# Constraint-based metabolic control analysis for rational strain engineering

Sophia Tsouka[1,*], Meric Ataman[1,2,*], Tuure Hameri[1], Ljubisa Miskovic[1], Vassily Hatzimanikatis[1,†]

* equally contributing authors

[1] Laboratory of Computational Systems Biology (LCSB), EPFL, CH-1015 Lausanne, Switzerland.

[2] current address: Computational and Systems Biology, Biozentrum University of Basel, Klingelbergstrasse 50-70, 4056, Basel, Switzerland

[†] Corresponding author:

Vassily Hatzimanikatis,

Laboratory of Computational Systems Biotechnology (LCSB), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

**Email:** vassily.hatzimanikatis@epfl.ch , **Phone:** +41 (0)21 693 98 70, **Fax:** +41 (0)21 693 98 75

## Abstract

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 metabolic concentrations, it fails to account for the physiological limitations of the cellular environment and metabolic engineering design constraints. We report here a constraint-based framework based on MCA, Network Response Analysis (NRA), for the rational genetic strain design that incorporates biologically relevant constraints, as well as genome editing restrictions. The NRA core constraints being similar to the ones of Flux Balance Analysis, allow it to be used for a wide range of optimization criteria and with various physiological constraints. We 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

32 engineering that accommodates the incorporation of physiological data at metabolic flux,
33 metabolite concentration, and enzyme expression levels.
34
35

## 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 quantify the distribution of control over metabolic states in a network such as fluxes and concentrations (Kacser et al., 1995). In MCA, Control Coefficients (CCs) quantify how a given metabolic flux or metabolite concentration would respond to perturbations of the system parameters. 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 task using classical MCA tools, especially if the design requires

multiple genetic alterations. Moreover, MCA does not allow including explicitly any form 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 obstacles. 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); (1996b) 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 guarantied thermodynamic feasibility.

To illustrate how NRA can be used to efficiently analyze, enumerate, and propose alternative metabolic engineering strategies, we used a large-scale thermodynamically-curated, metabolic model of *E. coli* (Hameri et al., 2019c), 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 flux and concentration CCs consistent with the experimental observations. We then used the generated CCs to formulate with NRA the design strategies in two case studies (i) improvement of glucose uptake rate, 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.

## Results and Discussion

### NRA method

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 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.

*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). NRA accommodates 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.

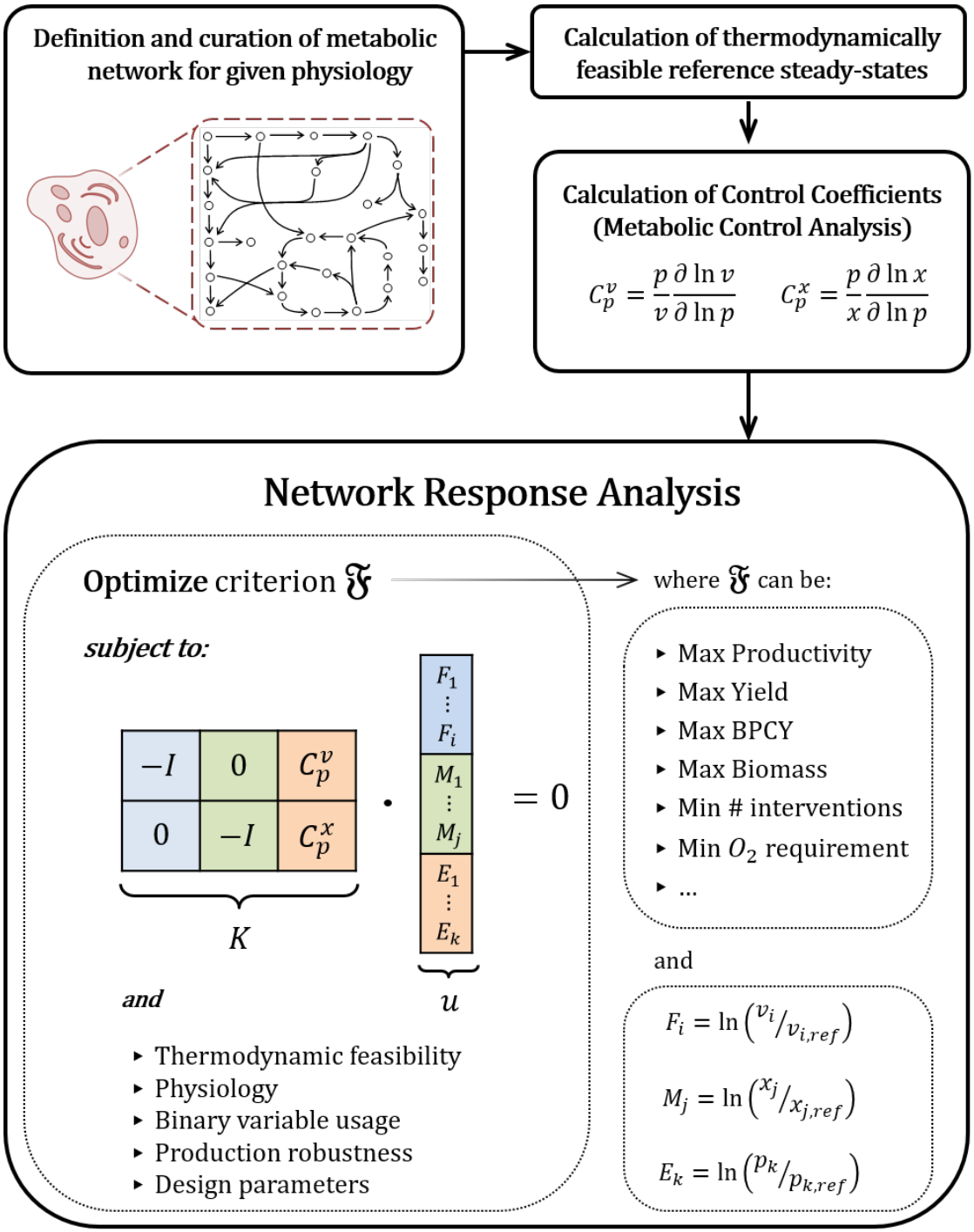



*Figure 1. The NRA workflow is organized in four main steps. In the first three steps, we formulate the stoichiometry, integrate available experimental data and compute the steady-state thermodynamically feasible fluxes and concentrations, and compute the flux and concentration control coefficients for 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.*

*Table 1. The NRA mathematical formulation together with a non-exhaustive selection of optimization objectives. The definition of indices, parameters, and variables is provided in Table 2.*

**Optimize Criterion $\mathfrak{F}$:**

| | *FBA* | *NRA* | |
|---|---|---|---|
| **Max productivity** | $\max\ v_{product}$ | $\max\ \tilde{F}_{product}$ | (1) |
| **Max yield** | $\max\ \frac{v_{product}}{v_{substrate}}$ | $\max\ \tilde{F}_{product} - \tilde{F}_{substrate}$ | (2) |
| **Max BPCY** | $\max\ v_{biomass} * \frac{v_{product}}{v_{substrate}}$ | $\max\ \tilde{F}_{biomass} + \tilde{F}_{product} - \tilde{F}_{substrate}$ | (3) |
| **Max biomass** | $\max\ v_{biomass}$ | $\max\ \tilde{F}_{biomass}$ | (4) |
| **Min # interventions** | — | $\min \sum_{k\in\mathcal{K}} (1 - z_k)$ | (5) |
| **Min $O_2$ requirement** | $\min\ v_{O_2,\,uptake}$ | $\min\ \tilde{F}_{O_2,\,uptake}$ | (6) |

**subject to constraints:**

| | *FBA* | *NRA* | |
|---|---|---|---|
| **Balance (Mass / Response)** | $N \cdot v = 0$ | $K \cdot u = 0 \Leftrightarrow \left\{ \begin{matrix} F_i - \sum_{k\in\mathcal{K}} \left(C_{E_k}^{v_i} * E_k\right) = 0 \\ M_j - \sum_{k\in\mathcal{K}} \left(C_{E_k}^{x_j} * E_k\right) \end{matrix} \right\}$ | (7) |
| **Variable bounds** | $lb \le v_i \le ub$ | $lb_F \le F_i \le ub_F$ | (8) |
| | | $lb_M \le M_j \le ub_M$ | (9) |
| | | $lb_E \le E_k \le ub_E \Leftrightarrow \left\{ \begin{matrix} 0 \le E_k^U \le ub_E \\ 0 \le E_k^D \le -lb_E \end{matrix} \right\}$ | (10) |
| **Binary variable usage** | — | $E_k^U + E_k^D + \xi * z_k \le \xi$ | (11) |
| | | $E_k^{UU} + E_k^{DU} \le 1$ | (12) |
| | | $E_k^U - \xi * E_k^{UU} < 0$ | (13) |
| | | $E_k^D - \xi * E_k^{DU} < 0$ | (14) |

| | | | |
|---|---|---|---|
| **Thermodynamic feasibility** | — | $\Delta_r G_i' < 0 \Leftrightarrow \Delta_r G_i'^{o} + RT \sum_{m \in \mathcal{M}} \tilde{M}_m^i < 0$ | (15) |

and

| | *FBA* | *NRA* | |
|---|---|---|---|
| **Production robustness (min productivity)** | $v_{product} \geq \alpha_0$ | $\tilde{F}_{product} \geq \beta_0$ | (16) |
| **Production robustness (min yield)** | $\frac{v_{product}}{v_{substrate}} \geq \alpha_1$ | $\tilde{F}_{product} - \tilde{F}_{substrate} \geq \beta_1$ | (17) |
| **Production robustness (min biomass)** | $v_{biomass} \geq \alpha_2$ | $\tilde{F}_{biomass} \geq \beta_2$ | (18) |
| **Design parameters (max # interventions)** | — | $\sum_{k \in \mathcal{K}} (1 - z_k) \leq \beta_3$ | (19) |
| **Design parameters (max $O_2$ requirement)** | $v_{O_2,uptake} \leq \alpha_3$ | $\tilde{F}_{O_2,uptake} \leq \beta_4$ | (20) |

where

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

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

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

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

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

138

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

| Index letter | Type | Refers to | Set or unit |
|---|---|---|---|
| $\boldsymbol{i}$ | Index | Reaction | $i \in \mathcal{I}$ |
| $\boldsymbol{j}$ | Index | Metabolite | $j \in \mathcal{J}$ |
| $\boldsymbol{k}$ | Index | Enzyme | $k \in \mathcal{K}$ |
| $\boldsymbol{m}$ | Index | Metabolite participating in reaction $i$ | $m \in \mathcal{M}$ |
| $\boldsymbol{v_i}$ | Variable | Flux of reaction $i$ | $mmol \cdot gDW^{-1} \cdot h^{-1}$ |
| $\boldsymbol{x_j}$ | Variable | Concentration of metabolite $j$ | $M$ |
| $\boldsymbol{p_k}$ | Variable | Catalytic activity of enzyme $k$ | $mmol \cdot h^{-1}$ |
| $\boldsymbol{\tilde{F}_i}$ | Variable | Scaled flux deviation of reaction $i$ | $\tilde{F}_i \in \mathbb{R}$ |
| $\boldsymbol{F_i}$ | Variable | Flux deviation of reaction $i$ | $F_i \in \mathbb{R} \cap [lb_F, ub_F]$ |
| $\boldsymbol{\tilde{M}_j}$ | Variable | Scaled concentration deviation of metabolite $j$ | $\tilde{M}_j \in \mathbb{R}$ |
| $\boldsymbol{M_j}$ | Variable | Concentration deviation of metabolite $j$ | $M_j \in \mathbb{R} \cap [lb_M, ub_M]$ |
| $\boldsymbol{E_k}$ | Variable | Catalytic activity deviation of enzyme $k$ | $E_k \in \mathbb{R} \cap [lb_E, ub_E]$ |
| $\boldsymbol{E_k^U}$ | Variable | Upregulation of catalytic activity of enzyme $k$ | $E_k^U \in \mathbb{R} \cap [0, ub_E]$ |
| $\boldsymbol{E_k^D}$ | Variable | Downregulation of catalytic activity of enzyme $k$ | $E_k^D \in \mathbb{R} \cap [0, -lb_E]$ |
| $\boldsymbol{\Delta_r G_i'}$ | Variable | Gibbs free energy change of reaction $i$ | $\Delta_r G_i' \in \mathbb{R}$ |
| $\boldsymbol{\Delta_r G_i'^o}$ | Variable | Standard Gibbs free energy change of reaction $i$ | $\Delta_r G_i'^o \in \mathbb{R} \cap \left[\Delta_r G_{i,est}'^o \pm err\right]$ |
| $\boldsymbol{u}$ | Variable | Vector of NRA variables | $u \in \mathbb{R}^{(\mathcal{I}+\mathcal{J}+2\mathcal{K})}$ |
| $\boldsymbol{E_k^{UU}}$ | Variable | Binary upregulation of catalytic activity of enzyme $k$ | $E_k^{UU} \in \{0,1\}$ |
| $\boldsymbol{E_k^{DU}}$ | Variable | Binary downregulation of catalytic activity of enzyme $k$ | $E_k^{DU} \in \{0,1\}$ |
| $\boldsymbol{z_k}$ | Variable | Deregulation of enzyme $k$ | $z_k \in \{0,1\}$ |
| $\boldsymbol{C_p^v}$ | Parameter | Flux control coefficient | $C_p^v \in \mathbb{R}$ |
| $\boldsymbol{C_p^x}$ | Parameter | Concentration control coefficient | $C_p^x \in \mathbb{R}$ |
| $\boldsymbol{v_{i,ref}}$ | Parameter | Reference flux of reaction $i$ | $mmol \cdot gDW^{-1} \cdot h^{-1}$ |
| $\boldsymbol{x_{j,ref}}$ | Parameter | Reference concentration of metabolite $j$ | $M$ |
| $\boldsymbol{p_{k,ref}}$ | Parameter | Reference catalytic activity of enzyme $k$ | $mmol \cdot h^{-1}$ |
| $\boldsymbol{N}$ | Parameter | Stoichiometric matrix | $N \in \mathbb{R}^{\mathcal{J} \times \mathcal{I}}$ |
| $\boldsymbol{K}$ | Parameter | NRA matrix | $K \in \mathbb{R}^{(\mathcal{I}+\mathcal{J}) \times (\mathcal{I}+\mathcal{J}+2\mathcal{K})}$ |
| $\boldsymbol{lb_F}$ | Parameter | Flux deviation lower bound | $lb_F \in \mathbb{R}$ |
| $\boldsymbol{ub_F}$ | Parameter | Flux deviation upper bound | $ub_F \in \mathbb{R}$ |
| $\boldsymbol{lb_M}$ | Parameter | Concentration deviation lower bound | $lb_M \in \mathbb{R}$ |
| $\boldsymbol{ub_M}$ | Parameter | Concentration deviation upper bound | $ub_M \in \mathbb{R}$ |
| $\boldsymbol{lb_E}$ | Parameter | Catalytic activity deviation lower bound | $lb_E \in \mathbb{R}$ |
| $\boldsymbol{ub_E}$ | Parameter | Catalytic activity deviation upper bound | $ub_E \in \mathbb{R}$ |
| $\boldsymbol{\Delta_r G_{i,est}'^o}$ | Parameter | Estimated standard Gibbs free energy change of reaction $i$ | $\Delta_r G_{i,est}'^o \in \mathbb{R}$ |
| $\boldsymbol{err}$ | Parameter | Associated error on the estimated standard Gibbs free energy change of reaction $i$ | $err \in \mathbb{R}$ |
| $\boldsymbol{R}$ | Parameter | Gas constant | $R = 1.9872 \times 10^{-3} kcal \cdot K^{-1} \cdot mol^{-1}$ |
| $\boldsymbol{T}$ | Parameter | Temperature | $K$ |
| $\boldsymbol{\xi}$ | Parameter | User-defined large constant | $\xi \in \mathbb{N} : \xi > max\{\lvert ub_E \rvert, \lvert lb_E \rvert\}$ |
| $\boldsymbol{\alpha_0}$ | Parameter | Minimum user-defined productivity | $mmol \cdot gDW^{-1} \cdot h^{-1}$ |
| $\boldsymbol{\alpha_1}$ | Parameter | Minimum user-defined yield | $\alpha_1 \in \mathbb{R}$ |
| $\boldsymbol{\alpha_2}$ | Parameter | Minimum user-defined growth rate | $h^{-1}$ |

| | | | |
|---|---|---|---|
| $\boldsymbol{\alpha_3}$ | Parameter | Maximum user-defined $O_2$ requirement | $mmol \cdot gDW^{-1} \cdot h^{-1}$ |
| $\boldsymbol{\beta_0}$ | Parameter | Minimum user-defined productivity | $\beta_0 \in \mathbb{R} \cap [lb_F, ub_F]$ |
| $\boldsymbol{\beta_1}$ | Parameter | Minimum user-defined yield | $\beta_1 \in \mathbb{R}$ |
| $\boldsymbol{\beta_2}$ | Parameter | Minimum user-defined growth rate | $\beta_2 \in \mathbb{R} \cap [lb_F, ub_F]$ |
| $\boldsymbol{\beta_3}$ | Parameter | Maximum user-defined number of interventions | $\beta_3 \in \mathbb{N}$ |
| $\boldsymbol{\beta_4}$ | Parameter | Maximum user-defined $O_2$ requirement | $\beta_4 \in \mathbb{R} \cap [lb_F, ub_F]$ |

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). We provide more details about the NRA formulation in Methods.

**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 metabolic 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.

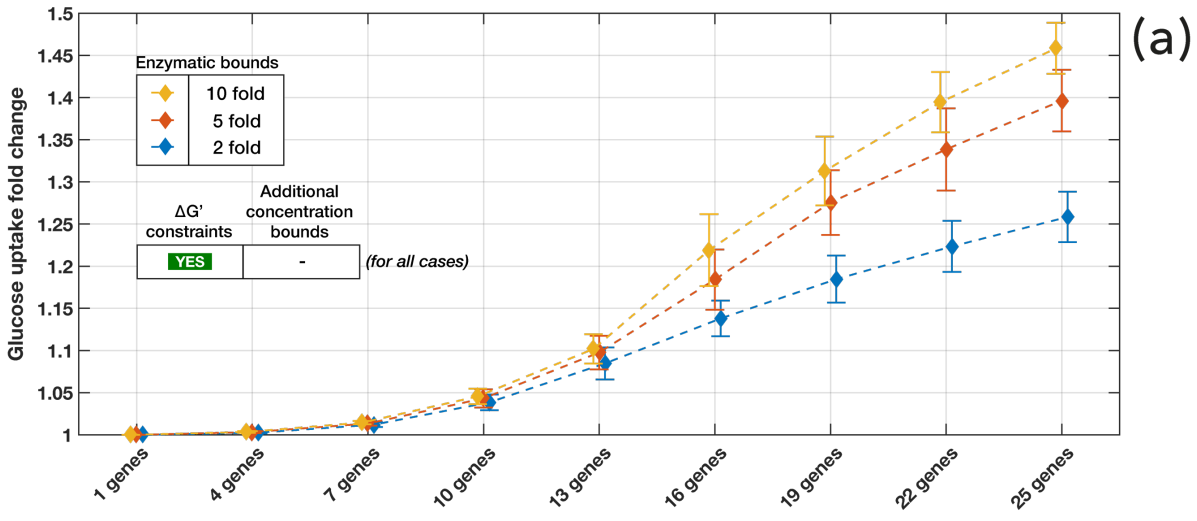
Here, we examined the effects of the imposed physiological and design constraints on the strain design for improved glucose uptake. To this end, we analyzed the achievable glucose uptake rates with 2-fold, 5-fold, and 10-fold maximum allowable deviation of enzyme activities from the reference level for a set of designs ranging from 1 to 25 gene manipulations (Figure 2a). The metabolite concentrations were subject to the thermodynamic feasibility constraints (Methods), and within the predefined physiological ranges (10nM - 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 (Figure 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 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 (Figure 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, 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 Hatzimanikatis, 2015). For this analysis, we have performed several studies by imposing different

198 concentration bounds together with and without thermodynamic constraints (Figure 2b). In
199 general, our results suggest that NRA without thermodynamic constraints tends to
200 overpredict the increase in glucose uptake (Figure 2b), meaning that thermodynamic
201 constraints are limiting factors of strain design. The notable exception was that, starting from
202 19 gene manipulations, the 2-fold constraints on concentrations are more limiting than the
203 thermodynamic ones (Figure 2b blue & orange lines). As expected, our results also show that
204 the tighter the concentration deviation bounds we impose, the less important improvements
205 of glucose uptake could be attained (Figure 2b). For example, the attainable increase of
206 glucose uptake rate with the thermodynamic and additional 2-fold and 10-fold constraints for
207 25 gene manipulations were 17% and 24%, respectively (Figure 2b, green and violet). We also
208 observed that the variance of glucose uptake increase was smaller as the concentration
209 bounds became more constrained. Similarly, we observed a trend that the variance in the
210 studies with the thermodynamic constraints was smaller than in the ones without
211 thermodynamic constraints.




212

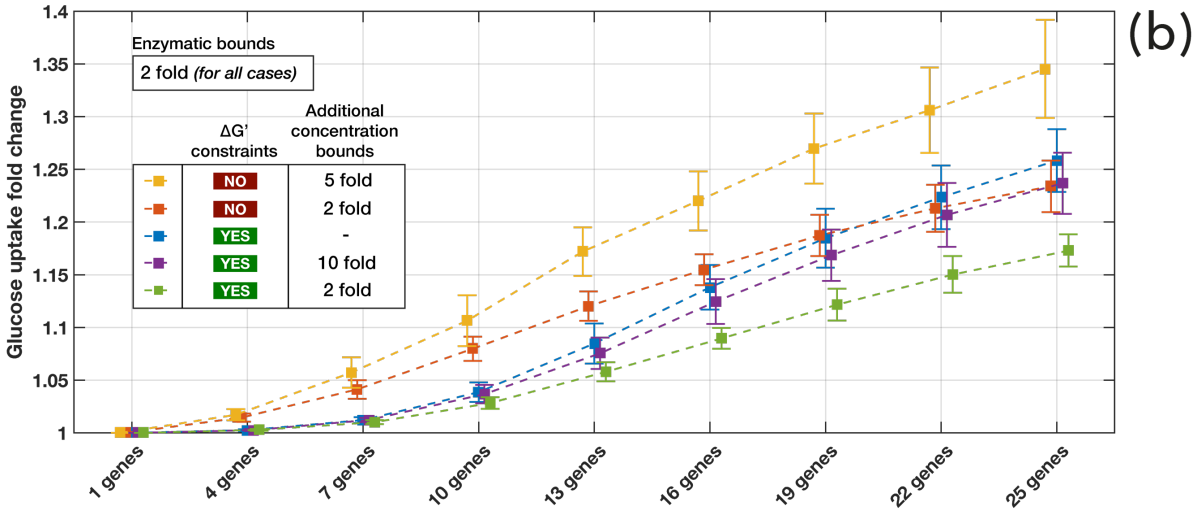



*Figure 2. Effects of the physiological and design constraints on glucose uptake rate for a set of designs with different number of gene manipulations. Effects of: (a) allowed 2-fold (blue), 5-fold (orange), and 10-fold (yellow) changes in enzyme perturbation magnitude, and (b) different imposed metabolite concentration bounds. The study was performed for one reference and 18 extreme sets of CCs selected using PCA (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). Blue lines correspond in both graphs.*

## 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 (Figure 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 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 (Figure 3). For instance, the seven gene manipulations design with ten concentration violations would result in 5.5% increase in glucose uptake.

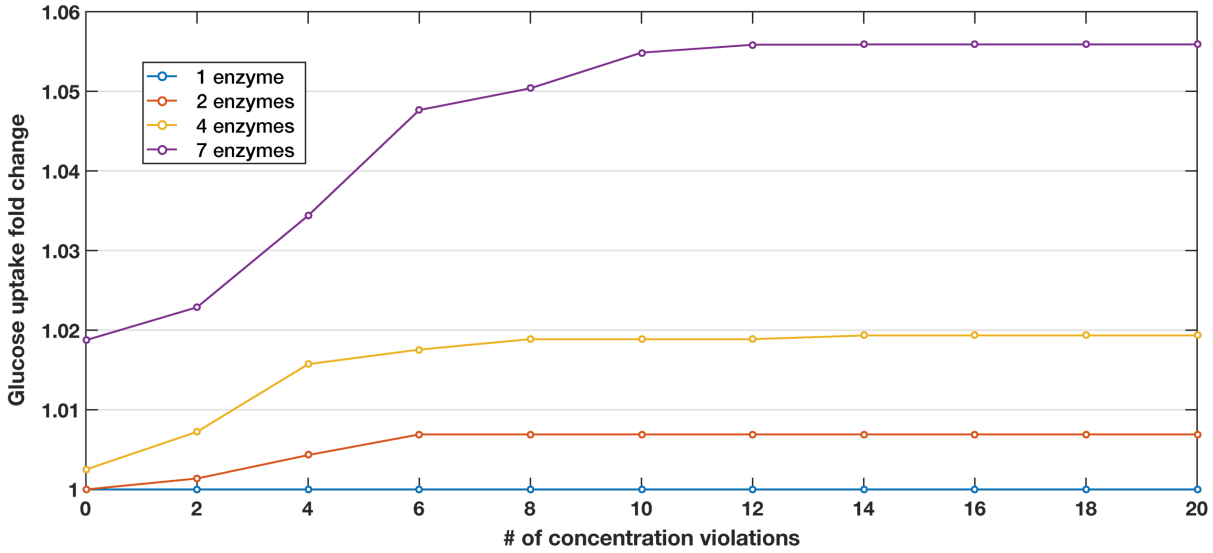



*Figure 3. Maximal attainable increase in glucose uptake as a function of a different number of metabolite concentration violations for one, two, four, and seven gene modifications. The fluxes were allowed to increase 10-fold and decrease to zero, the non-violated concentration bounds were subject to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

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 3a). The three gene manipulation studies involved, in total, seven species with concentrations violating the thermodynamic constraints. Among the seven species, peroxisomal protons appeared in all three studies. Moreover, irrespectively of the study, to achieve a higher glucose uptake, the concentrations of protons (both cytosolic and peroxisomal), 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.

*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 concentrations | # of parameter manipulations | | |
|---|---|---|---|
| | 2 | 4 | 7 |
| AMP (c) ↑ | X | - | X |
| CTP (c) ↓ | - | X | X |
| dCTP (c) ↓ | - | X | - |
| $H^+$ (c) ↑ | X | X | - |
| $H^+$ (p) ↑ | X | X | X |
| L-Glutamine (c) ↓ | - | - | X |
| L-Phenylalanine (c) ↑ | X | - | - |

**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 taking into account 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 (10nM - 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 (Figure 4), and (ii) predicted change in metabolic fluxes upon changes in enzyme activities (Figure 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 (Figure 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 downregulations) 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 (Figure 4). We also observed the knockout of PPS (Phosphoenolpyruvate synthase), which is associated with the conversion of pyruvate to phosphoenolpyruvate.

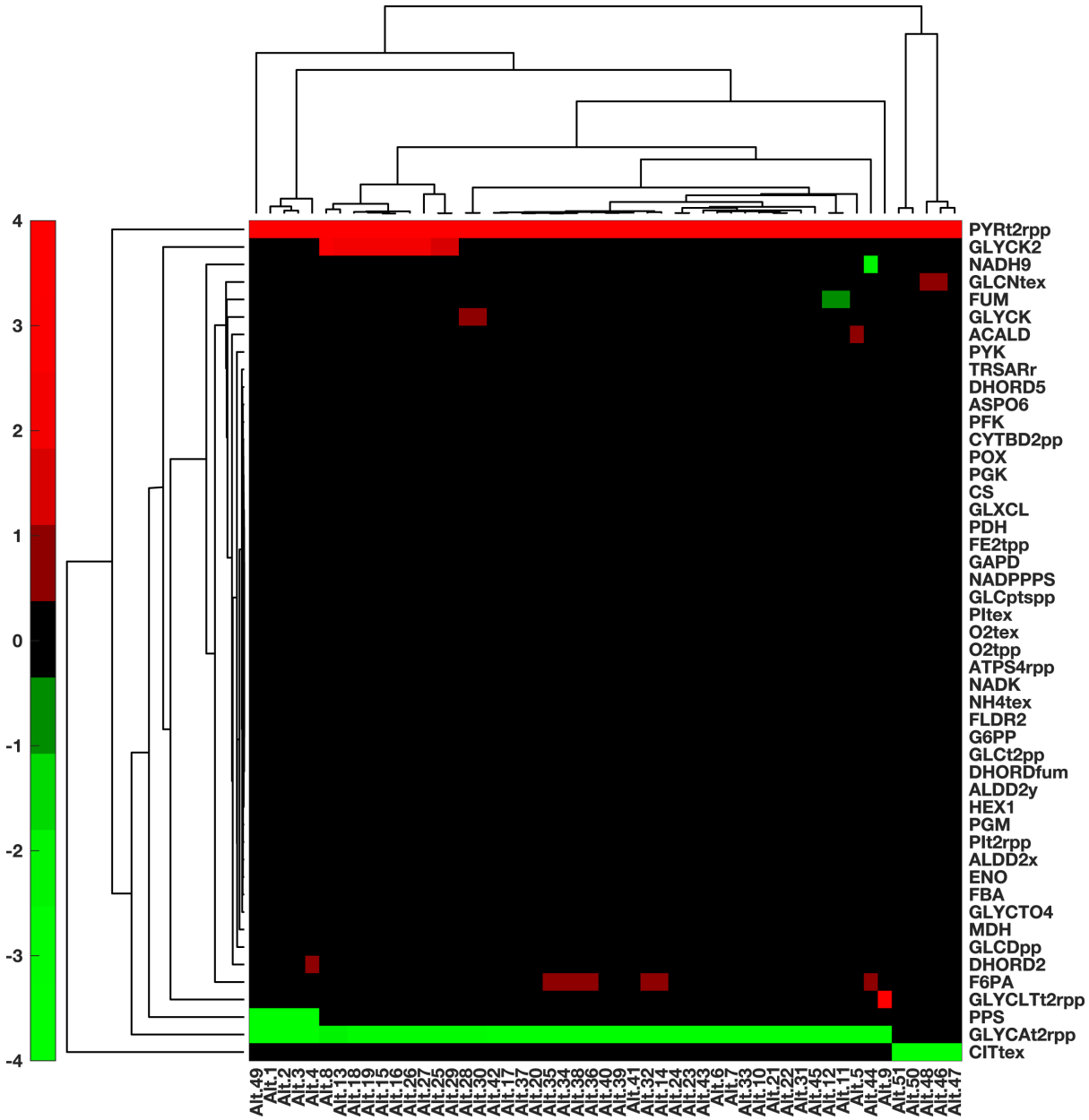



*Figure 4. Hierarchical clustering of the 51 alternative designs for the increase of pyruvate productivity, based on the suggested deregulation of individual enzymes.*

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 (Figures 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 GLYCAt2rpp; and (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); (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.

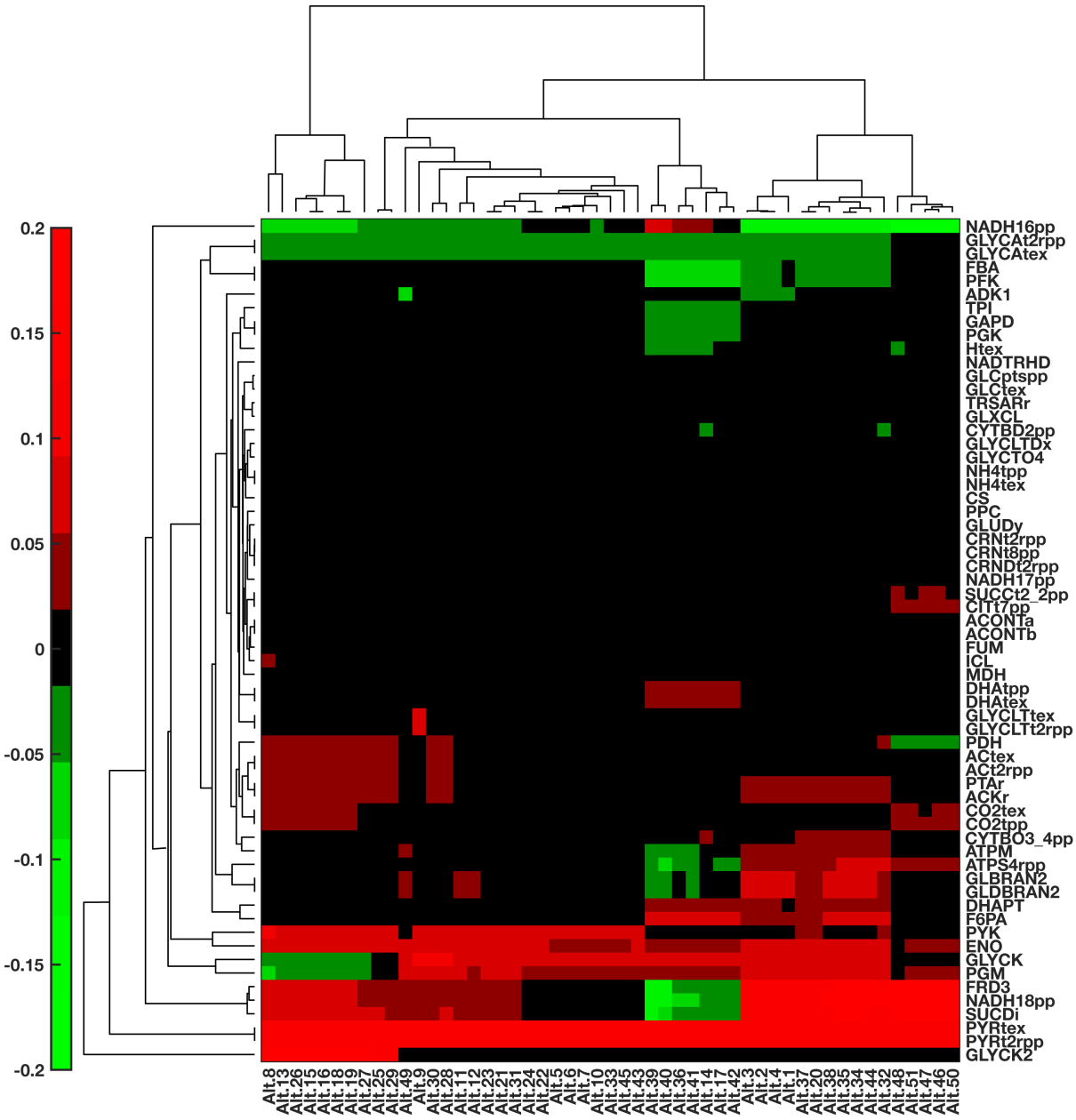



*Figure 5. Hierarchical clustering of the 51 alternative designs for the increase of pyruvate productivity, based on the absolute change in flux value of the 67 most affected reactions in the network.*

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 (Figure 6 and Supplementary Figure 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 CO2 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 (Figures 5, 6 and Supplementary figure 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 (Figure 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 (Figure 6 and Supplementary Figure S3). This was the only group with increased activity of the upper glycolysis. We have also observed a decrease in activity of PDH (Figures 5 and Supplementary Figure 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 (Figure 5, 6 and Supplementary Figure 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 $CO_2$ and acetate secretion pathways.

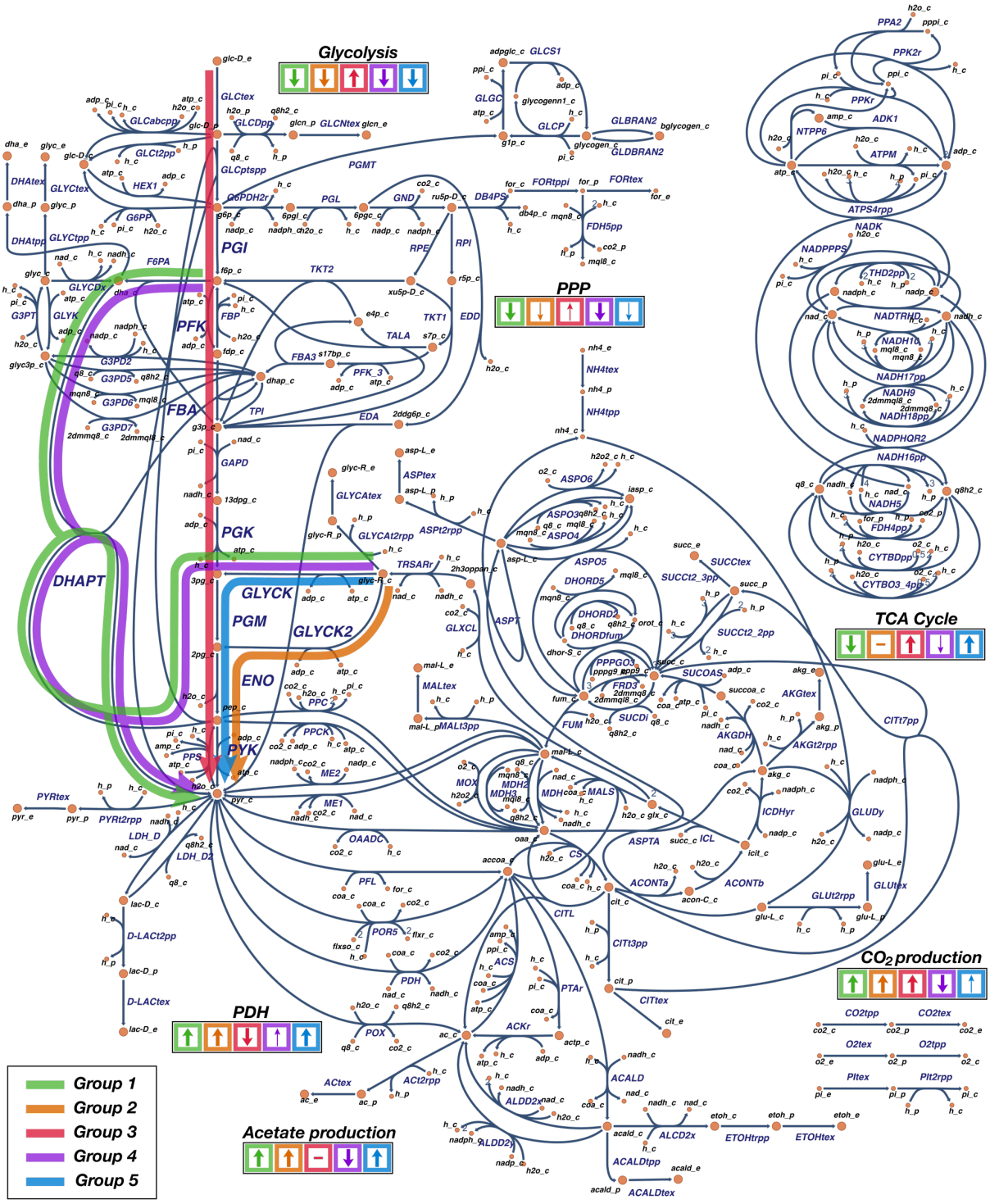



Figure 6. Overview of metabolic engineering strategies devised using NRA for the improved specific production rate of pyruvate while preserving the prespecified yield of pyruvate from glucose. 51 strategies devised with NRA were categorized in 5 distinct groups. The thick arrows on the graph denote the 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 $CO_2$ 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.

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) (Figure 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 Figure 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 (Figure 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.

*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, as well as glycolysis reactions leading to the synthesis of pyruvate (Table 4). 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.

*Table 4. Top ranked parameters based on their control over pyruvate production flux PYRtex. Ranking was computed based on the mean values of 50'000 sets of Control Coefficients.*

| Rank | Parameter Name | Control over PYRtex |
|---|---|---|
| 1 | NH4tex | **negative** |
| 2 | O2tpp | **negative** |
| 3 | NH4tpp | **negative** |
| 4 | ATPS4rpp | **positive** |
| 5 | GLCtex | **positive** |
| 6 | O2tex | **negative** |
| 7 | TPI | **positive** |
| 8 | PGI | **positive** |
| 9 | PFK | **positive** |
| 10 | RPI | **negative** |
| 11 | PItex | **positive** |
| 12 | PGM | **positive** |
| 13 | GLCptspp | **positive** |
| 14 | RPE | **positive** |
| 15 | PYRt2rpp | **positive** |

# Materials and Methods

## 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:

$$C_{p_k}^{v_i} = \frac{d\ lnv_i}{d\ lnp_k} = \frac{p_k\ dv_i}{v_i\ dp_k}$$

$$C_{p_k}^{x_j} = \frac{d\ lnx_j}{d\ lnp_k} = \frac{p_k\ dx_j}{x_j\ dp_k}$$

**Model description and calculation of control coefficients**

The stoichiometric model that was used in this study (Hameri et al., 2019c) was systematically reduced from the *E. coli* iJO1366 genome-scale model (Orth et al., 2011) around the originally defined reaction subsystems 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

446 building block contributions, and moieties were not considered in any study (Supplementary
447 Table S3).

448

449 **Addressing variability in control coefficients**

450 A common issue in MCA and in kinetic modeling is the uncertainty stemming from the scarcity
451 of knowledge concerning the kinetic properties of enzymes (Miskovic and Hatzimanikatis,
452 2011; Miskovic et al., 2015; Miskovic et al., 2019b; Wang et al., 2004). The usual approach in
453 addressing this issue involves the generation of a population of the CCs, and statistical analysis
454 thereof. To form the NRA models, we need to select sets of CCs that will be representative of
455 the generated population.

456 To select a representative set of CCs for our analysis, we took the population of 50'000 sets
457 of FCCs and CCCs computed with ORACLE for the aerobically grown *E. coli* in (Hameri et al.,
458 2019c). We first identified the vector of FCCs that was closest to the mean of the FCC
459 distribution with respect to glucose uptake and selected it as the representative set. Four
460 glucose uptake reactions in the model of *E. coli* exist with GLCptspp being responsible for
461 91.21% of the total flux through these reactions. We enforced this ratio in all performed NRA
462 studies.

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

471

472 **Confidence Intervals and Bonferroni correction**

473 For the computation of confidence intervals in Figure 2, we have used the Bonferroni
474 correction in order to account for the multivariate nature of our study. In univariate studies,
475 to account for the variability in samples, confidence intervals that contain the population

476 mean with the probability 1 – α (typically, α = 5%) are added around each sample mean
477 (Hameri et al., 2019a). However, univariate confidence intervals cannot be used when
478 multivariate problems are studied, instead the Bonferroni's correction of confidence intervals
479 is frequently applied. In Bonferroni's correction, for a problem with *p* variables, to ensure the
480 level 1 – α for all variables simultaneously, we need to choose level 1 – α/*p* for each of
481 individual variables. For instance, if we want to form confidence intervals for 10 variables with
482 an overall 95% confidence level, then we need to use individual 99.5% confidence intervals.

483

484 **Thermodynamic constraints**

485 To integrate thermodynamic constraints, we assumed that reactions operate in the
486 directionality determined by the computed reference steady state. Thus, the concentrations
487 of each metabolite in the respective cellular compartment need to be such as the $\Delta_r G_i'$ of
488 each reaction remains negative. These constraints are written as a function of the standard
489 Gibbs free energy change of the reaction ($\Delta_r G_i'^o$) and the logarithmic concentrations of the
490 participating metabolites, as introduced by (Henry et al., 2007). The $\Delta_r G_i'^o$ of each reaction is
491 computed using the Group Contribution Method (Mavrovouniotis, 1990; Mavrovouniotis,
492 1991). These values are further adjusted to take into account the thermodynamic properties
493 of the relevant cellular compartments; the pH gradient and electrochemical potential for
494 transport reactions, and ionic strength of dissociated metabolites (Henry et al., 2006).

495

496 **Constraints on enzyme activities**

497 Since the activity of an enzyme in the metabolic network could either be increased or
498 decreased, but not both at the same time, we made use of integer variables in the
499 formulation. Therefore, we split the catalytic activity deviations of our system, $E_k$, into the
500 continuous variables $E_k^U$ and $E_k^D$, which denote the upregulation and downregulation of the
501 gene encoding for enzyme *k*, respectively (Eqs. 11-14). As these should not have nonzero
502 values simultaneously, we define the integer binary variables $E_k^{UU}$ and $E_k^{DU}$. $E_k^{UU}$ equals one
503 if the gene catalyzing the enzyme *k* is upregulated and equals zero otherwise. In contrast,
504 $E_k^{DU}$ equals zero in the case of upregulation, and it is one for downregulation. As expressed in
505 Eq. 12, only one of these variables can be active at a time, since deregulation cannot occur in
506 both directions simultaneously, or they can both be inactive for the case of no change in the

507 respective enzyme's catalytic activity. To complete the formulation, these variables are
508 further coupled to the above defined split enzymatic deviation variables through Eqs. 13 and
509 14. The integer binary variable $z_k$ is equal to zero if the activity of enzyme k is modified in the
510 solution, and it equals to one otherwise (Eq. 11). ξ is a constant selected to be larger than the
511 absolute value of the largest enzymatic deviation constraints, $lb_E$ and $ub_E$, defined in Eq. 10.
512

513 **Software and optimization parameters**

514 The computations were made on a Mac Pro workstation running Mac OS X version 10.11.6,
515 equipped with a 2.7 GHz 12-Core Intel Xeon E5 processor and 32GB DDR3 memory, using
516 MATLAB version R2016a and the IBM CPLEX solver version 12.5.1. Time limits for the solver
517 were set as following: in Figure 2(a), for 2-fold (blue line) to 10 minutes, for 5-fold (orange
518 line) to 30 minutes, and for 10-fold (yellow line) to 3 hours; in Figure 2(b), for all cases to 10
519 mins; in Figure 3, for all cases to 30 minutes; in Table 2, for all cases to 30 minutes; in the
520 pyruvate case study (Figures 4-7), for all cases 3 hours.

521

## 522 Conclusions

523 The NRA framework enables the consistent and sophisticated design of metabolic engineering
524 strategies using MCA-based control coefficients. NRA is computationally faster and simpler
525 than other approaches since the derivation of control coefficients does not require the
526 numerical integration of non-linear kinetic models, and offers the implementation of a wide
527 variety of metabolic engineering criteria. To our knowledge, this type of approach has never
528 been applied to large or genome scale kinetic models of metabolism. Using a previously
529 published large-scale kinetic model of *E. coli*, we demonstrated that the NRA formulation can
530 be applied to large-scale metabolic networks. We used the PCA method to select a number
531 of representative sets of kinetic parameters among their population, in order to effectively
532 represent the uncertainty and flexibility of the kinetic model in respect to parametrization.
533 One of the main advantages of NRA is that, being a constraint-based modeling method, it can
534 accommodate the integration of biologically relevant bounds and constraints, which ensure
535 that the proposed strategies are consistent with the entire system capabilities and limitations
536 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. 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).

## Supplementary Material

Table S1: Metabolite concentration violation magnitudes for designs with two, four, and seven gene manipulations.

Table S2: List of the 51 generated alternative designs with the corresponding manipulations and magnitudes of manipulations, pyruvate productivity, and yield.

Table S3: List of aerobically grown *E.coli* model reactions, metabolites, and parameters considered in the study.

Figure S1: Absolute differences of fluxes in the network for the Alternative 1 design (Group 1). Blue/pink arrows and numbers denote an up-/down-regulation of the genes encoding for the respective enzyme and the corresponding fold-change value.

Figure S2: Absolute differences of fluxes in the network for the Alternative 25 design (Group 2). Blue/pink arrows and numbers denote an up-/down-regulation of the genes encoding for the respective enzyme and the corresponding fold-change value.

Figure S3: Absolute differences of fluxes in the network for the Alternative 51 design (Group 3). Blue/pink arrows and numbers denote an up-/down-regulation of the genes encoding for the respective enzyme and the corresponding fold-change value.

Figure S4: Absolute differences of fluxes in the network for the Alternative 40 design (Group 4). Blue/pink arrows and numbers denote an up-/down-regulation of the genes encoding for the respective enzyme and the corresponding fold-change value.

Figure S5: Absolute differences of fluxes in the network for the Alternative 45 design (Group 5). Blue/pink arrows and numbers denote an up-/down-regulation of the genes encoding for the respective enzyme and the corresponding fold-change value.

# References

Akita, H., Nakashima, N., Hoshino, T., 2016. Pyruvate production using engineered Escherichia coli. Amb Express. 6.

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. Metabolic Engineering. 35**,** 148-59.

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. Metabolic engineering. 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**,** 1668-75.

Causey, T. B., Shanmugam, K. T., Yomano, L. P., Ingram, L. O., 2004. Engineering Escherichia coli for efficient conversion of glucose to pyruvate. Proceedings of the National Academy of Sciences of the United States of America. 101**,** 2235-2240.

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. Biotechnology journal. 8**,** 1043-1057.

Costa, R. S., Hartmann, A., Vinga, S., 2016. Kinetic modeling of cell metabolism for microbial production. Journal of Biotechnology. 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. Current Opinion in Biotechnology. 19**,** 389-395.

Gombert, A. K., Nielsen, J., 2000. Mathematical modelling of metabolism. Current Opinion in Biotechnology. 11**,** 180-186.

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. Metabolic Engineering. 52**,** 29-41.

622 Hameri, T., Fengos, G., Hatzimanikatis, V., 2019c. The effects of model complexity and size on
623 metabolic flux distribution and control. Case study in E. coli. bioRxiv.
624 666859.
625 Hatzimanikatis, V., Bailey, J. E., 1996. MCA has more to say. Journal of Theoretical Biology.
626 182**,** 233-242.
627 Hatzimanikatis, V., Floudas, C. A., Bailey, J. E., 1996a. Analysis and design of metabolic
628 reaction networks via mixed-integer linear optimization. Aiche Journal. 42**,** 1277-
629 1292.
630 Hatzimanikatis, V., Floudas, C. A., Bailey, J. E., 1996b. Optimization of regulatory architectures
631 in metabolic reaction networks. Biotechnology and Bioengineering. 52**,** 485-500.
632 Henry, C., Broadbelt, L., Hatzimanikatis, V., 2007. Thermodynamics-Based Metabolic Flux
633 Analysis. Biophysical Journal.
634 Henry, C., Jankowski, M. D., Broadbelt, L. J., Hatzimanikatis, V., 2006. Genome-scale
635 thermodynamic analysis of Escherichia coli metabolism. Biophysical Journal. 90**,** 1453-
636 61.
637 Kacser, H., Burns, J. A., Fell, D. A., 1995. The Control of Flux. Biochem Soc T. 23**,** 341-366.
638 Kalman, D., Colker, C. M., Wilets, I., Roufs, J. B., Antonio, J., 1999. The effects of pyruvate
639 supplementation on body composition in overweight individuals. Nutrition. 15**,** 337-
640 340.
641 Klamt, S., Muller, S., Regensburger, G., Zanghellini, J., 2018. A mathematical framework for
642 yield (vs. rate) optimization in constraint-based modeling and applications in
643 metabolic engineering. Metabolic Engineering. 47**,** 153-169.
644 Li, Y., Chen, J., Lun, S. Y., 2001. Biotechnological production of pyruvic acid. Appl Microbiol
645 Biot. 57**,** 451-459.
646 Long, M. R., Ong, W. K., Reed, J. L., 2015. Computational methods in metabolic engineering
647 for strain design. Current Opinion in Biotechnology. 34**,** 135-141.
648 Maleki, N., Eiteman, M. A., 2017. Recent Progress in the Microbial Production of Pyruvic Acid.
649 Fermentation-Basel. 3.
650 Mavrovouniotis, M. L., 1990. Group Contributions for Estimating Standard Gibbs Energies of
651 Formation of Biochemical-Compounds in Aqueous-Solution. Biotechnology and
652 Bioengineering. 36**,** 1070-1082.
653 Mavrovouniotis, M. L., 1991. Estimation of Standard Gibbs Energy Changes of
654 Biotransformations. The Journal of Biological Chemistry. 266**,** 14440-14445.
655 McCloskey, D., Gangoiti, J. A., King, Z. A., Naviaux, R. K., Barshop, B. A., Palsson, B. O., Feist,
656 A. M., 2014. A model‐driven quantitative metabolomics analysis of aerobic and
657 anaerobic metabolism in E. coli K‐12 MG1655 that is biochemically and
658 thermodynamically consistent. Biotechnology and bioengineering. 111**,** 803-815.
659 Miskovic, L., Alff-Tuomala, S., Soh, K. C., Barth, D., Salusjärvi, L., Pitkänen, J.-P., Ruohonen, L.,
660 Penttilä, M., Hatzimanikatis, V., 2017. A design–build–test cycle using modeling and
661 experiments reveals interdependencies between upper glycolysis and xylose uptake
662 in recombinant S. cerevisiae and improves predictive capabilities of large-scale kinetic
663 models. Biotechnol Biofuels. 10**,** 166.
664 Miskovic, L., Béal, J., Moret, M., Hatzimanikatis, V., 2019a. Uncertainty Reduction in
665 Biochemical Kinetic Models: Enforcing Desired Model Properties. Plos Comput Biol.
666 15**,** e1007242.
667 Miskovic, L., Hatzimanikatis, V., 2010. Production of biofuels and biochemicals: in need of an
668 ORACLE. Trends in biotechnology. 28**,** 391-397.

669 Miskovic, L., Hatzimanikatis, V., 2011. Modeling of uncertainties in biochemical reactions.
670 Biotecnology and Bioengineering. 108**,** 413-23.
671 Miskovic, L., Tokic, M., Fengos, G., Hatzimanikatis, V., 2015. Rites of passage: requirements
672 and standards for building kinetic models of metabolic phenotypes. Current Opinion
673 in Biotechnology. 36**,** 1-8.
674 Miskovic, L., Tokic, M., Savoglidis, G., Hatzimanikatis, V., 2019b. Control Theory Concepts for
675 Modeling Uncertainty in Enzyme Kinetics of Biochemical Networks. Industrial &
676 Engineering Chemistry Research. 58**,** 13544-13554.
677 Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., Palsson, B. O., 2011. A
678 comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011.
679 Molecular Systems Biology. 7.
680 Orth, J. D., Thiele, I., Palsson, B. O., 2010. What is flux balance analysis? Nat Biotechnol. 28**,**
681 245-8.
682 Patil, K. R., Rocha, I., Forster, J., Nielsen, J., 2005. Evolutionary programming as a platform for
683 in silico metabolic engineering. BMC Bioinformatics. 6**,** 308.
684 Reder, C., 1988. Metabolic control theory: a structural approach. Journal of theoretical
685 biology. 135**,** 175-201.
686 Salvy, P., Fengos, G., Ataman, M., Pathier, T., Soh, K. C., Hatzimanikatis, V., 2019. pyTFA and
687 matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux
688 Analysis. Bioinformatics. 35**,** 167-169.
689 Schneider, P., Klamt, S., 2019. Characterizing and ranking computed metabolic engineering
690 strategies. Bioinformatics. 35**,** 3063-3072.
691 Soh, K. S., Miskovic, L., Hatzimanikatis, V., 2012. From network models to network responses:
692 integration of thermodynamic and kinetic properties of yeast genome-scale metabolic
693 networks. FEMS Yeast Research. 12**,** 129-143.
694 Tokic, M., Hatzimanikatis, V., Miskovic, L., 2020. Large-scale kinetic metabolic models of
695 Pseudomonas putida KT2440 for consistent design of metabolic engineering
696 strategies. Biotechnol Biofuels. 13.
697 Varma, A., Boesch, B. W., Palsson, B. O., 1993. Biochemical Production Capabilities of
698 Escherichia-Coli. Biotechnology and Bioengineering. 42**,** 59-73.
699 Wang, L., Birol, I., Hatzimanikatis, V., 2004. Metabolic Control Analysis under Uncertainty:
700 Framework Development and Case Studies. Biophysical Journal. 87**,** 3750-3763.
701 Wang, L., Dash, S., Ng, C. Y., Maranas, C. D., 2017. A review of computational tools for design
702 and reconstruction of metabolic pathways. Synth Syst Biotechno. 2**,** 243-252.
703 Zhang, X. L., Tervo, C. J., Reed, J. L., 2016. Metabolic assessment of E. coli as a Biofactory for
704 commercial products. Metabolic Engineering. 35**,** 64-74.
705 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
706 method for N-acetyl-d-neuraminic acid production using coupled bacterial cells with
707 a safe temperature-induced system. Appl Microbiol Biot. 86**,** 481-489.
708 Zhu, Y. H., Eiteman, M. A., Altman, R., Altman, E., 2008. High Glycolytic Flux Improves Pyruvate
709 Production by a Metabolically Engineered Escherichia coli Strain. Applied and
710 Environmental Microbiology. 74**,** 6649-6655.

711