Supplementary Materials: In Silico Identification of Microbial Partners to Form Consortia with Anaerobic Fungi

St. Elmo Wilken [1], Mohan Saxena [1], Linda R. Petzold [2] and Michelle A. O'Malley [1,*]

1. Dynamic Flux Balance Analysis Algorithm

In this section we provide more details regarding the dFBA formulation we used. The primary difference between our algorithm and that of [1] is that we impose only a single secondary LP on the system. We assume only that the flux derivative change between time steps is small. In [1], two additional assumptions per integrated variable need to be imposed on the system i.e. the sense of each LP (maximizing or minimizing) as well as the order in which the LPs are solved. If one has a good understanding of the system this can be justified, but for exploratory work these extra assumptions are difficult to motivate.

In Section 1.1 we show how the constraints for our algorithm are formulated. In Section 1.2 we illustrate that our algorithm matches the results of [1] for the *E. coli* example. Results not shown here indicate that our algorithm is not as accurate for consortia—the multiple solution issue becomes more apparent in the Yeast-Algae example in the same work. In Section 1.3 we motivate the form of the constraints we use.

1.1. Objective Function Derivation

We begin by assuming that μ^*, the optimal growth rate found by the FBA optimization problem, has already been found. Now we wish to minimize the flux derivative between the current time, t, and the previous time, $t-1$. This is expressed in equation (1), following the same nomenclature as that of the paper.

$$\begin{aligned} \min_{\mathbf{v}} \quad & \sum_i \left| 1 - \frac{dv_i}{dt}_t \Big/ \frac{dv_i}{dt}_{t-1} \right| \text{ for } i \in \mathcal{M}, \\ \text{s.t.} \quad & \mathbf{Sv} + \mathbf{s_1} - \mathbf{v_2} = \mathbf{b}, \\ & \mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max}, \\ & \mu(\mathbf{v}) = \mu^*. \end{aligned} \tag{1}$$

By making use of the Taylor expansion $\frac{dF}{dt} \approx \frac{F_t - F_{t-1}}{\Delta t}$ (neglecting higher order terms) we can rewrite equation (1) as,

$$\begin{aligned} \min_{\mathbf{v}} \quad & \sum_i \left| 1 - \frac{v_{i,t} - v_{i,t-1}}{v_{i,t-1} - v_{i,t-2}} \right| \text{ for } i \in \mathcal{M}, \\ \text{s.t.} \quad & \mathbf{Sv} + \mathbf{s_1} - \mathbf{v_2} = \mathbf{b}, \\ & \mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max}, \\ & \mu(\mathbf{v}) = \mu^*. \end{aligned} \tag{2}$$

Next we can rewrite the objective function by making use of a dummy variable γ_i,

$$\begin{aligned} \min_{\mathbf{v}} \quad & \sum_i \gamma_i \text{ for } i \in \mathcal{M}, \\ \text{s.t.} \quad & \mathbf{Sv} + \mathbf{s_1} - \mathbf{v_2} = \mathbf{b}, \\ & \mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max}, \\ & \mu(\mathbf{v}) = \mu^*, \\ & \left| 1 - \frac{v_{t-1,i}}{v_{t-1,i} - v_{t-2,i}} - \frac{v_{t,i}}{v_{t-1,i} - v_{t-2,i}} \right| \leq \gamma_i \text{ for } i \in \mathcal{M}. \end{aligned} \tag{3}$$

Finally, we can replace the absolute value function and recognize that $v_{i,t-1}$ and $v_{i,t-2}$ are known values from the previous iterations. This results in the optimization problem given by,

$$\begin{aligned} \min_{\mathbf{v}} \quad & \sum_i \gamma_i \text{ for } i \in \mathcal{M}, \\ \text{s.t.} \quad & \mathbf{Sv} + \mathbf{s_1} - \mathbf{v_2} = \mathbf{b}, \\ & \mathbf{v}_{\min} \leq \mathbf{v} \leq \mathbf{v}_{\max}, \\ & \mu(\mathbf{v}) = \mu^*, \\ & -\gamma_i \leq 1 - \frac{v_{t-1,i}}{v_{t-1,i} - v_{t-2,i}} - \frac{v_{t,i}}{v_{t-1,i} - v_{t-2,i}} \leq \gamma_i \text{ for } i \in \mathcal{M}. \end{aligned} \tag{4}$$

1.2. Validation of Algorithm and Comparison to Previous Work

Figure S1 shows that our algorithm differs minimally from the one proposed by [1].

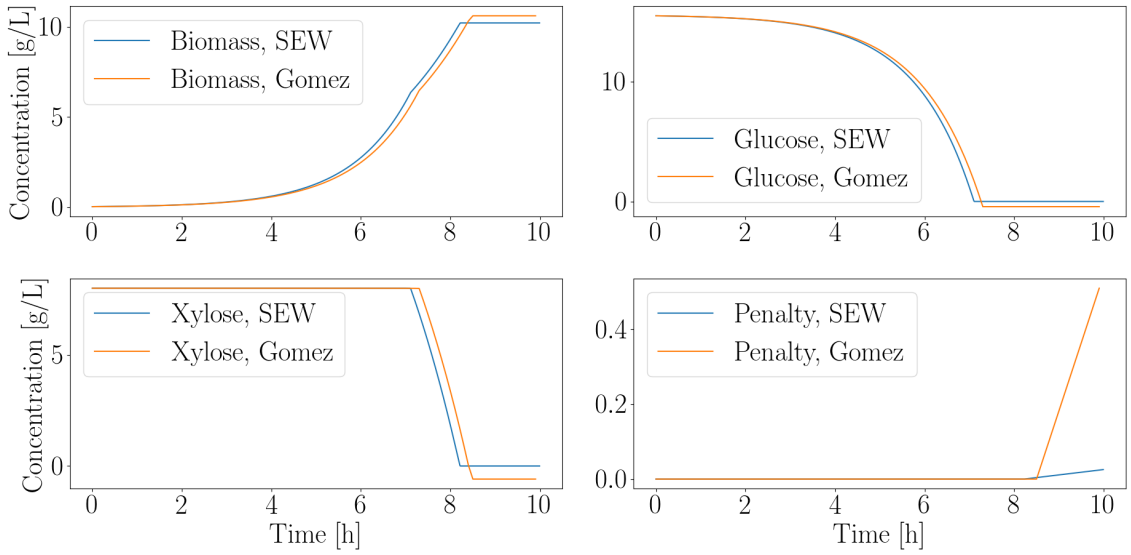

Figure S1. Comparison of our algorithm and the hierarchal optimization approach of [1] using a genome-scale model of *E. coli* from [2]. This model is the first example given in [1]. The same dynamic parameters were used in both simulations. Biomass refers to the organism's dry weight.

1.3. Constraint Bounds Justification

Typically in dFBA applications, Michaelis-Menten kinetics are assumed to constrain the uptake fluxes of the carbon sources. This requires two experimentally determined values, $v_{\max}$ and K_{M}. Typically in the FBA literature, an estimate of $v_{\max}$ is supplied as the flux used to validate the growth rate results. For some of the models we used, meaningful K_{M} values were not available. To compare the models as fairly as possible, we derived an empirical relation that bounds the flux. It is based on the assumption that the maximum uptake flux is bounded by the flux that would deplete the carbon source. Let G be the concentration of the carbon source. Then equation (5) relates the current concentration of G to the next time step by using the flux, v_G, and the external production rate, f_G, of the carbon source.

$$G_{t+1} = G_t + \Delta t v_G m_G X + \Delta t f_G^{\text{produced}} \tag{5}$$

Assuming that $G_{t+1} = 0$ (the organism attempts to consume as much carbon as possible), we can rewrite equation (5) as equation (6) using the same nomenclature,

$$v_G = -\frac{G_t + \Delta t f_G^{\text{produced}}}{\Delta t X m_{\text{glucose}}}. \tag{6}$$

Note that equation (6) does not upper bound the flux although it lower bounds it when the carbon source is depleted. To ensure that a realistic upper bound is imposed, we simply choose $\min(v_G, v_{\max,G})$ as the flux bound (see the main text). Figure S2 illustrates that there is virtually no difference (at least for *E. coli*) between using our constraints and Michaelis-Menten type constraints.

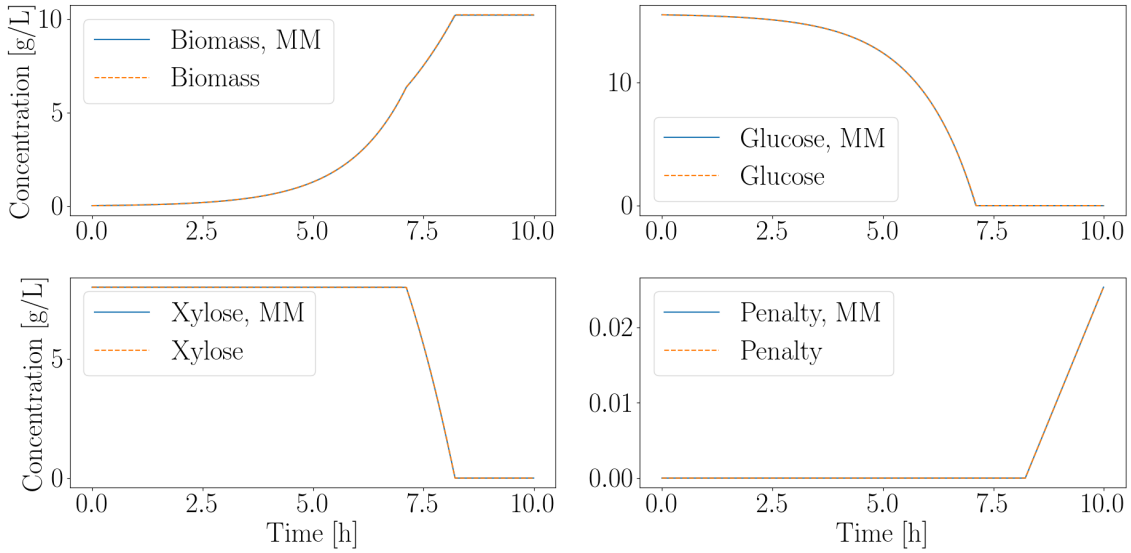

Figure S2. Comparison of the difference between Michaelis-Menten constraints and our single parameter uptake constraints for *E. coli* using the model of [2]. Michaelis-Menten (MM) parameters from [1] were used to model the base case. Biomass refers to the organism's dry weight.

2. Simulation Results

Because only a single secondary LP is required to be evaluated per time step our algorithm is more computationally efficient than both [1] and [3]. Simulation results of the other organisms are shown here, Figures S3–S6, although the data is summarized in the main text.

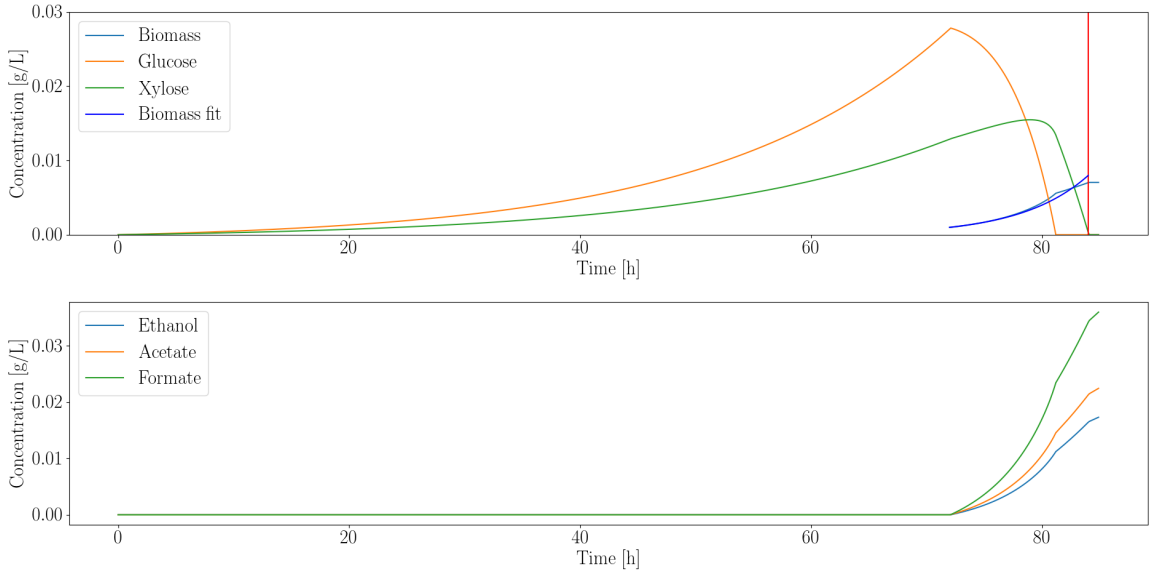

Figure S3. *E. coli* MG1655 simulated biomass (dry weight basis) growth rate on a glucose and xylose medium. The biomass fit curve, $\frac{dX}{dt} = \mu X$, was used to determine the average growth rate, μ, after inoculation subject to the simulation parameters. Metabolic waste products, ethanol, acetate and formate, accumulation were also modeled.

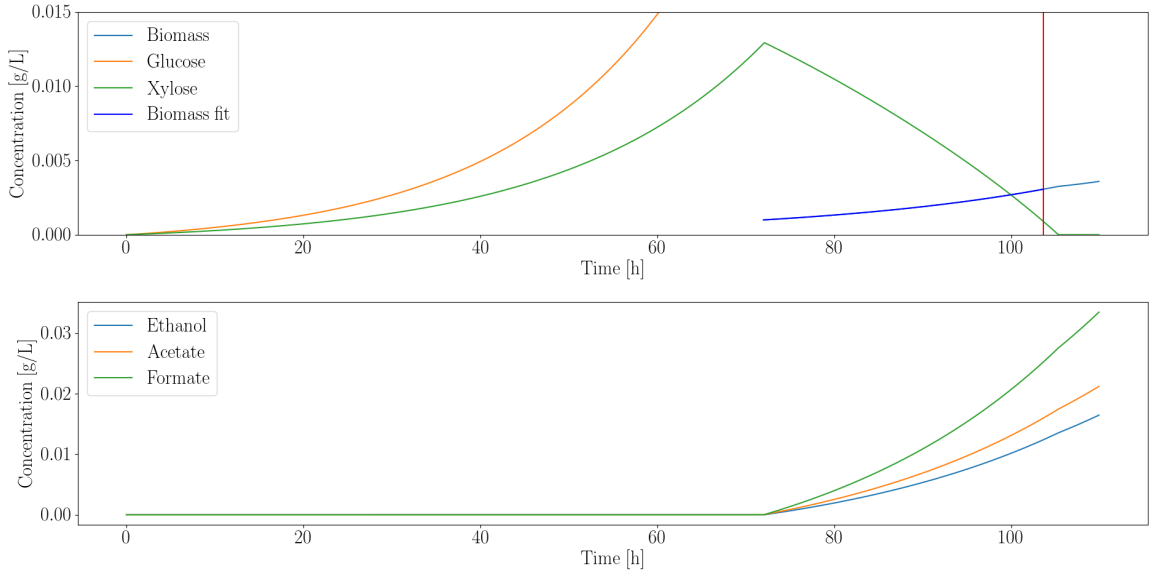

Figure S4. *E. coli* ZSC113 simulated biomass (dry weight basis) growth rate on a glucose and xylose medium. The biomass fit curve, $\frac{dX}{dt} = \mu X$, was used to determine the average growth rate, μ, after inoculation subject to the simulation parameters. Metabolic waste products, ethanol, acetate and formate, accumulation were also modeled.

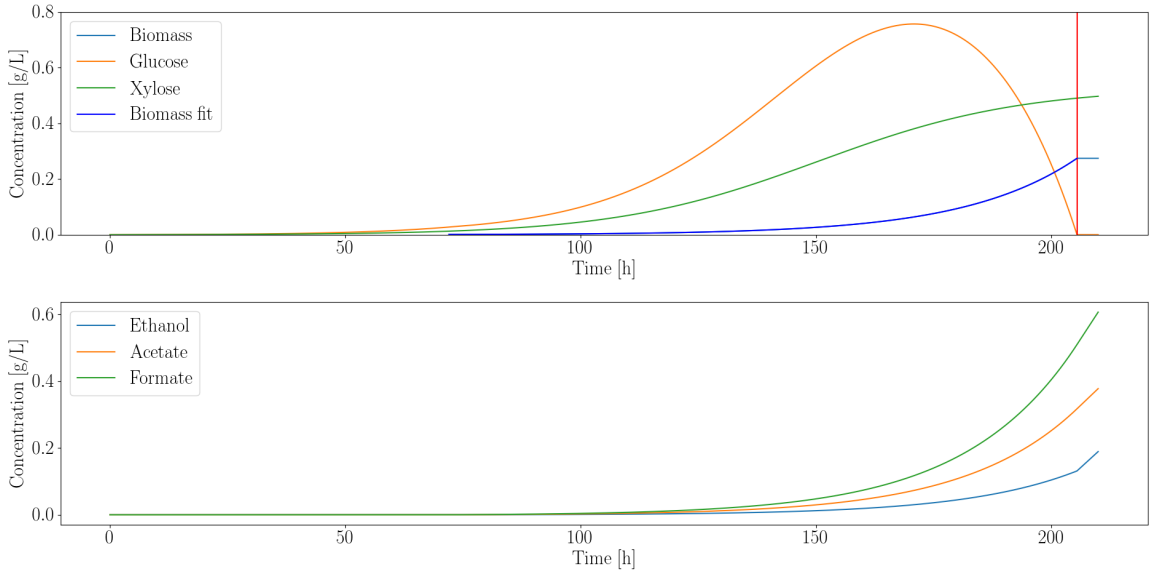

Figure S5. *L. lactis* simulated biomass (dry weight basis) growth rate on a glucose and xylose medium. The biomass fit curve, $\frac{dX}{dt} = \mu X$, was used to determine the average growth rate, μ, after inoculation subject to the simulation parameters. Metabolic waste products, ethanol, acetate and formate, accumulation were also modeled.

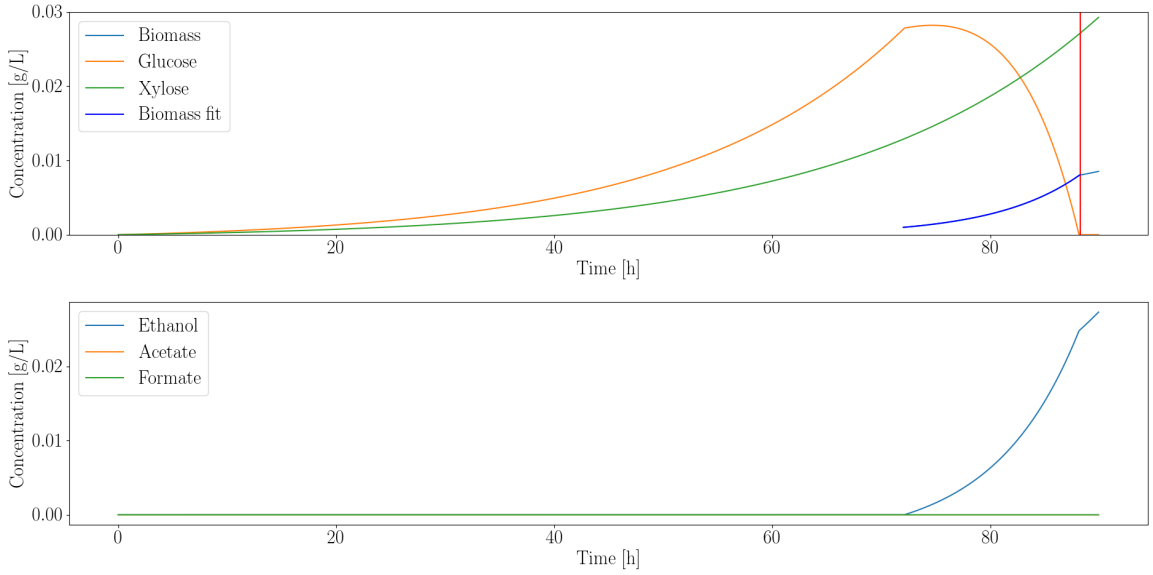

Figure S6. *S. cerevisiae* simulated biomass (dry weight basis) growth rate on a glucose and xylose medium. The biomass fit curve, $\frac{dX}{dt} = \mu X$, was used to determine the average growth rate, μ, after inoculation subject to the simulation parameters. Metabolic waste products, ethanol, acetate and formate, accumulation were also modeled.

3. *Neocallimastix* sp. S1 Isolation

Neocallimastix sp. S1 is a fungal strain isolated from sheep fecal matter through single colony isolation procedures detailed in [4].

References

1. Gomez, J.A.; Höffner, K.; Barton, P.I. DFBAlab: A fast and reliable MATLAB code for dynamic flux balance analysis. *BMC Bioinform.* **2014**, *15*, 409.
2. Reed, J.L.; Vo, T.D.; Schilling, C.H.; Palsson, B.O. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol.* **2003**, *4*, R54.
3. Sánchez, B.J.; Pérez-Correa, J.R.; Agosin, E. Construction of robust dynamic genome-scale metabolic model structures of Saccharomyces cerevisiae through iterative re-parameterization. *Metab. Eng.* **2014**, *25*, 159–173.
4. Henske, J.K. Engineering Regulation in Anaerobic Gut Fungi during Lignocellulose Breakdown. Ph.D. Thesis, University of California Santa Barbara, Santa Barbara, CA, USA, 2017.