In this section, we present the basic components of our kinetic model designed to simulate the flexible life strategies of human gut bacteria. The model is based on experiments conducted in monocultures and cocultures of a synthetic community of three species. However, it was crafted with a broad focus, allowing one to explore and simulate the expected behavior of the system under different conditions and to, in the future, expand it to other species.

Model States:

1. Subpopulations

Bacteria dynamically change their environments by consuming nutrients and excreting metabolic by-products. Within a community, species are often subjected to changing environments. For example, gut bacteria can be challenged by their host's feeding habits and circadian rhythms, as well as by the metabolic byproducts of neighboring species. To cope with these variations, cells within a population can alter their physiology, leading to phenotypic heterogeneity. This results in the emergence of subpopulations within a single species, which may or may not coexist in space and time.

Subpopulations materialize due to changes in cellular physiology in responses to environmental variations (either triggered by intrinsic cellular programs or, often, as an unavoidable consequence of being propelled towards intolerable conditions). In our model, we represent these subpopulations as distinct model structures for a given species, each having its own set of kinetic parameters. These structures are linked through transition rates and transition functions, which encode the likelihood that cells transition from one subpopulation to another in response to environmental cues.

Transition rates (r)

Cellular transitions can result from a single process, multiple independent processes, or multiple dependent processes. These transitions may occur rapidly, seemingly affecting cells concurrently or they may take place gradually, leading to the formation of co-existing subpopulations. The velocity at which a population transitions from one state to another is modulated by the transition rate (r).

Transition functions (Z)

In our model, transition functions encode the environmental cues that trigger transitions between subpopulations. These functions operate like sensors, responding to changes in factors such as pH, metabolite concentrations, or other environmental variables. Transitions can occur independently or involve activation or inhibition mechanisms.

Independent transitions

These transitions occur at a constant rate, unaffected by any environmental conditions. In this case, Z equals one, indicating a fixed transition rate. This transition is commonly assumed in scenarios such as when there is a constant death rate.

The following equation exemplifies an independent transition rate. Here, subpopulation A contributes towards the derivative of subpopulation B, given the counts N_A and N_B for each subpopulation, respectively:

$$\frac{dN_B}{dt} = (\underset{N_A \to N_B}{r})N_A + (\dots)$$

The first term on the right-hand side accounts for the contribution of subpopulation A to the growth of subpopulation B, where r is the transition rate from subpopulation A to subpopulation B. The ellipsis $(\dots)$ represents any additional terms that may affect the time derivative of subpopulation B, such as its growth rates or the reactor's dilution rate.

Activation-like transition

An activation-like transition is used when a specific molecule or condition triggers the transition between different subpopulations. This transition occurs when the concentration of the molecule or the strength of the condition increases to a certain threshold. For example, a subpopulation of cells in a slow growth mode (subpopulation A) may transition to a fast growth mode (subpopulation B) when the concentration of a metabolite, such as glucose, increases in the environment, and the subpopulation can sense it in a concentration-dependent manner (measured by S_M).

We use a Hill function (Nguyen & Kulasiri 2009) to modulate the transition rate based on the metabolite concentration. The Hill function is a sigmoidal curve that varies from 0 to 1 and represents the fraction of cells that transition from subpopulation A to subpopulation B at a given metabolite concentration. The transition rate is multiplied by the Hill function to obtain the effective transition rate, which is then used to compute the contribution towards the derivative of subpopulation B.

$$\underset{N_A \to N_B}{Z} = \frac{S_M^h}{K^h + S_M^h}$$

K is the half-max or Monod constant that is in the same unit as S_M (mM if we are talking about a metabolite) and defines the concentration where the transition function equals to 0.5 (the subpopulation A transitions to B at half the transition rate).

h is the Hill coefficient that determines how steep the transition will occur around the Monod constant (its slope).

The contribution towards the derivative is then:

$$\frac{dN_B}{dt} = \left(\frac{S_M^h}{K^h + S_M^h}\right)\left(\underset{N_A \to N_B}{r}\right)N_A + (\dots)$$

The model can also support a pure logical rule instead of a Hill function (equivalent to a Hill function with an infinite exponent), for instance:

$$\underset{(N_A \to N_B)}{Z} = \begin{cases} \underset{N_A \to N_B}{r}, & \text{if } S_M > \text{some value} \\ 0, & \text{otherwise} \end{cases}$$

This can be used initially to fit the model to data, as it requires less parameters. It can then easily be replaced by the smooth activation function described above.

Inhibition

Used when the transition is triggered by the 'lack' or 'depletion' of something. For example, when the depletion of a metabolite M that is used as an energy source leads to an increase in cell death rate, which is represented by an increase in the rate by which cells in subpopulation A (live) transition to subpopulation B (dead). The transition function used for inhibition is similar to the activation function, but is inverse:

$$\underset{N_A \to N_B}{Z} = \frac{K^h}{K^h + S_M^h}$$

The pure logical rule would be:

$$\underset{(N_A \to N_B)}{Z} = \begin{cases} \underset{N_A \to N_B}{r}, & \text{if } S_M < \text{some value} \\ 0, & \text{otherwise} \end{cases}$$

Metabolites

Species interactions and survival within an environment are largely shaped by the production and consumption of metabolites. In our model, we identified some of the nutrients available in the growth medium used, namely the Wilkins-Chagrin anaerobic broth. This medium contains glucose and pyruvate as added carbon sources. In addition, we measured the time series concentration of trehalose, a disaccharide produced by yeast and introduced into the medium through yeast extract. We also confirmed the presence of glutamate and mannose through single measurements (not time series).

Based on these observation and certain justified assumptions, we included in the model the concentration of several compounds, including glucose, pyruvate, trehalose, lactate, acetate, succinate, butyrate, formate, glutamate, and mannose. These compounds are consumed or produced by at least one of the three species in our model, as we explained in the page dedicated to the model state equations.

Feeding terms

Active subpopulations utilize nutrients for growth and maintenance, leading to the secretion of by-products. This process can be conceptualized as a directed graph, with nodes representing metabolites and edges representing biochemical reactions that transform one metabolite into another. The graph initiates with nutrients present in the medium and follows the flow of metabolites through the network until the by-products are secreted back into the medium, as illustrated below:

Mathematically, the consumption of metabolite M_a and production of metabolites M_d and M_c are connected by the consumption rate of metabolite M_a. For consumption, we employ the Monod equation (notably, with a slight abuse of notation, we denote to the concentration of metabolites measured in mM and the subpopulation counts measured in $10^{-5} cells \mu L^{-1}$ by their identifying names).

$$\frac{dM_a}{dt} = -\gamma_{C,M_a}\left(\frac{M_a}{K_{(C,M_a)} + M_a}\right) C \mu max_C + (\dots)$$

$$\frac{dM_d}{dt} = \gamma_{C,M_d}\left(\frac{M_a}{K_{(C,M_a)} + M_a}\right) C \mu max_C + (\dots)$$

$$\frac{dM_c}{dt} = \gamma_{C,M_c}\left(\frac{M_a}{K_{(C,M_a)} + M_a}\right) C \mu max_C + (\dots)$$

Where the $\gamma_{\text{subpopulation,metabolite}}$ terms represent the relative contribution (weight) of the subpopulation growth towards the consumption (if negative) or production (if positive) of the respective metabolite and k is the Monod constant.

In our model, we represented each species' core energy metabolism as an independent subnetwork that we extracted from its respective genome-scale metabolic model. We focused on the fraction of the network that could explain the changes in metabolite concentration that were measured during growth in monocultures. This approach was supported by the analysis of gene expression and differential expression of metabolic genes.

Analyzing the core energy graphs allows us to link subpopulations to sets of metabolites that contribute to their growth. These sets of metabolites are represented as feeding terms, which can display either an "or" (additive) relationship or an "and" (multiplicative) relationship between metabolites, based on the topology of the metabolic network.

Sometimes, a feeding term may require more than one metabolite to contribute to growth, while in other cases, a second metabolite may boost growth but not independently contribute to it. Below, we provide artificial toy examples of these two cases; Keep in mind, however, that the relationships among co-utilized metabolites are primarily determined by the topology of the metabolic network (see the model's state equations).

AND

The network requires the influx of both M_a and M_b to function (otherwise the intracellular metabolite that is highlighted in purple cannot be produced)

$$\frac{dM_a}{dt} = -\gamma_{C,M_a}\left(\frac{M_a}{K_{C,M_a}+M_a}\right)\left(\frac{M_b}{K_{C,M_b}+M_b}\right)C\mu max_C + (\dots)$$
$$\frac{dM_b}{dt} = -\gamma_{C,M_b}\left(\frac{M_a}{K_{C,M_a}+M_a}\right)\left(\frac{M_b}{K_{C,M_b}+M_b}\right)C\mu max_C + (\dots)$$
$$\frac{dM_c}{dt} = \gamma_{C,M_c}\left(\frac{M_a}{K_{C,M_a}+M_a}\right)\left(\frac{M_b}{K_{C,M_b}+M_b}\right)C\mu max_C + (\dots)$$
$$\frac{dM_d}{dt} = \gamma_{C,M_d}\left(\frac{M_a}{K_{C,M_a}+M_a}\right)\left(\frac{M_b}{K_{C,M_b}+M_b}\right)C\mu max_C + (\dots)$$

Boost

M_b may increase the flux through the network, but only in the presence of M_a. If M_b is absent, the network can still be functional, provided thar M_a is present, but not the other way around.

$$\frac{dM_a}{dt} = -\gamma_{C,M_a}\left[\left(\frac{M_a}{K_{C,M_a}+M_a}\right) + \left(\frac{M_a}{K_{C,M_a}+M_a}\right)\left(\frac{M_b}{K_{C,M_b}+M_b}\right)\right]C\mu max_C + (\dots)$$
$$\frac{dM_b}{dt} = -\gamma_{C,M_b}\left[\left(\frac{M_a}{K_{C,M_a}+M_a}\right) + \left(\frac{M_a}{K_{C,M_a}+M_a}\right)\left(\frac{M_b}{K_{C,M_b}+M_b}\right)\right]C\mu max_C + (\dots)$$
$$\frac{dM_c}{dt} = \gamma_{C,M_c}\left[\left(\frac{M_a}{K_{(M_a,C)}+M_a}\right) + \left(\frac{M_a}{K_{(M_a,C)}+M_a}\right)\left(\frac{M_b}{K_{(M_b,C)}+M_b}\right)\right]C\mu max_C + (\dots)$$
$$\frac{dM_d}{dt} = \gamma_{C,M_d}\left[\left(\frac{M_a}{K_{C,M_a}+M_a}\right) + \left(\frac{M_a}{K_{C,M_a}+M_a}\right)\left(\frac{M_b}{K_{C,M_b}+M_b}\right)\right]C\mu max_C + (\dots)$$

Environment (reactor)

Environmental conditions substantially influence microbial growth and interactions. Changes in the environment can lead to shifts in both the taxonomic composition and the function of the microbial system. A common approach to understanding these impacts is through collecting experimental data under varying conditions. However, we can take this a step further by encoding our knowledge into a dynamic model that explicitly accounts for environmental conditions. Once calibrated with parameters best fitting the experimental data, this model helps us predict the impact of environmental conditions on our system. These predictions, can be validated through further experiments, and if the results diverge, they serve to highlight key knowledge gaps.

Our model considers the impact of different environmental conditions by encoding the system's nutrient flows and pH dynamics. Specifically, we consider the consumption and production of metabolites by growing subpopulations combined with their responses to environmental changes. The composition of the environment affects the pH, which, in turn, plays a role in the growth and survival of the different subpopulations.

Environment pH

To predict pH changes in the environment, we developed an elastic net model using paired measurements of pH and fermentation acid concentrations. The model considers the concentrations of lactate, acetate, formate, and butyrate to estimate the environmental pH. We trained it using a randomly selected subset constituting 67% of all our experiments, encompassing both monocultures and cocultures. We evaluated the model's performance by comparing predicted pH values with actual measurements from the remaining 33% of the test cases, which were not part of the model training process.

pH limitation

In our model, the pH value at each time point is determined by the elastic net model, which is based on the concentrations of fermentation acids. The pH varies as the subpopulations consume and produce these acids. The fluctuating pH influences the growth rates of the subpopulations, according to their preferred pH and their sensitivity to pH variations. It is important to note that our model is only concerned with pH fluctuations below the initial anaerobic pH of the system (6.7), as we did not observe any pH increases beyond this value.

To model the impact of pH on growth, we used the probability density function of the gamma distribution, with some modifications. The gamma distribution is commonly used in statistics to model positive random variables with a skewed distribution. Since we aim to multiply the pH sensitivity by the growth rate, we require a function (referred to as $\xi_A(pH)$) that takes the current pH as an input and returns a value between zero and one. This output depends on the subpopulation's sensitivity to deviations from its preferred pH. The function must output a value of one at the optimal pH (or zero if we refer to pH sensitivity, where sensitivity = $1 - \xi_A(pH)$).

Starting from the probability distribution function of the gamma distribution:

$$\varphi(pH;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\ [pH]^{\alpha-1}\ e^{-\beta[ph]}$$

Where α and β are the shape and rate parameters of the gamma distribution and Γ is the gamma function. To obtain the properties we desire, we parameterized the above function with the optimal pH instead of the rate (β), and we wanted its maximum value to be exactly one. To achieve this, we set the mode (m) of the gamma distribution to be the optimal pH and solve for β :

$m = \frac{\alpha-1}{\beta}$

$pHopt = \frac{\alpha-1}{\beta}$

$\beta = \frac{\alpha-1}{pHopt}$

Next, for the optimal pH to return a value of one, we divide the function by its maximum value. Note that the maximum value of the function is exactly at the optimal pH. Tying it together, the impact of pH on the growth of subpopulation A is modelled by:

$$\xi_A(pH; pHopt_A, \alpha_A) = \frac{1}{\varphi(pHopt_A; \alpha_A, (\frac{\alpha_A-1}{pHopt_A}))} \varphi(pH; \alpha_A, (\frac{\alpha_A-1}{pHopt_A}))$$

φ is explained here

Below we show the pH sensitivity of a subpopulation that has $pHopt = 7.0$ for different values of α (note that α needs to be greater than one):

Growth

We now have all the ingredients for our general growth model. Our general growth model comprises a kinetic equation for each subpopulation, describing changes in its concentration in units of $10^{-5} cells \mu L^{-1}$, comparable to flow cytometer counts. These growth equations have four components:

a) Maximum intrinsic growth rate (μmax)

This represents the maximum growth rate that a subpopulation's physiology supports. In the absence of any limiting factors, cells would grow at a rate equal to μmax.

b) pH limitation (ξ)

This component takes into account the impact of pH on growth, using a modified probability density function of the gamma distribution. The function $\xi_A(pH)$ returns a value between zero and one, depending on how sensitive the subpopulation is to deviations from its preferred pH.

c) Nutrient consumption

This is based on the additive combination of the feeding terms that are actively used by a subpopulation.

d) Subpopulation transitions

This component takes into account incoming and outgoing subpopulation transitions.

Next, we illustrate the model equations for the following toy diagram where subpopulation A grows on M_1 and secretes M_2. When M_1 is depleted from the media, if both M_3 and M_4 are available, it switches to a new growth mode, summarized by subpopulation B that grows by simultaneously consuming M_3 and M_4 while still secreting M_2.

- Growth of subpopulation A

$$\frac{dA}{dt} = A \left(\xi_A(pH)\ \mu max_A \underbrace{\frac{M_1}{K_{A,M_1} + M_1}}_{\substack{\text{increase due to} \\ \text{metabolism}}} - \underbrace{Z_{T1}(M_1, M_3, M_4)\ r_{T1}}_{\text{transition to } B} \right) + (\dots)$$

$\xi_A(pH)$ is explained here

The transition function is:

$$Z_{T1}(M_1, M_3, M_4) = \underbrace{\frac{K_{A,M_1}^{h_{A,M_1}}}{K_{A,M_1}^{h_{A,M_1}} + M_1^{h_{A,M_1}}}}_{M_1 \text{ is depleted}} \underbrace{\frac{M_3^{h_{A,M_3}}}{K_{A,M_3}^{h_{A,M_3}} + M_3^{h_{A,M_3}}}}_{M_3 \text{ is available}} \underbrace{\frac{M_4^{h_{A,M_4}}}{K_{A,M_4}^{h_{A,M_4}} + M_4^{h_{A,M_4}}}}_{M_4 \text{ is available}}$$

- Growth of subpopulation B

$$\frac{dB}{dt} = B\, \xi_B(pH)\ \mu max_B \frac{M_3}{K_{B,M_3} + M_3} \frac{M_4}{K_{B,M_4} + M_4} + A\ Z_{T1}(M_1, M_3, M_4)\ r_{T1} + (\dots)$$

- Consumption of metabolite M_1

$$\frac{dM_1}{dt} = -\gamma_{A,M_1} \left(\frac{M_1}{K_{A,M_1} + M_1} \right) A \mu max_A + (\dots)$$

- Consumption of metabolite M_3

$$\frac{dM_3}{dt} = -\gamma_{B,M_3}(\frac{M_3}{K_{B,M_3}+M_3})(\frac{M_4}{K_{B,M_4}+M_4})B\mu max_B + (\dots)$$

- Consumption of metabolite M_4

$$\frac{dM_4}{dt} = -\gamma_{B,M_4}(\frac{M_3}{K_{B,M_3}+M_3})(\frac{M_4}{K_{B,M_4}+M_4})B\mu max_B + (\dots)$$

- Production of metabolite M_2

$$\frac{dM_2}{dt} = \gamma_{A,M_2}(\frac{M_1}{K_{A,M_1}+M_1})A\mu max_A + \gamma_{B,M_2}(\frac{M_3}{K_{B,M_3}+M_3})(\frac{M_4}{K_{B,M_4}+M_4})B\mu max_B + (\dots)$$

- pH

$$pH = \beta M_2$$

where β is the elastic net weight attributed to M_2.

Pulses

In addition to subpopulations growing by the exchange of metabolites and influenced by the environmental pH, the community is also influenced by the environmental regime, particularly by the way in which matter flows in and out of the system. In our model, the environmental regime is controlled by subdividing the simulation into arbitrary time intervals, which we refer to as "pulses". Within a "pulse", the following events are supported:

- Noncontinuous inflow (v_{in}) or outflow (v_{out}) of volume, which is performed at a single step at the beginning of the "pulse". Multiple "pulses" are required to simulate multiple influx and outflux events as in a serial passage experiment. The content of the influx volume, defined by the user, could include fresh metabolites or even spent media from another culture. The user may also set its pH and, if necessary, a specific concentration of bacteria to simulate migration.

- Continuous inflow (q_{in}) and outflow (q_{out}) of volume per time unit, added or removed from the reactor during a "pulse". Similar to the non-continuous events, the user defines the metabolome, microbiome, and pH of the feed.

With these events, a wide range of community regimes can be simulated. For example, one may start a pulse in a batch culture, then transition to a chemostat with a nutrient-rich medium, then move back to a batch culture. Simulations could include serial passages or even the mass transfer between separate reactors, allowing the exploration of regimes that lead to multistability and trigger alternative community states.