



## Constraint-based metabolic control analysis for rational strain engineering

Sophia Tsouka<sup>1</sup>, Meric Ataman<sup>1,2</sup>, Tuure Hameri, Ljubisa Miskovic, Vassily Hatzimanikatis<sup>\*</sup>

Laboratory of Computational Systems Biology (LCSB), EPFL, CH-1015, Lausanne, Switzerland

### A B S T R A C T

The advancements in genome editing techniques over the past years have rekindled interest in rational metabolic engineering strategies. While Metabolic Control Analysis (MCA) is a well-established method for quantifying the effects of metabolic engineering interventions on flows in metabolic networks and metabolite concentrations, it does not consider the physiological limitations of the cellular environment and metabolic engineering design constraints. We report here a constraint-based framework, Network Response Analysis (NRA), for rational genetic strain design. NRA is cast as a Mixed-Integer Linear Programming problem that integrates MCA, Thermodynamically-based Flux Analysis (TFA), biologically relevant constraints, as well as genome editing restrictions into a comprehensive platform for identifying metabolic engineering targets. We show that the NRA formulation and its core constraints are equivalent to the ones of Flux Balance Analysis (FBA) and TFA, which allows it to be used for a wide range of optimization criteria and with various physiological constraints. We also show how the parametrization and introduction of biological constraints enhance the NRA formulation compared to the classical MCA approach, and we demonstrate its features and its ability to generate multiple alternative optimal strategies given several user-defined boundaries and objectives. In summary, NRA is a sophisticated alternative to classical MCA for rational metabolic engineering that accommodates the incorporation of physiological data at metabolic flux, metabolite concentration, and enzyme expression levels.

### 1. Introduction

Recent improvements in genome editing techniques have paved the way for more sophisticated and performant metabolic engineering designs for achieving desired physiological states of host organisms. Two approaches for reaching the targeted states exist: (i) integrating heterologous pathways to disruptively overcome native control patterns, and (ii) modifying the endogenous regulatory architecture by removal of the existing control loops (Bailey, 1991). The former method can be rather arduous because it requires testing if the integration of DNA fragments into the original genome sequence perturbs cellular regulation in the desired fashion. The latter technique demands knowledge about cellular control so that the DNA sequence can be modified effectively and without unwanted side effects.

Mathematical models are nowadays becoming an indispensable part of strain design. Available gene-protein-reaction associations of various organisms provide invaluable information about cellular metabolism and enable the elaboration of these models. The models can be studied computationally to interrogate and analyze cellular behavior and derive metabolic engineering strategies for improved cellular performance (Gombert and Nielsen, 2000). Strain design requires the identification

and engineering of pathways toward the production of desired compounds (Hadadi and Hatzimanikatis, 2015), and mathematical models can provide an invaluable insight in the process of selection of deletions, insertions, and up- and down-regulation of genes encoding for metabolic enzymes. Reviews of the most prominent computational tools and workflows for the strain design are provided elsewhere (Costa et al., 2016; Long et al., 2015; Wang et al., 2017).

Metabolic control analysis (MCA) is a mathematical formalism that uses models to determine the distribution of control over metabolic states such as fluxes and concentrations (Hatzimanikatis and Bailey, 1996; Heinrich and Schuster, 1996; Kacser et al., 1995). In MCA, control on these metabolic states is quantified with Control Coefficients (CCs). CCs provide information on how a given metabolic flux or metabolite concentration would respond to increase or decrease of an enzymatic activity or to environmental changes. This information is used in traditional rational metabolic design to identify the rate-limiting steps of the network and select potential targets for engineering. Strain engineering typically requires a holistic approach where one simultaneously analyzes the effects of genetic manipulations on specific productivity of desired molecules, maximum achievable yield, energetic and redox requirements, etc. Simultaneous analysis of these effects is a cumbersome

<sup>\*</sup> Corresponding author. Laboratory of Computational Systems Biotechnology (LCSB), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015, Lausanne, Switzerland.

E-mail address: [vassily.hatzimanikatis@epfl.ch](mailto:vassily.hatzimanikatis@epfl.ch) (V. Hatzimanikatis).

<sup>1</sup> equally contributing authors.

<sup>2</sup> current address: Computational and Systems Biology, Biozentrum University of Basel, Klingelbergstrasse 50-70, 4056, Basel, Switzerland.

<https://doi.org/10.1016/j.ymben.2021.03.003>

Received 25 November 2020; Received in revised form 10 February 2021; Accepted 2 March 2021

Available online 22 April 2021

1096-7176/© 2021 The Authors. Published by Elsevier Inc. on behalf of International Metabolic Engineering Society. This is an open access article under the CC

BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

task using classical MCA tools, especially if the design requires multiple genetic alterations. Moreover, MCA does not allow explicit inclusion of physiological or design constraints, which can lead to unrealistic predictions.

We present here Network Response Analysis (NRA), a constraint-based workflow that aims to tackle these challenges. NRA utilizes populations of CCs to consistently derive metabolic engineering strategies and trace the effects of multiple parameter perturbations. The advantage of this method is that physiologically relevant bounds and constraints can be imposed to the system, as opposed to the classical MCA. NRA is inspired by the work by Hatzimanikatis et al. (1996a,1995b) who proposed a Mixed Integer Linear Programming (MILP) formulation for querying cellular responses upon enzymatic perturbations that uses MCA-based flux and concentration CCs. Therein, the authors applied their formulation on simple linear and branched pathways to propose metabolic engineering strategies. Here, we extend this formulation to allow for studying larger scale metabolic systems with guaranteed thermodynamic feasibility. The presented optimization formulation is

equivalent to the ones of FBA and TFA (Henry et al., 2007; Salvy et al., 2019), which allows its application to a large number of design objectives.

To illustrate how NRA can be used to efficiently analyze, enumerate, and propose alternative metabolic engineering strategies, we used a large-scale thermodynamically-curated, stoichiometric model of *E. coli* metabolism (Hameri et al., 2021), which describes the central carbon pathways in aerobic growth conditions. Using the stoichiometric model as a scaffold, we employed the ORACLE framework (Andreozzi et al., 2016a; Chakrabarti et al., 2013; Hameri et al., 2019b; Miskovic et al., 2017; Miskovic and Hatzimanikatis, 2010; Soh et al., 2012; Tokic et al., 2020) to generate populations of flux and concentration CCs. A set of ORACLE computational procedures allow us to integrate omics data, thermodynamics, and other physicochemical constraints into large-scale kinetic models of metabolism and use these models to compute CCs consistent with the experimental observations. We used the generated CCs to formulate a MILP optimization problem and devise the design strategies in two case studies (i) improvement of glucose uptake rate,

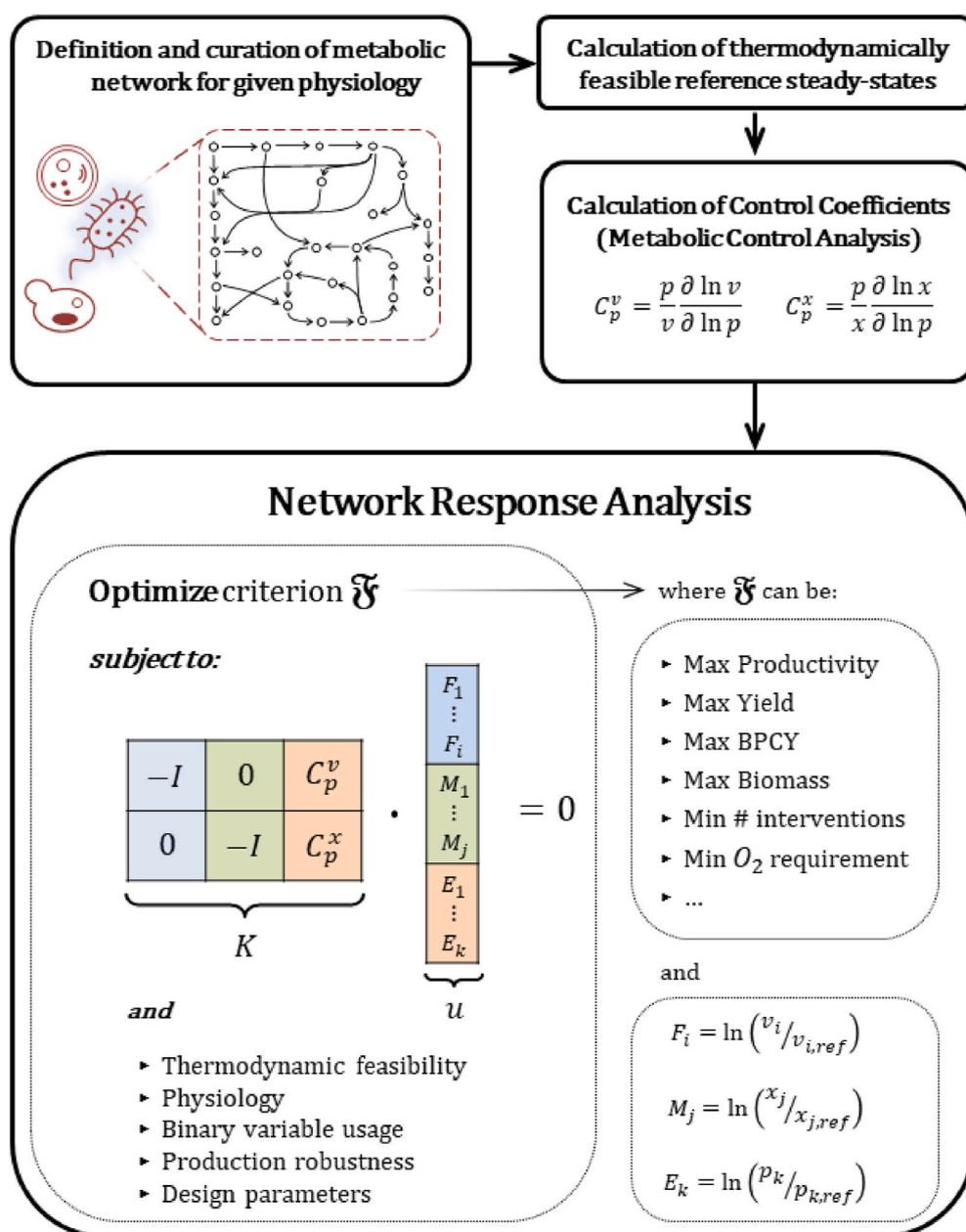


Fig. 1. The NRA workflow is organized in four main steps. In the first three steps, we formulate the stoichiometry, integrate available experimental data and compute the steady-state thermodynamically feasible fluxes and concentrations, and compute the flux and concentration control coefficients for the studied physiological condition. In the fourth step, metabolic engineering strategies are devised by solving a MILP. Criterion  $\mathfrak{F}$  and additional constraints can be chosen from a set of metabolic engineering criteria such as the ones provided in Table 1. Variables  $F_i$ ,  $M_j$  and  $E_k$  are the logarithmic deviations in flux, metabolite concentration and parameter with respect to their respective reference steady states (Eq. (23)), and their bounds define the solution space of the optimization problem (Eqs. 8-10). The definition of the other optimization variables and parameters is given in Table 2.

and (ii) maximization of specific productivity rate of pyruvate while preserving a pre-specified yield of pyruvate from glucose. These studies clearly show the potential, flexibility, and ease of use of NRA when realistic, multi-objective requirements for the strain design should be met.

## 2. Methods

### 2.1. NRA method for designing metabolic engineering strategies

#### 2.1.1. Overview of NRA

The first step of the NRA method is the selection and curation of a metabolic network that captures the physiology of a studied organism (Fig. 1). Then, we employ the Thermodynamic-based Flux Analysis (TFA) to integrate available metabolomics and fluxomics data and compute the thermodynamically consistent metabolite concentrations and metabolic fluxes in the network (Ataman and Hatzimanikatis, 2015; Henry et al., 2007; Salvy et al., 2019; Soh and Hatzimanikatis, 2010). We next calculate the relevant flux and concentration CCs (FCCs and CCCs) that describe the network's responses to parameter perturbations such as modifications of enzymatic activities with the ORACLE framework, which makes use of Monte Carlo sampling (Miskovic and Hatzimanikatis, 2011; Wang et al., 2004). Finally, we use the computed sets of CCs along with the user-defined requirements and additional physiological constraints to construct a constraint-based MILP optimization problem (Fig. 1).

The user-defined inputs depend on the studied problem and design limitations, and these typically include the number of desired gene manipulations, minimal allowable specific productivity, minimum allowable yield, etc. From experimental measurements or assumptions

on physiology, we can infer physiological constraints such as allowable (or desired) bounds on fluxes and concentrations in the metabolic network.

The outcome of the NRA optimization are sets of alternative combinations of genes that should be engineered to improve the cellular performance given the imposed user-defined inputs and physiological constraints. A principal advantage of the MILP formulation is that it allows the user to introduce constraints on metabolic states and additional relevant design constraints to the system, thus simultaneously offering flexibility and tight control over the rational strain design.

#### 2.1.2. NRA formulation

The NRA core equations can be expressed in a matrix-vector form (Table 1, Eq. 7) similar to the ones of Flux Balance Analysis (FBA) (Orth et al., 2010) and Thermodynamics-based Flux Analysis (TFA) (Henry et al., 2007; Salvy et al., 2019). The similarities and differences between FBA and NRA formulations are outlined in Table 1. The used MILP formulation allows NRA to accommodate a wide gamut of design objectives, such as the maximization of productivity or product yield (Eqs. 1-2), biomass-product coupled yield (BPCY) (Eq. 3), the maximization of biomass formation (Eq. 4), the minimization of required genome-editing interventions (Eq. 5), and the minimization of oxygen requirements (Eq. 6) (Klamt et al., 2018; Patil et al., 2005; Schneider and Klamt, 2019; Varma et al., 1993). Since we have defined the NRA variables in logarithmic form, we can express the otherwise nonlinear objectives like yield or BPCY in a linear form, rendering the solution of the mathematical problem easier to attain than with formulations such as FBA, where

**Table 1**

The NRA mathematical formulation together with a non-exhaustive selection of optimization objectives and its comparison to FBA. The definition of indices, parameters, and variables is provided in Table 2.

Optimize Criterion $\mathfrak{S}$ :	FBA	NRA	
<b>Max productivity</b>	$\max v_{product}$	$\max \tilde{F}_{product}$	(1)
<b>Max yield</b>	$\max \frac{v_{product}}{v_{substrate}}$	$\max \tilde{F}_{product} - \tilde{F}_{substrate}$	(2)
<b>Max BPCY</b>	$\max v_{biomass} * \frac{v_{product}}{v_{substrate}}$	$\max \tilde{F}_{biomass} + \tilde{F}_{product} - \tilde{F}_{substrate}$	(3)
<b>Max biomass</b>	$\max v_{biomass}$	$\max \tilde{F}_{biomass}$	(4)
<b>Min # interventions</b>	–	$\min \sum_{k \in \mathcal{R}} (1 - z_k)$	(5)
<b>Min <math>O_2</math> requirement</b>	$\min v_{O_2, uptake}$	$\min F_{O_2, uptake}$	(6)
<b>subject to constraints:</b>			
<b>Balance (Mass /Response)</b>	$N \cdot v = 0$	$K \cdot u = 0 \Leftrightarrow \begin{cases} F_i - \sum_{k \in \mathcal{R}} (C_{E_k}^M * E_k) = 0 \\ M_j - \sum_{k \in \mathcal{R}} (C_{E_k}^X * E_k) = 0 \end{cases}$	(7)
<b>Variable bounds</b>	$lb \leq v_i \leq ub$	$lb_F \leq F_i \leq ub_F$	(8)
		$lb_M \leq M_j \leq ub_M$	(9)
		$lb_E \leq E_k \leq ub_E \Leftrightarrow \begin{cases} 0 \leq E_k^U \leq ub_E \\ 0 \leq E_k^D \leq -lb_E \end{cases}$	(10)
<b>Binary variable usage</b>	–	$E_k^U + E_k^D + \xi * z_k \leq \xi$	(11)
		$E_k^{UU} + E_k^{DU} \leq 1$	(12)
		$E_k^U - \xi * E_k^{UU} < 0$	(13)
		$E_k^D - \xi * E_k^{DU} < 0$	(14)
<b>Thermodynamic feasibility</b>	–	$\Delta_r G_i^c < 0 \Leftrightarrow \Delta_r G_i^o + RT \sum_{m \in \mathcal{R}} \tilde{M}_m^i < 0$	(15)
<b>and</b>			
<b>Production robustness (min productivity)</b>	$v_{product} \geq \alpha_0$	$\tilde{F}_{product} \geq \beta_0$	(16)
<b>Production robustness (min yield)</b>	$\frac{v_{product}}{v_{substrate}} \geq \alpha_1$	$\tilde{F}_{product} - \tilde{F}_{substrate} \geq \beta_1$	(17)
<b>Production robustness (min biomass)</b>	$v_{biomass} \geq \alpha_2$	$\tilde{F}_{biomass} \geq \beta_2$	(18)
<b>Design parameters (max # interventions)</b>	–	$\sum_{k \in \mathcal{R}} (1 - z_k) \leq \beta_3$	(19)
<b>Design parameters (max <math>O_2</math> requirement)</b>	$v_{O_2, uptake} \leq \alpha_3$	$F_{O_2, uptake} \leq \beta_4$	(20)

$$K = \begin{bmatrix} -I [i \times i] & 0 [i \times j] & C_p^v [i \times k] & -C_p^v [i \times k] \\ 0 [j \times i] & -I [j \times j] & C_p^x [j \times k] & -C_p^x [j \times k] \end{bmatrix} \quad (21)$$

$$u = \begin{bmatrix} F_i [i \times 1] \\ M_j [j \times 1] \\ E_k^U [k \times 1] \\ E_k^D [k \times 1] \end{bmatrix} \quad (22)$$

$$\begin{aligned} F_i &= \ln\left(\frac{v_i}{v_{i,ref}}\right) \\ M_j &= \ln\left(\frac{x_j}{x_{j,ref}}\right) \\ E_k &= \ln\left(\frac{p_k}{p_{k,ref}}\right) \end{aligned} \quad (23)$$

$$\tilde{F}_i = F_i + \ln(v_{i,ref}) \quad (24)$$

$$\tilde{M}_j = M_j + \ln(x_{j,ref}) \quad (25)$$

Importantly, the NRA formulation allows us to prevent thermodynamically infeasible designs because it naturally includes thermodynamic constraints regarding the Gibbs free energy change ( $\Delta_r G_i'$ ) of each reaction (eq. 15). Furthermore, the proposed formulation allows imposing additional design criteria such as production robustness and operational parameters (Eqs. 16–19). The NRA optimization problems can be solved with the TFA toolbox (Salvy et al., 2019).

## 2.2. Constraints on enzyme activities

Since the activity of an enzyme in the metabolic network could either be increased or decreased, but not both at the same time, we made use of integer variables in the formulation. Therefore, we split the catalytic activity deviations of our system,  $E_k$ , into the continuous variables  $E_k^U$  and  $E_k^D$ , which denote the upregulation and downregulation of the gene encoding for enzyme  $k$ , respectively (Eqs. 11–14). As these should not have nonzero values simultaneously, we define the integer binary variables  $E_k^{UU}$  and  $E_k^{DU}$ .  $E_k^{UU}$  equals one if the gene catalyzing the enzyme  $k$  is upregulated and equals zero otherwise. In contrast,  $E_k^{DU}$  equals zero in the case of upregulation, and it is one for downregulation. As expressed in Eq. 12, only one of these variables can be active at a time, since deregulation cannot occur in both directions simultaneously, or they can both be inactive for the case of no change in the respective enzyme's catalytic activity. To complete the formulation, these variables are further coupled to the above defined split enzymatic deviation variables through Eqs. 13 and 14. The integer binary variable  $z_k$  is equal to zero if the activity of enzyme  $k$  is modified in the solution, and it equals to one otherwise (Eq. 11).  $\xi$  is a constant selected to be larger than the absolute value of the largest enzymatic deviation constraints,  $lb_E$  and  $ub_E$ , defined in Eq. 10.

## 2.3. Thermodynamic constraints

To integrate thermodynamic constraints, we assumed that reactions operate in the directionality determined by the computed reference steady state. Thus, the concentrations of each metabolite in the respective cellular compartment need to be such as the  $\Delta_r G_i'$  of each reaction remains negative. These constraints are written as a function of the standard Gibbs free energy change of the reaction ( $\Delta_r G_i^{\circ}$ ) and the logarithmic concentrations of the participating metabolites, as introduced by (Henry et al., 2007). The  $\Delta_r G_i^{\circ}$  of each reaction is computed using the Group Contribution Method (Mavrovouniotis, 1990, 1991).

These values are further adjusted to take into account the thermodynamic properties of the relevant cellular compartments; the pH gradient and electrochemical potential for transport reactions, and ionic strength of dissociated metabolites (Henry et al., 2006).

## 2.4. Metabolic control analysis notions

In MCA, the CCCs,  $C_p^x$ , and the FCCs,  $C_p^v$ , are defined as the fractional change of metabolite concentrations  $x$  and metabolic fluxes  $v$ , respectively, in response to a fractional change of system parameters  $p$  (Hatzimanikatis and Bailey, 1996; Kacser et al., 1995). These CCs serve as measurable outputs that provide information about the levels of control that system parameters have on the studied biological system and physiology. From the log(linear) formalism (Hatzimanikatis et al., 1996a; Reder, 1988),  $C_p^x$  and  $C_p^v$  can be derived through the following expressions:

$$C_p^x = - (NVE)^{-1} NV \Pi$$

$$C_p^v = EC_p^x + \Pi$$

where  $N$  is the stoichiometric matrix,  $V$  is the diagonal matrix whose elements are the steady-state fluxes,  $E$  is the elasticity matrix with respect to metabolites and  $\Pi$  is the matrix of elasticities with respect to parameters.

Hence, flux and concentration control coefficients are computed for each reaction flux  $i$  and metabolite concentration  $j$  with respect to the system parameter  $k$  as:

$$C_{pk}^{v_i} = \frac{d \ln v_i}{d \ln p_k} = \frac{p_k}{v_i} \frac{dv_i}{dp_k}$$

$$C_{pk}^{x_j} = \frac{d \ln x_j}{d \ln p_k} = \frac{p_k}{x_j} \frac{dx_j}{dp_k}$$

## 2.5. Model description and calculation of control coefficients

The stoichiometric model that was used in this study (Hameri et al., 2021) was systematically reduced from the *E. coli* iJO1366 genome-scale model (Orth et al., 2011) around the originally defined reaction sub-systems of glycolysis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, glyoxylate cycle, pyruvate metabolism and the electron transport chain (ETC), and describes the aerobically grown physiology of *E. coli* (Supplementary Table S3). The reduction was performed through the redGEM and the lumpGEM algorithms (Ataman et al., 2017; Ataman and Hatzimanikatis, 2017), thus ensuring preservation of as much information as possible as well as that thermodynamic feasibility constraints are respected. This model constitutes of 337 metabolites participating in 647 reactions, which are in turn associated with 271 enzymes that serve as parameters in the NRA formulation. The model was curated with thermodynamic feasibility constraints using TFA (Henry et al., 2007; Salvy et al., 2019) and relevant fluxomics data (McCloskey et al., 2014). The representative steady state profiles of the metabolite concentrations and metabolic fluxes were chosen with Principal Component Analysis (PCA) as detailed in (Hameri et al., 2019b). Then, the populations of control coefficients were built using the ORACLE workflow (Andreozzi et al., 2016a; Miskovic et al., 2017; Miskovic and Hatzimanikatis, 2010; Tokic et al., 2020).

The CCs of the analyzed quantities (glycose uptake, pyruvate production, yield of pyruvate from glucose) with respect to the lumped reactions, exchange reactions, individual biomass building block contributions, and moieties were not considered in any study (Supplementary Table S3).

**Table 2**  
Indices, variables, and parameters used in the NRA formulation.

Index letter	Type	Refers to	Set or unit
$i$	Index	Reaction	$i \in \mathcal{I}$
$j$	Index	Metabolite	$j \in \mathcal{J}$
$k$	Index	Enzyme	$k \in \mathcal{K}$
$m$	Index	Metabolite participating in reaction $i$	$m \in \mathcal{M}$
$v_i$	Variable	Flux of reaction $i$	$mmol \cdot gDW^{-1} \cdot h^{-1}$
$x_j$	Variable	Concentration of metabolite $j$	$M$
$p_k$	Variable	Catalytic activity of enzyme $k$	$mmol \cdot h^{-1}$
$\tilde{F}_i$	Variable	Scaled flux deviation of reaction $i$	$\tilde{F}_i \in \mathbb{R}$
$F_i$	Variable	Flux deviation of reaction $i$	$F_i \in \mathbb{R} \cap [lb_F, ub_F]$
$\tilde{M}_j$	Variable	Scaled concentration deviation of metabolite $j$	$\tilde{M}_j \in \mathbb{R}$
$M_j$	Variable	Concentration deviation of metabolite $j$	$M_j \in \mathbb{R} \cap [lb_M, ub_M]$
$E_k$	Variable	Catalytic activity deviation of enzyme $k$	$E_k \in \mathbb{R} \cap [lb_E, ub_E]$
$E_k^U$	Variable	Upregulation of catalytic activity of enzyme $k$	$E_k^U \in \mathbb{R} \cap [0, ub_E]$
$E_k^D$	Variable	Downregulation of catalytic activity of enzyme $k$	$E_k^D \in \mathbb{R} \cap [0, -lb_E]$
$\Delta_r G_i'$	Variable	Gibbs free energy change of reaction $i$	$\Delta_r G_i' \in \mathbb{R}$
$\Delta_r G_i'^o$	Variable	Standard Gibbs free energy change of reaction $i$	$\Delta_r G_i'^o \in \mathbb{R} \cap [\Delta_r G_{i,est}' \pm err]$
$u$	Variable	Vector of NRA variables	$u \in \mathbb{R}^{(\mathcal{J} + \mathcal{J} + 2\mathcal{K})}$
$E_k^{UU}$	Variable	Binary upregulation of catalytic activity of enzyme $k$	$E_k^{UU} \in \{0, 1\}$
$E_k^{DU}$	Variable	Binary downregulation of catalytic activity of enzyme $k$	$E_k^{DU} \in \{0, 1\}$
$z_k$	Variable	Deregulation of enzyme $k$	$z_k \in \{0, 1\}$
$C_p^v$	Parameter	Flux control coefficient	$C_p^v \in \mathbb{R}$
$C_p^x$	Parameter	Concentration control coefficient	$C_p^x \in \mathbb{R}$
$v_{i,ref}$	Parameter	Reference flux of reaction $i$	$mmol \cdot gDW^{-1} \cdot h^{-1}$
$x_{j,ref}$	Parameter	Reference concentration of metabolite $j$	$M$
$p_{k,ref}$	Parameter	Reference catalytic activity of enzyme $k$	$mmol \cdot h^{-1}$
$N$	Parameter	Stoichiometric matrix	$N \in \mathbb{R}^{\mathcal{J} \times \mathcal{J}}$
$K$	Parameter	NRA matrix	$K \in \mathbb{R}^{(\mathcal{J} + \mathcal{J}) \times (\mathcal{J} + \mathcal{J} + 2\mathcal{K})}$
$lb_F$	Parameter	Flux deviation lower bound	$lb_F \in \mathbb{R}$
$ub_F$	Parameter	Flux deviation upper bound	$ub_F \in \mathbb{R}$
$lb_M$	Parameter	Concentration deviation lower bound	$lb_M \in \mathbb{R}$
$ub_M$	Parameter	Concentration deviation upper bound	$ub_M \in \mathbb{R}$
$lb_E$	Parameter	Catalytic activity deviation lower bound	$lb_E \in \mathbb{R}$
$ub_E$	Parameter	Catalytic activity deviation upper bound	$ub_E \in \mathbb{R}$
$\Delta_r G_{i,est}'$	Parameter	Estimated standard Gibbs free energy change of reaction $i$	$\Delta_r G_{i,est}' \in \mathbb{R}$
$err$	Parameter	Associated error on the estimated standard Gibbs free energy change of reaction $i$	$err \in \mathbb{R}$
$R$	Parameter	Gas constant	$R = 1.9872 \times 10^{-3} kcal \cdot K^{-1} \cdot mol^{-1}$
$T$	Parameter	Temperature	$K$

**Table 2 (continued)**

Index letter	Type	Refers to	Set or unit
$\xi$	Parameter	User-defined large constant	$\xi \in \mathbb{N} : \xi > \max\{ ub_E ,  lb_E \}$
$\alpha_0$	Parameter	Minimum user-defined productivity	$mmol \cdot gDW^{-1} \cdot h^{-1}$
$\alpha_1$	Parameter	Minimum user-defined yield	$\alpha_1 \in \mathbb{R}$
$\alpha_2$	Parameter	Minimum user-defined growth rate	$h^{-1}$
$\alpha_3$	Parameter	Maximum user-defined $O_2$ requirement	$mmol \cdot gDW^{-1} \cdot h^{-1}$
$\beta_0$	Parameter	Minimum user-defined productivity	$\beta_0 \in \mathbb{R} \cap [lb_F, ub_F]$
$\beta_1$	Parameter	Minimum user-defined yield	$\beta_1 \in \mathbb{R}$
$\beta_2$	Parameter	Minimum user-defined growth rate	$\beta_2 \in \mathbb{R} \cap [lb_F, ub_F]$
$\beta_3$	Parameter	Maximum user-defined number of interventions	$\beta_3 \in \mathbb{N}$
$\beta_4$	Parameter	Maximum user-defined $O_2$ requirement	$\beta_4 \in \mathbb{R} \cap [lb_F, ub_F]$

## 2.6. Addressing variability in control coefficients

A common issue in MCA and in kinetic modeling is the uncertainty stemming from the scarcity of knowledge concerning the kinetic properties of enzymes (Miskovic and Hatzimanikatis, 2011; Miskovic et al., 2015, 2019b; Wang et al., 2004). The usual approach in addressing this issue involves the generation of a population of the CCs, and statistical analysis thereof. To form the NRA models, we need to select sets of CCs that will be representative of the generated population.

To select a representative set of CCs for our analysis, we took the population of 50'000 sets of FCCs and CCCs computed with ORACLE for the aerobically grown *E. coli* in (Hameri et al., 2021). We first identified the vector of FCCs that was closest to the mean of the FCC distribution with respect to glucose uptake and selected it as the representative set. Four glucose uptake reactions in the model of *E. coli* exist with GLCptspp being responsible for 91.21% of the total flux through these reactions. We enforced this ratio in all performed NRA studies.

Since the model is constrained to grow on minimal media with glucose as its sole carbon source, the choice of the representative set will have a strong impact on the design criteria we wish to explore. To investigate the variability in results that this choice can induce, we additionally selected several “extreme” CC-sets through the use of PCA. We used nine principal components to describe the space of CCs with respect to glucose uptake, which lead to a coverage of 96.63% of the space variance. We selected the minimum and maximum corresponding CC-sets for each component (2 x 9), leading to a total of 19 sets. We then constructed 19 NRA models with these CC-sets and used them in the performed studies.

## 2.7. Confidence intervals and Bonferroni correction

For the computation of confidence intervals, we have used the Bonferroni correction to account for the multivariate nature of our study. In univariate studies, to account for the variability in samples, confidence intervals that contain the population mean with the probability  $1 - \alpha$  (typically,  $\alpha = 5\%$ ) are added around each sample mean (Hameri et al., 2019a). However, univariate confidence intervals cannot be used when multivariate problems are studied, instead the Bonferroni's correction of confidence intervals is frequently applied. In Bonferroni's correction, for a problem with  $p$  variables, to ensure the level  $1 - \alpha$  for all variables simultaneously, we need to choose level  $1 - \alpha/p$  for each of individual variables. For instance, if we want to form confidence intervals for 10 variables with an overall 95% confidence level, then we need to use individual 99.5% confidence intervals.

## 2.8. Software and optimization parameters

The computations were made on a Mac Pro workstation running Mac OS X version 10.11.6, equipped with a 2.7 GHz 12-Core Intel Xeon E5 processor and 32 GB DDR3 memory, using MATLAB version R2016a and the IBM CPLEX solver version 12.5.1. Time limits for the solver were set as following: in Fig. 2a, for 2-fold (blue inverted triangles) to 10 min, for 5-fold (orange diamonds) to 30 min, and for 10-fold (yellow squares) to 3 h; in Fig. 2b, for all cases to 10 min; in Fig. 3, for all cases to 30 min; in Table 2, for all cases to 30 min; in the pyruvate case study (Figs. 4–6), for all cases 3 h.

## 2.9. Data and software availability

The code used to produce results of the manuscript is available at: <https://github.com/EPFL-LCSB/NRA> and <https://gitlab.com/realLCSB/nra>. Models and datasets produced within this work are deposited at: <http://doi.org/10.5281/zenodo.4501189>.

## 3. Results and discussion

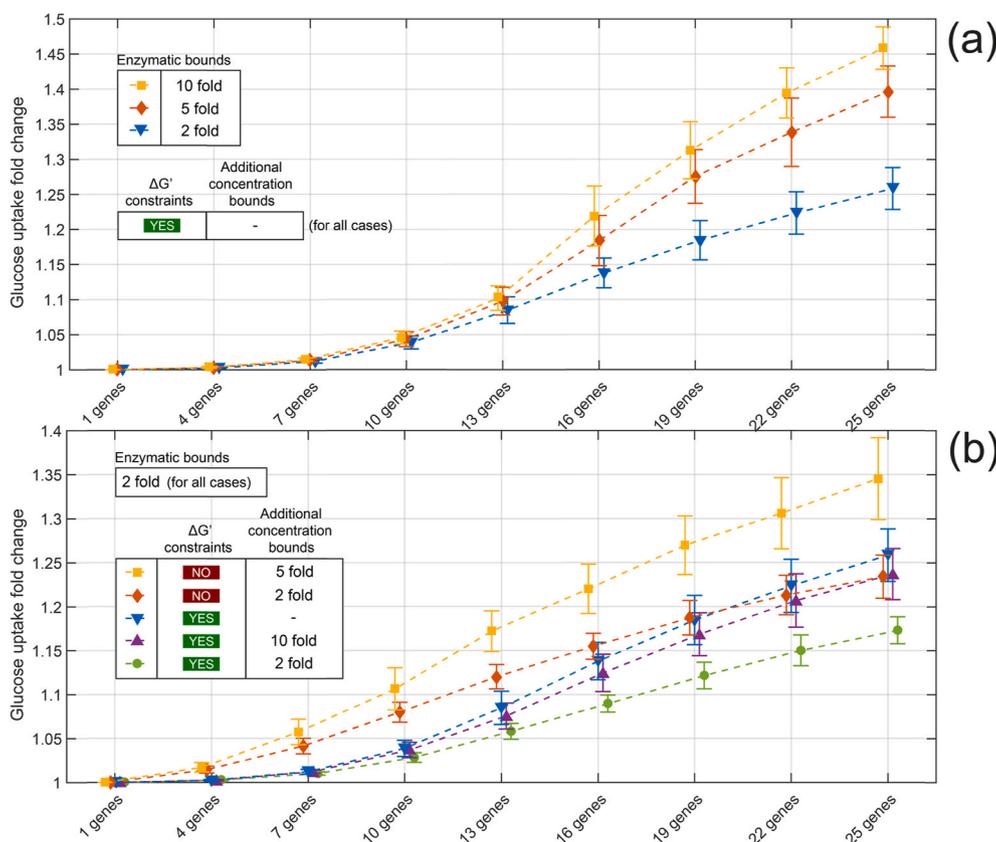
### 3.1. Strain design with physiological and design constraints for improved glucose uptake

Metabolic engineering interventions on pathways inevitably result in altered reaction rates as well as metabolite concentration levels. NRA, being a constraint-based method, allows for setting appropriate constraints on these quantities. Both fluxes and concentrations need to be constrained within realistic physiological bounds, conditional to each case study. For instance, severe changes in metabolite concentrations upon metabolic engineering interventions could significantly influence the organism's growth or even lead to an excess of toxic byproducts. The strain design should likewise consider that enzyme expression levels

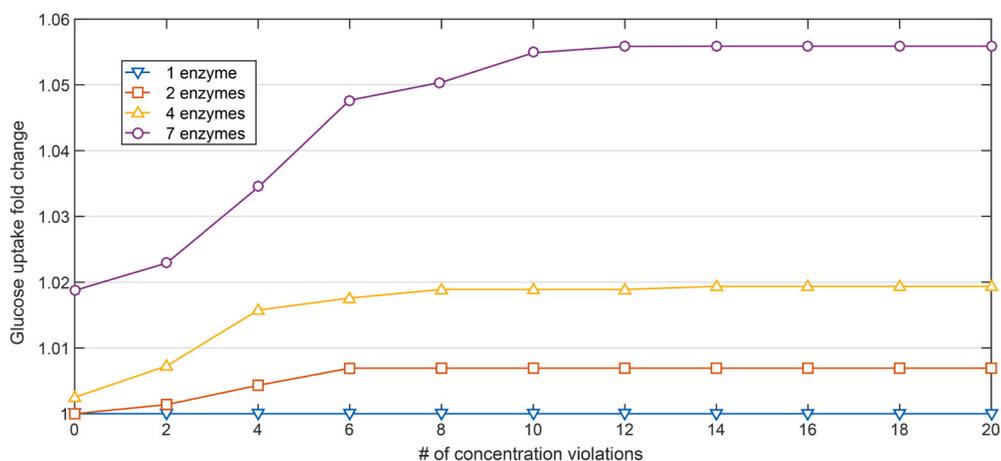
cannot increase beyond the currently reported experimentally achievable levels, and it cannot allow an infinite increase of reaction fluxes in the network. In contrast, the design should also be able to model gene knockouts by allowing both enzyme activities and reaction fluxes to decrease close to zero.

Here, we examined the effects of the imposed physiological and design constraints on the strain design for improved glucose uptake. For the studied physiology of *E. coli* growing under aerobic conditions, the experimentally observed glucose uptake rate was 8.09 mmol/gDW/h. Previous studies report *E. coli* strains with glucose uptake rates up to 12 mmol/gDW/h for aerobic (Jain and Srivastava, 2009) and even higher for anaerobic conditions (Causey et al., 2003; Gonzalez et al., 2017), suggesting that the rate of glucose consumption of the studied *E. coli* strain could still be increased. To this end, we analyzed the achievable glucose uptake rates with 2-fold, 5-fold, and 10-fold maximum allowable deviation of enzyme activities from the reference level for a set of designs ranging from 1 to 25 gene manipulations (Fig. 2a). The metabolite concentrations were subject to the thermodynamic feasibility constraints (Methods), and within the predefined physiological ranges (10 nM - 0.1M) for each cellular compartment. We allowed the fluxes to increase up to 10-fold of their reference level, and both fluxes and enzyme activities could reduce to zero. The latter means that solutions can include potential gene knockouts. As a mean to investigate the sensitivity of obtained solutions, we repeated the study for one reference and 18 extreme sets of control coefficients (Methods).

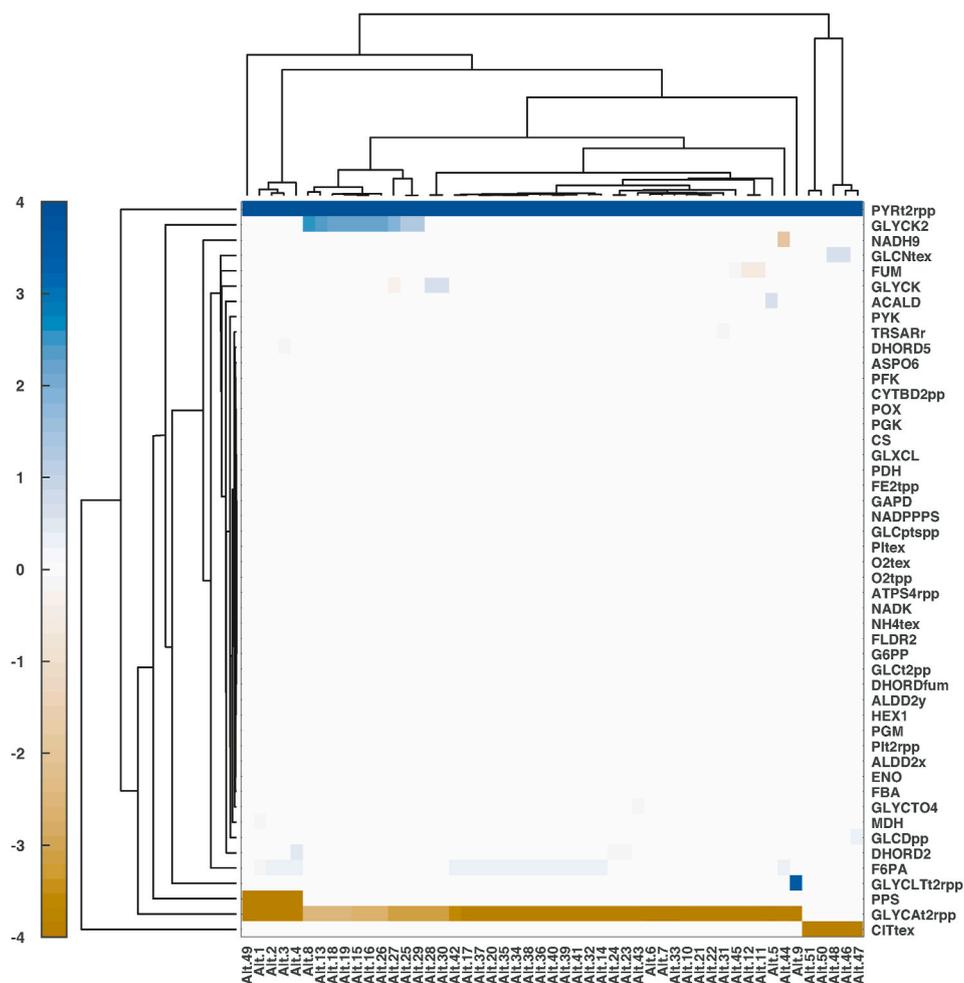
As the allowable enzyme activity change (Eqs. (10) and (23)) increased from 2- to 10-fold, the predicted attainable glucose uptake rate was about the same for up to 10 gene manipulations, indicating that for a small number of gene manipulations the upper limits on enzyme activity were not a limiting factor (Fig. 2a). However, starting from 13 gene manipulations, the difference between the predictions increased considerably. As expected, the higher limits on enzyme activity, the larger predicted improvement of glucose uptake was observed. For



**Fig. 2.** Effects of the physiological and design constraints on glucose uptake rate for a set of designs with different number of gene manipulations. Effects of: (a) allowed 2-fold (blue inverted triangles), 5-fold (orange diamonds), and 10-fold (yellow squares) changes in enzyme perturbation magnitude, and (b) different imposed metabolite concentration bounds. The study was performed for one reference and 18 extreme sets of CCs selected using Principal Component Analysis (Methods). In all cases, the fluxes were allowed to increase 10-fold and decrease to zero. The whiskers and the respective symbols indicate the interquartile ranges and the means of the considered CC-sets, respectively, as adjusted by the Bonferroni correction (Methods). Curves with blue inverted triangles correspond in the two graphs.



**Fig. 3.** Maximal attainable increase in glucose uptake as a function of a different number of metabolite concentration violations for one, two, four, and seven gene modifications. The fluxes were allowed to increase 10-fold and decrease to zero, the non-violated concentration bounds were subject to the thermodynamic constraints, and the enzymatic bounds were set to 2-fold. The reference model was used for all cases.



**Fig. 4.** Hierarchical clustering of the 51 alternative designs for the increase of pyruvate productivity, based on the suggested deregulation of individual enzymes.

example, NRA predicted for 25 gene manipulations that glucose uptake rate would increase by 26%, 39%, and 46% for 2-, 5-, and 10-fold change in enzyme activity, respectively. Interestingly, the predicted fold change of the glucose uptake across the nineteen studied reference and extreme CC-sets varied similarly for the designs with 13 or more gene manipulations (Fig. 2a whiskers). This rather constant variability as we go

toward a higher number of gene manipulations suggests that variability among 19 sets is primarily determined by the activity of a relatively small number of enzymes such as Glucose transport via diffusion, extracellular to periplasm (GLCtex), glyceraldehyde-3-phosphate dehydrogenase (GAPD), proton transport via diffusion, extracellular to periplasm (Htex), acetyl coenzyme A synthetase (ACS), malate synthase

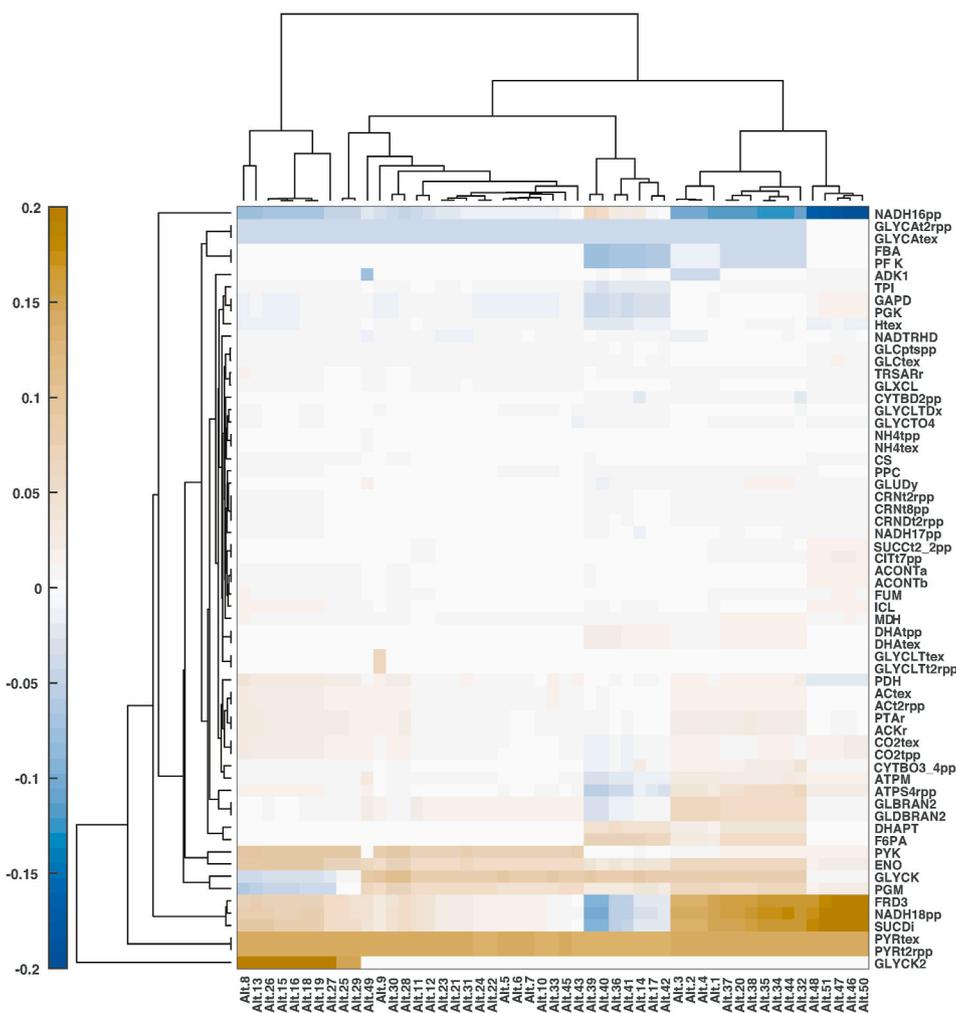


Fig. 5. Hierarchical clustering of the 51 alternative designs for the increase of pyruvate productivity, based on the absolute change in flux value of the 67 most affected reactions in the network.

(MLS), formate dehydrogenase (FDH4pp), and 2-dehydro-3-hydroxy phosphogluconate aldolase (EDA), which predominantly have control over the glucose uptake rate. This finding is in line with previous studies of metabolic systems demonstrating that just a few enzymes in the network (or corresponding parameters) determine the key metabolic properties such as system stability (Andreozzi et al., 2016b) or control over production fluxes (Miskovic et al., 2019a). A similar observation was reported in a more general context of biological systems (Daniels et al., 2008; Gutenkunst et al., 2007).

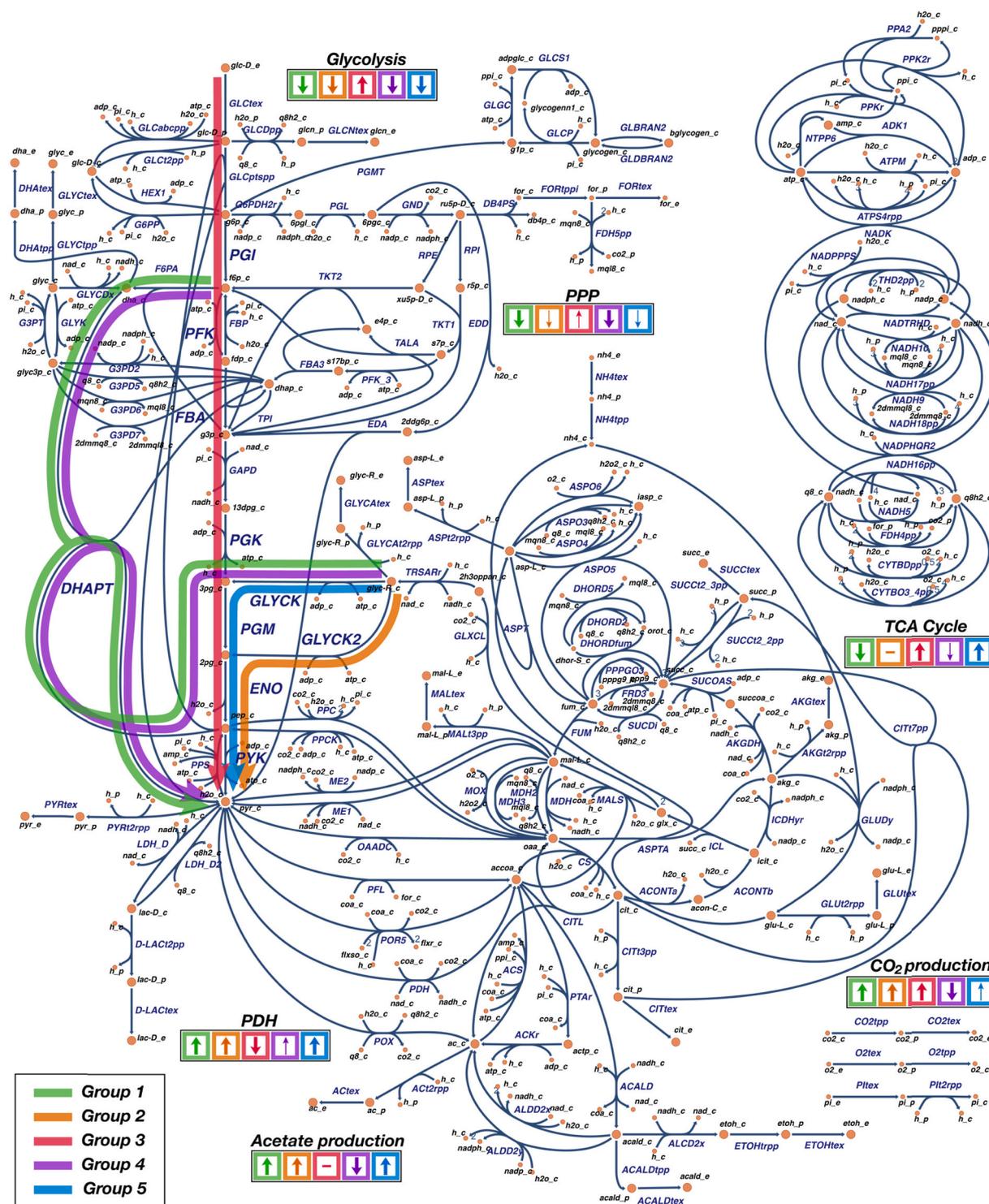
Next, we investigated how constraints on concentration deviations (Eqs. (9) and (23)) affect the attainable glucose uptake. This is a salient aspect of strain design because metabolic engineers have to ensure that metabolite concentrations remain within physiological bounds. For instance, it is vital not to exceed toxicity levels for some compounds. The studies on the effects of metabolite concentration constraints have also to consider thermodynamics because it is well known that the standard free Gibbs energy change of reactions couples the reaction directionalities and the metabolite concentrations (Ataman and Hatzi-manikatis, 2015). For this analysis, we have performed several studies by imposing different concentration bounds together with and without thermodynamic constraints (Fig. 2b). In general, our results suggest that NRA without thermodynamic constraints tends to predict higher increases in glucose uptake than NRA with thermodynamic constraints (Fig. 2b), meaning that thermodynamic constraints are limiting factors of strain design. The notable exception was that, starting from 19 gene manipulations, the 2-fold constraints on concentrations are more

limiting than the thermodynamic ones (Fig. 2b blue inverted triangles & orange diamonds). As expected, our results also show that the tighter the concentration deviation bounds we impose, the less important improvements of glucose uptake could be attained (Fig. 2b). For example, the attainable increase of glucose uptake rate with the thermodynamic and additional 2-fold and 10-fold constraints for 25 gene manipulations were 17% and 24%, respectively (Fig. 2b, green circles and violet triangles). We also observed that the variance of glucose uptake increase was smaller as the concentration bounds became more constrained. Similarly, we observed a trend that the variance in the studies with the thermodynamic constraints was smaller than in the ones without thermodynamic constraints.

### 3.1.1. Metabolite concentrations limiting the glucose uptake

Having demonstrated that limits on metabolite concentrations, either thermodynamic constraints or physiological limitations, significantly affect the attainable glucose uptake, we investigated how many and which metabolite concentrations should violate the thermodynamic constraints to achieve a higher glucose uptake. For simplicity and clarity of exposition, we allowed designs with one, two, four, and seven gene manipulations (Fig. 3).

In the cases of one and two gene manipulations, the flux through glucose uptake could not be modified with the thermodynamically feasible concentrations (zero violations). For a larger number of gene manipulations, a small increase in glucose uptake could be achieved even without violating the thermodynamics. For example, the



**Fig. 6.** Overview of metabolic engineering strategies devised using NRA for the improved specific production rate of pyruvate while preserving the pre-specified yield of pyruvate from glucose. 51 strategies devised with NRA were categorized in 5 distinct groups. The thick arrows on the graph denote the principal ways of carbon re-direction from the wild-type strain steady-state fluxes. The arrows in the colored boxes denote if the activity of the corresponding metabolic subsystem (glycolysis, pentose phosphate pathway (PPP), TCA cycle, acetate production, and CO<sub>2</sub> production) or reaction (PDH) was increased (arrow up), decreased (arrow down), or remained unchanged (dash). The thicker arrow in the colored boxes, the higher change in the activity occurred.

manipulation of seven genes would yield ~2% of glucose uptake increase for zero violations. However, when we allowed some concentration deviations to exceed their bounds, the potential violations pushed the attainable glucose uptake to higher values (Fig. 3). For instance, the seven gene manipulations design with ten concentration violations would result in 5.5% increase in glucose uptake.

Next, we focused on finding which were the metabolites whose concentration constraints should be violated to improve glucose uptake. To this end, we studied the case of four violations and two, four, and seven gene manipulations. For each gene manipulation study, we obtained the unique sets of four metabolite concentrations violating constraints (Table 3). The three gene manipulation studies involved, in

**Table 3**

Sets of four metabolite species with concentrations violating thermodynamic constraints for designs with two, four, and seven gene manipulations. The arrows indicate should a metabolite concentration be increased or decreased to improve glucose uptake. c: cytosol, p: periplasm.

Metabolite concentration	# of parameter manipulations		
	2	4	7
AMP (c) ↑	×	–	×
CTP (c) ↓	–	×	×
dCTP (c) ↓	–	×	–
H <sup>+</sup> (c) ↑	×	×	–
H <sup>+</sup> (p) ↑	×	×	×
L-Glutamine (c) ↓	–	–	×
L-Phenylalanine (c) ↑	×	–	–

total, seven species with concentrations violating the thermodynamic constraints. Among the seven species, periplasmic protons appeared in all three studies. Moreover, irrespectively of the study, to achieve a higher glucose uptake, the concentrations of protons (both cytosolic and periplasmic), AMP, and phenylalanine needed to be increased, while the ones for CTP, dCTP and glutamine needed to be decreased. The violations ranged from 2% for the case of cytosolic hydrogen to 57% for the case of CTP (Supplementary Table S1).

This analysis provides an opportunity to focus on each of the identified molecules, draw hypothesis about their role in the system limitations, and investigate these interplays and ways to overcome them *in vitro*. For example, it suggested that the pH value in compartments can be a limiting factor for metabolic design.

### 3.2. NRA design for Pyruvate production considers together specific production rate and yield

Pyruvate (pyruvic acid) is widely used in the food, chemical, and pharmaceutical industries. It is a precursor for the synthesis of various amino acids, and has been used for the production of antioxidants, food additives and supplements, pharmaceutical precursors, and biofuels (Atsumi et al., 2008; Kalman et al., 1999; Li et al., 2001; Zhang et al., 2010). The microbial production of pyruvate has been largely explored, and has involved both strain and process engineering and development (Maleki and Eiteman, 2017). In *E. coli*, pyruvate has been identified as one of the main hubs for the production of non-native commercial products (Zhang et al., 2016). The most common approach in microbial engineering for the overproduction of pyruvate is through deletions of the downstream utilization of pyruvate towards byproducts such as acetate, acetyl-CoA, and ethanol among others (Akita et al., 2016; Causey et al., 2004; Zhu et al., 2008).

To illustrate the features and flexibility of the NRA method, we showcase design for the improved specific productivity rate of pyruvate while simultaneously considering the yield of pyruvate from glucose, design constraints, and thermodynamic feasibility. We imposed the following design and physiology constraints: (i) up to five gene/enzyme activity manipulations, (ii) the genes encoding for metabolic enzymes could either be upregulated up to 50-fold or downregulated down to a knockout, (iii) the fluxes could increase up to 100-fold for upregulation and decrease down to zero for knockouts, and (iv) the concentration values were subject to the thermodynamic feasibility constraints and the physiological ranges (10 nM - 0.1M). Given these constraints, we first performed an optimization to determine the maximum yield of pyruvate from glucose. Then, we added the pyruvate yield to be at least 90% of this value to the set of constraints and maximized the specific pyruvate productivity rate. In this manner, we were able to implicitly account for the potential tradeoffs of yield and productivity that can occur in such designs.

We generated 51 alternative designs with five gene manipulations providing at least 99% of the maximum specific productivity rate of pyruvate and fulfilling the imposed constraints. The alternative designs

involved the manipulation of genes corresponding to 48 distinct enzymes (Supplementary Table S2). All cases provided over a 22-fold increase in both the pyruvate yield and specific productivity rate compared to the reference state. To understand better the mechanisms and identify metabolic patterns behind improved pyruvate production and yield, we performed clustering analysis over 51 designs with respect to (i) the 48 enzyme activity manipulations (Fig. 4), and (ii) predicted change in metabolic fluxes upon changes in enzyme activities (Fig. 5). For the clustering based on the absolute change in fluxes, we used the set of 67 reactions that had an absolute flux change of more than 0.01 mmol/gDW/h.

The transport of pyruvate from the cytosol to the periplasm (PYRt2rpp) appeared as a target in all designs with 50-fold upregulation of the PYRt2rpp encoding gene (Fig. 4). The upregulation of glycolytic enzymes and enzymes leading to pyruvate synthesis would also improve pyruvate production, with the most prominent target being glycerate kinase (GLYCK2). We also observed knockouts (or significant down-regulations) with the majority of downregulated genes involving the consumption of pyruvate towards the formation of byproducts. Among these, the periplasmic transport of glycerate (GLYCAT2rpp) was present in most generated sets, being replaced by the extracellular transport of citrate (CITtex) in a few cases (Fig. 4). We also observed the knockout of Phosphoenolpyruvate synthase (PPS), which is associated with the conversion of pyruvate to phosphoenolpyruvate.

A closer cross-inspection of the two figures together with the Supplementary Table S2 reveals that there are five groups of alternative ways to satisfy design specifications. Alternatives 1-4, 20, 32, 34, 35, 37, 38, 44, and 49 (Figs. 4 and 5, Supplementary Table S2) constituted the first group that improves pyruvate production while maintaining at least 90% of the yield by: (i) a strong upregulation of pyruvate transport PYRt2rpp; (ii) a strong downregulation of D-glycerate transport GLYCAT2rpp; (iii) a slight upregulation of glycolysis either via enolase (ENO) for Alternative 49 or via fructose 6-phosphate aldolase (F6PA) for other alternatives in this group (Supplementary Table S2); and (iv) a knockout of PPS for alternatives 1-4, 49 or a slight downregulation of fructose-bisphosphate aldolase (FBA) for alternatives 20, 32, 34, 35, 37, 38, and 44.

As a result of these manipulations, the carbon flow was re-directed from the secretion of (R)-glycerate toward the production of phosphoenolpyruvate through glycerate kinase (GLYCK), phosphoglycerate mutase (PGM), and ENO (Fig. 6 and Supplementary Fig. S1). Downstream, phosphoenolpyruvate is converted to pyruvate through dihydroxyacetone phosphotransferase (DHAPT), whose activity was also increased. This group is further characterized by a slight increase in acetate production and CO<sub>2</sub> secretion, and a deregulation of the ATP metabolism such as an increase of the ATP non-growth associated maintenance (ATPM) or a decrease in activity of adenylate kinase (ADK1) for alternatives 1-4, 49. Moreover, the conversion of fructose-6-phosphate to glyceraldehyde-3-phosphate instead through FBA and phosphofructokinase (PFK) was diverted through F6PA.

The second group consisting of alternatives 8, 13, 15, 16, 18, 19, 25-27, and 29 shared the manipulations (i) and (ii) with the first group. In addition, this group involved: (iii) an upregulation of glycerate kinase GLYCK2; and (iv) a slight upregulation of pyruvate kinase (PYK). The observed effects of these manipulations were similar to the ones of the first group with the increased activity of lower glycolysis and acetate secretion pathway (Figs. 5 and 6 and Supplementary Fig. S2). The notable difference was that the carbon diverted from glycerate secretion was channeled through GLYCK2, ENO, and PYK to pyruvate. Furthermore, we observed a slight increase in activity of the TCA cycle and pyruvate dehydrogenase (PDH), whereas the ATP metabolism remained mostly unchanged.

The third group formed by alternatives 46-48, 50, and 51 was distinct from the other groups because it involved strategy to knockout citrate transport CITtex instead of GLYCAT2rpp (Fig. 4). Additional manipulations in these group were a slight downregulation of citrate

synthase (CS) and a slight upregulation of glycolytic enzymes PGM (Alternatives 46, 47, 50, 51) or ENO (Alternative 48). Overall, these manipulations resulted in increased activity of the upper and lower glycolysis, pentose phosphate pathway, and the TCA cycle (Fig. 6 and Supplementary Fig. S3). This was the only group with increased activity of the upper glycolysis. We have also observed a decrease in activity of PDH (Fig. 5 and Supplementary Fig. S3).

The fourth group constituted by alternatives 14, 17, 36, 39–42 had a distinct pattern in the network flux distributions while sharing manipulations (i)–(iii) with the first group (Figs. 5 and 6 and Supplementary Fig. S4). A slight downregulation of PFK together with manipulations (i)–(iii) had a considerable impact by reducing the activity of the reactions in the upper glycolysis (PFK, FBA, triose-phosphate isomerase (TPI), glyceraldehyde-3-phosphate dehydrogenase (GAPD), phosphoglycerate kinase (PGK)), the ETC chain (NADH dehydrogenase (NADH18pp), Cytochrome oxidase bo3 (CYTBO3\_4pp) and the ATP metabolism (ATPM and ATP synthase (ATPS4rpp)). We have also observed, in contrast to other groups, a reduced activity in CO<sub>2</sub> and acetate secretion pathways.

The fifth group, composed of alternatives 5–7, 9–12, 21–24, 28, 30, 31, 33, 43, 45, and the second group have in common manipulations (i), (ii), and (iv) (Fig. 4 and Supplementary Table S2). Additionally, the fifth group involved either a very slight upregulation of glycolytic enzymes PGM, ENO, and PGK (alternatives 11, 12, 21–24, 31, 43, and 45) or a very slight downregulation of PDH (alternatives 5–7, 10, 28, and 30). As expected, the resulting flux distribution was similar to the one of the second group (Supplementary Fig. S5). The difference was that in this group the carbon from (R)-glycerate was diverted to 2-phosphoglycolate through GLYCK and PGM instead through GLYCK2 as it was done in the second group. Overall, compared to other groups, the manipulations of this group have changed the least the network flux distribution (Fig. 5).

Once the principal strategies are determined, the final decision is made by experts based on the comparative analysis of the proposed alternative groups and on considerations about the practical implementation of the designs.

### 3.2.1. Comparison with targets determined by looking only at unconstrained specific productivity

We proceeded by examining how different are the targets obtained with the NRA design from the ones determined by looking only the specific productivity rate of pyruvate. This comparison will reveal how the physiology and design constraints affect our design decisions. To this end, we computed the mean values of the control coefficient of the specific productivity rate of pyruvate with respect to network enzyme activities, and then ranked them according to their absolute value. Most of the top 15 enzymes represent either extracellular transports such as

oxygen uptake and ammonium secretion or glycolysis reactions leading to the synthesis of pyruvate (Table 4). We also observed that pentose phosphate pathway reactions Ribose-5-phosphate isomerase (RPI) and Ribulose 5-phosphate 3-epimerase (RPE) have a strong control on the pyruvate specific productivity. These reactions are around the branching point for purine nucleotides and histidine biosynthesis pathway, since ribose-5-phosphate is the main precursor for the production of those biomass components. Flux through RPE brings the carbon back to glycolysis, which reinforces pyruvate production. In contrast, RPI has a negative control over pyruvate production because it directs the carbon through purine nucleotides and histidine biosynthesis. Uptake of ammonia has a similar effect as RPI on pyruvate production since the main route to metabolize ammonia is through amino acid biosynthesis, which carries the carbon flux away from pyruvate.

Interestingly, the majority of these enzymes do not appear as targets in any of the NRA alternatives (Table 4 and Supplementary Table S2). Some of these enzymes exhibit a large control over multiple fluxes and concentrations across the metabolic network. These are, therefore, severely constrained by the imposed specifications in the constrained NRA design. This suggests that the NRA formulation will favor parameters that have less control over the network, ensuring that cellular balance will not be excessively perturbed.

## 4. Conclusions

The NRA framework enables the consistent and sophisticated design of metabolic engineering strategies using MCA-based control coefficients. NRA is computationally faster and simpler than other approaches since the derivation of control coefficients does not require the numerical integration of non-linear kinetic models and enables the implementation of a wide variety of metabolic engineering criteria. In particular, NRA allows to metabolic engineers to simultaneously consider the desired yield and specific productivity of target chemicals, which is a ubiquitous design consideration in metabolic engineering.

To our knowledge, this type of approach has never been applied to large or genome scale kinetic models of metabolism. Using a previously published large-scale kinetic model of *E. coli*, we demonstrated that the NRA formulation can be applied to large-scale metabolic networks. We used the PCA method to select a number of representative sets of kinetic parameters among their population, in order to effectively represent the uncertainty and flexibility of the kinetic model in respect to parameterization. One of the main advantages of NRA is that, being a constraint-based modeling method, it can accommodate the integration of biologically relevant bounds and constraints, which ensure that the proposed strategies are consistent with the entire system capabilities and limitations thereof. Since the NRA model predictions can be sensitive to the user-defined bounds on the allowable reaction flux, metabolite concentration and enzymatic expression deviations, the importance of including relevant physiological constraints, such as thermodynamic feasibility constraints, was discussed extensively. Focusing on the case of pyruvate production, a compound of great industrial interest, viable metabolic engineering strategies were shown to be readily derived using this formulation. Alternative solutions could also be generated and evaluated on their efficiency and potential implementation.

Information about the kinetic properties of enzymes in the NRA formulation stems from the control coefficients. Control coefficients are log-linear approximations of metabolic responses to genetic modifications, and one can overestimate the effects of increasing enzyme activities on the metabolic responses for larger perturbations (Schuster, 1999). Therefore, one might expect that NRA devised design strategies can have similar shortcomings for larger perturbations. However, it is worth mentioning that NRA designs, being constrained by the physiological and genome editing restrictions, will, in general, require smaller perturbations of enzyme activities than ones proposed by the classical MCA, thus alleviating this shortcoming. While metabolic engineering designs based on control coefficients have shown success in

**Table 4**

Top ranked parameters based on their control over pyruvate production flux PYR<sub>tex</sub>. Ranking was computed based on the mean values of 50'000 sets of Control Coefficients.

Rank	Parameter Name	Control over PYR <sub>tex</sub>
1	NH4 <sub>tex</sub>	negative
2	O2 <sub>tp</sub>	negative
3	NH4 <sub>tp</sub>	negative
4	ATPS4 <sub>rpp</sub>	positive
5	GLC <sub>tex</sub>	positive
6	O2 <sub>tex</sub>	negative
7	TPI	positive
8	PGI	positive
9	PFK	positive
10	RPI	negative
11	PI <sub>tex</sub>	positive
12	PGM	positive
13	GLC <sub>ptspp</sub>	positive
14	RPE	positive
15	PYR <sub>t2rpp</sub>	positive

experimental implementations (Andreozzi et al., 2016a; Miskovic et al., 2017), a step further in obtaining reliable predictions for strain designs requiring large perturbations of enzyme activities would be to use large-scale nonlinear kinetic models.

We believe that this formulation will provide a refined alternative to computational genetic design due to its simplicity and modularity, and that it will continue to be enhanced through the introduction of ever-growing omics data, and additional specialized constraints and objectives.

## Acknowledgements

S.T. and T.H. were supported by the Swiss National Science Foundation grant [315230\_163423]. M.A. was supported by the RTD grant MicroscapesX within SystemsX.ch, the Swiss Initiative for Systems Biology evaluated by the Swiss National Science Foundation, and RobustYeast within ERA net project via SystemsX.ch. L.M. and V.H. were supported by the Ecole Polytechnique Fédérale de Lausanne (EPFL).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymben.2021.03.003>.

## References

- Akita, H., Nakashima, N., Hoshino, T., 2016. Pyruvate Production Using Engineered *Escherichia coli*, vol. 6. *Amb Express*.
- Andreozzi, S., Chakrabarti, A., Soh, K.C., Burgard, A., Yang, T.H., Van Dien, S., Miskovic, L., Hatzimanikatis, V., 2016a. Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant *E. coli* using large-scale kinetic models. *Metab. Eng.* 35, 148–159.
- Andreozzi, S., Miskovic, L., Hatzimanikatis, V., 2016b. iSCHRUNK—in silico approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks. *Metab. Eng.* 33, 158–168.
- Ataman, M., Gardiol, D.F.H., Fengos, G., Hatzimanikatis, V., 2017. redGEM: systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. *PLoS Comput. Biol.* 13.
- Ataman, M., Hatzimanikatis, V., 2015. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. *Curr. Opin. Biotechnol.* 36, 176–182.
- Ataman, M., Hatzimanikatis, V., 2017. lumpGEM: systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites. *PLoS Comput. Biol.* 13, e1005513.
- Atsumi, S., Hanai, T., Liao, J.C., 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451, 86–U13.
- Bailey, J.E., 1991. Toward a science of metabolic engineering. *Science* 252, 1668–1675.
- Casey, T.B., Shanmugam, K.T., Yomano, L.P., Ingram, L.O., 2004. Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2235–2240.
- Casey, T.B., Zhou, S., Shanmugam, K.T., Ingram, L.O., 2003. Engineering the metabolism of *Escherichia coli* W3110 for the conversion of sugar to redox-neutral and oxidized products: homoacetate production. *Proc. Natl. Acad. Sci. U. S. A.* 100, 825–832.
- Chakrabarti, A., Miskovic, L., Soh, K.C., Hatzimanikatis, V., 2013. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol. J.* 8, 1043–1057.
- Costa, R.S., Hartmann, A., Vinga, S., 2016. Kinetic modeling of cell metabolism for microbial production. *J. Biotechnol.* 219, 126–141.
- Daniels, B.C., Chen, Y.J., Sethna, J.P., Gutenkunst, R.N., Myers, C.R., 2008. Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol.* 19, 389–395.
- Gombert, A.K., Nielsen, S., 2000. Mathematical modelling of metabolism. *Curr. Opin. Biotechnol.* 11, 180–186.
- Gonzalez, J.E., Long, C.P., Antoniewicz, M.R., 2017. Comprehensive analysis of glucose and xylose metabolism in *Escherichia coli* under aerobic and anaerobic conditions by <sup>13</sup>C metabolic flux analysis. *Metab. Eng.* 39, 9–18.
- Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., Sethna, J.P., 2007. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* 3, 1871–1878.
- Hadadi, N., Hatzimanikatis, V., 2015. Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways. *Curr. Opin. Chem. Biol.* 28, 99–104.
- Hameri, T., Boldi, M.O., Hatzimanikatis, V., 2019a. Statistical inference in ensemble modeling of cellular metabolism. *PLoS Comput. Biol.* 15, e1007536.
- Hameri, T., Fengos, G., Ataman, M., Miskovic, L., Hatzimanikatis, V., 2019b. Kinetic models of metabolism that consider alternative steady-state solutions of intracellular fluxes and concentrations. *Metab. Eng.* 52, 29–41.
- Hameri, T., Fengos, G., Hatzimanikatis, V., 2021 Mar 20. The effects of model complexity and size on metabolic flux distribution and control: case study in *Escherichia coli*. *BMC Bioinformatics* 22 (1), 134. <https://doi.org/10.1186/s12859-021-04066-y>.
- Hatzimanikatis, V., Bailey, J.E., 1996. MCA has more to say. *J. Theor. Biol.* 182, 233–242.
- Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996a. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE J.* 42, 1277–1292.
- Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996b. Optimization of regulatory architectures in metabolic reaction networks. *Biotechnol. Bioeng.* 52, 485–500.
- Heinrich, R., Schuster, S., 1996. *The Regulation of Cellular Systems*.
- Henry, C., Broadbelt, L., Hatzimanikatis, V., 2007. Thermodynamics-based metabolic flux analysis. *Biophysical Journal* 92 (5), 1792–1805. <https://doi.org/10.1529/biophysj.106.093138>.
- Henry, C., Jankowski, M.D., Broadbelt, L.J., Hatzimanikatis, V., 2006. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.* 90, 1453–1461.
- Jain, R., Srivastava, R., 2009. Metabolic investigation of host/pathogen interaction using MS2-infected *Escherichia coli*. *BMC Syst. Biol.* 3, 1–11.
- Kacser, H., Burns, J.A., Fell, D.A., 1995. The control of flux. *Biochem. Soc. Trans.* 23, 341–366.
- Kalman, D., Colker, C.M., Wilets, I., Roufs, J.B., Antonio, J., 1999. The effects of pyruvate supplementation on body composition in overweight individuals. *Nutrition* 15, 337–340.
- Klamt, S., Muller, S., Regensburger, G., Zanghellini, J., 2018. A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. *Metab. Eng.* 47, 153–169.
- Li, Y., Chen, J., Lun, S.Y., 2001. Biotechnological production of pyruvic acid. *Appl. Microbiol. Biotechnol.* 57, 451–459.
- Long, M.R., Ong, W.K., Reed, J.L., 2015. Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol.* 34, 135–141.
- Maleki, N., Eiteman, M.A., 2017. Recent progress in the microbial production of pyruvic acid. *Ferment. Basel* 3.
- Mavrouniotis, M.L., 1990. Group contributions for estimating standard Gibbs energies of formation of biochemical-compounds in aqueous-solution. *Biotechnol. Bioeng.* 36, 1070–1082.
- Mavrouniotis, M.L., 1991. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* 266, 14440–14445.
- McCloskey, D., Gangoiti, J.A., King, Z.A., Naviaux, R.K., Barshop, B.A., Palsson, B.O., Feist, A.M., 2014. A model-driven quantitative metabolomics analysis of aerobic and anaerobic metabolism in *E. coli* K-12 MG1655 that is biochemically and thermodynamically consistent. *Biotechnol. Bioeng.* 111, 803–815.
- Miskovic, L., Alf-Tuomala, S., Soh, K.C., Barth, D., Salusjärvi, L., Pitkänen, J.-P., Ruohonen, L., Penttilä, M., Hatzimanikatis, V., 2017. A design–build–test cycle using modeling and experiments reveals interdependencies between upper glycolysis and xylose uptake in recombinant *S. cerevisiae* and improves predictive capabilities of large-scale kinetic models. *Biotechnol. Biofuels* 10, 166.
- Miskovic, L., Béal, J., Moret, M., Hatzimanikatis, V., 2019a. Uncertainty reduction in biochemical kinetic models: enforcing desired model properties. *PLoS Comput. Biol.* 15, e1007242.
- Miskovic, L., Hatzimanikatis, V., 2010. Production of biofuels and biochemicals: in need of an ORACLE. *Trends Biotechnol.* 28, 391–397.
- Miskovic, L., Hatzimanikatis, V., 2011. Modeling of uncertainties in biochemical reactions. *Biotechnol. Bioeng.* 108, 413–423.
- Miskovic, L., Tokic, M., Fengos, G., Hatzimanikatis, V., 2015. Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes. *Curr. Opin. Biotechnol.* 36, 1–8.
- Miskovic, L., Tokic, M., Savoglidi, G., Hatzimanikatis, V., 2019b. Control theory concepts for modeling uncertainty in enzyme kinetics of biochemical networks. *Ind. Eng. Chem. Res.* 58, 13544–13554.
- Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.O., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Mol. Syst. Biol.* 7, 535.
- Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248.
- Patil, K.R., Rocha, I., Forster, J., Nielsen, J., 2005. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinf.* 6, 308.
- Reder, C., 1988. Metabolic control theory: a structural approach. *J. Theor. Biol.* 135, 175–201.
- Salvy, P., Fengos, G., Ataman, M., Pathier, T., Soh, K.C., Hatzimanikatis, V., 2019. pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis. *Bioinformatics* 35, 167–169.
- Schneider, P., Klamt, S., 2019. Characterizing and ranking computed metabolic engineering strategies. *Bioinformatics* 35, 3063–3072.
- Schuster, S., 1999. Use and limitations of modular metabolic control analysis in medicine and biotechnology. *Metab. Eng.* 1, 232–242.
- Soh, K.S., Hatzimanikatis, V., 2010. Network thermodynamics in the post-genomic era. *Curr. Opin. Microbiol.* 13, 350–357.
- Soh, K.S., Miskovic, L., Hatzimanikatis, V., 2012. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. *FEMS Yeast Res.* 12, 129–143.
- Tokic, M., Hatzimanikatis, V., Miskovic, L., 2020. Large-scale kinetic metabolic models of *Pseudomonas putida* KT2440 for consistent design of metabolic engineering strategies. *Biotechnol. Biofuels* 13.
- Varma, A., Boesch, B.W., Palsson, B.O., 1993. Biochemical production capabilities of *escherichia-coli*. *Biotechnol. Bioeng.* 42, 59–73.
- Wang, L., Birol, I., Hatzimanikatis, V., 2004. Metabolic control analysis under uncertainty: framework development and case studies. *Biophys. J.* 87, 3750–3763.

- Wang, L., Dash, S., Ng, C.Y., Maranas, C.D., 2017. A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol.* 2, 243–252.
- Zhang, X.L., Tervo, C.J., Reed, J.L., 2016. Metabolic assessment of *E. coli* as a Biofactory for commercial products. *Metab. Eng.* 35, 64–74.
- Zhang, Y.N., Tao, F., Du, M.F., Ma, C.Q., Qiu, J.H., Gu, L.C., He, X.F., Xu, P., 2010. An efficient method for N-acetyl-d-neuraminic acid production using coupled bacterial cells with a safe temperature-induced system. *Appl. Microbiol. Biotechnol.* 86, 481–489.
- Zhu, Y.H., Eiteman, M.A., Altman, R., Altman, E., 2008. High glycolytic flux improves pyruvate production by a metabolically engineered *Escherichia coli* strain. *Appl. Environ. Microbiol.* 74, 6649–6655.