MULTISCALE MODEL. SIMUL.
Vol. 18, No. 2, pp. 1025–1052

MULTISCALE FLUX-BASED MODELING OF BIOFILM COMMUNITIES*

T. ZHANG†, A. PARKER†‡, R. P. CARLSON‡, P. S. STEWART‡, AND I. KLAPPER§

Abstract. Models of microbial community dynamics generally rely on a subscale description for microbial metabolisms. In systems such as distributed multispecies communities like biofilms, where it may not be reasonable to simplify to a small number of limiting substrates, tracking the large number of active metabolites likely requires measurement or estimation of large numbers of kinetic and regulatory parameters. Alternatively, a largely kinetics-free framework is proposed combining cellular level constrained, steady state flux analysis of metabolism with macroscale microbial community models. This multiscale setup naturally allows coupling of macroscale information, including measurement data, with cell scale metabolism. Further, flexibility in methodology is stressed: choices at the microscale (e.g., flux balance analysis or elementary flux modes) and at the macroscale (e.g., physical-chemical influences relevant to biofilm or planktonic environments) are available to the user. Illustrative computations in the context of a biofilm, including comparisons of systemic and Nash equilibration as well as an example of coupling experimental data into predictions, are provided.

Key words. microbial communities, metabolisms, biofilms, multiscale, flux balance

AMS subject classifications. 92C42, 35Q92, 92C45

DOI. 10.1137/18M1234096

1. Introduction.

The study of microbial community dynamics has in large part employed tools of classical population ecology. Even with the increasingly sophisticated methodologies that have become available, though, classical methods may not be well suited for exploiting the new depth of microbiological coverage because of their reliance on ad hoc constitutive law-based kinetics and regulation of those kinetics which effectively attempt to precompute constitutive relations between metabolism and environment. Further, classical methods rely on large numbers of reaction kinetics and regulation parameters that are unmeasured and possibly effectively unmeasurable (because they are themselves functions of environmental conditions), also making usage of detailed metabolic information problematic. The systems biology community has, in response to these difficulties, introduced largely kinetics-free formulations at the cellular level, termed metabolic pathway analysis or constraint-based flux balance analysis [3, 5, 48, 59] based on genomics and other omics (e.g., transcriptomics) derived data. Here we present a framework to incorporate kinetics-free methods into full scale microbial communities, biofilm models in particular, in part building on related methods previously applied to compartment models of eukaryotic cells [7, 8]. The key revision is to replace classical kinetics functions entirely by cell-level steady state metabolic pathway models—this is where omics meets community.

We focus here on biofilms, communities of microbes living and interacting at close quarters in self-secreted polymeric matrices [10]. Individual organisms take up sub-

*Received by the editors December 18, 2018; accepted for publication (in revised form) March 30, 2020; published electronically May 27, 2020.
https://doi.org/10.1137/18M1234096
Funding: The authors wish to acknowledge support from NSF-DMS awards 1517100, 1516951, and 1361420 as well as The Fields Institute.
†Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717 (tianyu.zhang@montana.edu, albert.parker@montana.edu, rossc@montana.edu).
‡Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717 (phil_s@montana.edu).
§Department of Mathematics, Temple University, Philadelphia, PA 19122 (klapper@temple.edu).

strates as locally available and utilize them to extract energy and build biomaterial, and also releasing waste in the form of metabolic byproducts. Processing is accomplished by, essentially, metabolic assembly lines termed pathways. Large numbers of internal and external metabolites can be involved [32, 70]. In many instances, limits on transport rates of chemical substrates into the biofilm and of byproducts out of the biofilm are important [64, 65]. On the other hand, substrate fluxes in and out of cells appear as sinks and sources at the large scale. As a result, microscale and macroscales become closely coupled and hence a biofilm model provides a good test of a multiscale method.

Whether in a biofilm or in another type of community, individual microbes, and communities of microbes generally have many available metabolic pathways, and have extensive libraries of enzymes and regulatory mechanisms available to deploy in response to local environmental conditions. From a modeling point of view, a consequence is that kinetics parameters are functions of environmental conditions (by a variety of mechanisms, e.g., enzyme affinities and copy numbers). Hence determining rate parameters for metabolic models is difficult as rates are responsive to internal and external concentrations of, in principle, many metabolites. In response, two main modeling strategies have been employed.

First, at the community level, kinetic models can be replaced by empirical constitutive laws relating, for example, external concentrations of a few limiting substrates to, for example, biomass production rates. This strategy amounts to precomputing metabolic function and can be effective when metabolism is relatively simple. However, spatially distributed microbial systems, like gut or soil communities, or biofilms, are often rather complicated with different metabolic pathways and networks active in different locations and times. Consequently, anticipating metabolic activity with empirical constitutive response functions is likely to be difficult for models meant to address the increasingly detailed data that are becoming available.

A second strategy, stoichiometric metabolic analysis, has been applied to cellular level models of metabolism. The key idea is to assume that kinetic dynamics generally operate in a quasi-steady-state. This assumption is justified in the case that environmental changes are slow in comparison to time scales for metabolic systems to reach steady state, i.e., kinetics time scales are relatively short in comparison to system times, as is often supposed to be the case. Mathematically, the quasi-steady assumption amounts to constraining internal reaction flux vectors to the null space of the stoichiometric matrix (though thermodynamics considerations such as restriction of some reactions as unidirectional as well as reaction flux bounds may further restrict the viable solution space to a cone within the null space). That is, solution vectors satisfy mass flux balance—total material fluxes into and out of all internal metabolites must exactly balance. This balance condition is entirely independent of the kinetic rate parameters; hence there is no necessity to measure or estimate their values. However, the choice of solution is generally underdetermined, even with the extra constraints—cells still have freedom to deploy available resources to different metabolic pathways. Thus some additional selection principle is required in order to distinguish a particular solution. This principle must rely on that information which is available, namely, internal reaction fluxes (internal and external concentrations and other environmental information that are not accessible to a cellular level model). Environmental conditions are represented by imposed maximum allowable influxes of metabolites. A typical optimization strategy is to choose from the admissible flux vectors the one that maximizes biomaterial production, though other optimization principles are also employed.

Steady state metabolic analysis thus has the important advantages that it does not require values for kinetic rates and, via optimization, can avoid need for details of regulatory mechanisms which, like kinetic parameters, are problematic to characterize. A cost, though, is the requirement for a selection principle that can be evaluated without access to important system level data. In fact, metabolic regulation should be able to respond to external environmental conditions [35, 40]. Likewise, such models are limited in their capability to generate system level information and so limited in ability to inform larger length and time scales of the community level which is often the goal for microbial ecology and dynamics models. On the other hand, a community scale model with constitutive metabolic laws does, as a matter of course, produce much of that information, but its metabolic component is ultimately tied to kinetic rates. Thus combining the two scales is complementary and hence the interest in combining metabolic pathway analysis with community scale models.

Some efforts have been made in this direction based on implicit or explicit iteration between a community level model which updates local external chemical concentrations based on physical processes, e.g., diffusion, and given fluxes of chemicals into cells, and a cellular level model which updates metabolic fluxes based on new influx bounds arising from updated chemical concentrations [11, 16, 30, 31, 34, 44, 50, 57, 63, 68]. This approach, though a convenient way to utilize already developed varieties of metabolic flux models, can have some disadvantages. First, as generally practiced, optimization targets are largely inherited from the underlying cellular flux modeling so external information may not be utilized beyond setting flux bounds, though there are variants that look to combine metabolic analyses in a multispecies community [25, 67, 73]. Even excepting this issue, these methods are generally built around a particular flux modeling approach, restricting user flexibility. Second, this iteration procedure, for time dependent problems, often matches cellular and system time scales and, if so, may be inconsistent with the foundational principle of metabolic flux analysis that metabolism is at quasi-steady-state with respect to the environment. Further, from a practical point of view, in many cases the environmental time scales are of interest so that computing on the typically much shorter metabolic relaxation time scale is inefficient.

Here we propose to follow the standard multiscale practice of computing on the longer, system scale using a steady state cellular level assumption. It is then natural, and in fact this is a prerequisite, to incorporate steady state metabolic models. Many choices for cellular level and system level modeling strategies can be employed, and, as a bonus, all available environmental information can be made available to the cellular model and vice versa. To demonstrate, we model biofilm (a slowly growing system) dynamics and employ external, environmental information to inform a cellular scale metabolism encoded by an elementary flux mode representation. Biofilms are, generally, relatively complex systems from both an ecological and mathematical viewpoint and hence serve as a useful vehicle to demonstrate methods [72].

2. Kinetics-free, flux-balanced modeling.

2.1. Biofilm model. We briefly describe a general biofilm community model [38, 45] consisting of a domain Ω, a set of N material species (which may be actual biological species or other biological groupings or other particulate materials, e.g., water or inert organic material) with volume fractions $X_j(\mathbf{x}, t)$, $j = 1, \ldots, N$, and a set of M dissolved (exterior to microbial cells) metabolite (or substrate) or other chemical concentrations $C_k(\mathbf{x}, t)$, $k = 1, \ldots, M$. The material species satisfy volume constraint $X_1 + X_2 + \cdots + X_N = 1$. The aim, then, is to describe how, within the domain Ω,

chemicals and materials are produced/consumed and are transported. Note that Ω itself may change in time as a consequence of material changes, e.g., production of new biomaterial.

The dissolved concentrations $C_k(\mathbf{x}, t)$ satisfy transport equations, typically reaction-diffusion equations of the sort

$$D_k \nabla^2 C_k = \sum_{j=1}^{N} \frac{r_{jk}(\mathbf{C})}{Y_{jk}} X_j, \quad k = 1, 2, \ldots, M, \tag{1}$$

where r_{jk} is a consumption/production rate (units of time^{-1}) and Y_{jk} is a yield coefficient (units of volume fraction per concentration) [38]. On larger scales, advective transport in a bulk flow may also be of importance and can be included [12, 14, 22, 23, 28, 51, 66] but will not be considered here for simplicity. $\mathbf{C}$ is the vector with components of all dissolved concentrations C_k exterior to cells. Note the assumption of quasi-equilibrium—for a relatively thin biofilm we suppose that the diffusion-reaction equilibration time is short in comparison to other system time scales such as biological growth. This assumption is again not critical but is common in biofilm models [51] and in some ways simplifies what follows.

The rate functions r_{jk} are typically supposed to be empirically determined, for example, based on a Monod product form $r_{jk} = \mu_{jk} \prod_{\ell \in \mathcal{O}} C_\ell/(K_\ell + C_\ell)$, where chemical constituents are listed by index in $\mathcal{O}$, half-saturations K_ℓ indicate uptake saturation of dissolved quantity ℓ, and μ_{jk} provides a maximum specific usage rate of dissolved quantity k by species j. Note that the complexity of system (1) can be expected to grow rapidly with increasing number of dissolved quantities in which case parameter estimation becomes problematic. In response, in many studies the vector $\mathbf{C}$ is restricted to a small number of "limiting" quantities, and neglected concentrations are assumed to be saturated. This simplification can be effective but may fail in spatially or temporally variable systems where it is not clear in advance which quantities should be considered the limiting ones at which locations and at which times. Even when a limiting approximation is appropriate, however, parameterization is still often uncertain.

Supplementing transport equations (1) for dissolved quantities, we also have dynamical equations for volume fractions of the form

$$\frac{\partial}{\partial t} X_j + \nabla \cdot (\mathbf{u} X_j) = g_j(\mathbf{C}) X_j, \quad j = 1, 2, \ldots, N, \tag{2}$$

where g_j is a growth/decay rate and $\mathbf{u}$ is a material velocity. The rate function g_j will generally be a function of the usage rates r_{jk}, $k = 1, 2, \ldots, M$, as, among these, some might relate to biomass production or maintenance. Note, also, we omit in (2) some effects that are sometimes included in biofilm models: conversion terms (e.g., conversion between active and inactive material), diffusive or other forms of mobility, etc. [18, 19, 39], as well as the possibility for different material types to move with different velocities [15, 37, 52].

Upon summing (2), we obtain the volume conservation constraint

$$\nabla \cdot \mathbf{u} = \sum_{j=1}^{N} g_j(\mathbf{C}) X_j, \tag{3}$$

that is, the biofilm expands/contracts locally so as to accomodate local net material expansion/contraction. Later we will consider only variation in one spatial dimension

(the coordinate z transverse to the biofilm) in which case, together with, say, an immovable wall condition $\mathbf{u}|_{z=0} = (0, 0, w(0)) = \mathbf{0}$ at a solid boundary, constraint (3) is sufficient to determine deformation of the material everywhere. In more than one dimension, generally, system (2) is supplemented by a balance of forces equation in addition to constraint (3) in order to determine $\mathbf{u}$ [38]. In any case, once $\mathbf{u}$ is determined, then at the free boundary $\boldsymbol{\gamma}(t)$

$$\frac{d}{dt}\boldsymbol{\gamma} = (\mathbf{n} \cdot \mathbf{u}|_{\boldsymbol{\gamma}})\, \mathbf{n}, \tag{4}$$

where $\mathbf{n}$ is an outward unit normal.

In what follows, we will limit consideration to a single material species, i.e., $N = 1$, and a short list of dissolved substrates ($M = 3$). In the case that $N = 1$, the volume fraction $X = X_1 = 1$ so that (2) becomes extraneous. Also, constraint (3) simplifies to $\nabla \cdot \mathbf{u} = g_1(\mathbf{C}) = g(\mathbf{C})$ so that, for a one dimensional system, (4) reduces to

$$\frac{d}{dt}L = \int_0^L g(\mathbf{C}(z))dz, \tag{5}$$

where $L(t)$ is the biofilm thickness.

2.2. Metabolic model. We again, briefly, introduce a general metabolic model consisting of the interior of a well-mixed cell, a set of m dissolved interior metabolite concentrations $c_k(t)$, $k = 1, 2, \ldots, m$, and a set of n reactions among the metabolites. Note that m, the number of internal concentrations, and M, the number of external concentrations, are not necessarily the same. A particular reaction can be represented by a chemical equation of the form

$$\alpha A + \beta B \rightleftharpoons \gamma C + \delta D, \tag{6}$$

where A, B, C, D are metabolite labels (unitless) and $\alpha, \beta, \gamma, \delta$ are stoichiometric coefficients (metabolites per reaction, measured, for example, in moles). The number of metabolites involved in any one reaction may of course vary from reaction to reaction.

The entire set of n reactions can be encoded in an m×n stoichiometric matrix S where the jth column of S corresponds to the jth reaction and the kth row corresponds to the kth metabolite. That is, if the kth metabolite is involved in the jth reaction, then S_{kj} is set to the corresponding stoichiometric coefficient (with minus sign if the metabolite appears on the left and plus sign if it appears on the right of the chemical equation). Otherwise, S_{kj} is set to 0. Given S, then

$$\frac{d}{dt}\mathbf{c} = S\mathbf{v}, \tag{7}$$

where $\mathbf{c}$ is the vector of interior metabolite concentrations (to be distinguished from the vector $\mathbf{C}$ of exterior concentrations) and $\mathbf{v}$ is the n-vector of fluxes (reactions per volume per time) through the n reactions. Generally, $\mathbf{v} = \mathbf{v}(\mathbf{c})$ through a constitutive law which depends on kinetics, and kinetics parameters, of the metabolic reactions. However, using the steady state assumption of metabolic pathway analysis, (7) reduces to

$$S\mathbf{v} = \mathbf{0}. \tag{8}$$

The idea is solve (8) for $\mathbf{v}$ in place of solving (7) plus constitutive law $\mathbf{v} = \mathbf{v}(\mathbf{c})$ for $\mathbf{c}$. Note the considerable advantage that (8) is parameter free other than known

stoichiometric coefficients in S, and knowledge of reaction fluxes is sufficient to characterize biomass production as well as influx and outflux of metabolites from cells. As a tradeoff though, internal metabolite concentrations $\mathbf{c}$, even at steady state, are unknown.

However, there is a caveat: often there are more reactions than metabolites in which case system (8) is underdetermined. Note that the choice of null space vector is not entirely unconstrained as thermodynamical considerations make some vectors in the null space of S very unlikely to occur in practice [53, 54]. However, these restrictions involve metabolite concentrations $\mathbf{c}$ which are unknown so are generally difficult to apply. In some cases, though, thermodynamic constraints can be imposed anyway, one important example being reactions that proceed in only one direction except at unrealistically extreme values of the reaction component concentrations, i.e., reactions for which the flux can effectively assumed to be of one sign only (with the reaction conventionally set so that its flux is nonnegative) at physiologically reasonable metabolite concentrations. Consequently (8) is often supplemented by a number of nonnegativity constraints of the form $v_\ell \geq 0$, where ℓ ranges over the set of effectively single-direction reactions. Still, even after imposition of these extra constraints, the system remains underdetermined. Hence some criteria must be adopted by which to choose a designated flux vector $\mathbf{v}$ from the solution space of (8). Much attention has been given to developing such principles and there are a number of options available; we do not need to assume a restriction to use of any one such methodology here.

In preparation for connecting to a macroscale community model, we replace (8) by a set of such equations,

$$S^{(j)}\mathbf{v}^{(j)} = \mathbf{0}, \quad j = 1, 2, \ldots, N, \tag{9}$$

one for each of the N material species. That is, each active material species will have its own set of internal reactions and metabolites, and so its own stoichiometric matrix $S^{(j)}$ and flux vector $\mathbf{v}^{(j)}$ (with $S = 0$ for inactive material species). Conversely, one can think of the stoichiometric matrix $S^{(j)}$ as defining species j. Note that $S^{(j)}$ is independent of location $\mathbf{x}$, whereas the distinguished solution $\mathbf{v}^{(j)} = \mathbf{v}^{(j)}(\mathbf{x})$ generally. Given, then, flux vectors $\mathbf{v}^{(j)}$, the growth rates g_j in the biomaterial equations (2) can also be replaced by largely parameter-free functions of biomass-producing (or removing) reaction fluxes, i.e.,

$$\frac{\partial}{\partial t}X_j + \nabla \cdot (\mathbf{u}X_j) = Y(h_j(\mathbf{v}^{(j)}) - d_j)X_j, \quad j = 1, 2, \ldots, N, \tag{10}$$

where $\mathbf{v}^{(j)}$ is the vector of fluxes involved with species j, $h_j(\mathbf{v}^{(j)}(\mathbf{x}))$ is the net specific biomass production rate for species j, d_j is a decay (or maintenance) coefficient, and Y converts biomass to volume.

2.3. Exchange fluxes. Internal and external (to cells) chemical concentrations are connected through exchanges of the form

$$c_k^{(j)} \rightleftharpoons C_k \tag{11}$$

between internal concentrations $c_k^{(j)}$ (kth substrate concentration, internal to species j) and external concentrations C_k. When convenient, we can regard (11) to be a reaction, as in (6), in a general sense. Exchange reactions serve as the couplings between the cellular metabolic model and microbial community model. That is, if

reaction (11) proceeds with flux $e_{jk}(\mathbf{x})$ (units of concentration per time per volume fraction) between the interior and exterior of material species j at macroscale location $\mathbf{x}$, then system (1) can be replaced by

$$D_k \nabla^2 C_k = -\sum_{j=1}^{N} e_{jk} X_j, \quad k = 1, 2, \ldots, M, \tag{12}$$

eliminating the need for the empirical rate functions r_{jk}. Note that $e_{jk}(\mathbf{x})$ measures the specific flux of metabolite k produced by material species j into the exterior domain. Conversely, exchange reactions appear in the stoichiometric matrix S as monocomponent reactions

$$c_k^{(j)} \rightleftharpoons ( ),$$

where the notation ( ) denotes exit from the internal cellular metabolic system. Each exchange reaction corresponds to a column of S containing all 0s except for a single 1 entry. The fluxes e_{jk} are then the components of solutions $\mathbf{v}^{(j)}$ to (9) corresponding to those same columns.

Fixing a flux for exchange (11) sets a constraint for the metabolic model: internal fluxes in and out of species j must balance this exchange flux. In practice, metabolic modelers do not necessarily fix exchange fluxes but rather might set upper bounds on them. We will do the same here. Choosing the value for an upper bound on flux into a cell is a physical problem depending on material transport properties of the particular species and also, possibly, biologically regulated transport mechanisms. Here we suppose that flux out of cells is unbounded or, less stringently, that the flux out of cells is never rate limiting, following metabolic modeling convention. Thus, flux e_{jk} for exchange (11) would be required to satisfy inequalities of the form

$$-E_{\mathrm{jk,max}} \leq e_{jk} < \infty. \tag{13}$$

Then, in practical terms, all of the rate functions r_{jk} and their attendant parameters are replaced by a set of maximum realizable influxes $E_{\mathrm{jk,max}}$, possibly determined as functions of exterior concentrations $\mathbf{C}$.

2.4. Optimization. Equations (9) for the metabolic models together with (10) and (12) for the macroscale biofilm model make up the foundations of a kinetics-free system. As already noted, though, system (9) is underdetermined and requires a selection principle for choosing a particular flux vector out of the full solution space. Flux balance analysis will typically make this choice based on a cellular level optimization principle such as maximum biomass production rate or maximum energy usage rate, though other methodologies are also possible [2, 58, 60]. Such cellular level optimization principles can be used here as well, though we are free to also incorporate in part or in full external, environmental information.

As an example of an environmentally based principle, and one that we will use below, we could choose fluxes in order to minimize external resource availability in some prescribed manner, with motivation that an ecologically stable community would find itself in an environment which is sparse enough that invasion by outside microbes cannot succeed. As a particular measure of resource availability, we will choose the enthalpy of combustion (heat released when chemical constituents are fully combusted with oxygen) in the exterior environment; other measures are certainly possible. Combustion enthalpy provides a rough estimate of available catabolic potential and, for the simple metabolic models considered here, a proxy for growth potential.

Two additional notes: first, with multispecies and spatially distributed communities comes extra freedom. Different species in one location as well as species (different or not) in different locations will have different choices of flux vectors and perhaps different optimization goals. As one example, one could ask that all flux vectors are chosen in concert so as to minimize the system-wide objective using L_1, or systemic, optimization. Alternatively, one could ask that independent species groups choose their own flux vectors so as to optimize their own, local objective, independently of strategies of other such groups. This latter strategy is, essentially, Nash optimization [46]: the system-wide flux vector profile is such that it is nowhere the case that a local perturbation would locally (only) improve optimality. Other minimization methodologies, beyond systemic and Nash, are also possible and, in general, different optimization schemes will lead to different results. An example of such is provided later.

Second, it is also possible to employ various sorts of optimization objectives. As one alternative, rather than (or in addition to) physically motivated objectives, fitting of empirical data might be used. Specifically, we can ask for flux vectors that result in system solutions that, in addition to or alternatively to resource optimization, best match a given criterion such as measured data. For instance, we can suppose that additional boundary concentration or flux measurements are available through observation, and ask that solutions of (9) are selected so as to produce model boundary concentrations or fluxes that best match these data; again, see computations presented later for an example.

2.5. Kinetic models. As an alternative to steady state assumptions, direct modeling of kinetics is also possible. Modeling kinetics involves tracking internal (to cells) metabolite concentrations and connecting them to reaction fluxes through constitutive relations; recall (7). Kinetics modeling has a number of advantages over steady state models, notably including the lack of requirement for optimization objectives which are at least oversimplifications and at worst lead to incorrect conclusions. And kinetics models have been applied to study of microbial systems including biofilms, e.g., [1].

Kinetics models are generally concentration based, and so connect naturally into large scale models like classical continuum systems such as (1)–(4), which are also concentration based, without explanation needed. For that reason, though they also fit the multiscale format used here, we do not discuss them in detail (though see [1] for an example). It should be noted, though, that (7) involves constitutive laws $\mathbf{v} = \mathbf{v}(\mathbf{c})$, which while often well characterized, require parameterization that can be problematic. Kinetics modeling has been implemented, even for fairly large systems, in special circumstances (e.g., hundreds of metabolites and reactions in a well-studied *E. coli* well-mixed, single-species system [36]). However, full microbial systems generally exhibit large, complex metabolisms distributed among complex multispecies communities and for which kinetics parameters can vary in unexpected ways in time and environment (not to mention the spatial variations that occur within biofilms); direct measurement of kinetics parameters is likely impractical in many such cases.

Estimating metabolic kinetic parameters from data is an active area of research at present, including uncertainty quantification, e.g., [6]. These methods are only recently being applied in microbial community models [17, 33, 56], despite the clear need, possibly due to the complexity of multispecies microbial interactions as well as the sparsity of appropriate data. We do not utilize them below—the methods presented here do not address and are not aimed at parameter estimation—though

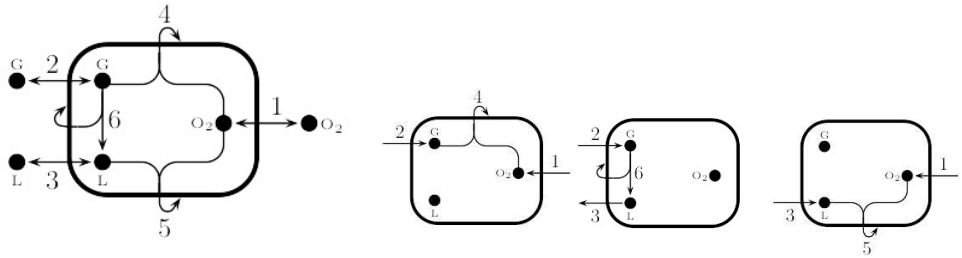

FIG. 1. *(Left) Metabolic model consisting of three metabolites: oxygen O_2, glucose G, and lactate L, and six reactions: exchange reactions* 1–3 *connecting interior and exterior pools of oxygen, glucose, and lactate respectively, as well as interior paths for glucose respiration (reaction* 4*), lactate respiration (reaction* 5*), and glucose fermentation (reaction* 6*). Reactions* 1–3 *are reversible and reactions* 4–6 *are nonreversible. Hooked arrows indicate biomass production. Stoichiometric coefficients are given in* (15)*. (Right) Elementary flux modes for the metabolic model on the left. From left to right: glucose respiration, glucose to lactate fermentation, and lactate respiration. See* (16) *for the corresponding flux vectors including stoichiometry.*

we do briefly illustrate the possibility of using available data to modify optimization objectives, a somewhat related issue.

3. Example system.

3.1. Metabolic model. To illustrate, we consider a system consisting of a single species biofilm grown on top of an agar base (providing an energy source, supposed here to be glucose) with exposure from above to air or bulk fluid (providing an oxygen source). The microbe is provided with a simplified metabolism (see Figure 1), in which glucose (Glu) converts to ATP either through full respiration in combination with oxygen (O_2) or, without oxygen, through fermentation. In the latter case, a fermentative product, supposed here to be lactate (Lac), is produced as a byproduct. Lactate itself can also be fully respired with oxygen to produce ATP. Note that this metabolic model is intended as a coarse graining of a true metabolic map which would include, generally, many more internal metabolites and internal reactions.

For this system, in the notation introduced above, $N = 1$ and $M = 3$, and the exchange and internal chemical equations are

$$\begin{aligned} \mathrm{O_2} &\rightleftharpoons (\ ), \\ \mathrm{Glu} &\rightleftharpoons (\ ), \\ \mathrm{Lac} &\rightleftharpoons (\ ), \\ \mathrm{O_2} + \mathrm{Glu} &\rightharpoonup \alpha_1 \mathrm{ATP}, \\ \mathrm{O_2} + \mathrm{Lac} &\rightharpoonup \alpha_2 \mathrm{ATP}, \\ \mathrm{Glu} &\rightharpoonup \mathrm{Lac} + \alpha_3 \mathrm{ATP}, \end{aligned} \tag{14}$$

where the first set of three equations denotes exchange between internal and external metabolites, and the second set of three equations denotes internal metabolism in lumped form. For convenience, we measure carbon containing components on a one carbon mole (Cmole) basis: glucose is scaled as $C_1H_2O_1$ and lactate is also scaled as $C_1H_2O_1$. "ATP" refers to extraction of energy (in the form of ATP generated from ADP) which serves here as a proxy for generation of new biomass [49]; see below. Note that a proportional relationship between ATP generation and biomass growth, though a common modeling tool, can be misleading particularly in slow growth situations. We make this assumption only to keep our metabolic model as simple as possible

and don't look to argue for or against its use in more complicated models. The stoichiometric coefficients α_ℓ, $\ell = 1, 2, 3$, do not need to be specified in this simple system; rather, we provide aggregate yield rates for new biomass; again, see below.

Note that system (14) assumes growth in a nutritionally complete media, such as a LB medium, which uses glucose (and lactate) as an energy source for, e.g., polymerization of monomers found in the medium. Additional sources (such as a nitrogen source) and products (e.g., H_2O) are suppressed in (14) because they are supposed to not be transport limited. Including them, and others, can allow the reactions to be elementally and electron balanced. We don't attempt to represent any actual microorganism though; rather, the stoichiometry and yields are only intended to be illustrative of the methodology. In fact, reactions (14) are heavily simplified in some ways—they implicitly connect ATP production with biomass yield.

The stoichiometric matrix for reactions (14) and for the metabolic model in Figure 1 is

$$S = \begin{pmatrix} 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{pmatrix} \tag{15}$$

with the six columns encoding the stoichiometry of the six reactions above and as numbered in Figure 1, and the three rows corresponding to the three interior metabolites oxygen, glucose, and lactate, respectively. A row for "biomass" is not included in S; rather, biomass-related stoichiometric coefficients enter into the external rate h; see (10) and below. A solution vector $\mathbf{v}$ of $S\mathbf{v} = \mathbf{0}$ has as its components fluxes through the six reactions. Some of the reactions are supposed unidirectional; we allow a solution vector $\mathbf{v}$ only if all of its unidirectional components are nonnegative. For this example, then, we require components 4–6 of $\mathbf{v}$ to be nonnegative. We also follow convention and, for the reversible exchange fluxes 1–3, denote flux from interior to exterior as positive.

Note that there is a large space of admissible flux vectors $\mathbf{v}$. This is typical and not special to this example. There are a variety of methods that have been used for distinguishing a particular solution $\mathbf{v}$ from this space. We choose one method, based on elementary flux modes [9, 13, 61, 71], that happens to be convenient for the example we present, but other methodologies are equally deployable.

Graphically, a solution of $S\mathbf{v} = \mathbf{0}$ can be viewed as a pathway through the metabolic map. Examples can be seen in the elementary flux modes (see Figure 1 right), for the example system. Elementary flux modes (EFMs) are a set of allowable pathways through the metabolic model (i.e., pathways that correspond to vectors in the null space of the stoichiometric matrix that satisfy nonnegativity constraints) which are minimal in the sense that an EFM contains no subpathway which is also an EFM. Here, the set of EFMs (unique up to normalization) correspond to the three metabolic modes of glucose respiration, glucose fermentation with lactate byproduct, and lactate respiration. Denoting them by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, and normalizing by input carbon content, we can write them in vector form as

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}. \tag{16}$$

For a general metabolic map, any allowable pathway vector can be shown to be a (generally nonunique) nonnegative linear combination of EFMs. For the simplified system here, it happens though that each allowable pathway vector does have a unique representation

$$\mathbf{v}(\mathbf{x}) = \beta_1(\mathbf{x})\mathbf{v}_1 + \beta_2(\mathbf{x})\mathbf{v}_2 + \beta_3(\mathbf{x})\mathbf{v}_3 \tag{17}$$

with flux amplitudes $\beta_j \geq 0$. Note that the number of EFMs generally increases rapidly with the size of the metabolic model [32] making an EFM representation increasingly complicated in more detailed systems.

Underdetermination of (8) manifests here in the freedom in (17); any nonnegative set of β_j's results in an admissible solution $\mathbf{v}$. Thus a scheme, including additional constraints such as (13), must be introduced to distinguish particular choices. This will be addressed below. Once chosen, however, the biomass production flux can be computed as $h_j(\mathbf{v}^{(j)}) = h(\beta)$; see (10). For the example here,

$$h(\boldsymbol{\beta}(\mathbf{x})) = Y_{\mathrm{GR}}\beta_1(\mathbf{x}) + Y_{\mathrm{GF}}\beta_2(\mathbf{x}) + Y_{\mathrm{LR}}\beta_3(\mathbf{x}) \tag{18}$$

with aggregate yield-rate coefficents indicating Cmoles of biomass produced per carbon normalized pathway reaction per time. We use values $Y_{\mathrm{GR}} = 19$, $Y_{\mathrm{GF}} = 0.5$, and $Y_{\mathrm{LR}} = 16$. Note that they are effectively arbitrary here, because time units are essentially arbitrary so that rates of generation of new biomass are also essentially arbitrary. However, their relative values matter—we have set glucose respiration to be a bit more efficient than lactate respiration, and both are significantly more efficient than glucose fermentation.

3.2. Exchange fluxes and the biofilm model. In this system there are three exchange fluxes of the form (11) between internal and external oxygen ($c_1 \rightleftharpoons C_1$), internal and external glucose ($c_2 \rightleftharpoons C_2$), and between internal and external lactate ($c_3 \rightleftharpoons C_3$). These three fluxes are denoted, since this is a one species system ($N = 1$), by $e_{1k} = e_k(\mathbf{x})$, $k = 1, 2, 3$. We bound inflow following (13), with $j = 1$, by

$$E_{k,\max} = \frac{\eta_k C_k}{K_k + C_k}, \tag{19}$$

where η_k determines a maximum transport rate and K_k determines transport rate saturation.

In terms of the flux amplitudes β_j (see (17)), the exchange flux functions $e_{jk}(\mathbf{v}(\mathbf{x})) = e_k(\mathbf{v}(\mathbf{x}))$, $k = 1, 2, 3$ (see (12)), are given by the components $e_k(\mathbf{v}) = -\mathbf{v} \cdot \mathbf{i}_k$ of $\mathbf{v}$ corresponding to exchange reactions, where $\mathbf{i}_k$ is a 6-vector with a 1 in the kth place and 0's elsewhere. In this example,

$$\begin{aligned} e_1(\mathbf{x}) &= -\beta_1(\mathbf{x}) - \beta_3(\mathbf{x}), \\ e_2(\mathbf{x}) &= -\beta_1(\mathbf{x}) - \beta_2(\mathbf{x}), \\ e_3(\mathbf{x}) &= \beta_2(\mathbf{x}) - \beta_3(\mathbf{x}), \end{aligned} \tag{20}$$

where the coefficients indicate pathway stoichiometry: moles of substrate k (here oxygen, glucose, and lactate) produced per pathway ℓ (here glucose respiration, glucose fermentation, and lactate respiration). In (20), e_1 is flux of oxygen, e_2 is flux of glucose, and e_3 is flux of lactate (following the sign convention that $c \rightharpoonup C$ is considered positive).

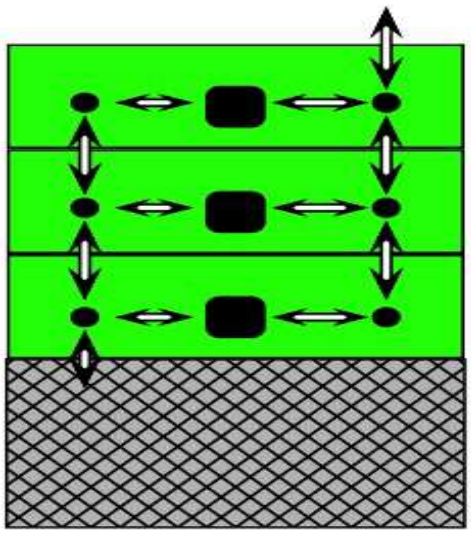

FIG. 2. *Macroscale biofilm (green solid) on solid substratum (gray hatched), open to air or bulk fluid above. Quantities vary in the vertical direction only, and each green layer corresponds to a different spatial location. Black boxes represent internal, microscale metabolic models (see Figire 1) in different layers and black circles represent exterior chemical concentrations in different layers. Horizontal arrows are exchange fluxes; vertical arrows are diffusive fluxes. Chemical quantities may or may not be able penetrate into or out of the substratum (permeability) or into/out of the air (volatility), encoded in the boundary conditions.*

With these specifications for exchange fluxes, the one dimensional biofilm model (1)–(5) reduces for this example to

$$D_k \frac{d^2}{dz^2} C_k = -e_k X, \quad k = 1, 2, 3, \tag{21}$$

with $z \in (0, L(t))$, where, from (2), (5), and (10), the height/thickness of the biofilm changes according to

$$\frac{d}{dt} L = \int_0^L Y(h(\mathbf{v}(z)) - d) dz \tag{22}$$

with e_k and h as described in (17), (18), and (20) above. Together with boundary conditions (given below), this system is complete up to determination of β_1, β_2, β_3. A cartoon is shown in Figure 2: the right-hand sides of (21) are represented, essentially, by horizontal arrows (exchange fluxes), and the left-hand sides are represented, essentially, by vertical fluxes (diffusive and other physical fluxes). In fact, physical fluxes could be considered as additional chemical reactions [21] and added in with the others, though such fluxes generally have well established constitutive relations with concentrations. The dark rectangles in Figure 2 are black boxes for purposes of the biofilm model; what happens inside is the domain of the metabolic model and is communicated to the biofilm model through exchange fluxes and biomass production. Biomass production results in pressure-driven expansion over long times (see (22)), which causes the biofilm to gradually expand.

To complete the model (biofilm plus metabolism), choice of a metabolic flux vector $\mathbf{v}$ satisfying $S\mathbf{v} = \mathbf{0}$, with S given in (15), is required. This amounts to fixing values

of the flux amplitudes β_1, β_2, β_3, which are arbitrary up to the requirement that all constraints be satisfied. In this example, the three EFMs are all unidirectional, so we require $\beta_j(\mathbf{x}) \geq 0$. We also assume that the exchange fluxes satisfy maximum influx constraints of the form (13). As $e_{jk} = e_k$ here, with bounds $E_{k,\max}$ as formulated in (19), then constraints (13) become

$$-e_k(\mathbf{x}) \leq \frac{\eta_k C_k}{K_k + C_k}, \quad k = 1, 2, 3. \tag{23}$$

Note that the right-hand sides of inequalities (23) depend on values of $\boldsymbol{\beta}(\mathbf{x})$ from the entire domain. Given inequality constraints (23), then, the optimization procedure is described next.

3.3. Optimization. Two issues need to be addressed: an optimization principle must be chosen, and a measurement procedure must be introduced. For the first, we can apply physically based optimization principles such as maximization of rate of biomass production or maximizing entropy production [24, 69]. Alternatively, a more directly biologically based principle, like maximization of biomass production can be used. Typically in metabolic analyses, these principles are implemented at the interior cellular level. Here, however we can optimize at the cellular interior or exterior, or system-wide levels, or any combination. As an example, we determine $\boldsymbol{\beta}$ by minimizing external enthalpy of combustion. A possible motivation is ecological—in order for a community to be stable against invasion, it might reduce the available (i.e., exterior to cells) energy to as low a level as possible, such as has effectively been observed in chemostat systems [27, 47, 62]. However, we make this choice of objective for illustrative purposes without claiming it to be better or worse than other possibilities, and it is not a necessary choice. External enthalpy of combustion density $H(\mathbf{x})$ is given by

$$H(\mathbf{x}) = \left(\sum_{k=1}^{M} \epsilon_k C_k(\mathbf{x}) \right), \tag{24}$$

where ϵ_k is the standard enthalpy of combustion per mole for substrate k. In the example system, $M = 3$ with $\epsilon_1 = 0$, $\epsilon_2 = 2805$, and $\epsilon_3 = 1344$, in units of kJ/mol, for substrates oxygen, glucose, and lactate, respectively [42]. The vector $\boldsymbol{\beta}(\mathbf{x})$ is chosen to minimize a measurement of $H(\mathbf{x})$.

In addition, we also present an entirely different type of optimization principle in which determination of $\boldsymbol{\beta}(\mathbf{x})$ is made so as to best match a given set of data. For example, given a profile of, say, $C_1(\mathbf{x})$, we can ask for a determination of $\boldsymbol{\beta}(\mathbf{x})$ such that the solution of the biofilm problem (at a fixed time t for simplicity) best matches that profile. The motivation comes from a sort of inverse problem: suppose that we know something of the genetic capability of a microbial population and, in addition, we have some measured environmental data, e.g., an oxygen profile. To what extent can we determine the metabolic activity of the system? At the same time, we may be able to predict unmeasured environmental data, e.g., the glucose and lactate profiles.

Whatever the optimization principle, some method of measurement of optimality is also required. Optimality of the argument of a spatially dependent function like that in (24) requires choice of some statistic. We illustrate with two different strategies. First, optimization is measured via Nash equilibration where the choice of $\boldsymbol{\beta}(\mathbf{x})$ is made at each location $\mathbf{x}$ without regard to its effects at other locations. That is, we consider $\hat{\boldsymbol{\beta}}(\mathbf{x})$ to be a Nash optimizer if, given any location $\mathbf{x}_0$, any allowable

perturbation to $\hat{\boldsymbol{\beta}}(\mathbf{x})$ in a neighborhood of $\mathbf{x}_0$ results in an increase in $H(\mathbf{x}_0)$. Second, as an alternative strategy, we optimize over the entire biofilm by choosing $\boldsymbol{\beta}(\mathbf{x})$ to minimize $\int_0^L H(z)dz = \|H\|_1$. Here the entire microbial community works together to minimize the total, system-wide enthalpy of combustion. Systemic optimization might be less biologically plausible than Nash equilibration (though perhaps there can be cooperation of some sort through signaling or other mechanisms [41, 43], and system-wide optima might suggest consortial strategies). However, the main motivation is illustrative; again we make no claims as to what method is best and only intend to illustrate the possible.

4. Parameters and methods. We present in section 5 below results of simulations based on the example biofilm system described in section 3. Technical details are described here.

4.1. Boundary conditions. Boundary conditions for the system are chosen to be representative of what might be used to simulate a biofilm with tissue interface ($z = 0$) and air interface $z = L$, where L is the coordinate of the top boundary of the biofilm: for oxygen (C_1), Dirichlet BC at the top $z = L$ and at the bottom $z = 0$; for glucose (C_2), Dirichlet BC at the bottom $z = 0$, no-flux BC at the top $z = L$; for lactate (C_3), zero Dirichlet BC at the bottom $z = 0$, no-flux BC at the top $z = L$. In particular, at the bottom boundary $z = 0$,

$$C_1|_{z=0} = 0, \quad C_2|_{z=0} = C_2^0, \quad C_3|_{z=0} = 0, \tag{25}$$

and, at the top boundary,

$$C_1|_{z=L} = C_1^L, \quad \left.\frac{\partial C_2}{\partial z}\right|_{z=L} = 0, \quad \left.\frac{\partial C_3}{\partial z}\right|_{z=L} = 0. \tag{26}$$

Effectively, glucose is supplied to the layer from below ($z = 0$) and oxygen from above ($z = L$). Lactate enters the system only internally via fermentation, though it can exit the system via the $z = 0$ boundary. It is assumed that glucose and lactate are unable to cross the upper boundary into the "air" region. At the bottom, the "tissue" interface, lactate and oxygen are removed from the system.

4.2. Model parameters. Lengths are scaled in units of 100 μm, e.g., $z = 1$ means $z = 100\,\mu$m; times are scaled in units of hours, e.g., $t = 1$ means $t = 1$ hr; and concentrations are scaled by a reference concentration 4×10^{-4} mol $\cdot$ l^{-1}, e.g., $C = 1$ means $C = 4\times10^{-4}$mol$\cdot$l^{-1}. See Table 1 for a list of parameter values used. In addition to parameter values already discussed elsewhere, diffusion coefficients are chosen based on measurements made in biofilm environments [64], conversion coefficient value Y is based an assumption for live cells of specific volume = 0.91 cm$^3\cdot$g^{-1}. As the "microbe" we simulate is essentially artificially constructed for illustrative purposes, we don't attempt to match parameters in constraint bounds (23) to the literature. Rather, values for these parameters are chosen to result in significant growth, for substrate concentrations present in simulations, on the designated system time scale of hours. Similarly, the decay rate d is not chosen based on literature reports (and in fact is quite large in comparison to measured values) but rather in order that the time scale for transition from thin to approximately steady thick biofilm regime also occurs on that system time scale.

4.3. Computational methods. Equations (21) are defined on a domain $z \in [0, L]$ with a moving upper boundary at $z = L(t)$. For convenience, a new spatial

TABLE 1

Parameter values used in computations. Other than K_js, η_js, and d, values are either free, given by stoichiometry, or measurable at the macroscale.

Symbol	Parameter	Value
C^0	Reference concentration of chemicals	4×10^{-4} mol $\cdot$ l^{-1}
C_1^L	Oxygen upper boundary condition	4×10^{-4} mol $\cdot$ l^{-1}
C_2^0	Glucose upper boundary condition	4×10^{-4} mol $\cdot$ l^{-1}
d	Biomass decay rate	2.5×10^{-4} s^{-1}
D_1	Oxygen diffusion coefficient	1×10^{-9} m$^2 \cdot$ s^{-1}
D_2	Glucose diffusion coefficient	6×10^{-10} m$^2 \cdot$ s^{-1}
D_3	Lactate diffusion coefficient	8×10^{-10} m$^2 \cdot$ s^{-1}
h_0	Characteristic length scale	10^{-4} m
K_1	Oxygen transport rate saturation	4×10^{-4} mol $\cdot$ l^{-1}
K_2	Glucose transport rate saturation	4×10^{-4} mol $\cdot$ l^{-1}
K_3	Lactate transport rate saturation	4×10^{-4} mol $\cdot$ l^{-1}
t_0	Characteristic time scale	3600 s
Y	Yield of volume per biomass	9.1×10^{-7} m$^3 \cdot$ g^{-1}
Y_{GF}	Yield of biomass per reaction	0.5 mol/mol
Y_{GR}	Yield of biomass per reaction	19 mol/mol
Y_{LR}	Yield of biomass per reaction	16 mol/mol
$\alpha_{n-1,1}$	Oxygen flux weight	0.7 kJ$\cdot$mol^{-1}
$\alpha_{n,3}$	Lactate concentration weight	0.7 kJ$\cdot$mol^{-1}
ϵ_1	Oxygen enthalpy coefficient	0 kJ$\cdot$mol^{-1}
ϵ_2	Glucose enthalpy coefficient	2805 kJ$\cdot$mol^{-1}
ϵ_3	Lactate enthalpy coefficient	1344 kJ$\cdot$mol^{-1}
η_1	Oxygen max transport rate	1.5×10^{-6} mol $\cdot$ l$^{-1}\cdot$s^{-1}
η_2	Glucose max transport rate	7×10^{-7} mol $\cdot$ l$^{-1}\cdot$s^{-1}
η_3	Lactate max transport rate	9×10^{-7} mol $\cdot$ l$^{-1}\cdot$s^{-1}

coordinate $\zeta = z/L(t)$ is introduced in order to obtain a fixed computational domain $\zeta \in [0,1]$, and (21) is written in terms of variable ζ. These equations are solved for $\zeta \in [0,1]$, and results are reported on the original spatial domain $z \in [0, L]$. Standard finite difference methods are employed with the spatial domain $[0,1]$ discretized into n subintervals of size $\Delta\zeta = 1/n$ ($n = 20$ in the computations shown below). See [72] for more details. Flux amplitudes and concentration values are tracked at the center of each subinterval, and denoted as follows: $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T)^T$, $\mathbf{C} = (\mathbf{C}_1^T, \mathbf{C}_2^T, \mathbf{C}_3^T)^T$, $\boldsymbol{\beta}_j = (\beta_{j,1}, \ldots, \beta_{j,n})^T$, $1 \le j \le 3$, $\mathbf{C}_k = (C_{k,1}, \ldots, C_{k,n})^T$, $1 \le k \le 3$. A given flux $\boldsymbol{\beta}$ determines the right-hand side of (21), and the corresponding concentration profiles $\mathbf{C}$ are obtained by solving (21). We denote this mapping from $\boldsymbol{\beta}$ to $\mathbf{C}$ by a function $\mathbf{C} = G(\boldsymbol{\beta})$.

(1) For a fixed biofilm thickness L, the flux amplitudes $\boldsymbol{\beta}$ and the corresponding concentration profiles $\mathbf{C}$ within the biofilm are computed by solving the optimization problem discussed earlier, with details given below. (2) In the case of growing biofilm simulations, an initial biofilm thickness L_0, a total simulation time t_F, and a growth time interval Δt are chosen. (Results shown below are for $L_0 = 0.1, t_F = 15, \Delta t = 0.02$.) Then, at each time $t_m = m\Delta t$ and biofilm thickness L_m (at that time), the flux $\boldsymbol{\beta}_m$ is computed by optimization as in the fixed thickness computations, the rate of change of the thickness $\dot{L}_m$ is computed by (22), and then the thickness is updated by $L_{m+1} = L_m + \dot{L}_m \Delta t$. Time stepping continues until the total simulation time t_F is reached.

Enthalpy of combustion density given by (24) is defined in each discretization subinterval as $H(\vec{C}_i) = \sum_{k=1}^{3} \epsilon_k \, C_{k,i}, \ 1 \le i \le n$, where $\vec{C}_i = (C_{1,i}, C_{2,i}, C_{3,i})^T$. For Nash optimization, there are n objective functions $F_{\mathrm{Nash},i}, 1 \le i \le n$, one for each subinterval. The input of $F_{\mathrm{Nash},i}$ is the value of flux vector $\boldsymbol{\beta}$ on the ith subinterval

while its values on all other subintervals are fixed and denoted by $\vec{\beta}_i$, and the output of $F_{\text{Nash},i}$ is the enthalpy of combustion density on the ith subinterval, $H(\vec{C}_i)$, and so

$$F_{\text{Nash},i}(\vec{\beta}_i) = H(\vec{C}_i), \quad \text{where } \vec{\beta}_i = (\beta_{1,i}, \beta_{2,i}, \beta_{3,i})^T, \ \mathbf{C} = G(\boldsymbol{\beta}),\ 1 \le i \le n. \tag{27}$$

In the case of system optimization, there is one objective function F_{Sys} for the entire domain. The input for F_{Sys} is the entire flux vector $\boldsymbol{\beta}$, and the output of F_{Sys} is the midpoint quadrature approximation of the L_1 norm of the enthalpy of combustion density $H(\mathbf{x})$. Thus,

$$F_{\text{Sys}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} H(\vec{C}_i)\Delta\zeta. \tag{28}$$

The discrete version of the influx constraints given by (20) and (23) on the ith subinterval is

$$A \cdot \vec{\beta}_i \le B_i(\boldsymbol{\beta}), \tag{29}$$

where

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \quad \vec{\beta}_i = \begin{bmatrix} \beta_{1,i} \\ \beta_{2,i} \\ \beta_{3,i} \end{bmatrix}, \quad B_i(\boldsymbol{\beta}) = \begin{bmatrix} B_{1,i} \\ B_{2,i} \\ B_{3,i} \end{bmatrix}, \quad B_{k,i} = \frac{\eta_k C_{k,i}}{K_k + C_{k,i}}.$$

Since the right-hand sides of the constraints (29) also depend on the unknown $\boldsymbol{\beta}$'s, an iterative procedure is applied to compute $\boldsymbol{\beta}$. The following algorithm is used to find the optimal flux $\boldsymbol{\beta}$ and the associated chemical concentrations $\mathbf{C}$ for a fixed biofilm thickness.

Pick initial value $\boldsymbol{\beta}^{(0)}$, set Error = 1, tol = 10^{-6}
while Error > tol
  Obtain $\boldsymbol{\beta}^{(k+1)}$ from $\boldsymbol{\beta}^{(k)}$ by the optimization step (details given below)
  Error = $|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}|/|\boldsymbol{\beta}^{(k)}|$
end

In the case of Nash optimization, the optimization step involves looping over all subintervals. Namely,

$\tilde{\boldsymbol{\beta}}^{(0)} = \boldsymbol{\beta}^{(k)}$
for $i = 1$ to n % loop over subintervals
  $\underset{\vec{\beta}_i}{\text{minimize}} \quad F_{\text{Nash},i}(\vec{\beta}_i)$
  subject to $A \cdot \vec{\beta}_i \le B_i(\tilde{\boldsymbol{\beta}}^{(i-1)}),\ \vec{\beta}_i \ge 0$
  set $\tilde{\boldsymbol{\beta}}^{(i)}$ as $\tilde{\boldsymbol{\beta}}^{(i-1)}$ with ith component replaced by $\vec{\beta}_i$
end
$\boldsymbol{\beta}^{(k+1)} = \tilde{\boldsymbol{\beta}}^{(n)}$

For systemic optimization, the optimization step is

$$\underset{\boldsymbol{\beta}^{(k+1)}}{\text{minimize}} \quad F_{\text{Sys}}(\boldsymbol{\beta}^{(k+1)})$$
$$\text{subject to} \quad A_s \cdot \boldsymbol{\beta}^{(k+1)} \le B(\boldsymbol{\beta}^{(k)}),\ \boldsymbol{\beta}^{(k+1)} \ge 0$$

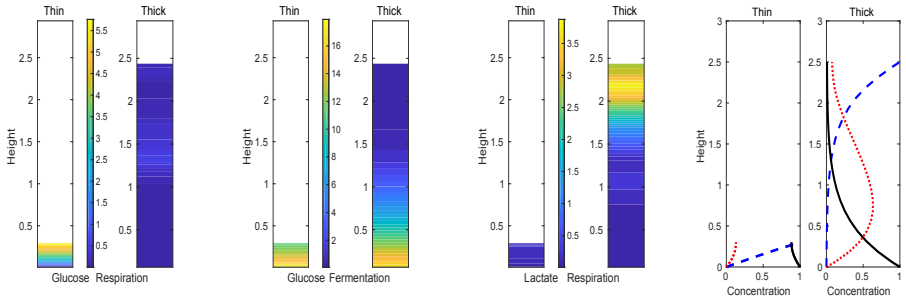

FIG. 3. *Snapshots of thin and thick biofilms. Bar plots show amplitudes of the EFMs—each pair has the thin biofilm on the left and the thick biofilm on the right. Color bar in-between indicates amplitude. Far left: glucose respiration. Middle left: glucose fermentation. Middle right: lactate respiration. Far right: line plots show chemical concentrations (dimensionless units, scaled by* $4 \times 10^{-4} mol \cdot l^{-1}$*), respectively, the thin and thick biofilm snapshots. (Blue dashed) oxygen concentration. (Black solid) glucose concentration. (Red dotted) lactate concentration. Note that the thin biofilm shows a profile combining glucose respiration and fermentation. The thick biofilm predominantly ferments glucose to lactate near the* $z = 0$ *interface and respires lactate near the* $z = 2.5$ *interface.*

with

$$A_s = A \otimes I_n, \quad B(\boldsymbol{\beta}^{(k)}) = (B_{1,1}, \ldots B_{1,n}, B_{2,1}, \ldots, B_{2,n}, B_{3,1}, \ldots, B_{3,n})^T(\boldsymbol{\beta}^{(k)}),$$

where $\otimes$ denotes the Kronecker product and I_n is the identity matrix of size n. In both cases, the constrained minimization problem is solved using the MATLAB function `fmincon`.

5. Results.

5.1. Fixed layers. We begin with fixed time snapshots of thin and thick biofilm layers; see Figure 3. In each, we consider a biofilm layer of fixed size with microbial metabolism as described in Figure 1 and section 3.1 at steady state. Oxygen, glucose, and lactate diffuse through the layer and exchange internally/externally as described in section 3.2, and microbes at each spatial location optimally allocate their available resources locally, meaning that chemical flux into the local microbes is allocated to the three available metabolic pathways (glucose respiration, glucose fermentation, and lactate respiration) in such a way as to minimize local exterior (to cells) enthalpy of combustion, i.e., computations reflect Nash equilibria based on environmental chemical concentrations as measured through enthalpy of combustion. Note: we have also used, in place of minimization of enthalpy of combustion, maximization of biomass production rate (18) as the optimization objective [26], with little change in results. For systems in which growth rate and enthalpy reduction are less closely linked, this might not be the case.

In the thin layer, relatively little consumption of resources occurs so that glucose penetrates fully. The same is essentially true for oxygen, though there is diffusive transport from the "air" boundary ($C_1|_{z=L} = 1$) to the "tissue" boundary ($C_1|_{z=0} = 0$) so that oxygen concentration is forced to 0 at $z = 0$ by the boundary condition. We see, notably, pathways for both respiration and fermentation of glucose are active throughout. Lactate respiration is mostly absent as oxygen is used preferentially for glucose respiration. Note that the model predicts glucose fermentation even in the presence of oxygen, consistent with observations (e.g., [29]), and with the overflow effect [4]. Fermentation in oxygen rich conditions occurs here as a consequence of a mismatch between maximal uptake rates for oxygen and glucose; after allocation

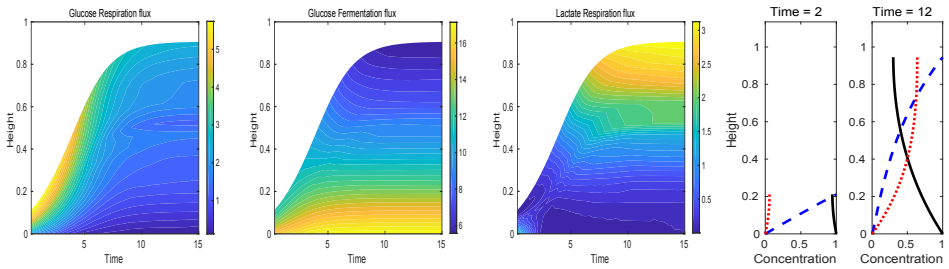

FIG. 4. *Growing biofilm. Left three panels are amplitudes of the three EFMs, horizontal axes are time, vertical axes are height; height* 0 *is the substratum (the "tissue" interface), and the boundary between shaded and unshaded areas is top of the biofilm (the "air" interface). Shading indicates amplitude of flux, unshaded regions are exterior to the biofilm. Far left: flux through the glucose respiration pathway. Middle left: flux through the glucose fermentation pathway. Middle right: flux through the lactate fermentation pathway. Far right: plots show external chemical concentrations (dimensionless units, scaled by* 4×10^{-4} *mol* $\cdot$ l^{-1}*) at* $t = 2$ *and* $t = 12$*: (blue dashed) oxygen concentration, (black solid) glucose concentration, (red dotted) lactate concentration. Note: at early times, the flux profile resembles that of a thin biofilm as in Figure* 3 *dominated by a combination of glucose respiration and glucose fermentation. At later times, the flux profile resembles that of a thick biofilm as in Figure* 3 *dominated by glucose fermentation near the "tissue" interface where oxygen levels are low and dominated by lactate respiration near the "air" interface where oxygen levels are high. At intermediate times, where the flux profile changes from thin biofilm to thick biofilm, the flux of new biomaterial also changes, reflecting the system-wide drop in metabolic efficiency and slowing of biofilm growth.*

of all available oxygen influx to respiration, excess capacity for influx of glucose is directed to fermentation. We don't claim that this mechanism is the same as the one responsible for observations.

By contrast, in the thick biofilm computation, we observe predominantly glucose fermentation metabolism in a lower sublayer where oxygen is unable to penetrate, and lactate respiration in an upper sublayer where oxygen is available but glucose is not. Note that lactate is produced in the lower layers and transferred to the upper layer by diffusive transport where it is consumed. Consequently, glucose respiration is mostly absent because glucose is largely depleted before it reaches the oxygen rich region. Similar profiles have been observed in single species biofilm observations [55].

5.2. Growing biofilm. Next, we consider a growing biofilm (see Figure 4), allowing the biofilm height to change according to (22). The left three panels of Figure 4 show both the biofilm height as well as EFM amplitudes in time (horizontal axes are time; vertical cross sections roughly correspond to the bar graphs in Figure 3). Note that at early times, the distribution of fluxes in the biofilm is similar to that seen in the thin biofilm snapshot, Figure 3, and at later times, the fluxes in the biofilm transition to a distribution similar to that seen in the thick biofilm snapshot, Figure 3.

Observe that growth rate slows at the transition between thin and thick regimes, where glucose respiration is replaced by glucose fermentation to lactate combined with lactate respiration. In our simplified model, the latter is not significantly less efficient than the former. However, there is a penalty paid for diffusive transport of lactate from where it is produced (near $z = 0$) to where it is consumed (near $z = L$). This transport penalty increases as biofilm height L increases, eventually approaching a cost at which growth and loss balance.

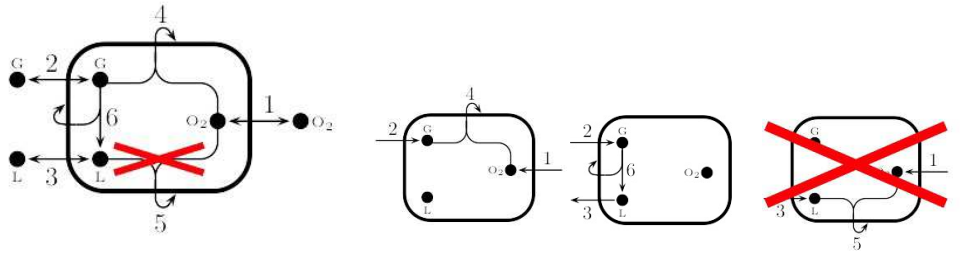

FIG. 5. *Reduced metabolism, compared to Figure* 1*, without ability to respire lactate. Stoichiometric matrix is the same in* (15) *except with column* 5 *deleted. (Left) Metabolic model with suppressed reaction crossed out. (Right) EFMs for the metabolic model on the right, identical to those shown in Figure* 1 *except that the lactate respiration pathway is unavailable.*

5.3. Nash versus systemic optimization. To illustrate the importance of choice of optimization measure, we present an example involving a "mutant" organism which has lost the ability to respire lactate; see Figure 5. That is, the mutant can only respire glucose or ferment it to lactate. In this reduced metabolism, glucose fermentation is unfavorable in the sense that energy in lactate, the fermentation product, becomes lost to the system. Thus, if the system were to somehow coordinate its metabolism over the entire biofilm, for example, via signaling of some sort, it might choose to avoid fermentation.

We consider again a fixed time biofilm snapshot and compute the metabolic distribution to minimize enthalpy of combustion, as always, but using two different measures of optimality: (1) Nash optimality (as applied in other computations as well as in this subsection) where metabolism is chosen based on local, pointwise, enthalpy without regards to nonlocal consequences, and (2) systemic optimality (used only in this subsection) where metabolism is chosen based on total system enthalpy (L_1 norm). In the systemic case, microbes in different locations cooperate to produce a globally efficient biofilm.

See Figure 6 for an example computation. Note that the metabolic distributions are qualitatively different for the two optimization measures. In the Nash case, glucose is consumed as it enters the biofilm from the substratum $z = 0$ mostly via fermentation in the lower region, where oxygen is depleted. In the systemic case, fermentation is highly unfavorable in comparison to respiration, so glucose is allowed to be diffusively transported toward the top of the biofilm where more oxygen results in increased respiration and, despite the transport penalty, overall decrease in system-wide enthalpy. Looking at a comparison of the local enthalpy density in Figure 6, we see that the Nash system chooses a small local advantage from fermenting glucose to lactate near the substratum leading, though, to (comparatively) significant excess enthalpy density above from unused lactate.

We remark that a contrast between Nash and systemic equilibrium is clear in this mutant system lacking lactate respiration capability. In the full metabolism, though, including lactate respiration capability, Figure 1, the two equilibria are similar somewhat coincidentally: Nash optimized results are already reported in Figure 3, the systemic optimization results (not shown) are similar. In particular, it is not perhaps obvious why systemic optimization chooses to ferment glucose near the bottom of the biofilm. The reason to do so is to minimize the effect of the diffusive transport penalty. Systemic optimization chooses to ferment glucose near the substratum and then diffusively transport lactate to the upper, oxygenated region, paying the diffusion transport penalty in lower enthalpy lactate rather than in glucose.

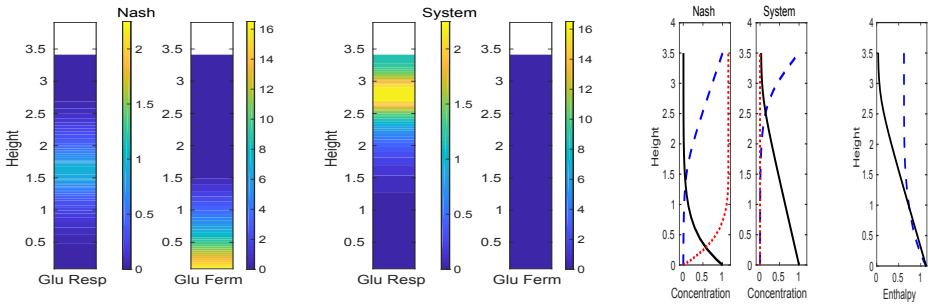

FIG. 6. *Snapshots of Nash-optimized and systemic-optimized biofilms with lactate respiration inhibited as in Figure* 5. *Bar plots show amplitudes of the active EFMs. Left: Nash optimized glucose respiration and fermentation flux amplitude. Middle: systemic optimized glucose respiration and fermentation flux amplitude. Right: first two line plots show chemical concentrations (dimensionless units, scaled by* $4 \times 10^{-4}\, mol \cdot l^{-1}$*) for, respectively, the Nash and systemic optimizations, with (blue dashed) oxygen concentration, (black solid) glucose concentration, (red dotted) lactate concentration. Third line plot shows density of enthalpy of combustion in (blue dashed) Nash optimization and (black solid) systemic optimization. Note that the Nash optimized biofilm is dominated by glucose fermentation near the* $z = 0$ *interface with a small region above of where the remaining glucose is respired, while, in the systemic optimized biofilm, glucose is transported diffusively to a respiration zone near the top of the biofilm. The plot of enthalpy density shows that there is a small local advantage to fermenting glucose near* $z = 0$ *at the cost of a large increase in global enthalpy due to unused lactate.*

5.4. Data integration. In some situations, additional data are available beyond metabolic capabilities, and in such cases it may be desirable to incorporate this information into a model. For example, given a set of observations of a microbial system and a hypothesized set of metabolic capabilities for its microbes, how do or don't the observations constrain the actual metabolic activity, i.e., constrain what the microbes are doing and where they are doing it? This is an inverse problem mathematically, and difficult generally, but useful. A natural additional question: can we identify more or less helpful additional measurements, that is, measurements that have more or less constraining impact on our model results? This latter constitutes something of a sensitivity analysis.

Short of exploring these questions in depth here, we propose a method to begin moving in those directions by modifying the optimization objective (24) to account for additional known data. Particularly, if concentration data are known or inferred at certain given locations y_ℓ, $\ell = 1, 2, \ldots, \mathcal{L}$, then we look to minimize using an augmented energy function of the form

$$\begin{aligned}\widehat{H}(C_1(x), C_2(x), \ldots, C_M(x), x) &= H(C_1(x), C_2(x), \ldots, C_M(x)) \\ &\quad + \sum_{\ell=1}^{\mathcal{L}} \sum_{m=1}^{M} \alpha_{\ell,m}(x) |C_m(y_\ell) - \widetilde{C}_m(y_\ell)|, \end{aligned} \tag{30}$$

where $\widetilde{C}_m(y_\ell)$ are the additional concentration data to be integrated and $\alpha_{\ell,m}(x)$ are weights (set to zero where data are not available). For example, see below, we may have additional information on concentrations at the boundary of a biofilm and wish to understand how that might constrain the metabolic profile inside the biofilm. By varying the magnitude of the weights $\alpha_{\ell,m}$, the relative importance of the two

components of $\widehat{H}$ can be varied, though we do not explore this possibility here. Note that time dependence can also be allowed in (30) but is not included here for simplicity.

To illustrate, we employ a synthetic example based on the full metabolism, Figure 1 (with lactate respiration capability, designated LR positive, or LRP), and on the reduced metabolism, Figure 5 (without lactate respiration capability, designated LR negative, or LRN), as follows: we first compute biofilm snapshot concentration profiles for the LR negative metabolism (5), identical in fact to the Nash optimized solutions shown in Figure 6, and pretend that these profiles are actually a "real-life" experimental system from which we are able to extract only certain restricted bits of data via measurements of some sorts. Next, we provide that limited data, in this case, some information on boundary concentrations or fluxes at the top $z = 2.5$ for oxygen and/or lactate, to the fully capable LR positive model (1). That is, we pretend to take a measurement of oxygen and lactate concentration or flux at the top of a real-life biofilm using the computed LR negative profiles as values for those measurements, then feed them to the LR positive model. The aim is to investigate how "data" taken from the real-life LR-negative profiles and provided to the full-capability, LR-positive solver through objective (30) might impact predictions of a model with full, LR-positive capability. In particular, can the full-capability model recognize, with the help of experimental data, a restricted capability organism?

For all computations we use the same set of boundary conditions including, in particular, conditions (26) at the top of the biofilm layer. We provide "measured data" from the LR negative profiles to the LR positive model as follows. For lactate we can suppose that in addition to the no-flux condition on lactate at the top boundary, we also can measure the lactate concentration there (taken to be the LR negative concentration at $z = 2.5$). For oxygen we suppose that in addition to the given Dirichlet condition at the top, we also can measure the oxygen flux there (taken to be the top boundary oxygen flux from the LR negative computation). Mathematically, using the optimization of (30), we try to force an additional lactate Dirichlet condition onto the already imposed Neumann one, and/or for oxygen we try to force an additional Neumann condition onto the already imposed Dirichlet one.

We consider the following cases for fitting:

1. LR positive without any additional LR negative data (LRP);
2. LR positive supplemented by the upper interface oxygen flux value "measured" from the LR negative profile (LRP+Oxy);
3. LR positive supplemented by the upper interface lactate concentration value measured from the LR negative profile (LRP+Lac);
4. LR positive supplemented by both the upper interface lactate concentration and oxygen flux values measured from the LR negative profile (LRP+Oxy +Lac).

As described, in case 1 we use boundary conditions as already imposed in (25)–(26) while in cases 2–4 we supplement those boundary conditions with the extra boundary data as follows. Prescription of lactate concentration is implemented numerically by setting the lactate concentration at the upper boundary node. Prescription of oxygen flux is equivalent to prescribing the derivative of oxygen concentration. Numerically this amounts to setting the oxygen concentration at the first discretization node into the domain from the boundary (oxygen concentration at the boundary node is already fixed by the Dirichlet boundary condition). In terms of a discretized version of (30), these amount to supplying $\widetilde{C}_3(y_n)$ and/or $\widetilde{C}_1(y_{n-1})$, where y_n is the location of the

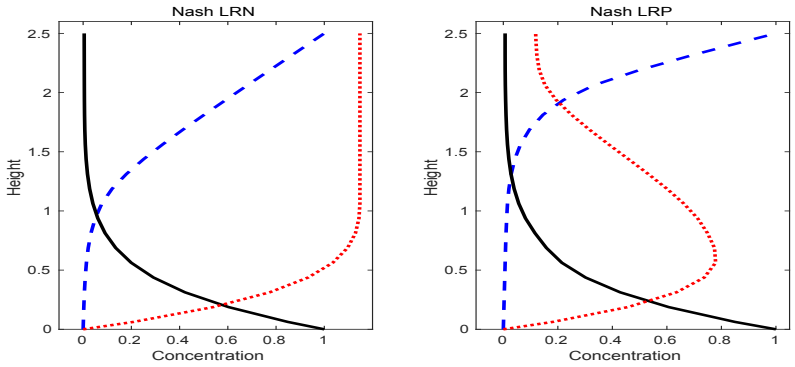

FIG. 7. *Substrate concentrations (dimensionless units, scaled by* 4×10^{-4} *mol* $\cdot$ l^{-1}*) in the LRN biofilm (left). and the LRP biofilm (right). (Blue dashed) oxygen concentration. (Black solid) glucose concentration. (Red dotted) lactate concentration.*

top discretized point in the biofilm. Augmented objective (30) becomes

$$\begin{aligned}\widehat{H}(C_1(x), C_2(x), C_3(x), x) = H(C_1(x), C_2(x), C_3(x)) \\ + \alpha_{n-1,1}(x)|C_1(y_{n-1}) - \widetilde{C}_1(y_{n-1})| \\ + \alpha_{n,3}(x)|C_3(y_n) - \widetilde{C}_3(y_n)|.\end{aligned} \tag{31}$$

For weights we use $\alpha_{n-1,1}$, $\alpha_{n,3} = 0.7$ or 0 depending on the case.

See Figure 7 for concentration profiles for LR negative (left) and LR positive (right) profiles for reference. Note the most evident difference is in the lactate concentration profiles. For both LR-negative and LR-positive, lactate is produced in approximately the same region near the lower biofilm boundary. However, the LR-positive biofilm respires it above in the region where oxygen is available, while the LR-negative biofilm is unable to respire it at all. The aim, then, is to supplement the LR-positive computations with the extra data described above in order to, using an optimization objective (31), investigate how the new predictions resemble those for the LR-negative system particularly with respect to lactate concentration profile.

We show predicted concentration profiles for fitting cases 1–4 in Figure 8, columns 1–4, respectively. First, for reference again, column 1 compares LR-positive (without additional data, i.e., $\alpha_{n-1,1} = \alpha_{n,3} = 0$) profiles to LR-negative profiles. Next, the effects of adding oxygen boundary-flux data to (31) ($\alpha_{n-1,1} = 0.7$, $\alpha_{n,3} = 0$) are shown in column 2. Observe that the extra information about oxygen flux has little effect on predicted profiles. Though one might hypothesize that net oxygen use (here largely determined by oxygen influx at the top boundary) is important and hence the extra oxygen flux data would be informative, in this example there is in fact relatively little difference in total oxygen use between the LR negative and positive systems so that the extra data have little effect on predictions. In contrast see column 3 where the LR positive system is augmented with lactate concentration data from the top boundary ($\alpha_{n-1,1} = 0$, $\alpha_{n,3} = 0.7$). Perhaps surprisingly, this extra information results in very close fits. Note the lactate concentration fits in the bottom row of Figure 8. Both LR negative and LR positive systems generate lactate from glucose near the bottom of the biofilm where oxygen concentration is low (lower left subplot of Figure 8). But it seems that forcing the LR positive system to, at the top boundary of the biofilm, match lactate concentration with the LR negative system effectively disallows lactate

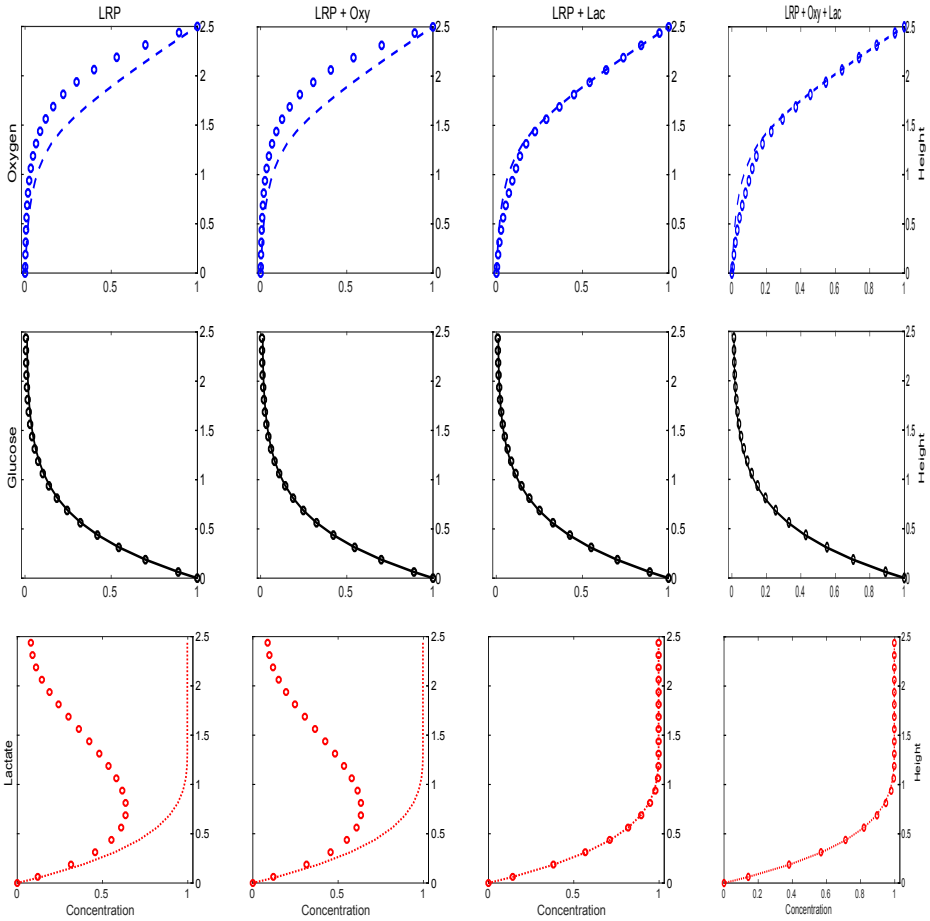

FIG. 8. *Comparisons between LR negative concentration profiles and data-augmented LR positive concentration profiles. Concentration units are dimensionless, with scaling by* $4 \times 10^{-4}\, mol \cdot l^{-1}$. *Top row: oxygen. Middle row: glucose. Bottom row: lactate. First column: LR negative versus LR positive. Second column: LR negative versus LR positive + extra oxygen flux data. Third column: LR negative versus LR positive + extra lactate concentration data. Fourth column: LR negative versus LR positive + extra oxygen flux data + extra glucose concentration data. In each plot, the line style curve is the LR negative concentration profile and the circle symbols are the data-augmented LR positive concentration profile. Note that oxygen data augmentation does not improve the augmented LR positive predictions, while lactate data (with or without the oxygen data) does.*

loss via respiration. Finally, for completeness, column 4 has plots from case 4, adding both the oxygen and lactate boundary data ($\alpha_{n-1,1} = \alpha_{n,3} = 0.7$), showing little change from the column 3 profiles. Overall, this example, as artificial as it might be, still suggests the utility of the methodology for assisting in experimental design.

The quality of concentration fits does not extend to pathway flux amplitude fitting; see, for example, Figure 9 where we show profiles of lactate respiration flux amplitudes in the various cases 1–4 of the data-augmented LR positive system. The noise in the predicted pathway flux amplitudes is a consequence of the underdetermined linear relation between pathway fluxes β_j and exchange fluxes e_k (see (20)) which have a one dimensional null space; recall that exchange fluxes determine chemical concentrations through (21). That is, there are many different combinations of pathway fluxes which can lead to the same exchange fluxes. Hence fitting concentration data is inde-

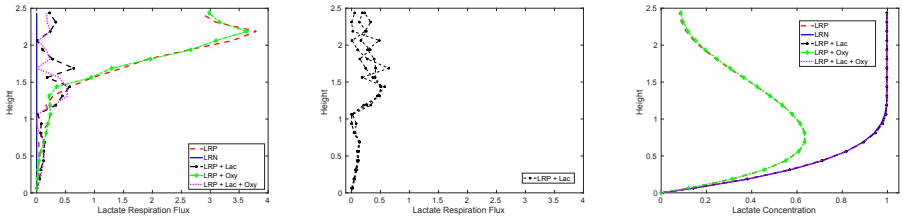

FIG. 9. *Amplitude of lactate respiration pathway flux for fitting cases LRP, LRP+Oxy, LRP+Lac, LRP+Oxy+Lac; see left panel (fluxes are in dimensionless units, scaled by* $1.1 \times 10^{-7}\, mol/(l \cdot s)$*). The solid blue line at* 0 *is the amplitude of lactate respiration pathway flux for the LR negative system. Middle panel: amplitude of lactate respiration pathway flux for three different realizations of the same LRP+Lac system. The noise is a consequence of indeterminancy in the solution. Noise also appears in all other pathway flux profiles (not shown). Right panel: for reference, lactate concentration (dimensionless units, scaled by* $4 \times 10^{-4}\, mol \cdot l^{-1}$*) profiles resulting from pathway fluxes (left and middle panels) are shown. The concentration curves for LRP and LRP+Oxy lie nearly on top of each other, as do the concentration curves for LRP+Lac, LRP+Lac+Oxy, and LRN, regardless of variation in the pathway fluxes. For example, all three of the lactate flux realizations shown in the middle panel result in nearly the same LRP+Lac concentration fit in the right panel.*

terminant. To illustrate, the middle panel of Figure 9 shows different realizations of the lactate respiration flux predicted by the LR positive system augmented by lactate boundary data (glucose respiration and fermentation pathway fluxes not shown); each realization, however, results in concentration profiles nearly identical to those seen in Figure 9, right panel. As a remark, though the metabolism used here is very simplified, this sort of indeterminancy is likely a general feature, meaning that (exterior) concentration data may not be an effective way to determine pathway amplitudes.

The significance of the close match in columns 3 and 4 of Figure 8 should not be overstated as both the LR negative and LR positive profiles arise from the same computational system except with different metabolic models. That is, the data are generated from a similar system as that used for fitting. Adding artificial noise (not shown) does not change the conclusions in any substantial way, probably for the same reason. Nevertheless, this example demonstrates the ability for easily connecting to experimental data and design. The pathway amplitude noise seen in Figure 9 is also illustrative, as it is likely generally the case the concentration data are not sufficient to uniquely determine the pathway flux profile.

6. Discussion. Population level models of microbial communities are burdened by often difficult to measure kinetics and regulatory parameters. Further, the numbers of such parameters can rapidly increase as demands, fueled by availability of increasingly detailed molecular data, are made for more detailed descriptions of microbial community function. The result can be uncertainty in a high dimensional parameter space. However, at the cellular level and based on a steady state assumption, the metabolic modeling community has already independently recognized and developed methodologies to mitigate kinetic parameter overload and regulatory ambiguity. Our aim here has been to propose a procedure to marry steady state metabolic methods to population scale modeling, and to do so in a general fashion that allows flexibility in the mixing and matching of different strategies both at the cellular and population levels (we employ an EFM-based method in examples, but such is not by any means required). At the same time, as we couple cellular scale metabolic models to the large community and environmental scale, we have been able to incorporate that

large scale environmental information back to the cellular level. Though not addressed here, it should further be possible to also incorporate molecular information beyond genomics-derived metabolic maps. Also not addressed here, because we use artificial examples without inherent noisiness, are the challenges of uncertainty quantification which become important upon introduction of real data.

Steady state metabolic modeling itself generally requires some selection principle for choosing an identified solution from an underdetermined set of possible solutions. We can employ standard strategies like maximization of biomaterial production. However, one potentially useful feature of the multiscale approach we propose is the possibility of naturally coupling larger scale environmental conditions to this selection, including thermodynamic or empirical data, and at local or nonlocal scale. Along the same lines, integration and inversion of data are also natural, either from large scale to small scale or vice versa. As an aside, in the process of exploring even only a few possibilities, we've observed that results can depend qualitatively on choice of optimization principle, at least within the simple and artifical metabolic model used here. For example, choice of localized (Nash) optimization vs nonlocalized (systemic) optimization (section 5.3) can lead to qualitatively important macroscale differences, as can small variations in data integration (section 5.4). Even seemingly minor choices such as optimizing for ATP production versus mass/ATP yield (not shown) can have consequences on the macroscale. Such structural instability is already familiar to the metabolic modeling community, and not surprising in retrospect. It can be considered both a danger and an opportunity (as it raises the possibility of experimental testing—indeed, one of our main aims is to facilitate the comparison of community scale with microscale metabolism). In any case, macroscale modelers should be aware of this issue.

We have illustrated our methodology in the context of a one dimensional biofilm model, a moderately complicated system mathematically, but expect that the principles extend to more (e.g., multidimensional, multispecies biofilms in fluid or viscoelastic fluid environments) or less (e.g., batch or chemostat reactors) complex setups and, indeed, to microbial communities beyond biofilms specifically; to change, essentially, replace, the left-hand sides of (10) and (12) with different system scale physics. Indeed, as long as the system time scale is long compared to the time scale for metabolisms to approach steady state, then the multiscale approach presented here seems applicable. In the absence of a separation of time scales, the validity of steady state metabolic modeling might be questioned.

On the metabolic side, we have used a model characterized by EFMs. Any other flux-driven model platform could be used instead. Further, we have supposed a heavily simplified, single organism metabolism. We have no reason to expect that moderately more complicated metabolisms cannot be readily substituted. However full (e.g., genome scale) metabolic models combined with more complex systems, like biofilms, would likely pose computational challenges requiring more attention to efficient numerical algorithms.

REFERENCES

[1] E. Alpkvist and I. Klapper, *A new multidimensional multispecies continuum model for heterogeneous biofilm development*, Bull. Math. Biol., 69 (2007), pp. 765–789.

[2] M.R. Antoniewicz, *Methods and advances in metabolic flux analysis: A mini-review*, J. Ind. Microbiol. Biotechnol., 42 (2015), pp. 317–325.

[3] J.E. Bailey, *Complex biology with no parameters*, Nature Biotech., 19 (2001), pp. 503–504.

[4] M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J.R. Williamson, and T. Hwa, *Overflow metabolism in* E. coli *results from efficient proteome allocation*, Nature, 528 (2015), pp. 99–104.

[5] A. Bordbar, J.M. Monk, Z.A. King, and B.Ø. Palsson, *Constraint-based models predict metabolic and associated cellular functions*, Nat. Rev. Genet., 15 (2014), pp. 107–120.
[6] D. Calvetti, Y. Cheng, and E. Somersalo, *Uncertainty quantification in flux balance analysis of spatially lumped and distributed models of neuron-astrocyte metabolism*, J. Math. Biol., 73 (2016), pp. 1823–1849.
[7] D. Calvetti, J. Heino, E. Somersalo, and K. Tunyan, *Bayesian stationary state flux balance analysis for a skeletal muscle metabolic model*, Inverse Probl. Imaging, 1 (2007), pp. 247–263.
[8] D. Calvetti and E. Somersalo, *Quantitative* in silico *analysis of neurotransmitter pathways under steady state conditions*, Front. Endocrin., 4 (2013), 137.
[9] R.P. Carlson, *Decomposition of complex microbial behaviors into resource-based stress responses*, Bioinformatics, 25 (2009), pp. 90–97.
[10] W.G. Characklis, *Fouling biofilm development: A process analysis*, Biotech. Bioeng., 23 (1981), pp. 1923–1960.
[11] J. Chen, J.A. Gomez, K. Höffner, P. Phalak, P.I. Barton, and M.A. Henson, *Spatiotemporal modeling of microbial metabolism*, BMC Systems Biol., 10 (2016), 21.
[12] C. Chen, S. Hou, D. Ren, M. Ren, and Q. Wang, 3-*D spaio-temporal structures of biofilms in a water channel*, Math. Method. Appl. Sci., 38 (2015), pp. 4461–4478.
[13] B.L. Clarke, *Stoichiometric network analysis*, Cell Biophys., 12 (1988), pp. 237–253.
[14] N.G. Cogan, *Two-fluid model of biofilm disinfection*, Bull. Math. Biol., 70 (2008), pp. 800–819.
[15] N.G. Cogan and J.P. Keener, *The role of the biofilm matrix in structural development*, Math. Med. Biol., 2 (2004), pp. 147–166.
[16] J.A. Cole, L. Kohler, J. Hedhli, and Z. Luthey-Schulten, *Spatially-resolved metabolic cooperativity within dense bacterial colonies*, BMC Systems Biol., 9 (2015), 15, https://doi.org/10.1186/s12918-015-0155-1.
[17] O.K. Oyebamiji, D.J. Wilkinson, B. Li, P.G. Jayathilake, P. Zuliani, and T.P. Curtis, *Bayesian emulation and calibration of an individual-based model of microbial communities*, J. Comput. Sci., 30 (2019), pp. 194–208.
[18] B. D'Acunto, L. Frunzo, I. Klapper, and M.R. Mattei, *Modelling multispecies biofilms including new bacterial species invasion*, Math. Biosci., 259 (2015), pp. 20–26.
[19] B. D'Acunto, L. Frunzo, I. Klapper, M.R. Mattei, and P. Stoodley, *Mathematical modeling of dispersal phenomenon in biofilms*, Math. Biosci., 307 (2019), pp. 70–87.
[20] J. Dockery and I. Klapper, *Finger formation in biofilm layers*, SIAM J. Appl. Math., 62 (2002), pp. 853–869.
[21] N.C. Duarte, M.J. Herrgård, and B.Ø. Palsson, *Reconstruction and validation of* Saccharomyces cerevisiae *iND*750*, a fully compartmentalized genome-scale metabolic model*, Genome Res., 14 (2004), pp. 1298–1309.
[22] R. Duddu, D.L. Chopp, and B. Moran, *A two-dimensional continuum model of biofilm growth incorporating fluid flow and shear stress based detachment*, Biotech. Bioeng., 103 (2009), pp. 92–104.
[23] H.J. Eberl and R. Sudarsan, *Exposure of biofilms to slow flow fields: The convective contribution to growth and disinfection*, J. Theoret. Biol., 253 (2008), pp. 788–807.
[24] J.S. Edwards, R.U. Ibarra, and B.Ø. Palsson, *In silico predictions of* Escherichia coli *metabolic capabilities are consistent with experimental data*, Nat. Biotech., 19 (2001), pp. 125–130.
[25] Y. Fang, T.D. Scheibe, R. Mahadevan, S. Garg, P.E. Long, and D.R. Lovley, *Direct coupling of a genome-scale microbial* in silico *model and a groundwater reactive transport model*, J. Contam. Hydrol., 122 (2011), pp. 96–103.
[26] A.M Feist and B.Ø. Palsson, *The biomass objective function*, Curr. Opin. Microbiol., 13 (2010), pp. 344–349.
[27] G.F. Gause, *Experimental studies on the struggle for existence:* 1. *Mixed population of two species of yeast*, J. Exp. Biol., 9 (1932), pp. 389–402.
[28] J.F. Hammond, E.J. Stewart, J.G. Younger, M.J. Solomon, and D.M. Bortz, *Variable viscosity and density biofilm simulations using an immersed boundary method, part* I: *Numerical scheme and convergence results*, Comput. Model. Eng. Sci., 98 (2014), pp. 295–340.
[29] K. Han, H.C. Lim, and J. Hong, *Acetic acid formation in* Escherichia coli *fermentation*, Biotech. Bioeng., 39 (1991), pp. 663–671.
[30] W.R. Harcombe, W.J. Riehl, I. Dukovski, B.R. Granger, A. Betts, A.H. Lang, G. Bonilla, A. Kar, N. Leiby, P. Mehta, C.J. Marx, and D. Segrè, *Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics*, Cell Rep., 7 (2014), pp. 1104–1115.

[31] M.A. Henson and P. Phalak, *Byproduct cross feeding and community stability in an* in silico *biofilm model of the gut microbiome*, Processes, 5 (2017), 13, https://doi.org/10.3390/pr5010013.

[32] K.A. Hunt, J.P. Folsom, R.L. Taffs, and R.P. Carlson, *Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition*, Bioinformatics, 30 (2014), pp. 1569–1578.

[33] B.D. Jackson, J.M. Connolly, R. Gerlach, I. Klapper, and A.E. Parker, *Bayesian Estimation and Uncertainty Quantification in Models of Urea Hydrolysis by* E. coli *Biofilms*, manuscript.

[34] N. Jayasinghe, A. Franks, K.P. Nevin, and R. Mahadevan, *Metabolic modeling of spatial heterogeneity of biofilms in microbial fuel cells reveals substrate limitations in electrical current generation*, Biotech. J., 9 (2014), pp. 1350–1361.

[35] M. Kargeti and K.V. Venkatesh, *The effect of global transcriptional regulators on the anaerobic fermentative metabolism of* Escherichia coli, Mol. Biosystems, 13 (2017), pp. 1388–1398.

[36] A. Khodayari and C.D. Maranas, *A genome-scale* Escherichia coli *kinetic metabolic model k-ecoli*457 *satisfying flux data for multiple mutant strains*, Nature Comm., 7 (2016), 13806.

[37] I. Klapper and J. Dockery, *Role of cohesion in material description of biofilms*, Phys. Rev. E (3), 74 (2006), 031902.

[38] I. Klapper and J. Dockery, *Mathematical description of microbial biofilms*, SIAM Rev., 52 (2010), pp. 221–265.

[39] I. Klapper and B. Szomolay, *An exclusion principle and the importance of mobility for a class of biofilm models*, Bull. Math. Biol., 73 (2011), pp. 2213–2230.

[40] N. Klitgord and D. Segré, *Environments that induce synthetic microbial ecosystems*, PLoS Comput. Biol., 6 (2010), e1001002.

[41] J.-U. Kreft, *Biofilms promote altruism*, Microbiology, 150 (2004), pp. 2751–2760.

[42] P.J. Linstrom and W.G. Mallard, eds., *NIST Chemistry WebBook, NIST Standard Reference Database Number* 69, National Institute of Standards and Technology, Gaithersburg MD, 2018 https://doi.org/10.18434/T4D303.

[43] J. Liu, A. Prindle, J. Humphries, M. Gabalda-Sagarra, M. Asally, D.D. Lee, S. Ly, J. Garcia-Ojalvo, and G.M. Süel, *Metabolic co-dependence gives rise to collective oscillations within biofilms*, Nature, 523 (2015), pp. 550–554.

[44] R. Mahadevan, J.S. Edwards, and F.J. Doyle III, *Dynamic flux balance analysis of diauxic growth in* Escherichia coli, Biophys. J. 83 (2002), pp. 1331–1340.

[45] M.R. Mattei, L. Frunzo, B. D'Acunto, Y. Pechaud, F. Pirozzi, and G. Esposito, *Continuum and discrete approach in modeling biofilm development and structure: A review*, J. Math. Biol., 76 (2018), pp. 945–1003.

[46] N. Nisan, T. Roughgarden, É. Tardos, and V.V. Vazirani, eds., *Algorithmic Game Theory*, Cambridge University Press, Cambridge, 2007.

[47] S. Nowack and I. Klapper, *Exclusion in a temporally varying chemostat system: Dependence on trade-offs*, SIAM J. Appl. Math., 78 (2018), pp. 2819–2839.

[48] J.D. Orth, I. Thiele, and B.Ø. Palsson, *What is flux balance analysis?*, Nature Biotech., 28 (2010), pp. 245–248.

[49] W.J. Payne and W.J. Wiebe, *Growth yield and efficiency in chemosynthetic microorganims*, Annu. Rev. Microbiol., 32 (1978), pp. 155–183.

[50] P. Phalak, J. Chen, R.P. Carlson, and M.A. Henson, *Spatiotemporal metabolic modeling of a chronic wound biofilm consortium*, IFAC-PapersOnLine, 49 (2016), pp. 32–37.

[51] C. Picioreanu, M.C.M. van Loosdrecht, and J.J. Heijnen, *Effect of diffusive and convective substrate transport on biofilm structure formation: A two-dimensional modeling study*, Biotech. Bioeng., 69 (2000), pp. 504–515.

[52] B. Polizzi, O. Bernard, and M. Ribot, *A time-space model for the growth of microalgae biofilms for biofuel production*, J. Theoret. Biol., 432 (2017), pp. 55–79.

[53] N.D. Price, J.L. Reed, and B.Ø. Palsson, *Genome-scale models of microbial cells: Evaluating the consequences of constraints*, Nature Rev. Microbiol., 2 (2004), pp. 886–897.

[54] H. Qian and D.A. Beard, *Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium*, Biophys. Chem., 114 (2004), pp. 213–220.

[55] S.A. Rani, B. Pitts, H. Beyenal, R.A. Veluchamy, Z. Lewandowski, W.M. Davison, K. Buckingham-Meyer, and P.S. Stewart, *Spatial patterns of DNA replication, protein synthesis, and oxygen concentration within bacterial biofilms reveal diverse physiological states*, J. Bacteriol., 189 (2007), pp. 4223–4233.

[56] A. Trucchia, M.R. Mattei, V. Luongo, L. Frunzo, and M.C. Rochoux, *Surrogate-based*

uncertainty and sensitivity analysis for bacterial invasion in multi-species biofilm modeling, Commun. Nonlinear Sci., 73 (2019), pp. 403–424.

[57] T.D. Scheibe, R. Mahadevan, Y. Fang, S. Garg, P.E. Long, and D.R. Lovley, *Coupling a genome-scale metabolic model with a reactive transport model to describe* in situ *uranium bioremediation*, Microb. Biotech., 2 (2009), pp. 274–286.

[58] J. Schellenberger and B.O. Palsson, *Use of randomized sampling for analysis of metabolic networks*, J. Biol. Chem., 284 (2009), pp. 5457–5461.

[59] C.H. Schilling, S. Schuster, B.Ø. Palsson, and R. Heinrich, *Metabolic pathway analysis: Basic concepts and scientific applications in the post genomic era*, Biotech. Progr., 15 (1999), pp. 296–303.

[60] R. Schuetz, L. Kuepfer, and U. Sauer, *Systematic evaluation of objective functions for predicting intracellular fluxes in* Escherichia coli, Mol. Systems Biol., 3 (2007), 119.

[61] S. Schuster and S. Hilgetag, *On elementary flux modes in biochemical reaction systems at steady state*, J. Biol. Systems, 2 (1994), pp. 165–182.

[62] H.L. Smith and P. Waltman, *The Theory of the Chemostat*, Cambridge University Press, Cambridge, 1995.

[63] H.-S. Song, D.G. Thomas, J.C. Stegen, M. Li, C. Liu, X. Song, X. Chen, J.K. Fredrickson, J.M. Zachara, and T.D. Scheibe, *Regulation-structured dynamic metabolic model provides a potential mechanism for delayed enzyme response in denitrification process*, Front Microbiol., 8 (2017), 01866.

[64] P.S. Stewart, *Diffusion in biofilms*, J. Bacteriol., 185 (2003), pp. 1485–1491.

[65] P.S. Stewart and M.J. Franklin, *Physiological heterogeneity in biofilms*, Nature Rev. Microbiol., 6 (2008), 199210.

[66] J.A. Stotsky, J.F. Hammond, L. Pavlovsky, E.J. Stewart, J.G. Younger, M.J. Solomon, and D.M. Bortz, *Variable viscosity and density biofilm simulations using an immersed boundary method, part* II: *Experimental validation and the heterogeneous rheology-IBM*, J. Comput. Phys., 317 (2016), pp. 204–222.

[67] R. Taffs, J.E. Aston, K. Brileya, Z. Jay, C.G. Klatt, S. McGlynn, N. Mallette, S. Montross, R. Gerlach, W.P. Inskeep, D.M. Ward, and R.P. Carlson, *In silico approaches to study mass and energy flows in microbial consortia: A syntrophic case study*, BMC Systems Biol., 3 (2009), 114.

[68] G.D. Tartakovsky, A.M. Tartakovsky, T.D. Scheibe, Y. Fang, R. Mahadevan, and D.R. Lovley, *Pore-scale simulation of microbial growth using a genome-scale metabolic model: Implications for Darcy-scale reactive transport*, Adv. Water Res., 59 (2013), pp. 256–270.

[69] P. Unrean and F. Srienc, *Metabolic networks evolve towards states of maximum entropy production*, Metab. Eng., 13 (2011), pp. 666–673.

[70] J.B. van Klinken and K.W. van Dijk, *FluxModeCalculator: An efficient tool for large-scale flux mode computation*, Bioinformatics, 32 (2016), pp. 1265–1266.

[71] S.J. Wiback, R. Mahadevan, and B.Ø. Palsson, *Reconstructing metabolic flux vectors from extreme pathways: Defining the α-spectrum*, J. Theoret. Biol., 224 (2003), pp. 313–324.

[72] T. Zhang, B. Pabst, I. Klapper, and P. S. Stewart, *General theory for integrated analysis of growth, gene, and protein expression in biofilms*, PloS ONE, 8 (2013), e83626.

[73] A.R. Zomorrodi and C.D. Maranas, *OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities*, PLOS Comp. Biol., 8 (2012), e1002363.