

Dynamic optimization of metabolic networks coupled with gene expression

Steffen Waldherr^a, Diego A. Oyarzún^b, Alexander Bockmayr^c

^a*Institute for Automation Engineering, Otto von Guericke University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany, steffen.waldherr@ovgu.de*

^b*Department of Mathematics, Imperial College London, SW7 2AZ London, United Kingdom*

^c*DFG Research Center MATHEON, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany*

Abstract

The regulation of metabolic activity by tuning enzyme expression levels is crucial to sustain cellular growth in changing environments. Metabolic networks are often studied at steady state using constraint-based models and optimization techniques. However, metabolic adaptations driven by changes in gene expression cannot be analyzed by steady state models, as these do not account for temporal changes in biomass composition.

Here we present a dynamic optimization framework that integrates the metabolic network with the dynamics of biomass production and composition. An approximation by a timescale separation leads to a coupled model of quasi steady state constraints on the metabolic reactions, and differential equations for the substrate concentrations and biomass composition. We propose a dynamic optimization approach to determine reaction fluxes for this model, explicitly taking into account enzyme production costs and enzymatic capacity. In contrast to the established dynamic flux balance analysis, our approach allows predicting dynamic changes in both the metabolic fluxes and the biomass composition during metabolic adaptations. Discretization of the optimization problems leads to a linear program that can be efficiently solved.

We applied our algorithm in two case studies: a minimal nutrient uptake network, and an abstraction of core metabolic processes in bacteria. In the minimal model, we show that the optimized uptake rates reproduce the empirical Monod growth for bacterial cultures. For the network of core metabolic processes, the dynamic optimization algorithm predicted

commonly observed metabolic adaptations, such as a diauxic switch with a preference ranking for different nutrients, re-utilization of waste products after depletion of the original substrate, and metabolic adaptation to an impending nutrient depletion. These examples illustrate how dynamic adaptations of enzyme expression can be predicted solely from an optimization principle.

Keywords: flux optimization, constraint-based methods, dynamic optimization, metabolic-genetic networks

1. Introduction

A key aspect of cellular dynamics is the ability to adapt metabolic activity to changing environments. This involves a dynamic re-organization of enzyme expression levels, in order to accommodate for variability in nutrient abundance and environmental shocks that have a deleterious impact on growth. These adaptations emerge from a complex array of regulatory interactions between metabolism and the genetic machinery. Since many of these interactions are unknown or incompletely understood, a fully mechanistic grasp of how they control metabolic adaptations is currently beyond our reach. Moreover, the analysis of large-scale mechanistic models is typically hampered by the high number of molecular species and parameters involved.

An alternative approach to predict metabolic adaptations is to assume an underlying optimality principle [1, 2, 3]. Numerous studies have considered metabolic adaptations in microbes by computing optimal metabolic fluxes in a stoichiometric model under a suitable objective function [4, 5, 6, 7]. Stoichiometric models are a structural description of a metabolic network and cannot provide information on the enzyme concentrations. Several approaches have attempted to overcome this by integrating gene regulation with stoichiometric models, either by modelling enzyme expression qualitatively with Boolean variables describing regulatory effects [8, 5], or by explicitly including enzyme capacity constraints in the optimization problems [9]. An alternative approach are the cybernetic models [10], where regulation is explicitly modelled and assumed to optimize a cellular objective. While the classical cybernetic approach explicitly includes reaction kinetics in the model, a hybrid approach has been suggested [11], where the rates are determined from flux balance analysis.

Models that integrate metabolism and gene expression can potentially

yield better predictions than those focused on metabolism in isolation. This can be particularly helpful in metabolic adaptations caused by environmental fluctuations. To capture the dynamics of biomass and gene expression linked to metabolic activity, previous studies have mostly used ad-hoc combinations of various modeling frameworks. Examples are combinations of constraint-based steady state models with ordinary differential equations or Boolean regulatory logic [3, 5, 12]. This combination approach has been successful in proposing integrated models up to whole-cell dynamics [13].

In this paper, we propose a dynamic modeling framework for metabolic networks coupled with gene expression of enzymes and production of other macromolecules. We develop an optimization algorithm to predict optimal time courses for nutrient uptake, metabolic fluxes, and gene expression rates in such networks.

The classical approach to constraint-based optimization of metabolic fluxes, commonly called Flux Balance Analysis (FBA), relies on an optimization problem with algebraic constraints stemming from a steady state restriction [14, 15, 16]. Mathematically, the FBA approach in the simplest form leads to a linear program of the form

$$\max_v \{c^T v \mid Sv = 0, v_{\min} \leq v \leq v_{\max}\}, \quad (1)$$

where v is the reaction flux vector, c a biomass weighting vector, S the stoichiometric matrix, and v_{\min} , v_{\max} are lower and upper component-wise bounds on the fluxes, respectively. While the most common optimization objective is the maximization of biomass production, an experimental evaluation also highlighted additional biologically relevant objectives [17].

One point of critique to FBA is its coarse description of the biomass composition. While growth-dependent changes in the biomass composition have been taken into account in the past [18], constraints related to the actual biomass composition by enzymes or other cellular macromolecules are usually not considered. At least on the level of individual metabolic pathways, there is good evidence that the enzyme production cost is an important factor in the regulation of these pathways [19]. Thus, it seems plausible that the inclusion of biomass composition and enzyme costs in metabolic optimization can potentially improve the quality of its predictions. As an extension to FBA in this direction, the resource balance analysis (RBA) approach has been proposed [9]. This includes the conversion of metabolites into specific enzymes and other proteins in the network, and adds the enzymatic capacity

as constraint on metabolic fluxes for the optimization. RBA yields a linear optimization problem and can intrinsically describe changes in both the growth rate and biomass composition due to environmental changes from an optimization principle alone. A conceptually equivalent approach has been proposed independently in [20] under the term ME (metabolism and macromolecular expression) model. Both approaches are however limited to situations of steady exponential growth.

For batch processes or in changing environments, the model needs to go beyond the stationary approach and account for dynamic changes in metabolic activity. FBA has been used to predict dynamic changes in biomass and nutrients using iterative approaches [3]. However, the iterative optimization uses a steady state constraint and does not account for the model dynamics, and thus the predictions may not be optimal in changing environments. Dynamic effects are physiologically important, as is evidenced by the experimental observation that even in steady state, cells would show flux distributions which are slightly suboptimal, but which allow for easier transitions to other environmental conditions [21]. Also, the numerical accuracy of the iterative approach can at best be evaluated heuristically or by numerical experimentation, unless specialized numerical algorithms are applied [22].

By formulating an appropriate dynamic optimization problem, it is possible to compute optimal fluxes over the whole time range of interest. This approach has been proposed in dynamic flux balance analysis (dFBA) [23]. In dFBA, one can distinguish between a “static optimization approach (SOA)”, similar to the previously used iterative FBA [3], and a “dynamic optimization approach (DOA)”. The static approach is useful to get feasible nutrient and biomass dynamics under metabolic constraints, but it cannot resolve the optimization problem over the complete timescale of interest. The dynamic approach DOA directly considers an objective function which depends on the dynamics over the complete timescale, potentially under dynamic metabolic constraints, and thus provides a consistent solution to the dynamic optimization problem. However, in the same way as classical FBA, dynamic FBA uses only a coarse description of biomass composition. Biomass is captured only as one component, and different allocations of biomass to different metabolic tasks, such as considered in RBA, cannot be represented.

In the study described here, we developed a mathematical framework for dynamic models of coupled metabolism and gene expression. We denote such models with the term *metabolic-genetic networks*. From a rigorous timescale separation, we approximate this model by a quasi steady state, constraint-

Table 1: **Different flux optimization approaches**

Optimization approach	No enzyme cost	Enzyme cost included
static	FBA [14]	RBA [9], ME networks [20]
iterative	iFBA [3], dFBA (SOA) [23]	
dynamic	dFBA (DOA) [23]	deFBA (<i>this paper</i>)

based part for the intracellular metabolism, and a dynamic part for the evolution of biomass and substrate concentrations. For this model class, we developed a dynamic optimization approach, called *dynamic enzyme-cost FBA* (deFBA), that includes a detailed description of biomass and accounts for the enzyme cost. The deFBA method respects biophysical constraints motivated from resource balance analysis [9], and additionally includes dynamic changes in biomass composition and substrate concentrations.

In Table 1 we compare our deFBA approach to the established methods available in the literature. The distinction among the methods is based on two criteria: one for the type of optimization approach (static, iterative, or dynamic) and one for whether the enzyme production cost is taken into account for the optimization or not. The comparison highlights the improved generality of deFBA compared to established methods with respect to these two criteria.

We applied the deFBA method to two exemplary metabolic-genetic networks, a minimal nutrient uptake network, and a larger network modeled that describes core cellular processes in bacteria. For the example of a minimal metabolic-genetic network, we evaluated the dynamics resulting from an optimization with a set of biologically meaningful objective functionals. Using a Michaelis-Menten reaction rate for the substrate uptake, we showed that the minimal metabolic-genetic network is equivalent to the empirical Monod growth kinetics. We also observed a close similarity between optimal solutions and Monod kinetics when minimizing the time for substrate metabolism or maximizing the biomass integral with deFBA. We argue that this observation supports the biological validity of these objective functionals.

For the larger network of core cellular processes, we focussed on the discounted biomass integral as objective function. We analyzed different scenarios of nutrient availability which are relevant in a biotechnological setting, including the switch from one carbon source to another and growth under oxygen limitation. With the dynamic optimization approach, we observed

clearly distinguished growth phases, obtained biologically reasonable adaptation dynamics upon changes in nutrient availability, and could predict different dynamic biomass compositions of the cells depending on the growth conditions.

In summary, the proposed deFBA method for metabolic networks coupled with gene expression allows us to infer dynamic adaptations of the cellular metabolic state from biophysical capacity constraints under an optimality principle. The approach can predict metabolic changes occurring in cellular adaptations to a dynamic environment, without knowledge of the involved regulatory mechanisms.

2. Modeling and optimization of metabolic-genetic networks

2.1. Model construction

Our dynamic optimization algorithm is based on a dynamic mass balance model of a metabolic-genetic network. We modeled metabolic-genetic networks with three types of molecular species:

- Extracellular nutrients and waste, with the molar amount vector Y ;
- Intracellular metabolites, with the molar amount vector X ;
- Macromolecules like gene products or large metabolites forming cellular building blocks, with the molar amount vector P .

We split the network reactions accordingly into three classes:

- Exchange reactions, with fluxes V_y , between the cell and the environment;
- Metabolic reactions, with fluxes V_x , converting one set of metabolites into another one;
- Biomass reactions, with fluxes V_p , converting metabolites into macromolecules or vice versa, for example gene expression or anabolic reactions.

We denote the vector of all reaction fluxes as

$$V = (V_y, V_x, V_p)^T. \quad (2)$$

We assumed two general properties of such networks, lead to the time scale separation employed in this paper (described in the Section 2.2):

- Each macromolecule is composed of a large number of small metabolites. For example, a simple production reaction for a macromolecule P from a single metabolite X may be represented as $\alpha X \rightarrow P$, where α is a large stoichiometric coefficient.
- The biomass reactions, i.e., the production of macromolecules, get proportionally slower as the relative stoichiometry α of macromolecules to metabolites increases. This property is accounted for by scaling the biomass reaction fluxes V_p with a small dimensionless factor ε in the dynamic model.

For the purpose of timescale separation, we assumed that each reaction flux can be expressed as a time-varying function $V_i(t, y, x, P)$ of the species concentrations, where y and x are the concentrations of extracellular species and intracellular metabolites, respectively. The explicit time-dependence of V_i represents modulation of enzyme activity and gene expression due to cellular signalling. Assuming that the extracellular volume ϑ_e is constant in time and the cellular volume $\vartheta_c(t, P)$ is a time-varying function of the amount of macromolecules P , the concentrations x and y are defined as

$$y = \frac{Y}{\vartheta_e} \quad x = \frac{X}{\vartheta_c(t, P)}. \quad (3)$$

The reaction fluxes are considered to be given in units of molar amount per time, as is generally recommended for models with multiple compartments, for example in the SBML specification [24]. This is reflected by the assumption that the reaction flux $V_i(t, y, x, P)$ depends on the molar amount of the enzymes, and not on their concentration.

For a general network of this type, we derived the differential equations for the network dynamics from mass balancing as

$$\begin{aligned} \dot{Y} &= -S_y^y V_y \\ \dot{X} &= S_y^x V_y + S_x^x V_x - \alpha \varepsilon S_p^x V_p \\ \dot{P} &= \varepsilon S_p^p V_p, \end{aligned} \quad (4)$$

where the matrices S_j^i , $i, j \in \{x, y, p\}$ describe the stoichiometry of species i in reactions V_j .

2.2. Timescale approximations of metabolic-genetic networks

2.2.1. Transformation to the singular perturbation normal form

Based on the previous transformation (3) to units of concentration for the extracellular species y , we rewrote the metabolic-genetic network model (4) as

$$\begin{aligned} \dot{y}(t) &= -\frac{1}{\vartheta_e} S_y^y V_y(t, y, x, P) \\ \dot{X}(t) &= S_y^x V_y(t, y, x, P) + S_x^x V_x(t, x, P) - \alpha \varepsilon S_p^x V_p(t, x, P) \\ \dot{P}(t) &= \varepsilon S_p^p V_p(t, x, P), \end{aligned} \quad (5)$$

where $x = X/\vartheta_c(t, P)$. Importantly, we did not transform the intracellular metabolic state $X(t)$ to concentrations in the left hand side of (5), in order to avoid the non-linearity in the differential equations that would result from a time-varying volume during growth.

Based on the assumptions in Section 2.1, the time scale separation is expressed mathematically as the limit $\alpha \rightarrow \infty$, $\varepsilon \rightarrow 0$, with the product $\alpha\varepsilon$ staying constant. Since α describes the ratio of biomass molarity to the consumed nutrient molarity, when α tends to infinity with a finite initial nutrient supply $Y(0)$, the model would have a trivial solution where no biomass can be produced. This problem can be avoided by assuming that the extracellular volume $\vartheta_e \rightarrow \infty$, with both the product $\varepsilon\vartheta_e$ and the extracellular nutrient concentrations y remaining finite, i.e., the total nutrient amount is in proportion with the achievable biomass. In summary, we assume 1) a large ratio for the specific molecular mass of macromolecules to metabolites, 2) slow reactions that produce macromolecules, and 3) a large extracellular volume compared to the cellular volume.

A long time scale is then defined by

$$T = \varepsilon t. \quad (6)$$

Assuming that the reaction rates $V_i(t, y, x, P)$ and the cellular volume $\vartheta_c(t, P)$ are slowly varying with respect to their explicit dependence on t , we transformed the model (5) to the long time scale (see equation (A.2) in Appendix A). This assumption will generally be valid for regulation of enzymatic activity by slowly changing environmental conditions, gene expression, and changes in cellular morphology, as these are expected to act on a similar time scale as the accumulation of macromolecules P . Note that the framework can nevertheless accommodate for fast allosteric regulation,

since that would typically be modelled by time-invariant kinetics depending on x only. As an example, a typical rate model for an enzymatic reaction with enzyme P , substrate x_1 , and allosteric inhibitor x_2 is given by $V = V_{max}(P)x_1/((K_1 + x_1)(K_2 + x_2))$ [25], which does not depend explicitly on time and thus is trivially “slowly varying” in the sense required here.

With a slight abuse of notation, we will represent V_i and ϑ_c being slowly varying in the following by writing $V_i(T, y, x, P)$ and $\vartheta_c(T, P)$.

2.2.2. Derivation of a quasi steady state model

The model on the long time scale is in the standard form to perform singular perturbation by Tikhonov’s theorem [26]. To apply this theorem, the following conditions need to be verified:

Condition 1 The quasi steady state equation in the limit $\varepsilon \rightarrow 0$, given by setting the right hand side of \dot{X} to zero in (5), needs to have a locally unique solution

$$X = q(T, y, P) \tag{7}$$

for all admissible values of T , y , and P .

Condition 2 The quasi steady state (7) needs to be an exponentially stable steady state of the boundary layer model (i.e., the fast dynamics of (5)), which is given by

$$\dot{X} = S_y^x V_y(T, y, x, P) + S_x^x V_x(T, x, P) - \alpha \varepsilon S_p^x V_p(T, x, P), \tag{8}$$

uniformly in T , y , and P . Note that for stability analysis, T , y and P are considered as constants in (8).

For the technical details behind these conditions, we refer to [26, chapter 11].

If these conditions hold, one can approximate solutions of the original model by a reduced model, where the fast dynamics \dot{X} are considered to be in quasi steady state, and the variable X is replaced by its quasi steady state solution (7) in the dynamics of the slow variables y and P . For a rigorous presentation of the reduced model, see equation (A.5) in Appendix A. Solutions of the original model on the long time scale are denoted by $(y(T, \varepsilon), X(T, \varepsilon), P(T, \varepsilon))$, and solutions of the reduced model by $(y(T, 0), P(T, 0))$. Application of Tikhonov’s theorem [26, Theorem 11.1] then yields the following result.

Theorem 1. *If the metabolic-genetic network model satisfies Conditions 1 and 2 above and the reduced model (A.5) has a unique solution $(y(T, 0), P(T, 0))$, then there exists $\varepsilon^* > 0$ such that for all $\varepsilon < \varepsilon^*$, the original model (A.2) has a unique solution $(y(T, \varepsilon), X(T, \varepsilon), P(T, \varepsilon))$ on a finite time interval $0 \leq T \leq T^*(\varepsilon)$, and this solution satisfies the bound*

$$\|y(T, \varepsilon) - y(T, 0)\| + \|P(T, \varepsilon) - P(T, 0)\| = \mathcal{O}(\varepsilon). \quad (9)$$

*In addition, for any $T^{**} > 0$, there exists ε^{**} with $0 < \varepsilon^{**} < \varepsilon^*$ such that the solution in the fast variable satisfies*

$$\|X(T, \varepsilon) - q(T, y(T, 0), P(T, 0))\| = \mathcal{O}(\varepsilon) \quad (10)$$

*for all $T \in [T^{**}, T^*(\varepsilon)]$ and $\varepsilon < \varepsilon^{**}$.*

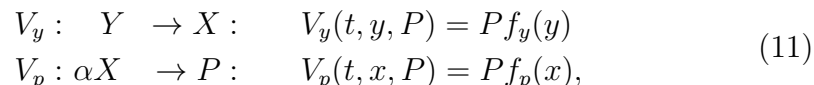
Theorem 1 says that solutions of the reduced model (A.5) can be used to approximate the slow variables of the original model (A.2) for small values of ε , and the quasi steady state solution $q(T, y, P)$ can be used to approximate the fast variable after a short transient. The approximation requires that the boundary layer model, i.e., the dynamics of the metabolic network, has a unique, exponentially stable steady state. Note that the approximation is generally valid on a finite time scale only, but this does not pose a problem here, since we are considering optimization over a finite time span.

Regarding Condition 1, explicit conditions for the existence of a unique steady state have been presented in [27, 28] for the special case of unbranched metabolic pathways with specific feedback regulatory mechanisms. Regarding Condition 2, the stability analysis of metabolic networks is a field of active research, with most stability conditions being only applicable to networks with simple stoichiometries [29, 27].

In order to show rigorously that a given metabolic-genetic network can be well approximated by the quasi steady state model derived here, one would have to check Conditions 1 and 2 based on a kinetic model. These conditions, however, are hard to check in realistic networks because enzyme kinetics are not always known and because it may not be possible to solve for the steady state of X . As is commonly done in constraint-based models, we typically need to assume existence and stability of a quasi steady state based on biophysical insight. However, there are cases where unstable dynamics have been shown for metabolic processes, for example oscillations in glycolysis [30].

2.2.3. *A minimal metabolic-genetic network for nutrient uptake*

As a minimal example for a metabolic-genetic network, we considered a nutrient uptake network composed of one nutrient Y , one intracellular metabolite X , and one gene product P . The minimal network consists of an uptake reaction V_y and a biomass reaction V_p as follows:



where each reaction rate V_y, V_p is split into the amount of gene product P acting as enzyme and a kinetic term f_y, f_p , respectively. With the classical Michaelis-Menten model, one would for example have $f_y(y) = k_{\text{cat}}y/(K_y + y)$.

The quasi steady state constraint on the long timescale obtained from setting dynamics of X to zero is given by

$$P f_y(y) = \alpha \varepsilon P f_p \left(\frac{X}{\vartheta_c(T, P)} \right) \quad (12)$$

for this network. If f_p is invertible in the relevant domain and $P \neq 0$, a quasi steady state solution (7) for X is computed as

$$q(T, y, P) = f_p^{-1} \left(\frac{f_y(y)}{\alpha \varepsilon} \right) \vartheta_c(T, P). \quad (13)$$

The boundary layer model (8) is

$$\dot{X} = P f_y(y) - \alpha \varepsilon P f_p \left(\frac{X}{\vartheta_c(T, P)} \right). \quad (14)$$

Since the boundary layer model describes the short time scale, P, y , and T can be considered as constants. A stability analysis of the linear approximation shows that the steady state is locally exponentially stable uniformly in T, y , and P if there exists a constant $\delta > 0$ independent of T, y , and P , such that

$$\frac{\alpha \varepsilon P}{\vartheta_c(T, P)} f_p' \left(\frac{q(T, y, P)}{\vartheta_c(T, P)} \right) > \delta. \quad (15)$$

This condition will be satisfied if $P/\vartheta_c(T, P)$, i.e., the gene product concentration, is bound away from zero and the function $f_p(x)$ is strictly increasing in x . Assuming a non-vanishing gene product concentration, the stability

condition is satisfied for any kinetics $f_p(x)$ that is strictly increasing in the substrate concentration x .

Finally, the approximate model for the minimal network (11) on a long time scale is given by

$$\begin{aligned}\dot{Y} &= -V_y \\ \dot{P} &= \varepsilon V_p,\end{aligned}\tag{16}$$

together with the quasi steady state constraint

$$V_y - \alpha \varepsilon V_p = 0.\tag{17}$$

Note that we have gone back to units of molar amounts for the extracellular metabolites Y and the original time scale t for easier biological interpretation.

2.3. Dynamic optimization in metabolic-genetic networks

We developed a new dynamic optimization approach to predict the time-courses of fluxes, substrate concentrations, and biomass in metabolic-genetic networks, which we call *dynamic enzyme-cost FBA* (deFBA). We included constraints on enzyme capacity and biomass composition as in resource balance analysis (RBA) [9] or ME networks [20], but considered a dynamic flux optimization problem as in the dynamic approach of dFBA [23].

While in principle it is possible to apply a constraint-based dynamic flux optimization directly on the full metabolic-genetic network model (4), the time scale separation presented in Section 2.2 can greatly help in reducing the complexity of the optimization problem and thus increase the efficiency of the solvers. First, the time scale separation reduces the dimensionality of the problem by removing the metabolic variables X ; second, optimizing only on the slow dynamics allows for a coarser time discretization grid.

In the reduced model (A.6), we still face the problem that we typically do not know the exact kinetics for the reaction fluxes $V(T, y, x, P)$. We can circumvent this problem by considering the fluxes V_i as free variables in the optimization, as is commonly done in constraint based models.

As shown in Section 2.2.2 and Appendix A, the reduced model to be used in the optimization is constructed as

$$\begin{aligned}\dot{Y} &= -S_y^y V_y \\ \dot{P} &= \varepsilon S_p^p V_p,\end{aligned}\tag{18}$$

together with the quasi steady state constraint

$$S_y^x V_y + S_x^x V_x - \alpha \varepsilon S_p^x V_p = 0. \quad (19)$$

In (18) and (19), the reaction flux vectors V_y , V_x , and V_p are considered as free time-dependent variables to be used in the dynamic optimization, subject to constraints described later in this section.

In order to develop the optimization problem, we need to further elaborate on the structure of the metabolic-genetic networks. Most reactions in the network will be catalyzed by an enzyme, which needs to be included in the vector of macromolecular species P . For ease of notation, we will assume that the first m components of P correspond to the network's enzymes, and the remaining components to non-enzymatic macromolecules. Each enzyme catalyzes a set of one or more reactions. The set of reactions catalyzed by enzyme P_i is denoted by the set of indices

$$\text{cat}(i) = \{j \in \mathbb{N} : P_i \text{ catalyzes } V_j\}. \quad (20)$$

We also define a vector b which contains the molecular weights of the macromolecules, such that the scalar product $b^T P$ is equal to the cells' dry weight.

We included the following biophysical constraints in the optimization problem:

- Enzyme capacity constraints: Generally, the reaction fluxes catalyzed by an enzyme P_i are limited by upper and lower bounds of the form

$$\sum_{j \in \text{cat}(i)} \left| \frac{V_j}{k_{\text{cat}, \pm j}} \right| \leq P_i, \quad (21)$$

where each $k_{\text{cat}, +j}$ ($k_{\text{cat}, -j}$) is the forward (backward) k_{cat} value for the reaction V_j . For irreversible fluxes, i.e., non-negatively constrained V_j , the constraint (21) can be written as

$$h_{c,i}^T V \leq e_i^T P, \quad (22)$$

where the j -th component of the vector $h_{c,i}$ is

$$(h_{c,i})_j = \begin{cases} k_{\text{cat}, +j}^{-1} & \text{if } j \in \text{cat}(i) \\ 0 & \text{otherwise,} \end{cases}$$

and e_i is a vector with a 1 in its i -th entry and zero elsewhere. If some of the fluxes catalyzed by the enzyme P_i are reversible, we need to account for positive and negative signs as well as forward and backward k_{cat} values independently, and the constraint (21) needs to be written as

$$H_{c,i}V \leq E_iP, \quad (23)$$

where the matrix $H_{c,i}$ is composed of rows such as the vector $h_{c,i}$ in the irreversible case (22), but covering all combinations of positive and negative signs for the components corresponding to the reversible fluxes, and all rows of the matrix E_i are equal to e_i^T . As an example, consider a case where enzyme P_1 catalyzes V_1 reversibly and V_2 irreversibly, such that (21) can be specified as

$$\left| \frac{V_1}{k_{\text{cat},\pm 1}} \right| + \frac{V_2}{k_{\text{cat},+2}} \leq P_1. \quad (24)$$

Then $H_{c,1}$ matrix is then

$$H_{c,1} = \begin{pmatrix} -k_{\text{cat},-1}^{-1} & k_{\text{cat},+2}^{-1} & 0 & \cdots & 0 \\ k_{\text{cat},+1}^{-1} & k_{\text{cat},+2}^{-1} & 0 & \cdots & 0 \end{pmatrix}, \quad (25)$$

so that both possible signs of V_1 in the constraint (24) are accounted for, and $E_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{pmatrix}$.

At the network level, this translates to the constraints

$$H_c V \leq H_E P, \quad (26)$$

where the matrices H_c and H_E are defined as vertical concatenations of the matrices $H_{c,1}$ to $H_{c,m}$ and E_1 to E_m , respectively, where m is the number of enzymes in P .

- Biomass-independent flux bounds, for example positivity of irreversible fluxes:

$$V_{\min} \leq V \leq V_{\max}. \quad (27)$$

- Positivity of molecular species:

$$P \geq 0, \quad Y \geq 0. \quad (28)$$

- Biomass composition constraints, related to the cell’s solvent capacity: In addition to enzymes, the macromolecular species P will in general also include molecules which form the structural support of the cells, for example membrane molecules. We assumed that this structural components need to be present in a minimal fraction of the total dry weight $b^T P$ of the cells. If P_s is a structural biomass component, such a constraint can be written as

$$\psi_s b^T P \leq P_s, \quad (29)$$

where ψ_s is the minimal fraction of the total dry weight for P_s . If there are several such constraints, resulting for example from different structural components, they can be written in matrix form as

$$H_B P \leq 0, \quad (30)$$

where each row of H_B is of the form $\varphi_s b^T - e_s^T$, for different indices s . With this constraint, the amount of structural cell components will impose an upper bound on the feasible enzyme amount. In conjunction with the enzyme capacity constraint (21), this also sets an upper constraint on the sum of fluxes as in the molecular crowding extension to FBA [31, 32].

For the dynamic optimization problem, these constraints are *path constraints*, i.e., they must be satisfied at every time point within the optimization horizon.

We combined the reduced model (18)–(19) for a metabolic-genetic network with the biophysical constraints into a dynamic optimization problem as follows. For notational convenience, we introduce the variable $Z = (Y, P)$. In addition to the biophysical constraints listed above, some objective functionals required to define a *terminal constraint* in the form of a target set \mathcal{Z} , which was imposed on the network’s state only at the terminal time. We then define the set of all dynamic fluxes V that take the network to the target set within an arbitrary non-negative time t_f as

$$\mathcal{V}(\mathcal{Z}, Z_0) = \bigcup_{t_f \geq 0} \{V \in \mathcal{M}[0, t_f] \mid Z(t_f, V, Z_0) \in \mathcal{Z}\}, \quad (31)$$

where $\mathcal{M}[0, t_f]$ is the set of measurable functions of appropriate dimension over the interval $[0, t_f]$, and $Z(t_f, V, Z_0)$ is the solution of the differential equation (18) with flux variables $V(t)$ and initial condition $Z(0) = Z_0$ [33].

The objective functional J in the optimization is given in general form as an integral over the dynamic variables plus a term for the terminal state:

$$J = \int_0^{t_f} \Phi(Y(t), P(t), V(t)) dt + \Psi(Z(t_f)). \quad (32)$$

The dynamic optimization problem is then constructed as

$$\begin{aligned} & \max_{\mathcal{V}(Z, Z_0)} \int_0^{t_f} \Phi(Y(t), P(t), V(t)) dt + \Psi(Z(t_f)) \\ & \text{s.t. } \dot{Y} = -S_y^y V_y \\ & \quad \dot{P} = \varepsilon S_p^p V_p \\ & \quad Z(0) = Z_0 \\ (19) \quad & S_y^x V_y(t) + S_x^x V_x(t) - \alpha \varepsilon S_p^x V_p(t) = 0 \\ (26) \quad & H_c V(t) \leq H_E P(t) \\ (27) \quad & v_{\min} \leq V(t) \leq v_{\max} \\ (28) \quad & Z(t) \geq 0 \\ (30) \quad & H_B P(t) \leq 0, \end{aligned} \quad (33)$$

where equation numbers refer to the derivation of the constraints above.

A variety of approaches are available to numerically solve such an optimization problem [34]. For this study, we applied collocation methods based on a time discretization of the dynamic variables $Y(t)$, $P(t)$, and $V(t)$ [35, 36, 37]. The discretized problem was then solved with the Python optimization package `cvxopt` (<http://abel.ee.ucla.edu/cvxopt/>). Details are given in Appendix B.

3. Analysis of a minimal metabolic-genetic network and the Monod growth kinetics

As a first case study to demonstrate the deFBA approach, we considered the minimal nutrient uptake network introduced in (11), with the approximated dynamics given by (16)–(17). We used the deFBA method to compare the optimal solutions for three different objective functionals. Biomass maximization is a common objective used in classical flux balance analysis

[16]. We incorporated it into our algorithm in two alternative ways: first, as biomass maximization at the end of the optimization horizon, and second, as discounted maximization of the biomass integrated over time. The maximization of terminal biomass has also been used in dFBA [23], whereas the discounted biomass objective has been used as an evolutionary fitness measure in a recent analysis of microbial metabolism [38]. As third objective, we considered the minimization of the time required to metabolize the available nutrient completely. The minimization of the substrate consumption time has also been used previously to predict enzyme concentrations in pathway activation [39, 40].

Formally, the three objective functionals were defined as follows.

- Maximization of biomass at the end of the considered time interval:

$$J_1 = P(t_f) \tag{34}$$

- Discounted maximization of the biomass integral:

$$J_2 = \int_0^{t_f} P(\tau) e^{-\varphi\tau} d\tau \tag{35}$$

with a discount parameter $\varphi \geq 0$. A positive discount parameter can in general be used to reduce the effect of the terminal time t_f on the objective function, since the objective function value is uniformly bounded for varying terminal times, provided that the discount parameter is larger than the maximal growth rate.

- Minimization of the time required to metabolize the nutrient completely:

$$J_3 = - \int_0^{t_f} d\tau = -t_f \tag{36}$$

with the terminal constraint

$$Y(t_f) = 0. \tag{37}$$

The biophysical inequality constraints for this example were given by positivity of the molecular species and irreversibility of the two fluxes:

$$\begin{aligned} 0 &\leq Y & 0 &\leq P \\ 0 &\leq V_y & 0 &\leq V_p \end{aligned} \tag{38}$$

We also constrained the enzymatic capacity according to (21), where P is considered as an enzyme catalyzing both the uptake reaction V_y and the biomass reaction V_p :

$$\frac{V_y}{k_{\text{cat},y}} + \frac{\varepsilon V_p}{k_{\text{cat},p}} \leq P, \quad (39)$$

where $k_{\text{cat},y}$ and $k_{\text{cat},p}$ are the catalytic constants for V_y and V_p , respectively.

We performed an analytical study of the optimization problem for the objective functionals J_2 and J_3 , which is described in Appendix C. The analysis showed that there is a unique optimal solution which is composed of an initial exponential growth phase until the substrate is completely metabolized, followed by a stationary phase of nil growth. This is in agreement with the typical growth kinetics of bacterial cultures [41]. We also found that the time t_s at which the culture switches from exponential to stationary growth is:

$$t_s = \left(\frac{\alpha}{k_{\text{cat},y}} + \frac{1}{k_{\text{cat},p}} \right) \log \left(1 + \frac{Y_0}{\alpha P_0} \right), \quad (40)$$

where Y_0 and P_0 are the initial concentrations of metabolites and biomass, respectively. The analytical study showed that the minimal network will reach the maximal biomass when the nutrient is depleted. Regarding the objective J_1 , this means that any solution where the nutrient is depleted at the terminal time t_f is an optimal solution for the objective J_1 . The solution for objective J_2 , for example, is also optimal for objective J_1 , suggesting that the optimum for J_1 may not be unique. We confirmed the non uniqueness numerically as described next.

We computed numerical solutions for all three objective functionals, using a discretization approach as described in Appendix B. The optimization results are shown in Figure 1. In case of objectives J_2 and J_3 , the numerical solution is in agreement with the analytically predicted biphasic growth kinetics.

We also computed the variability of the optimal solutions, using a similar concept as in flux variability analysis [42]: We solved an optimization problem as given in (33), but with the original objective functional as an additional constraint, and the minimization or maximization of a flux or state variable at one time point as new objective. By computing the variability of the optimal solutions point-wise in time, a significant variability of the optimal solutions is observed for the objective J_1 , whereas the uniqueness of the solutions for objectives J_2 and J_3 manifests itself in the fact that there is no point-wise

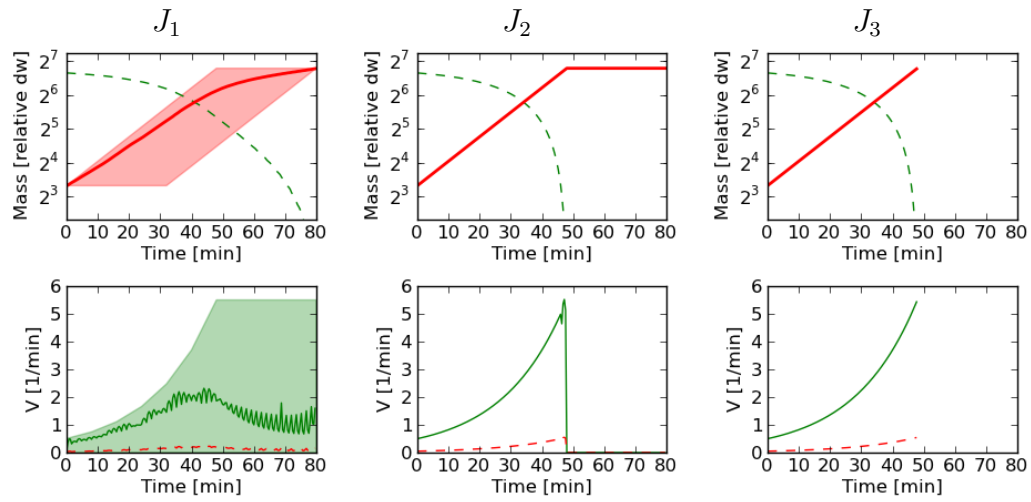


Figure 1: Dynamic optimization results for the minimal metabolic-genetic network. Top row: Optimal time courses of substrate (dashed green line) and biomass (red line) for objective functionals J_1 , J_2 , and J_3 . Bottom row: Optimal time courses of fluxes (V_y green, εV_p dashed red). Filled regions show the point-wise variability of the optimal solution. The following parameter values were used: $\alpha = 10$, $\varepsilon = 0.1$, $k_{\text{cat},y} = 1 \text{ min}^{-1}$, $k_{\text{cat},p} = 1 \text{ min}^{-1}$, $\varphi = 0.01 \text{ min}^{-1}$, $t_f = 80 \text{ min}$ (for objectives J_1 and J_2). The initial condition was taken as $Y(0) = 100$ and $P(0) = 1$.

variability in the numerical solutions (bottom row of Figure 1).

Since the proposed optimization method does not assume any specific reaction kinetics, it yields optimal solutions that, in principle, may not be realizable with a plausible biochemical mechanism. To check whether the optimal biphasic growth profile is consistent with a realistic kinetic law, we compared it to the simulated growth curves obtained with a typical kinetic model. We extended the minimal network to a kinetic model with reaction rates given by the Michaelis-Menten law:

$$\begin{aligned} V_y &= \frac{k_y P Y}{K_y \vartheta_e + Y} \\ \varepsilon V_p &= \frac{k_p P X}{K_p + X}. \end{aligned} \tag{41}$$

For a faithful comparison with the dynamic optimization results, the parameters k_y and k_p need to satisfy the enzymatic capacity constraint (39). We aimed for parameter values ensuring that the enzyme P operates at full capacity, corresponding to the constraint (39) being satisfied with equality. Together with the quasi steady state constraint (17), we obtained parameter values given by

$$\begin{aligned} k_y &= \left(\frac{1}{k_{\text{cat},y}} + \frac{1}{\alpha k_{\text{cat},p}} \right)^{-1} \\ k_p &= \left(\frac{\alpha}{k_{\text{cat},y}} + \frac{1}{k_{\text{cat},p}} \right)^{-1}. \end{aligned} \tag{42}$$

Simulations of the kinetic model (shown in Figure 2) were practically identical to the optimal solutions for objectives J_2 and J_3 , both of which were computed without presupposing any specific reaction kinetics. These simulation results are similar to the classical Monod model of bacterial growth [41]. In Appendix D, we rigorously prove the equivalence of the minimal nutrient uptake network (11) and the Monod model.

4. Dynamic optimization of a core carbon network

4.1. Network description

We constructed a metabolic-genetic network model (see Figure 3) as an abstraction of core processes relating carbon uptake and growth. It accounts for the uptake of different extracellular species as nutrients, including two

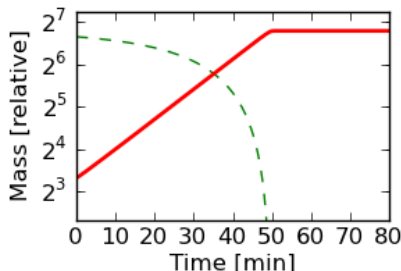


Figure 2: Simulation of the minimal metabolic-genetic network with Michaelis-Menten reaction rates. Parameter values were set as $k_y = 0.5 \text{ min}^{-1}$, $k_p = 0.05 \text{ min}^{-1}$, $K_y \vartheta_e = K_p = 1$ and initial condition $Y(0) = 100$, $X(0) = 0$, $P(0) = 1$.

carbon sources **Carb1** and **Carb2**, oxygen, fermentation products, and other organic molecules. The model includes the major anabolic and catabolic processes together with the translational mechanisms for ribosome and enzyme assembly.

The stoichiometry of the exchange and metabolic reactions were taken from the original model proposed in [8]. However, in contrast to the work in [8], we do not include regulatory interactions but instead added catalytic enzymes, structural macromolecules, and ribosomes as biomass components with appropriate biomass production reactions. We used deFBA to predict the adaptation dynamics by optimizing the reaction rates without the need to include regulatory interactions.

The model is detailed in Table 2–3, together with the utilized enzyme catalytic constants. To get reasonable flux bounds on reactions describing diffusive exchange across the plasma membrane, we defined the structural component **S** as the enzymatic macromolecule for these reactions, together with an appropriate rate constant for diffusion over the plasma membrane or through pore proteins. In Table 2, enzymes are denoted as **T_i** for transport enzymes and **E_j** for catalytic enzymes. The macromolecules for biomass reactions and diffusive transport are the ribosome **R** and the structural component **S**, respectively. The catalytic constants of the enzymes are based on typical k_{cat} values in metabolism [43, 44, 45]. While the original model [8] did not specify which reactions were modelled as reversible, our model includes reversible amino acid biosynthesis, and reversible synthesis and secretion of waste products **D** and **E**.

The catalytic constants of the biomass reactions, corresponding to protein translation, are based on the measured translation rate of 17 amino acids/s in

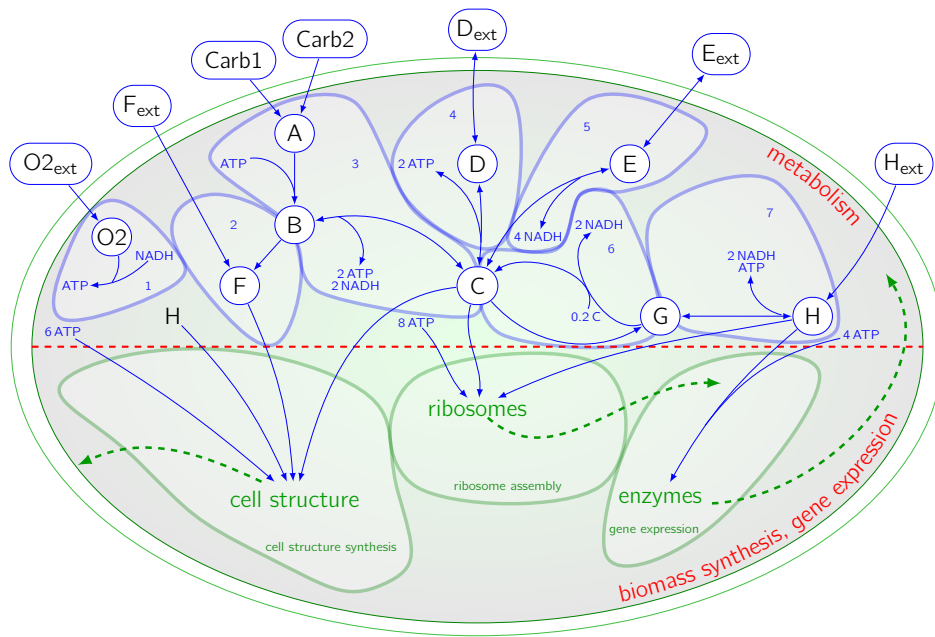


Figure 3: Schematic diagram of the abstract core metabolic-genetic network. Upper part of the diagram shows metabolic processes, the lower part biomass / gene expression processes. Abstract representations of core metabolic processes are labelled as follows: (1) Electron transport chain, (2) Lipid metabolism, (3) Glycolysis, (4) Acetate fermentation, (5) Ethanol fermentation, (6) Citric acid cycle, (7) Amino acid metabolism. Extracellular species correspond to different nutrients and waste products as detailed in the main text. Intracellular species A–H correspond to metabolite pools, and O₂ is intracellular oxygen.

E. coli [46]. The catalytic constant in the model for each protein translation reaction is the ratio of this translation rate and the number of amino acid residues in the specific protein. The ATP requirements for the gene expression reactions are deduced from a flux-based analysis of gene expression [47]. The stoichiometry of amino acids in the biomass reactions is close to the observed average protein length of 300 amino acids [48], but with a higher stoichiometry for reactions which are meant to be a condensed description of a complex pathway.

Despite its relative simplicity, the model includes a range of metabolic processes typically found in microbes, such as uptake of alternative carbon sources, aerobic and anaerobic glycolysis, and the uptake and synthesis of lipids and amino acids. The network includes fermentation products D and E, which can also be reutilized as nutrients when oxygen is available. The amino acid H can be used directly for the synthesis of proteins, or be diverted to the carbon metabolism. The network is parametrized such that **Carb1** should be expected to be the preferred carbon source as compared to **Carb2**, since the corresponding uptake reaction $\text{Carb1} \rightarrow \text{A}$ is modeled with a higher enzymatic efficiency and a lower enzyme production cost than $\text{Carb2} \rightarrow \text{A}$. The biosynthetic pathways $\text{B} \rightarrow \text{F}$ and $\text{G} \rightarrow \text{H}$ are modeled to require a substantially larger investment in terms of enzyme production cost compared to the uptake pathways $\text{F}_{\text{ext}} \rightarrow \text{F}$ and $\text{H}_{\text{ext}} \rightarrow \text{H}$. By the C-G cycle illustrated in Figure 3, the carbon sources can be completely catabolized under aerobic conditions. Otherwise, they can be catabolized to the fermentation products D and E.

4.2. Dynamic optimization of the core carbon network

We used our deFBA algorithm to predict the substrate concentrations, biomass composition, and growth kinetics under three metabolic adaptation scenarios: carbon switch, oxygen limitation and rich medium growth. Since the analysis of the minimal network suggested that maximizing the discounted biomass integral is a biologically reasonable objective in a dynamic setting, we used this objective functional for the analysis of the core carbon network. We modeled each scenario by manipulating the initial conditions of the model as detailed in Table 4.

We computed the initial biomass composition by resource balance analysis (RBA) [9] so as to yield maximal aerobic growth rate on **Carb1** alone. The initial biomass composition thus corresponded to a cellular state pre-adapted to such an environment. RBA determines the biomass distribution for the

Table 2: Exchange and metabolic reactions, together with rate constants for the core metabolic-genetic network. Reversible reactions use equal forward and backward k_{cat} values.

Reaction	Enzyme	k_{cat} (min^{-1})
Exchange reactions		
Carb1 \rightarrow A	T _{C1}	3000
Carb2 \rightarrow A	T _{C2}	2000
F _{ext} \rightarrow F	T _F	3000
O2 _{ext} \rightarrow O2	S	1000
D \leftrightarrow D _{ext}	S	1000
E \leftrightarrow E _{ext}	S	1000
H _{ext} \rightarrow A	T _H	3000
Metabolic reactions		
A + ATP \rightarrow B	E _B	1800
B \rightarrow C + 2 ATP + 2 NADH	E _C	1800
B \rightarrow F	E _F	1800
C \rightarrow G	E _G	1800
G \rightarrow 0.8 C + 2 NADH	E _N	1800
C \leftrightarrow 2 ATP + 3 D	E _D	1800
C + 4 NADH \leftrightarrow 3 E	E _E	1800
G + ATP + 2 NADH \leftrightarrow H	E _H	1800
NADH + O \rightarrow ATP	E _T	1800

Table 3: Biomass reactions, weight vector b and initial conditions for biomass product, together with the rate constants for the core metabolic-genetic network. All biomass reactions are catalyzed by the ribosome R .

Reaction	b	$P(0) / (\mu\text{gl}^{-1})$	$k_{\text{cat}} (\text{min}^{-1})$
$400 \text{ H} + 1600 \text{ ATP} \rightarrow \text{T}_{\text{C1}}$	4	17.0	2.5
$1500 \text{ H} + 6000 \text{ ATP} \rightarrow \text{T}_{\text{C2}}$	15	0.0	0.67
$400 \text{ H} + 1600 \text{ ATP} \rightarrow \text{T}_{\text{F}}$	4	0.0	2.5
$400 \text{ H} + 1600 \text{ ATP} \rightarrow \text{T}_{\text{H}}$	4	0.0	2.5
$500 \text{ H} + 2000 \text{ ATP} \rightarrow \text{E}_{\text{B}}$	5	28.3	2
$500 \text{ H} + 2000 \text{ ATP} \rightarrow \text{E}_{\text{C}}$	5	25.8	2
$1000 \text{ H} + 4000 \text{ ATP} \rightarrow \text{E}_{\text{D}}$	10	3.6	1
$1000 \text{ H} + 4000 \text{ ATP} \rightarrow \text{E}_{\text{E}}$	10	0.0	1
$1500 \text{ H} + 6000 \text{ ATP} \rightarrow \text{E}_{\text{F}}$	15	2.5	0.67
$500 \text{ H} + 2000 \text{ ATP} \rightarrow \text{E}_{\text{G}}$	5	29.7	2
$2500 \text{ H} + 10000 \text{ ATP} \rightarrow \text{E}_{\text{H}}$	25	14.7	0.4
$500 \text{ H} + 2000 \text{ ATP} \rightarrow \text{E}_{\text{N}}$	5	15.0	2
$500 \text{ H} + 2000 \text{ ATP} \rightarrow \text{E}_{\text{T}}$	5	52.1	2
$2000 \text{ H} + 4000 \text{ C} + 16000 \text{ ATP} \rightarrow \text{R}$	60	29.2	0.2
$250 \text{ H} + 250 \text{ C} + 250 \text{ F} + 1500 \text{ ATP} \rightarrow \text{S}$	7.5	233	3

Table 4: Initial nutrient conditions (mM), oxygen inflow V_{O} (mM s^{-1}), oxygen turnover γ_{O} (s^{-1}), initial biomass $b^{\text{T}}P(0)$ (g l^{-1}), and discount factor φ (min^{-1}) for scenarios 1–3

Scenario	Carb1	Carb2	O2 _{ext}	D _{ext}	E _{ext}	F _{ext}	H _{ext}	V_{O}	γ_{O}	$b^{\text{T}}P(0)$	φ
1	2	30	50	0	0	0	0	20	0.4	0.005	0.1
2	50	0	5	0	0	0	0	2	0.4	0.005	0.1
3	50	0	50	0	0	5	5	20	0.4	0.005	0.3

maximal growth rate in steady state exponential growth, thus avoiding spurious transients caused by a non-optimal initial composition of biomass.

We assumed the culture to be run as an aerated batch process. The extracellular oxygen dynamics were modeled by the differential equation

$$\frac{d}{dt} \text{O2}_{\text{ext}} = V_{\text{O}} - \gamma_{\text{O}} \text{O2}_{\text{ext}}, \quad (43)$$

where V_{O} is the oxygen inflow and γ_{O} the ventilation rate, with different values per Scenario as given in Table 4. Oxygen turnover described by (43) is added to the full biological model for the dynamic optimization.

The objective functional is the discounted biomass integral

$$J = \int_0^{t_f} b^{\text{T}} P(t) e^{-\varphi t} dt, \quad (44)$$

where b is the weight vector for the individual biomass components as given in Table 3, P contains the reaction products of the biomass reactions listed in Table 3, and φ is the discount factor as given in Table 4 for the three scenarios.

In addition to the enzyme capacity constraint, we used a biomass composition constraint to ensure that the structural component S makes up for at least 35 % of total biomass:

$$0.35b^{\text{T}}P \leq S. \quad (45)$$

4.3. Scenario 1: carbon switch

In this scenario, we studied cellular growth under both carbon sources, with a low concentration for a preferred carbon source **Carb1**, and a high concentration of the other carbon source **Carb2**. The optimization predicted four distinct growth phases, labelled a–d in Figure 4. In phase a, cells grew exclusively on **Carb1**. After its depletion, they switched to **Carb2** uptake (phase b). The optimization predicted a nutrient uptake pattern comparable to catabolite repression [49], where a preferred carbon source is completely consumed before cells switch to the non-preferred carbon source. In the growth phases a and b, cells produced the waste metabolite D. When both carbon sources were consumed, the optimization predicted re-consumption of the previously excreted waste metabolite D (phase c), thus being able to

sustain growth, though at a significantly lower rate. The stationary phase d was reached after complete consumption of the substrate.

The predictions seem to indicate a significant intracellular reorganization well before the complete depletion of the second carbon source (see Figure 4B–C). This means that the optimal response suggests a cellular adaptation to the impending nutrient depletion, reminiscent of changes in gene regulation observed in *S. cerevisiae* just before a glucose–gluconeogenic switch [50]. It is known that some microorganisms, for example *Bacillus subtilis*, monitor the ratio of population size to nutrient availability by quorum sensing [51], which could allow them to predict an upcoming nutrient depletion and appropriately adjust their gene expression levels before starvation.

Figure 4D illustrates the predicted reorganization of the metabolic network in terms of the enzyme activities. Glycolysis, represented by the enzyme E_B , was only active in phases a and b. The C–G cycle, involving enzymes E_G and E_N , was most active in phase c, where ATP production from glycolysis ceased and had to be substituted by respiration. There was a significant drop in activity of the amino acid synthesis pathway represented by E_H in phase c, presumably related to the lower enzyme and ribosome biomass fraction in that phase.

4.4. Scenario 2: oxygen limitation

In this scenario, we studied cell growth under limited oxygen availability. The deFBA predictions display five growth phases, labelled a–e in Figure 5. In phase a, cells grew aerobically on **Carb1**. In phase b, oxygen was depleted and cells continued to grow anaerobically, producing both waste metabolites **E** and **D**. There was also a significant drop in the growth rate during phase b. The growth phases c and d show that there is also a preferential order in the re-metabolization of waste products.

With the small continuous supply of oxygen that is considered in the model, cells first grew exclusively on **E** (phase c), consuming **D** only after **E** has been depleted (phase d). Phase e was then the stationary phase.

From the enzyme activities shown in Figure 5D, we got similar results as in scenario 1 concerning glycolysis and amino acid synthesis. According to the oxygen limitation, there was a significant drop in the respiratory activity, represented by the enzyme E_T , after phase a.

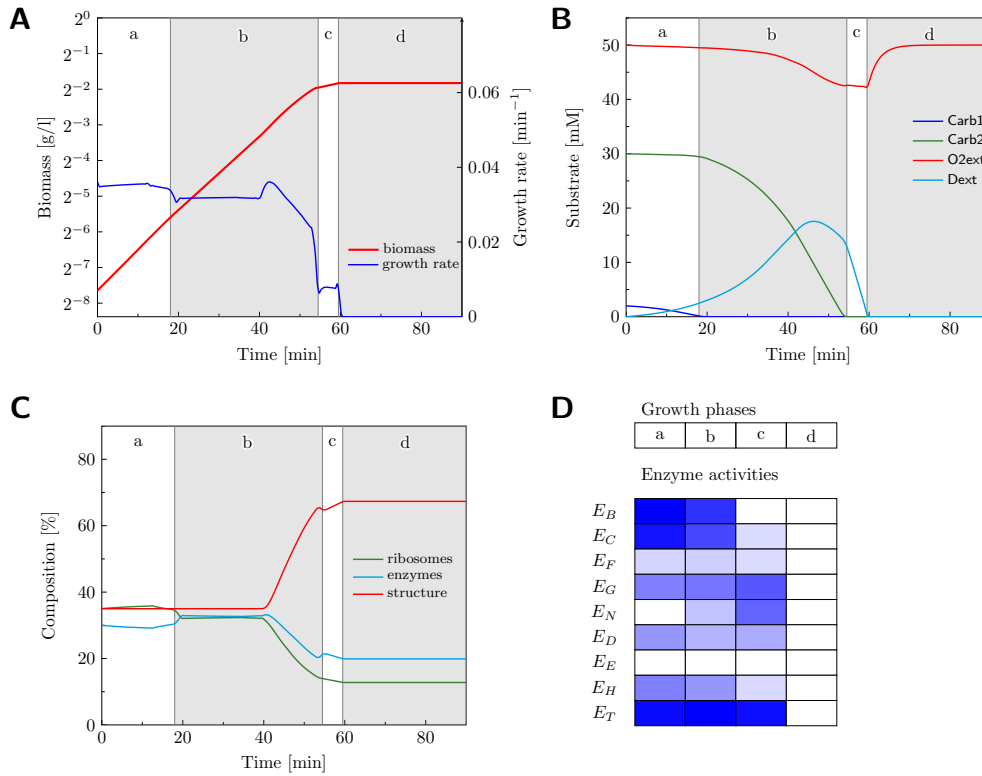


Figure 4: Dynamic optimization results for the core metabolic-genetic network in Scenario 1 (two carbon sources). Time intervals a-d show different growth regimes according to substrate availability. A: Biomass and growth rate. B: Concentrations of extracellular metabolites. C: Dry weight percentages for aggregate cellular components. D: Enzyme activity during growth phases a-d, averaged over each phase. Stronger blue denotes higher relative enzyme activity within a phase. See Table 2 for enzyme labels.

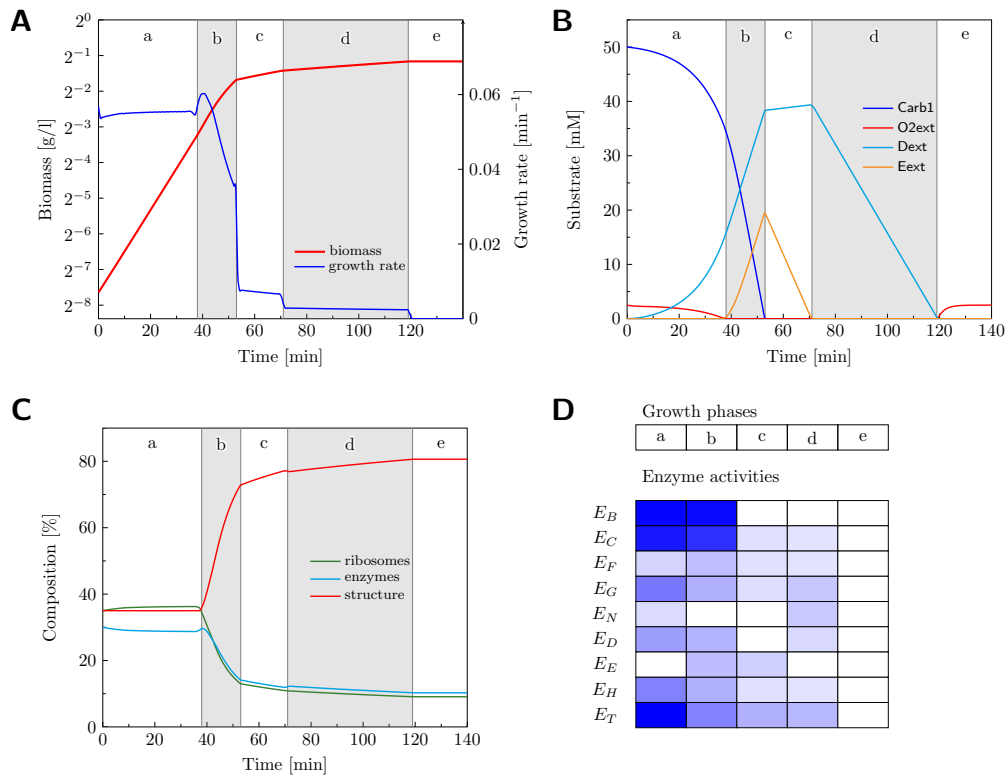


Figure 5: Dynamic optimization results for the core metabolic-genetic network in Scenario 2 (oxygen limitation). Phases a-e show different growth regimes according to substrate availability. A: Biomass and growth rate. B: Concentrations of extracellular metabolites. C: Dry weight percentages for aggregate cellular components. D: Enzyme activity during growth phases a-e, averaged over each phase. Stronger blue denotes higher relative enzyme activity within a phase.

4.5. Scenario 3: transition to and growth in rich medium

In this scenario, we studied the effect of adding amino acids (H_{ext}) and lipids (F_{ext}) to the medium for cells grown previously on **Carb1** alone. Initially, cells were assumed to be pre-adapted to a medium without amino acids and lipids, and the optimization predicted an initial transient where cells adapted to the new medium. The model predicts expression of amino acid and lipid transporters so as to shift from synthesis to uptake of amino acids and lipids from the medium. We then observed five growth phases, labelled a–e in Figure 6.

In phase a, where also the generic amino acid **H** is available in the medium, cells grew with a significantly higher rate than in the other scenarios. This higher growth rate presumably results from decreased enzyme cost for biosynthesis pathways during this phase. Also, during phases a and b the ribosomal biomass fraction was significantly increased (Figure 6 C), while the enzymatic biomass fraction was reduced. This agrees with previous studies about the dependence of the growth rate on global cellular parameters, where increased mRNA and decreased protein fractions have been associated with increased growth rate [52]. After depletion of external **F** and **H**, the ratio between ribosomes and enzymes returned to the original state, while the cells continued to grow on **Carb1** alone (phase c). When **Carb1** was depleted, cells re-metabolized the waste product **D** (phase d), before entering stationary phase (e).

An additional observation from the enzyme activities shown in Figure 6 D is that glycolysis was most active only in phase b, presumably due to the need of precursor molecules for amino acid synthesis after the extracellular supply of **H** had been used up.

5. Discussion

In this paper we have presented a dynamic optimization approach for metabolic networks coupled with gene expression. The proposed deFBA method can be used to predict the dynamic adaptation of intracellular metabolites and biomass composition under environmental perturbations. Based on a biophysically motivated approximation by separation of timescales, we obtained a reduced model for metabolic networks coupled with biomass dynamics. In the reduced model, the metabolic fluxes are computed from a quasi steady state approximation, while the nutrient concentrations, enzyme

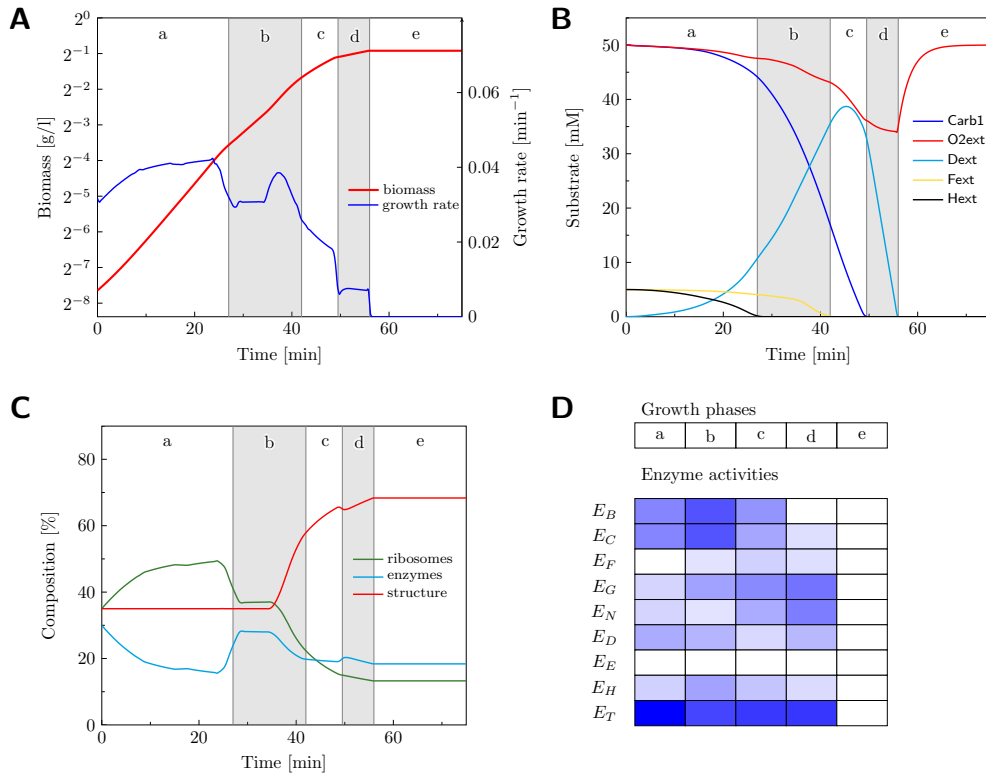


Figure 6: Dynamic optimization results for the core metabolic-genetic network in Scenario 3 (rich medium). Phases a-e show different growth regimes according to substrate availability. A: Biomass and growth rate. B: Concentrations of extracellular metabolites. C: Dry weight percentages for aggregate cellular components. D: Enzyme activity during growth phases a-e, averaged over each phase. Stronger blue denotes higher relative enzyme activity within a phase.

expression levels and other biomass components are described by a system of differential equations.

The inclusion of a detailed biomass composition with constraints on fluxes from enzyme capacity was previously proposed in RBA [9] and ME networks [20] for steady state models. We developed a dynamic optimization method, called dynamic enzyme-cost flux balance analysis (deFBA), which includes the biomass composition and related enzyme capacity constraints in a dynamic model. The algorithm allows predicting the time courses of metabolic fluxes and biomass concentrations from an optimality principle. Our approach provides a generalization to the currently established dynamic flux balance analysis (dFBA) reported in [23], and offers two advantages. Firstly, it can readily account for constraints on enzyme levels and their biosynthetic cost. Secondly, it is based on computationally efficient linear optimization techniques and thus seems well suited for large scale networks. Computational costs can become particularly limiting in dFBA, as it is based on non-linear optimization techniques that are hard to scale with network size. In addition, deFBA solves one optimization problem for the dynamics over the complete time interval, thus also avoiding numerical problems in differential equations constrained by linear programs [22], as encountered in iterative FBA or the static optimization approach of dFBA [3, 23].

The deFBA approach allows predicting changes in the enzyme expression levels from an optimization principle alone, without the need of explicit models for gene regulatory interactions as proposed in other FBA-related methods [8, 5]. We exploited this feature by using deFBA to predict metabolic adaptations in two biologically relevant metabolic-genetic systems: a minimal nutrient uptake network and a larger core carbon uptake system.

For the minimal network we showed numerically and analytically that the optimal growth kinetics are composed of an exponential and a stationary phase. These biphasic growth kinetics appear in the solutions that maximize the discounted biomass integral or minimize the nutrient metabolization time. We subsequently found that these optimal growth kinetics are essentially equivalent to the classical Monod model of bacterial growth [41] and can be accurately modeled with uptake kinetics following a Michaelis-Menten law. This result can be interpreted as a rigorous derivation of the Monod growth model, which was so far an empirical model based on observed growth dynamics. The close correspondence of the Monod growth model with the dynamic optimization results suggests that maximization of the discounted biomass integral or the minimization of the metabolization

time are biologically plausible objectives in a dynamic context, and in fact both have been previously used in evolutionary [38] or metabolic pathway studies [39, 40], respectively.

The plausibility of other objective functionals should also be examined in deFBA, similarly as has been done for static flux balance analysis [17]. However, some objectives which have been proposed there, such as the maximization of ATP production, seem to be more appropriate for networks focussing on the metabolic level alone, and may not be appropriate for metabolic-genetic networks. A dynamic optimization approach opens up the possibility of using other objective functions that do not apply in a static setting, such as the minimization of the metabolization time introduced in Section 3.

We also considered the maximization of the final biomass in the minimal network, as this was one of the original objectives proposed with dFBA [23]. The predictions, however, showed a significant variability due to non-uniqueness of the optimal solution. Non-unique optimal solutions are a common problem with all constraint-based models [53]. In networks with two equivalent parallel pathways, classical FBA would not be able to predict which of the two is active. This situation can be remedied by considering different enzyme costs for the two pathways, as first suggested in [9] for static optimization, and in our study for the dynamic case (Section 4.3). The use of static objective functionals in a dynamic setting can aggravate the problem of non-uniqueness. For example, as we observed in Figure 1, when maximizing the terminal biomass, a dynamic approach may not be able to decide whether a single pathway is active early or late within the considered time horizon. The alternative objective functionals that we proposed in this study remedied the problem of non-unique optimal solutions in the time domain, as growth at an earlier time would give a better objective functional value than growth at a later time.

Our study on the core carbon network used a larger model that includes the uptake of different extracellular species (including nutrients, oxygen, and organic precursor molecules) together with some of the main energy turnover processes, and the assembly of ribosomes and enzymes. We applied the deFBA method to predict the growth kinetics and time courses for substrates and biomass composition that maximized the discounted biomass integral in three scenarios. These scenarios were chosen from classical examples of metabolic adaptation processes: the switch from one carbon source to another, growth under reduced oxygen availability, and the transition of cells to a rich medium, which subsequently gets depleted of nutrients. The resulting

growth kinetics reproduce a number of known biological observations, such as the overall cellular composition under different growth conditions [52] or a hierarchy of preferences for different carbon sources [54]. The results show that a significant change in gene regulation before an impending nutrient depletion would be optimal for cellular growth. In view of recent experimental results on gene regulation during the glycolytic-gluconeogenic switch [50], we suggest growth optimality over the full time period as theoretical explanation for regulatory activity before the switching time point.

The metabolic part of the model in Section 4 was first analyzed with regulatory FBA (rFBA) in [8]. We compared these results to our results obtained by the deFBA method. Both approaches result in a sequence of growth phases with a distinct metabolic flux pattern. For the specific scenarios, both the carbon switch and the anaerobic growth have also been studied in [8] and gave similar results, for example the preference for one carbon source over the other. Importantly, in rFBA, this preference resulted from explicitly building it into the regulatory logic, while with deFBA, it followed implicitly from the different enzymatic efficiencies and metabolic costs of the alternative pathways. While our results predict the re-metabolization of fermentation products, similar to an observed acetate re-utilization in *E. coli* [3], which has also been reproduced in iterative FBA [3] and dynamic FBA [23], the corresponding reactions seemed to be modelled as irreversible in the rFBA study [8], which would have prevented the re-metabolization there.

An important general distinction between rFBA and deFBA concerns the biological knowledge required to build the network models. In rFBA, fluxes are constrained by Boolean rules modeling regulatory mechanisms, which have to be known in the modeling step. With deFBA, regulatory interactions are not included in the model, but the specific enzymes for each reaction together with their metabolic production costs are added to the network and thus need to be known. Importantly, these two approaches do not exclude each other: it should be well possible to construct network models with both regulatory constraints and constraints from the capacity and metabolic costs of individual enzymes.

Although we have focussed on the metabolic constraints relating to enzymatic capacity, the deFBA framework readily allows for inclusion of thermodynamic constraints on metabolic fluxes [55]. Moreover, a recent study has suggested that constraints on gene regulatory mechanisms may also be relevant and contribute to some mismatch between observed gene expression

and fitness levels predicted from the theoretical optimum based on metabolic constraints [56]. The inclusion of suitable constraints on the regulatory mechanisms in a metabolic optimization framework is an open problem, but we suggest that it may become tractable in a dynamic optimization setup such as deFBA, for example by including metabolic costs of gene regulation. In addition, the objective functions considered here do not take the robustness against unpredictable changes in the environment into account, which may also contribute to slower growth than predicted from a pure growth rate optimization [57].

In essence, the approach presented here can predict the temporal regulation of gene expression from an optimization principle, without requiring any knowledge of regulatory interactions. It yields predictions for biomass dynamics in metabolic adaptations, while respecting constraints of enzymatic capacity and mass conservation.

Appendix A. Derivation of the long time scale models

Since the reaction fluxes $V_i(t, y, x, P)$ and the cellular volume $\vartheta_c(t, P)$ are assumed to be slowly varying, we could rewrite them on the long time scale as

$$\begin{aligned} V_i(\varepsilon^{-1}T, y, x, P) &= \tilde{V}_i(T, y, x, P) \\ \vartheta_c(\varepsilon^{-1}T, P) &= \tilde{\vartheta}_c(T, P) \end{aligned} \quad (\text{A.1})$$

even in the limit $\varepsilon \rightarrow 0$.

On the long time scale, the metabolic-genetic network is then rewritten as

$$\begin{aligned} y' &= -\frac{1}{\varepsilon \vartheta_e} S_y^y \tilde{V}_y \left(T, y, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) \\ \varepsilon X' &= S_y^x \tilde{V}_y \left(T, y, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) + S_x^x \tilde{V}_x \left(T, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) \\ &\quad - \alpha \varepsilon S_p^x \tilde{V}_p \left(T, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) \\ P' &= S_p^p \tilde{V}_p \left(T, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right), \end{aligned} \quad (\text{A.2})$$

where we used $X' = dX/dT$ to denote the time derivative on the long time scale.

Based on the model (A.2), the quasi steady state equation for Condition 1 in Section 2.2.2 is given by

$$\begin{aligned} S_y^x \tilde{V}_y \left(T, y, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) + S_x^x \tilde{V}_x \left(T, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) \\ - \alpha \varepsilon S_p^x \tilde{V}_p \left(T, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) = 0, \end{aligned} \quad (\text{A.3})$$

and the boundary layer model (8) is more explicitly written as

$$\begin{aligned} \dot{X} = S_y^x \tilde{V}_y \left(T, y, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) + S_x^x \tilde{V}_x \left(T, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right) \\ - \alpha \varepsilon S_p^x \tilde{V}_p \left(T, \frac{X}{\tilde{\vartheta}_c(T, P)}, P \right), \end{aligned} \quad (\text{A.4})$$

where the slow variables T , y , and P are considered as constant in (A.3) and (A.4).

The reduced model is constructed as

$$\begin{aligned} y' &= -\frac{1}{\varepsilon \vartheta_e} S_y^y \tilde{V}_y \left(T, y, \frac{q(T, y, P)}{\tilde{\vartheta}_c(T, P)}, P \right) \\ P' &= S_p^p \tilde{V}_p \left(T, \frac{q(T, y, P)}{\tilde{\vartheta}_c(T, P)}, P \right), \end{aligned} \quad (\text{A.5})$$

where $X = q(T, y, P)$ is the solution to the quasi steady state constraint (A.3). Going back to the original t time scale and to units of molar amount for Y , the reduced dynamics are

$$\begin{aligned} \dot{Y} &= -S_y^y V_y \left(t, \frac{Y}{\vartheta_e}, \frac{q(\varepsilon t, Y/\vartheta_e, P)}{\vartheta_c(t, P)}, P \right) \\ \dot{P} &= \varepsilon S_p^p V_p \left(t, \frac{q(\varepsilon t, Y/\vartheta_e, P)}{\vartheta_c(t, P)}, P \right). \end{aligned} \quad (\text{A.6})$$

Appendix B. Numerical solution of the dynamic optimization problem by collocation

In this section, we discuss the numerical solution of the optimization problem (33) in more detail. As a first step, let us consider the case where

the terminal time t_f is fixed *a priori*. In the collocation scheme, the time interval $[0, t_f]$ is divided into N equally sized intervals, each of length

$$h = \frac{t_f}{N}. \quad (\text{B.1})$$

Within each interval, K collocation points are determined. All collocation points are given by the sequence

$$t_{1,1}, t_{1,2}, \dots, t_{1,K}, t_{2,1}, \dots, t_{N,K}. \quad (\text{B.2})$$

Within each interval, the q -th collocation point is at position r_q (relative to the interval $[-1, 1]$), where r_q is determined by the collocation scheme and order. In the computational experiments for this study, we used Radau collocation points of order 2 and 3, determined by zeros of the Legendre polynomials [37]. The collocation points are thus computed as

$$t_{i,q} = (i - 1)h + (r_q + 1)\frac{h}{2}. \quad (\text{B.3})$$

The flux variable and the derivative of the species variable are discretized by the following interpolation scheme:

$$\begin{aligned} V(t) &= \sum_{q=1}^K v_{i,q} L_q \left(\frac{2t - 2t_{i-1} - h}{h} \right), & t_{i-1} \leq t \leq t_i \\ \dot{Z}(t) &= \sum_{q=1}^K \dot{z}_{i,q} L_q \left(\frac{2t - 2t_{i-1} - h}{h} \right), & t_{i-1} \leq t \leq t_i, \end{aligned} \quad (\text{B.4})$$

where L_q , $q = 1, \dots, K$ are suitable interpolation functions defined on the interval $(-1, 1)$. In this study, we used the Lagrange polynomials

$$L_q(r) = \prod_{1 \leq i \leq K, i \neq q} \frac{r - r_i}{r_q - r_i} \quad (\text{B.5})$$

as interpolation functions [58]. The boundaries of the time intervals are given by $t_i = ih$, $i = 1, \dots, N$ and $t_0 = 0$.

The species variable Z is discretized at the boundaries of the N intervals in time, and its value within an interval is approximated by integrating over

the time derivative:

$$\begin{aligned}
Z(t_{i-1} + \tau) &= z_{i-1} + \int_0^\tau \dot{Z}(t_{i-1} + s) ds \\
&= z_{i-1} + \sum_{q=1}^K \dot{z}_{i,q} \int_0^\tau L_q(2s/h - 1) ds \\
&= z_{i-1} + h/2 \sum_{q=1}^K \dot{z}_{i,q} \int_{-1}^{r(\tau)} L_q(s) ds,
\end{aligned} \tag{B.6}$$

with $r(\tau) = 2\tau/h - 1$.

The continuous optimization problem (33) is now approximated by a finite-dimensional problem, in which the optimization is done over the vector $w \in \mathbb{R}^{NK(n+m)+Nn}$, defined by

$$w = (v_{1,1}, v_{1,2}, \dots, v_{N,K}, \dot{z}_{1,1}, \dot{z}_{1,2}, \dots, \dot{z}_{N,K}, z_1, \dots, z_N), \tag{B.7}$$

with $v_{i,q}$, $\dot{z}_{i,q}$, and z_i , $i = 1, \dots, N$, $q = 1, \dots, K$ corresponding to the interpolation coefficients in (B.4) and (B.6).

The discretized optimization problem is written as

$$\begin{aligned}
&\max_w c^T w + e \\
&\text{s.t. } M_e w = c_e \\
&\quad M_i w \leq c_i,
\end{aligned} \tag{B.8}$$

where the vector c and number e stem from the discretization of the the objective functional, the equality constraint matrix M_e and vector c_e from collocation of the differential equation and the initial condition, and the inequality constraint matrix M_i and vector c_i from the path and terminal constraints in (33). The optimization problem (B.8) is a linear program and can directly be solved by common optimization software.

Appendix C. Analytical results for the minimal metabolic-genetic network

Appendix C.1. Problem statement

In this section we show how to obtain the analytic solution of the optimization problems in Section 3. We seek to solve the following dynamic optimization problem:

$$\max_{V_y} J_i, \quad i = 2, 3. \quad (\text{C.1})$$

in the time interval $[0, t_f]$ subject to the model dynamics

$$\dot{Y} = -V_y, \quad (\text{C.2})$$

$$\dot{P} = \varepsilon V_p \quad (\text{C.3})$$

and the quasi steady state constraint

$$V_y = \alpha \varepsilon V_p, \quad (\text{C.4})$$

together with the positivity constraints

$$Y \geq 0, \quad P \geq 0, \quad V_y \geq 0, \quad V_p \geq 0 \quad (\text{C.5})$$

and the enzyme capacity constraint

$$\frac{V_y}{k_{\text{cat},y}} + \frac{\varepsilon V_p}{k_{\text{cat},p}} \leq P. \quad (\text{C.6})$$

The objective functions are the discounted biomass integral

$$J_2 = \int_0^{t_f} P(\tau) e^{-\varphi \tau} d\tau, \quad (\text{C.7})$$

and the time needed to metabolize all the nutrient

$$J_3 = -t_f. \quad (\text{C.8})$$

In the case of objective J_2 , the final time t_f is pre-specified. In the case of J_3 , the final time t_f is free and subject to an additional terminal constraint $Y(t_f) = 0$.

Using the quasi steady state approximation $V_y = \alpha\varepsilon V_p$, the model dynamics can be written as

$$\dot{Y} = -V_y, \tag{C.9}$$

$$\dot{P} = V_y/\alpha, \tag{C.10}$$

and we rewrite the enzyme capacity constraint as

$$V_y \leq P/K, \tag{C.11}$$

with

$$K = 1/k_{\text{cat},y} + 1/(\alpha k_{\text{cat},p}). \tag{C.12}$$

In the next sections we show how to obtain analytic solutions to both of these problems.

Appendix C.2. Solution for objective J_2

To avoid ambiguities, from now on the star * denotes optimal trajectories. We first note that $V_y \geq 0$ implies that $P \geq 0$, because $P_0 \geq 0$ and $\dot{P} = V_y/\alpha \geq 0$ for all $V_y \geq 0$ and $\alpha > 0$. In addition, by the quasi steady state constraint $V_y = \alpha\varepsilon V_p$, we have that $V_y \geq 0$ implies that $V_p \geq 0$, and therefore the optimization problem can be simplified to

$$\max J_2 \tag{C.13}$$

subject to the dynamics (C.9)–(C.10), and the constraints

$$V_y \leq P/K, V_y \geq 0, \text{ and } Y \geq 0. \tag{C.14}$$

The optimal nutrient dynamics are $\dot{Y}^* = -V_y^*$, which can be solved by integration

$$Y^*(t) = Y_0 - \int_0^t V_y^*(\tau) d\tau. \tag{C.15}$$

This means that for any initial condition $Y_0 > 0$, the optimal nutrient concentration will reach $Y^* = 0$ only if there exists $0 < t_s \leq t_f$ such that

$$\int_0^{t_s} V_y^*(\tau) d\tau = Y_0. \tag{C.16}$$

We will now obtain the optimal uptake rate assuming that equation (C.16) does not have a solution for t_s . We will then use this solution *a posteriori* to analyze the case when (C.16) does have a solution. If equation (C.16) does not have a solution, the nutrient does not deplete and therefore constraint $Y \geq 0$ never becomes active in the optimization interval $[0, t_f]$. We can thus ignore this constraint and reduce the optimization problem to

$$\max J_2 \tag{C.17}$$

subject to (C.9)–(C.10) and $0 \leq V_y \leq P/K$. Since J_2 grows as $P(t)$ grows (pointwise in time) and $\dot{P} = V_y/\alpha \geq 0$, the optimal uptake rate V_y^* must be maximal pointwise in time while respecting $0 \leq V_y \leq P/K$. Intuitively, this means that the optimal rate satisfies $V_y^* = P^*/K$, but we can also alternatively use a proof by contradiction as follows. Assume that in the optimal solution the constraint $V_y \leq P/K$ is not active, so that $V_y^* = P^*/K - \delta(t)$ with $\delta(t) > 0$ on a subinterval of $[0, t_f]$. Substituting this optimal rate in $\dot{P}^* = V_y^*/\alpha$ we get

$$\dot{P}^* = \frac{P^*}{K\alpha} - \frac{\delta}{\alpha}. \tag{C.18}$$

Equation (C.18) is a linear inhomogeneous differential equation with solution

$$P^*(t) = P_0 e^{\frac{t}{K\alpha}} - e^{\frac{t}{K\alpha}} \int_0^t e^{-\frac{\tau}{K\alpha}} \delta(\tau) d\tau. \tag{C.19}$$

Using the definition of $J_2 = \int_0^{t_f} P(t) e^{-\varphi t} dt$, we can compute the value of the corresponding optimal objective as

$$J_2^* = P_0 \int_0^{t_f} e^{\frac{t}{K\alpha}} e^{-\varphi t} dt - \underbrace{\int_0^{t_f} e^{\frac{t}{K\alpha}} e^{-\varphi t} \left(\int_0^t e^{-\frac{\tau}{K\alpha}} \delta(\tau) d\tau \right) dt}_{>0}, \tag{C.20}$$

$$< P_0 \int_0^{t_f} e^{\frac{t}{K\alpha}} e^{-\varphi t} dt, \tag{C.21}$$

which contradicts the optimality of J_2^* , and therefore we conclude that $\delta(t) = 0$ on the interval $[0, t_f]$, apart from a set of zero measure. The optimal uptake rate and biomass concentration are then

$$V_y^*(t) = \frac{P_0}{K} e^{\frac{t}{K\alpha}}, \text{ for } t \in [0, t_f], \tag{C.22}$$

$$P^*(t) = P_0 e^{\frac{t}{K\alpha}}, \text{ for } t \in [0, t_f]. \tag{C.23}$$

From these expressions we can *a posteriori* obtain a condition for equation (C.16) to have a solution for t_s . Substituting the optimal rate V_y^* in (C.16), we get an equation for t_s :

$$\alpha \left(e^{\frac{t_s}{K\alpha}} - 1 \right) = \frac{Y_0}{P_0}, \quad (\text{C.24})$$

and solving for t_s we get

$$t_s = K\alpha \log \left(1 + \frac{Y_0}{\alpha P_0} \right). \quad (\text{C.25})$$

Therefore, if $t_f < t_s$ the nutrient never depletes and the optimal solution are the exponentials in (C.22)–(C.23). Conversely, if $t_f \geq t_s$ the equation in (C.16) has a solution for t_s and we can establish that

$$Y^*(t) = 0, \text{ for } t \in [t_s, t_f], \quad (\text{C.26})$$

$$V_y^*(t) = 0, \text{ for } t \in [t_s, t_f]. \quad (\text{C.27})$$

Note that (C.26) is true because when Y^* reaches zero at $t = t_s$, it can only be made positive by a negative V_y^* , which would violate the positivity constraint $V_y \geq 0$. Similarly, if (C.27) is not true, then V_y^* must become positive for some non empty time interval after $t = t_s$, but this would imply that $Y^* < 0$ in that interval, thereby violating the positivity constraint $Y \geq 0$. We thus conclude that for $t \geq t_s$, the network enters stationary phase and the biomass remains constant:

$$P^*(t) = P^*(t_s), \text{ for } t \in [t_s, t_f]. \quad (\text{C.28})$$

Appendix C.3. Solution for objective J_3

In this case the optimal solution for objective J_3 can be obtained with similar arguments as the one for J_2 . Maximization of J_3 is equivalent to minimization of the time it takes the nutrient to deplete (t_f). Since $\dot{Y} = -V_y < 0$ it follows that t_f decreases as V_y grows pointwise in time. This essentially means that the constraint $V_y \leq P/K$ must be active for $t \in [0, t_f]$ and therefore the optimal solution is $V_y^* = P^*/K$. Following a similar procedure as in the case of J_2 , we have that the optimal uptake rate and biomass concentrations are exponentials

$$V_y^*(t) = \frac{P_0}{K} e^{\frac{t}{K\alpha}}, \text{ for } t \in [0, t_f], \quad (\text{C.29})$$

$$P^*(t) = P_0 e^{\frac{t}{K\alpha}}, \text{ for } t \in [0, t_f]. \quad (\text{C.30})$$

Note that, analogously to equation (C.16), in this case the time needed for nutrient depletion, t_f , satisfies

$$\int_0^{t_f} V_y^*(\tau) d\tau = Y_0. \quad (\text{C.31})$$

Substituting (C.29) in (C.31) and solving for the optimal t_f we get

$$t_f = K\alpha \log \left(1 + \frac{Y_0}{\alpha P_0} \right). \quad (\text{C.32})$$

Appendix D. Equivalence between the minimal metabolic-genetic network and the Monod growth kinetics

Since the nutrient uptake and metabolization is done by enzymes, it appears reasonable to assume a common enzymatic rate law such as the Michaelis-Menten rate here. With this assumption, we obtained the following result.

Proposition 2. *If the uptake reaction V_y in the minimal metabolic-genetic network (11) is given by the Michaelis-Menten rate law with P as an enzyme and Y as a substrate,*

$$V_y = \frac{k_{\text{cat}} P y}{K_m + y}, \quad (\text{D.1})$$

then the long timescale approximation (16) of the minimal metabolic-genetic network is equivalent to the Monod growth kinetics given by

$$\begin{aligned} \dot{Y} &= -\frac{1}{\varrho} \mu(y) P \\ \dot{P} &= \mu(y) P. \end{aligned} \quad (\text{D.2})$$

In the model (D.2), Y denotes the substrate, P is the biomass, $y = Y/\vartheta_e$ the substrate concentration,

$$\mu(y) = \frac{\mu_{\text{max}} y}{K_y + y} \quad (\text{D.3})$$

is the empirical growth rate, and $\varrho \geq 0$ is the yield coefficient [59].

Thus the Monod growth model is theoretically explained by a Michaelis-Menten type nutrient uptake reaction together with optimality of cell growth with respect to an objective of time-minimal growth or a discounted biomass integral.

Proof of Proposition 2. From the quasi-steady state condition

$$V_y = \alpha \varepsilon V_p, \tag{D.4}$$

we have

$$V_p = \frac{1}{\varepsilon \alpha} \frac{k_{\text{cat}} y}{K_m + y}. \tag{D.5}$$

Then the long timescale approximation of the minimal metabolic-genetic network is given by

$$\begin{aligned} \dot{Y} &= -\frac{k_{\text{cat}} P y}{K_m + y} \\ \dot{P} &= \frac{1}{\alpha} \frac{k_{\text{cat}} P y}{K_m + y}. \end{aligned} \tag{D.6}$$

By comparing this equation to (D.2), we see that the dynamics are identical when identifying the parameters with

$$\begin{aligned} \mu_{\text{max}} &= \frac{k_{\text{cat}}}{\alpha} \\ K_y &= K_m \\ \varrho &= \frac{1}{\alpha}. \end{aligned} \tag{D.7}$$

□

Acknowledgements

We thank Alexandra Reimers for helpful comments on a previous version of the paper. Diego Oyarzún acknowledges support from an Imperial College London Junior Research Fellowship.

References

References

- [1] M. R. Watson, A discrete model of bacterial metabolism, *Comput Appl Biosci* 2 (1) (1986) 23–27.
- [2] D. A. Fell, J. R. Small, Fat synthesis in adipose tissue. An examination of stoichiometric constraints, *Biochem J* 238 (3) (1986) 781–786.
- [3] A. Varma, B. Ø. Palsson, Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110, *Appl Environ Microbiol* 60 (10) (1994) 3724–3731.
- [4] N. A. van Riel, M. L. Giuseppin, C. T. Verrips, Dynamic optimal control of homeostasis: an integrative system approach for modeling of the central nitrogen metabolism in *Saccharomyces cerevisiae*, *Metab Eng* 2 (1) (2000) 49–68. doi:10.1006/mben.1999.0137.
URL <http://dx.doi.org/10.1006/mben.1999.0137>
- [5] M. W. Covert, B. Ø. Palsson, Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*, *J Biol Chem* 277 (31) (2002) 28058–28064. doi:10.1074/jbc.M201691200.
URL <http://dx.doi.org/10.1074/jbc.M201691200>
- [6] A. L. Meadows, R. Karnik, H. Lam, S. Forestell, B. Snedecor, Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation, *Metab Eng* 12 (2) (2010) 150–160. doi:10.1016/j.ymben.2009.07.006.
URL <http://dx.doi.org/10.1016/j.ymben.2009.07.006>
- [7] R. Steuer, H. Knoop, R. Machné, Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth, *J Exp Bot* 63 (6) (2012) 2259–2274. doi:10.1093/jxb/ers018.
URL <http://dx.doi.org/10.1093/jxb/ers018>
- [8] M. W. Covert, C. H. Schilling, B. Ø. Palsson, Regulation of gene expression in flux balance models of metabolism, *J Theor Biol* 213 (1) (2001) 73–88. doi:10.1006/jtbi.2001.2405.
URL <http://dx.doi.org/10.1006/jtbi.2001.2405>

- [9] A. Goelzer, V. Fromion, G. Scorletti, Cell design in bacteria as a convex optimization problem, *Automatica* 47 (6) (2011) 1210–1218. doi:10.1016/j.automatica.2011.02.038.
URL <http://www.sciencedirect.com/science/article/pii/S0005109811001397>
- [10] D. Ramkrishna, H.-S. Song, Dynamic models of metabolism: Review of the cybernetic approach, *AIChE J* 58 (2012) 986–997.
- [11] J. I. Kim, J. D. Varner, D. Ramkrishna, A hybrid model of anaerobic *E. coli* GJT001: combination of elementary flux modes and cybernetic variables, *Biotechnol Prog* 24 (5) (2008) 993–1006. doi:10.1002/btpr.73.
URL <http://dx.doi.org/10.1002/btpr.73>
- [12] M. W. Covert, N. Xiao, T. J. Chen, J. R. Karr, Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*, *Bioinformatics* 24 (18) (2008) 2044–2050. doi:10.1093/bioinformatics/btn352.
URL <http://dx.doi.org/10.1093/bioinformatics/btn352>
- [13] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, Jr, N. Assad-Garcia, J. I. Glass, M. W. Covert, A whole-cell computational model predicts phenotype from genotype, *Cell* 150 (2) (2012) 389–401. doi:10.1016/j.cell.2012.05.044.
URL <http://dx.doi.org/10.1016/j.cell.2012.05.044>
- [14] A. Varma, B. Ø. Palsson, Metabolic flux balancing: Basic concepts, scientific and practical use, *Nat Biotech* 12 (10) (1994) 994–998.
URL <http://dx.doi.org/10.1038/nbt1094-994>
- [15] J. L. Reed, B. Ø. Palsson, Thirteen years of building constraint-based in silico models of *Escherichia coli*, *J Bacteriol* 185 (9) (2003) 2692–2699.
- [16] J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis?, *Nat Biotechnol* 28 (3) (2010) 245–248. doi:10.1038/nbt.1614.
URL <http://dx.doi.org/10.1038/nbt.1614>
- [17] R. Schuetz, L. Kuepfer, U. Sauer, Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*, *Mol Syst*

Biol 3 (2007) 119. doi:10.1038/msb4100162.
URL <http://dx.doi.org/10.1038/msb4100162>

- [18] J. Pramanik, J. D. Keasling, Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements, *Biotechnol Bioeng* 56 (4) (1997) 398–421. doi:3.0.CO;2-J.
URL <http://dx.doi.org/3.0.CO;2-J>
- [19] F. Wessely, M. Bartl, R. Guthke, P. Li, S. Schuster, C. Kaleta, Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs, *Mol Syst Biol* 7 (2011) 515.
URL <http://dx.doi.org/10.1038/msb.2011.46>
- [20] J. A. Lerman, D. R. Hyduke, H. Latif, V. A. Portnoy, N. E. Lewis, J. D. Orth, A. C. Schrimpe-Rutledge, R. D. Smith, J. N. Adkins, K. Zengler, B. Ø. Palsson, In silico method for modelling metabolism and gene product expression at genome scale, *Nat Commun* 3 (2012) 929. doi:10.1038/ncomms1928.
- [21] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, U. Sauer, Multidimensional optimality of microbial metabolism, *Science* 336 (6081) (2012) 601–604. doi:10.1126/science.1216882.
URL <http://dx.doi.org/10.1126/science.1216882>
- [22] K. Höffner, S. M. Harwood, P. I. Barton, A reliable simulator for dynamic flux balance analysis, *Biotechnol Bioeng* 110 (3) (2013) 792–802.
- [23] R. Mahadevan, J. S. Edwards, F. J. Doyle, Dynamic flux balance analysis of diauxic growth in *Escherichia coli*, *Biophys J* 83 (3) (2002) 1331–1340.
- [24] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S.

- Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, S. B. M. L. F. , The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *Bioinformatics* 19 (4) (2003) 524–531.
- [25] C. P. Fall, *Computational Cell Biology*, Springer, New York, USA, 2002.
- [26] H. K. Khalil, *Nonlinear Systems*, 3rd Edition, Prentice Hall, Upper Saddle River, New Jersey, 2002.
- [27] D. A. Oyarzún, G.-B. V. Stan, Synthetic gene circuits for metabolic control: design trade-offs and constraints, *J R Soc Interface* 10 (2013) 87. doi:10.1098/rsif.2012.0671.
URL <http://dx.doi.org/10.1098/rsif.2012.0671>
- [28] A. Goelzer, V. Fromion, Towards the modular decomposition of the metabolic network, in: V. V. Kulkarni, G.-B. Stan, K. Raman (Eds.), *A Systems Theoretic Approach to Systems and Synthetic Biology I: Models and System Characterizations*, Springer Science+Business Media, Dordrecht, 2014, pp. 121–152. doi:10.1007/978-94-017-9041-3_{5}.
- [29] N. Meslem, V. Fromion, Lyapunov function for irreversible linear metabolic pathways with allosteric and genetic regulation, in: *Proc. of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, Orlando, FL, USA, 2011, pp. 5182–5187.
- [30] M. F. Madsen, S. Danø, P. G. Sørensen, On the mechanisms of glycolytic oscillations in yeast, *FEBS J* 272 (11) (2005) 2648–60. doi:10.1111/j.1742-4658.2005.04639.x.
URL <http://dx.doi.org/10.1111/j.1742-4658.2005.04639.x>
- [31] Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabási, Z. N. Oltvai, Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity, *Proc Natl Acad Sci U S A* 104 (31) (2007) 12663–12668. doi:10.1073/pnas.0609845104.
URL <http://dx.doi.org/10.1073/pnas.0609845104>
- [32] A. Vazquez, Q. K. Beg, M. A. Demenezes, J. Ernst, Z. Bar-Joseph, A.-L. Barabási, L. G. Boros, Z. N. Oltvai, Impact of the solvent capacity

- constraint on *E. coli* metabolism, BMC Syst Biol 2 (2008) 7. doi:10.1186/1752-0509-2-7.
URL <http://dx.doi.org/10.1186/1752-0509-2-7>
- [33] J. Macki, A. Strauss, Introduction to Optimal Control Theory, Springer-Verlag, New York, 1982.
- [34] C. Vilas, E. Balsa-Canto, M.-S. G. García, J. R. Banga, A. A. Alonso, Dynamic optimization of distributed biological systems using robust and efficient numerical techniques, BMC Syst Biol 6 (2012) 79. doi:10.1186/1752-0509-6-79.
URL <http://dx.doi.org/10.1186/1752-0509-6-79>
- [35] O. von Stryk, Numerical solution of optimal control problems by direct collocation, in: R. Bulirsch, A. Mielke, J. Stoer, K.-H. Well (Eds.), Optimal Control — Calculus of Variations, Optimal Control Theory and Numerical Methods, Birkhäuser, 1993, pp. 129–143. doi:10.1.1.53.9817.
- [36] M. Razzaghi, J. Nazarzadeh, K. Nikraves, A collocation method for optimal control of linear systems with inequality constraints, Math Probl Eng 3 (6) (1998) 503–515.
- [37] L. T. Biegler, An overview of simultaneous strategies for dynamic optimization, Chem Eng Process 46 (2007) 1043–1053.
- [38] S. A. Frank, The trade-off between rate and yield in the design of microbial metabolism, J Evol Biol 23 (2010) 609–613. doi:doi:10.1111/j.1420-9101.2010.01930.x.
- [39] E. Klipp, R. Heinrich, H.-G. Holzhütter, Prediction of temporal gene expression. Metabolic optimization by re-distribution of enzyme activities, Eur J Biochem / FEBS 269 (22) (2002) 5406–13.
URL <http://www.ncbi.nlm.nih.gov/pubmed/12423338>
- [40] D. A. Oyarzún, B. P. Ingalls, R. H. Middleton, D. Kalamatianos, Sequential activation of metabolic pathways: a dynamic optimization approach, Bull Math Biol 71 (8) (2009) 1851–72. doi:10.1007/s11538-009-9427-5.
URL <http://www.ncbi.nlm.nih.gov/pubmed/19412635>

- [41] J. Monod, The growth of bacterial cultures, *Annu Rev Microbiol* 3 (1949) 371–394.
- [42] A. P. Burgard, S. Vaidyaraman, C. D. Maranas, Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments, *Biotechnol Prog* 17 (5) (2001) 791–797. doi:10.1021/bp0100880.
URL <http://dx.doi.org/10.1021/bp0100880>
- [43] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, R. Milo, The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters, *Biochemistry* 50 (21) (2011) 4402–4410. doi:10.1021/bi2002289.
URL <http://dx.doi.org/10.1021/bi2002289>
- [44] I. Schomburg, A. Chang, D. Schomburg, BRENDA, enzyme data and metabolic information, *Nucleic Acids Res* 30 (1) (2002) 47–49.
- [45] R. Milo, P. Jorgensen, U. Moran, G. Weber, M. Springer, Bionumbers—the database of key numbers in molecular and cell biology, *Nucleic Acids Res* 38 (Database issue) (2010) D750–D753. doi:10.1093/nar/gkp889.
URL <http://dx.doi.org/10.1093/nar/gkp889>
- [46] R. Young, H. Bremer, Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate, *Biochem J* 160 (2) (1976) 185–194.
- [47] T. E. Allen, B. Ø. Palsson, Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes, *J Theor Biol* 220 (1) (2003) 1–18.
- [48] L. Brocchieri, S. Karlin, Protein length in eukaryotic and prokaryotic proteomes, *Nucleic Acids Res* 33 (10) (2005) 3390–3400. doi:10.1093/nar/gki615.
URL <http://dx.doi.org/10.1093/nar/gki615>
- [49] K. Bettenbrock, S. Fischer, A. Kremling, K. Jahreis, T. Sauter, E.-D. Gilles, A Quantitative Approach to Catabolite Repression in *Escherichia coli*, *J Biol Chem* 281 (5) (2006) 2578–2584. doi:10.1074/jbc.M508090200.
URL <http://www.jbc.org/content/281/5/2578.abstract>
<http://www.jbc.org/content/281/5/2578.full>

- [50] G. G. Zampar, A. Kümmel, J. Ewald, S. Jol, B. Niebel, P. Picotti, R. Aebbersold, U. Sauer, N. Zamboni, M. Heinemann, Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast, *Mol Syst Biol* 9 (2013) 651. doi:10.1038/msb.2013.11.
URL <http://dx.doi.org/10.1038/msb.2013.11>
- [51] I. B. Bischofs, J. A. Hug, A. W. Liu, D. M. Wolf, A. P. Arkin, Complexity in bacterial cell-cell communication: quorum signal integration and subpopulation signaling in the *Bacillus subtilis* phosphorelay, *Proc. Natl. Acad. Sci. USA* 106 (16) (2009) 6459–6464. doi:10.1073/pnas.0810878106.
URL <http://dx.doi.org/10.1073/pnas.0810878106>
- [52] S. Klumpp, Z. Zhang, T. Hwa, Growth rate-dependent global effects on gene expression in bacteria, *Cell* 139 (7) (2009) 1366–1375. doi:10.1016/j.cell.2009.12.001.
URL <http://dx.doi.org/10.1016/j.cell.2009.12.001>
- [53] R. Mahadevan, C. H. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models, *Metab Eng* 5 (4) (2003) 264–276.
- [54] D.-E. Chang, D. J. Smalley, D. L. Tucker, M. P. Leatham, W. E. Norris, S. J. Stevenson, A. B. Anderson, J. E. Grissom, D. C. Laux, P. S. Cohen, T. Conway, Carbon nutrition of *Escherichia coli* in the mouse intestine, *Proc Natl Acad Sci U S A* 101 (19) (2004) 7427–7432. doi:10.1073/pnas.0307888101.
URL <http://dx.doi.org/10.1073/pnas.0307888101>
- [55] C. S. Henry, L. J. Broadbelt, V. Hatzimanikatis, Thermodynamics-based metabolic flux analysis, *Biophys J* 92 (5) (2007) 1792–1805. doi:10.1529/biophysj.106.093138.
URL <http://dx.doi.org/10.1529/biophysj.106.093138>
- [56] M. N. Price, A. M. Deutschbauer, J. M. Skerker, K. M. Wetmore, T. Ruths, J. S. Mar, J. V. Kuehl, W. Shao, A. P. Arkin, Indirect and suboptimal control of gene expression is widespread in bacteria, *Mol Syst Biol* 9 (2013) 660. doi:10.1038/msb.2013.16.
URL <http://dx.doi.org/10.1038/msb.2013.16>

- [57] A. Goelzer, V. Fromion, Bacterial growth rate reflects a bottleneck in resource allocation, *Biochim Biophys Acta* 1810 (10) (2011) 978–988.
doi:10.1016/j.bbagen.2011.05.014.
URL <http://dx.doi.org/10.1016/j.bbagen.2011.05.014>
- [58] D. Garg, M. A. Patterson, W. W. Hager, A. V. Rao, D. Benson, G. T. Huntington, An overview of three pseudospectral methods for the numerical solution of optimal control problems, preprint AAS 09-332 (2009).
- [59] I. J. Dunn, E. Heinzle, J. Ingham, J. E. Pfenosil, *Biological Reaction Engineering: Dynamic Modelling Fundamentals with Simulation Examples*, Wiley-VCH, Weinheim, Germany, 2003.