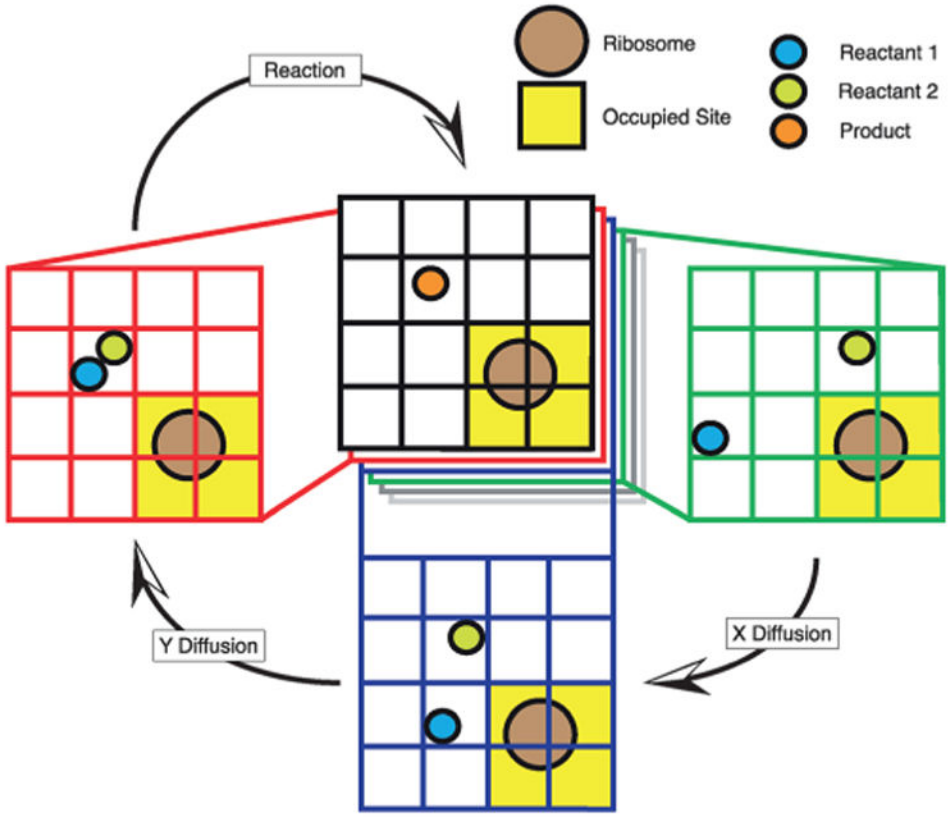

Figure 1.

Schematic of the MPD-RDME implementation for a small patch of cytoplasm within a single bacterial cell on a single GPU. In one time step, the state of the simulation transitions through three intermediates. The *t* state, shown in green, undergoes *x*-diffusion, transitioning to the blue state; the blue state undergoes *y*-diffusion, transitioning to the red state; and although not shown, the red state would transition through another state through *z*-diffusion before a reaction step finally updates the system to the $t+\Delta t$ state shown in black. Also indicated is the location of a ribosome and the corresponding occupied sites into which active particles cannot diffuse. Current versions of the Lattice Microbes code do not allow for the motion of obstacles, but planned future versions will.

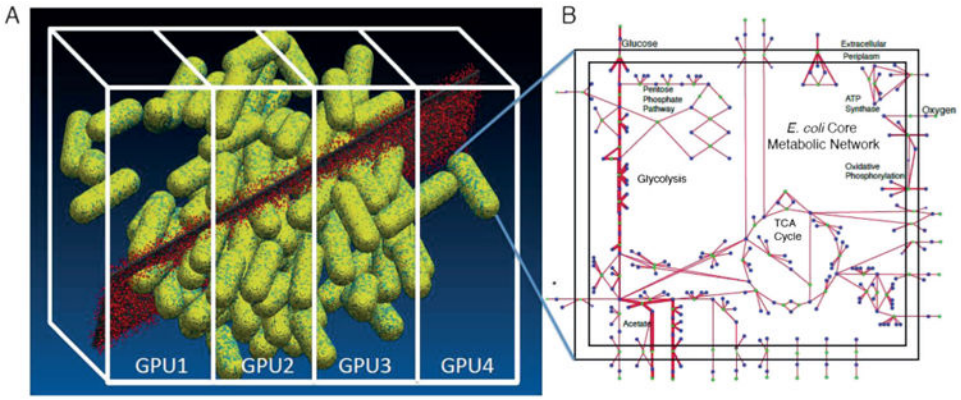

Figure 2.

A schematic highlighting the utility of the multi-GPU version of Lattice Microbes for simulating large volumes (A), and the hybrid use of Lattice Microbes with COBRApy[53] (B) within the PSE to model the metabolism of individual cells in a cluster of *E. coli*. Here, a volume including 100 cells is spread across four GPUs. Each cell can take up glucose (grey), and oxygen (red); the rates at which these metabolites enter a cell sets the upper bounds on the uptake rates for that cell's constraint-based model of metabolism. Here, although they pervade the entire simulation volume, the diffusing metabolites are depicted only within relatively narrow planes so as not to obscure the cells. The glucose, being in higher concentration than the oxygen (1×10^{-4} M and 1×10^{-5} M, respectively) is depicted in the narrower central plane. The blue points on the surface of the cells are glucose transporters. This hybrid RDME/FBA technique is described in greater detail in Section 4.

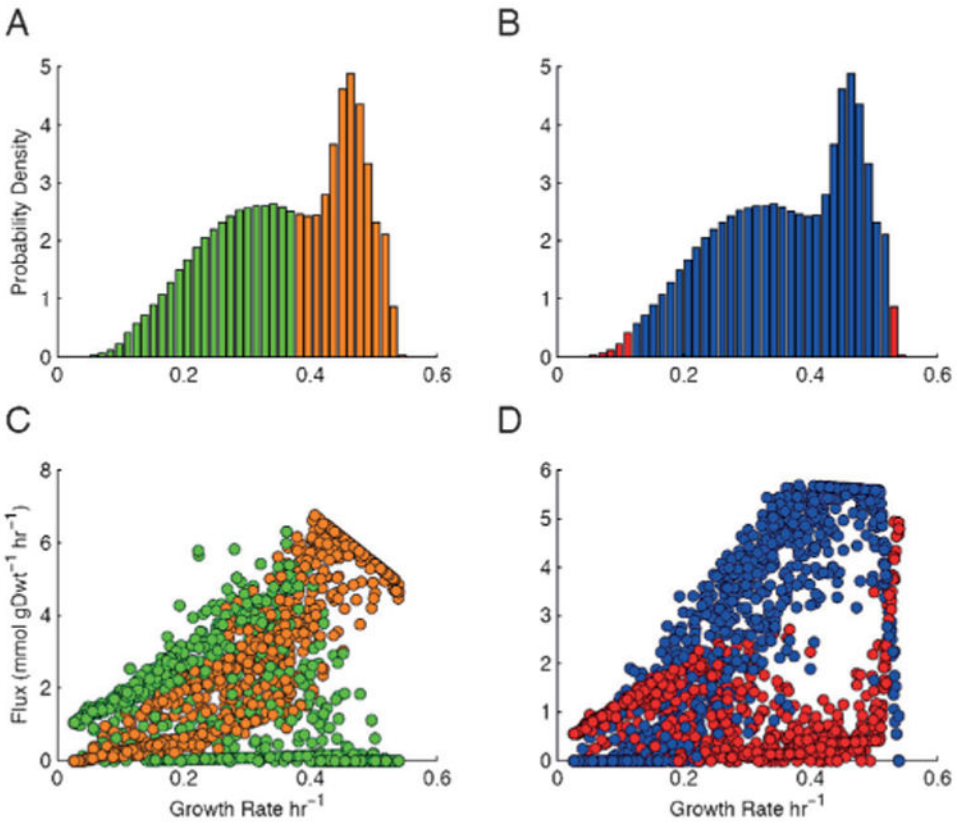

Figure 3.

Growth rate distributions (A and B), and metabolic reaction flux scatter plots (C and D). The distributions are colored to reflect which metabolic behavior is dominant in that growth rate regime. Cells growing at rates below approximately 0.38 hr^{-1} pursue a predominantly acetogenic metabolism (shown in green), whereas their faster-growing cousins make heavy use of the tricarboxylic acid (TCA) cycle (shown in orange). Similarly, cells growing at very low and very high growth rates tend to use the EMP (Embden-Meyerhof-Parnas) pathway (shown in red), whereas cells growing at intermediate rates tend to predominantly use the ED (Entner-Doudorof) pathway (shown in blue).

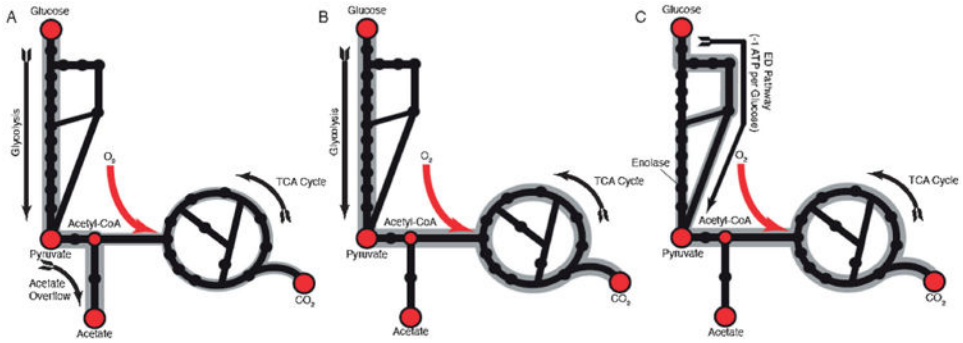

Figure 4.

Central metabolism cartoons. Key metabolites are shown as red circles, reactions are shown as black circles, and flux is indicated in grey. (A) Acetate overflow metabolism. Acetate generation is the predominant pathway utilized by our slow-growing subpopulation. (B) Utilization of the TCA cycle is more efficient when carbon sources are scarce, as is the case for our fast-growing cells. (C) The ED pathway requires half the enolase flux than the EMP pathway does, but generates one fewer ATP per glucose molecule; a large proportion of intermediate growing cells in our population makes use of this pathway.

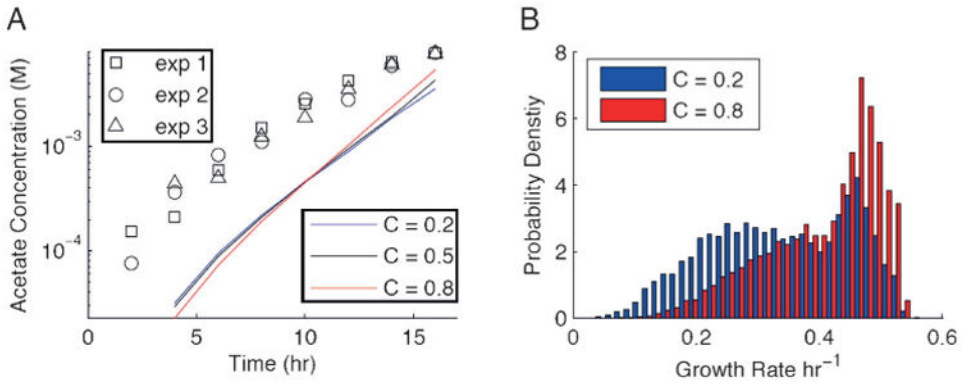

Figure 5.

(A) Predicted and observed supernatant acetate concentration for growing in vivo and in silico cultures. Black points represent experimental measurements performed in triplicate, while the lines represent theoretical predictions with varying generational growth rate correlations. (B) Histograms representing the distribution of growth rates seen in a growing population modeled with a high generational growth rate correlation coefficient (C = 0.8, shown in red), and low correlation coefficient (C = 0.2, shown in blue). As growth rate correlation is ramped up, probability density is shifted from the slow-growth tail to the fast-growth tail.

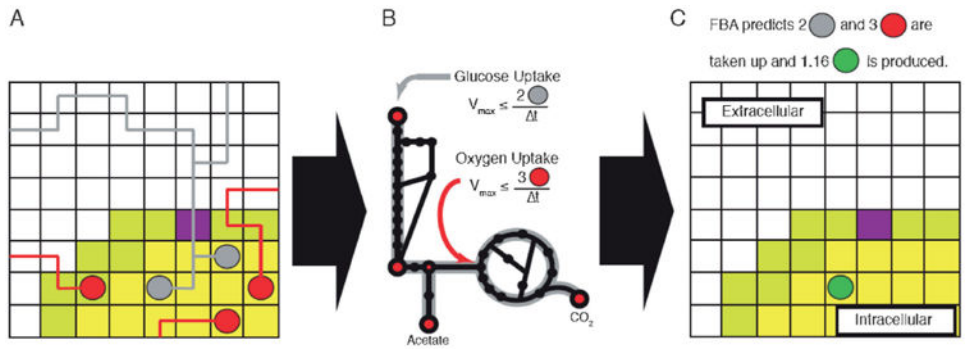

Figure 6.

Schematic of hybrid RDME/FBA simulations. A portion of a cell's cytoplasm is depicted in yellow and its membrane is depicted in light green; the extracellular environment in shown in white. A glucose transporter is visible in purple. (A) The state of a simulation after Δt of RDME simulation is shown. Glucose (grey) and oxygen (red) have diffused or been transported into the cell during the last Δt; their diffusive trajectories are shown as lines. Following the RDME simulation, all lattice sites within a given cell are looped through, and all particles of each type are summed to compute the average rates of substrate accumulation. (B) These rates are used to set upper bounds on the uptake rates for a genome-scale model of metabolism. FBA is then used to predict the substrate uptake and byproduct production rates. (C) The results of the FBA calculation are used to update the spatial lattice. In this case, all available glucose and oxygen molecules are predicted to have been taken up and 1.16 molecules of acetate are predicted to be produced. As such, all glucose and oxygen molecules inside the cell are removed and 1 (rounded to nearest whole number) acetate molecule is added to a random intracellular site. The lattice is then ready for the next round of RDME simulation.

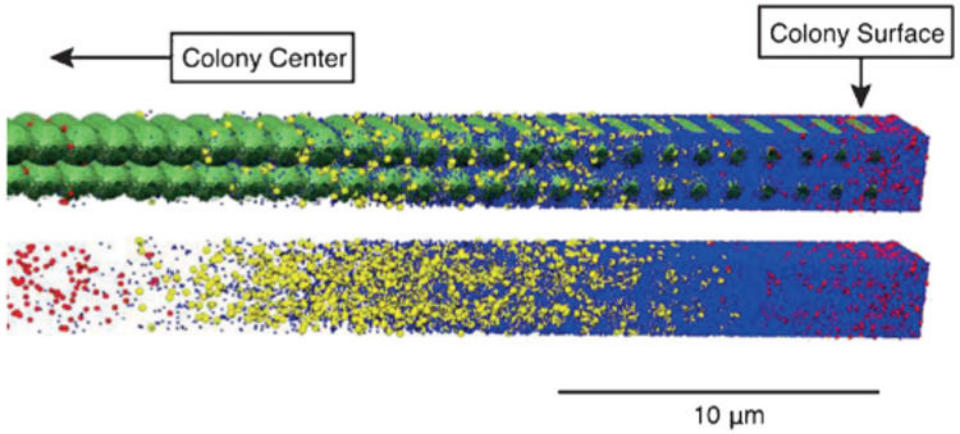

Figure 7.

Emergence of spatially dependent metabolic phenotypes within a simulation of a column of *E. coli*. Glucose (blue) and oxygen (red) diffuse from a constant-concentration boundary at the right. Cells (shown in green above, and hidden below) at the periphery quickly utilize the oxygen available to them, whereas those in the interior slip into anaerobic acetogenic (yellow) metabolism. Lattice Microbes trajectory imaged with VMD (Visual Molecular Dynamics).[64]