CommPhitting: A linear fitting model for discerning chemical parameters of microbial communities

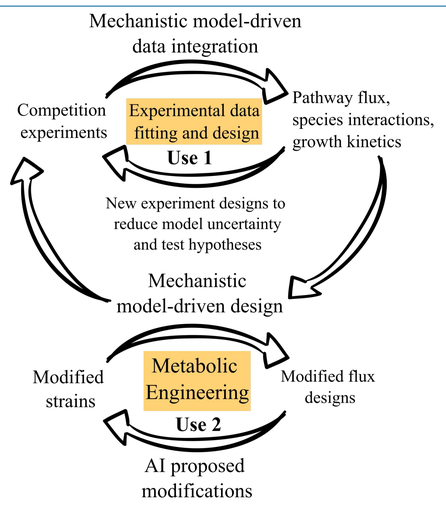
Andrew Freiburger[1] (afreiburger@anl.gov), Jeffrey Dewey,[2] Fatima Foflonker,[1] Gyorgy Babnigg,[2] Dionysios Antonopoulos[2], **Christopher Henry**[1] (chenry@anl.gov)

[1]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL ; [2]Biosciences Division, Argonne National Laboratory, Lemont, IL

COMMUNITY MODELING

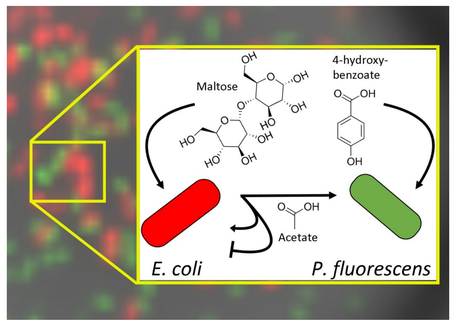
Microbial communities are ubiquitous in environments as diverse as the forest soils and animal intestines because they offer numerous practical efficiencies that encourage their establishment. These consortia consequently influence medicine, ecology, and are piquing industrial interest for bioproduction applications. Chemical dynamics within these communities are therefore imperative for research in the life sciences, yet the combinatorially complex interactions in these consortia remain intractable for experimentation. Simulation can capture aspects of community interactions, but existing methods are insufficiently flexible or mechanistically resolved for basic and applied uses.

We therefore developed a method (**CommPhitting**) to simultaneously integrate experimental data with genome-scale metabolic models that dynamically model molecular biology with mechanistic resolution. Our model predicts chemical interactions and growth kinetic coefficients among each metabolic phenotype. The second use case of our model, which is under development, can apply the fitted parameters from the first use case of basic understading to predict outcomes of rational engineering and community design (depicted at right).



BIOLOGICAL SYSTEM

An experimentally convenient 2-member community of *P. fluorescens* SBW25 and *E. coli* (minimally modified) MG1655 was assembled to exemplify our model. This community exhibits interesting cross-feeding of acetate from *E. coli* to *P. fluorescens*, which allows the latter to grow in maltose meda despite not consuming maltose (depicted at right). The growth characteristics of this community in maltose (consumed by *E. coli)*, 4-hydroxybenzoate (consumed by *P. fluorescens*), and acetate (consumed by both) media were examined via optical density (OD 590 nm) and fluorescence (544 nm excitation & 590 nm emission for *E. coli* and 485 nm excitation & 535 nm emission for *P. fluorescens*) via specialized reporter proteins for each member.



MODELING METHODS

DIMENSIONS

s	A species in the examined community.
k	A growth phenotype of species s.
t	An experimental time point.
j	Extracellular metabolite.
i	A biomass partition.

PARAMETERS

$E_{s,t}$	Float	The experimental growth signal for a species at instant t.
$es_{s,k}$	Boolean	A designation of truth for $k \in s$
Δt	Float	The seconds per timestep, which determines the amount of biomass growth per timest
$n_{k,i}$	Float	The exchange flux of each metabolite i in each strain k.
$cvct$ & $cvcf$	Float	Conversion coefficients of phenotype biomass to and from the stationary phase, respecti
bcv_k	Float	The greatest fraction of biomass ($0 < bcv < 1$) of strain k that can transition phenotypes in a timestep.
$cvmin$	Float	The minimal value of variable $cvt_{k,t}$.
$stat$	Float	The optimization penalty for the stationary phenotype of each species.
$kcat_z$	Float	The $kcat$ growth rate constant for the biomass partition z.
$kcat_k$	Float	The $kcat$ growth rate constant for the phenotype k.

VARIABLES

EC_k	Continuous	The conversion coefficient ($0 < EC < 1000$) from parameter $E_{s,t}$ into biomass, which is unique for each strain k.
$EB_{s,t}$	Continuous	The computed biomass from each experimental datum, as the product of EC_k & $E_{s,t}$.
bin_k^z	Binary	A binary switch that determines whether a given biomass partition of a phenotype is active and contributes to the total $kcat$ of the phenotype.
$b_{k,t}^z$	Continuous	The phenotype biomass partitions, which each exhibit a distinct $kcat$.
$b_{k,t}$	Continuous	The total predicted phenotype biomass from the fitting model.
$EV_{s,t}$	Continuous	The variance between the computed experimental biomass $EB_{s,t}$ and the predicted biomass $b_{k,t}$.
$g_{k,t}$	Continuous	The predicted growth rate for each strain at each datum.
$c_{t,i}$	Continuous	The concentration of metabolite i at an experimental datum.
$cvt_{k,t}$ &	Continuous	The quantity of strain k biomass that transitions to and from the stationary phase, respectively, at an experimental datum.

CommPhitting captures biological objects through the dimensions, variables, and parameters that are depicted above and at right.

The first use case of CommPhitting which investigates a community by fitting experimental data, minimizes variance between simulated biomass and derived biomass from the experimental data

$$\sum_{s}^{S}\left(\sum_{t,j}^{T,J}(EV_{s,t,j}^2)\right) - \sum_{t,j,k}^{T,J,K}(cvct * cvt_{t,j,k}) - \sum_{t,j,k}^{T,J,K}(cvcf * cvf_{t,j,k})$$

while minimizing the quantity of phenotypic transitions to mitigate overfitting. The biological processes of the community system are captured through the following system of linear constraints:

$$b_{t,j,k} + \frac{\Delta t}{2}(g_{t,j,k} + g_{t+1,j,k}) + cvf_{t,j,k} - cvt_{t,j,k} = b_{t+1,j,k} \quad \textbf{a}$$

$$b_{t,j,k} - \sum_{k,s}^{K,S}(es_{k,s} * cvf_{t,j,k}) + \sum_{k,s}^{K,S}(es_{k,s} * cvt_{t,j,k}) = b_{t+1,j,k} \quad \textbf{b}$$

$$c_{t,j,i} + \frac{\Delta t}{2}\sum_{k}^{K}(n_{i,k}(g_{t,j,k} + g_{t+1,j,k})) = c_{t+1,j,i} \quad \textbf{c}$$

$$cvt_{t,j,k} \leq bcv * b_{t,j,k} + cvmin \quad \textbf{d}$$

$$EB_{s,t,j} - \sum_{k}^{K}(es_k * b_{t,j,k}) = EV_{s,t,j} \quad \textbf{e}$$

$$E_{s,t,j} * EC = EB_{s,t,j} \quad \textbf{f}$$

g

$$b_{k,t}^z <= 1000 - 1000 * bin_k^z \quad \forall z \in Z$$

$$b_{k,t}^z <= kcat_z * b_{k,t} \quad \forall z \in Z$$

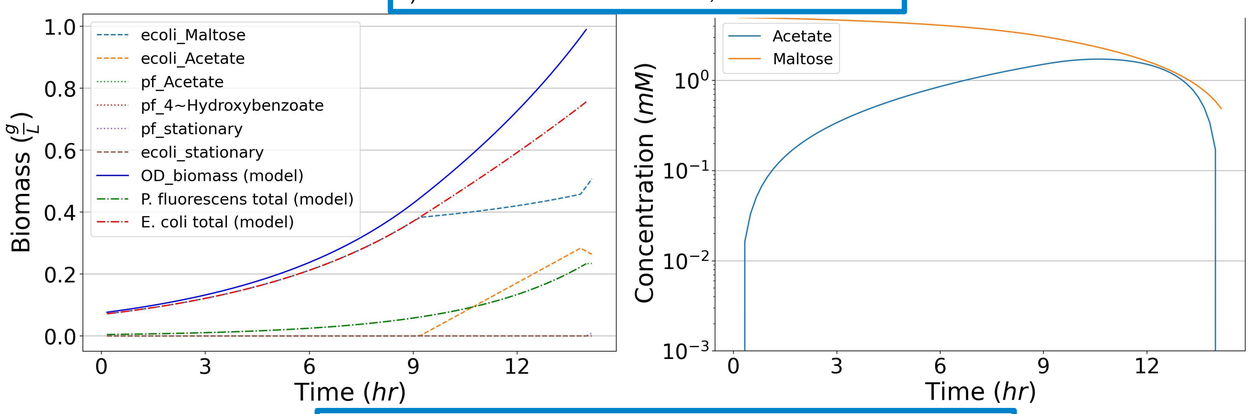
$$kcat_z * b_{k,t} - 1000 * bin_k^z <= 0 \quad \forall z \in Z$$

These constraints represent: biomass change in the a) non-stationary and b) stationary phenotypes; c) concentration change; d) the minimum biomass fraction that transitions its phenotype; e) biomass variance; f) the model prediction of biomass; and g) growth kinetic rates. respectively. Constraints a) & c) utilize Heun's method, which is a 2nd-order Runge-Kutta method.

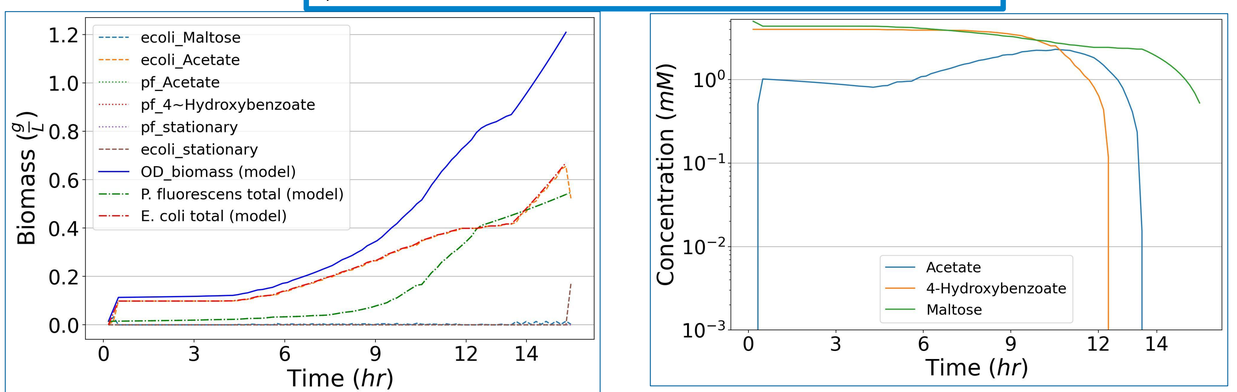
EXPERIMENTAL FIT

A few batch growth datasets of our 2-member community on **a)** Maltose, **b)** Maltose+4HB, or **c)** Acetate media, coupled with precise Acetate concentrations from metabolomics data, were fitted via CommPhitting to exemplify the first use case of investigating community dynamics from experimental data. A representative set of conditions for each media were simulated, from which results the phenotype biomass and media concentrations are depicted, in the following sets of figures.

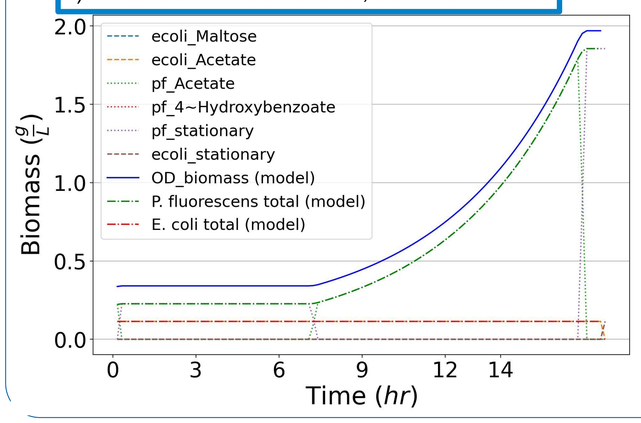
a) 1:10 *E. coli : P. fluorescens* , 5mM of Maltose



b) 1:10 *E. coli : P. fluorescens* , 5mM of Maltose and 4mM of 4HB



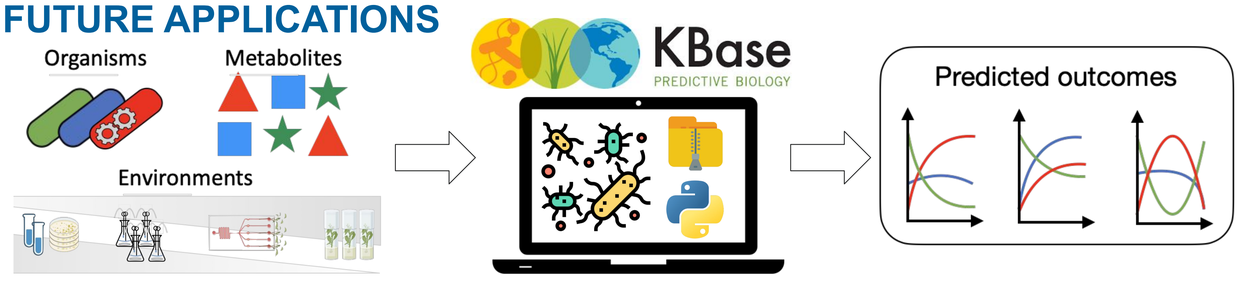
c) 1:1 *E. coli : P. fluorescens* , 50mM of Acetate



growth rate constants [seconds^{-1}]

Simulation	*E. coli* Acetate	*P. fluor.* Acetate	*E. coli* Maltose	*P. fluor.* 4-HB
Maltose (a)	6e-3	0.28	0.19	266
Maltose + 4HB (b)	5e-5	0.19	1	501
Acetate (c)	5e-5	0.22	129	0.86

FUTURE APPLICATIONS



We are beginning to apply our model to predict community behaviors in various environments, with goal of guiding the rational design of microbial communities and their members: e.g. *E. coli* modifications, such as new chemical pathways, that foster bioproduction of commodity chemicals or improve its competitive fitness relative to *P. fluorescens* in our community. This may be most succinctly achieved by simulating our BIOLOG growth data which investigates myriad environmental conditions.

The insights and techniques from our simplified 2-member community are generalizable for other, larger, community systems, such as a 10-member Plant-Microbe Interface community from our collaborators at ORNL and eventually complete microbiomes. We finally intend to offer all of our data, models, and methods in KBase as published Narratives and a Narrative Application for open-science use and assessment of our method towards rationally designing microbial communities.

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
BERKELEY LAB
JGI JOINT GENOME INSTITUTE
Northwestern University
THE UNIVERSITY OF CHICAGO
Argonne NATIONAL LABORATORY