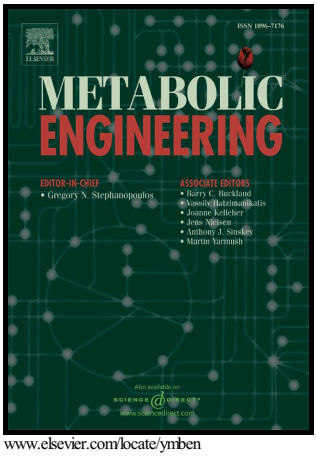

www.elsevier.com/locate/ymben

Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant *E. coli* using large-scale kinetic models

Stefano Andreozzi, Anirikh Chakrabarti, Keng Cher Soh, Anthony Burgard, Tae Hoon Yang, Stephen Van Dien, Ljubisa Miskovic, Vassily Hatzimanikatis

PII: S1096-7176(16)00021-5
DOI: http://dx.doi.org/10.1016/j.ymben.2016.01.009
Reference: YMBEN1087

To appear in: *Metabolic Engineering*

Received date: 26 July 2015
Revised date: 16 December 2015
Accepted date: 29 January 2016

Cite this article as: Stefano Andreozzi, Anirikh Chakrabarti, Keng Cher Soh Anthony Burgard, Tae Hoon Yang, Stephen Van Dien, Ljubisa Miskovic and Vassily Hatzimanikatis, Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant *E. coli* using large scale kinetic models, *Metabolic Engineering* http://dx.doi.org/10.1016/j.ymben.2016.01.009

Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant *E. coli* using large-scale kinetic models

Stefano Andreozzi[1,2,#], Anirikh Chakrabarti[1,2,#], Keng Cher Soh[1,2], Anthony Burgard[3], Tae Hoon Yang[3], Stephen Van Dien[3], Ljubisa Miskovic[1,2] and Vassily Hatzimanikatis[*1,2]

[1]Laboratory of Computational Systems Biotechnology (LCSB), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland

[2]Swiss Institute of Bioinformatics, CH-1015 Switzerland.

[3]Genomatica, Inc., San Diego, California, USA

Running Title: Rational design of 1,4-butanediol with Oracle

To be submitted to: Metabolic Engineering

[#]- these authors have contributed equally

[*]Corresponding author:

Vassily Hatzimanikatis,

Laboratory of Computational Systems Biotechnology (LCSB),

Swiss Federal Institute of Technology (EPFL),

CH-1015 Lausanne, Switzerland

Email: vassily.hatzimanikatis@epfl.ch

Phone: + 41 (0)21 693 98 70 Fax: +41 (0)21 693 98 75

Further instructions: http://www.elsevier.com/journals/metabolic-engineering/1096-7176/guide-for-authors#99000

Abstract

Rational metabolic engineering methods are increasingly employed in designing the commercially viable processes for the production of chemicals relevant to pharmaceutical, biotechnology, and food and beverage industries. With the growing availability of omics data and of methodologies capable to integrate the available data into models, mathematical modeling and computational analysis are becoming important in designing recombinant cellular organisms and optimizing cell performance with respect to desired criteria. In this contribution, we used the computational framework ORACLE (Optimization and Risk Analysis of Complex Living Entities) to analyze the physiology of recombinant *E. coli* producing 1,4-butanediol (BDO) and to identify potential strategies for improved production of BDO. The framework allowed us to integrate data across multiple levels and to construct a population of large-scale kinetic models despite the lack of available information about kinetic properties of every enzyme in the metabolic pathways. We analyzed these models and we found that the enzymes that primarily control the fluxes leading to BDO production are part of central glycolysis, the lower branch of tricarboxylic acid (TCA) cycle and the novel BDO production route. Interestingly, among the enzymes between the glucose uptake and the BDO pathway, the enzymes belonging to the lower branch of TCA cycle have been identified as the most important for improving BDO production and yield. We also quantified the effects of changes of the target enzymes on other intracellular states like energy charge, cofactor levels, redox state, cellular growth, and byproduct formation. Independent earlier experiments on this strain confirmed that the computationally

obtained conclusions are consistent with the experimentally tested designs, and the findings of the present studies can provide guidance for future work on strain improvement. Overall, these studies demonstrate the potential and effectiveness of ORACLE for the accelerated design of microbial cell factories.

Keywords: ORACLE | 1,4-butanediol | Metabolic Control Analysis | Kinetic Models | Thermodynamics | Flux Balance Analysis| Uncertainty.

List of abbreviations

ORACLE – Optimization and Risk Analysis of Complex Living Entities

TFA – Thermodynamics-based Flux Balance Analysis

GEM – GEnome-scale Model

LG – Low Growth phase

1 Introduction

Microbial cell factories are becoming a norm for commercially viable production of chemicals for pharmaceutical, biotechnology, food and beverage industries (Borodina et al., 2015; Chen and Nielsen, 2013; Choi et al., 2015; Jenkins et al., 1998; Lee et al., 2012; Pfleger et al., 2015). However, engineering of microbial cell factories requires a simultaneous optimization of several criteria such as productivity, yield, titer, stress tolerance, all the while retaining the efficient, cost-effective and robust process. Several metabolic engineering strategies capable of integrating available proteomics, transcriptomics, metabolomics, fluxomics, and other types of 'omics' data into a systems design have been developed to meet this kind of specifications (Chen and Nielsen, 2013; Kim et al., 2012; Lewis et al., 2012; Thomas et al., 2007). An essential part of these strategies are *in silico* tools that can help in: (i) improving the

production of the desired chemicals from the natural producers; (ii) identifying enzymes from other organisms capable of performing desired catalytic activity; or even (iii) synthetizing pathways not present in any known organism (Hatzimanikatis et al., 2005; Leonard et al., 2008; Soh and Hatzimanikatis, 2010a; Yim et al., 2011). One of the most prominent examples where a rational metabolic engineering strategy played a key role is the production of 1,4-butanediol (BDO) in *E. coli* (Yim et al., 2011). Therein, as no known organism does naturally produce BDO, the authors have used a variant of BNICE algorithm (Hatzimanikatis et al., 2005; Soh and Hatzimanikatis, 2010a) to discover all pathways from *E. coli* central metabolites to BDO. Based on this information, the authors constructed a new synthetic pathway using enzymes known in other organisms, and then they have used another *in silico* modeling tool, the OptKnock algorithm (Burgard et al., 2003), to optimize so obtained synthetic strain. However, strain optimization for improved specific productivity and yield will require kinetic models.

In this contribution, we used the kinetic modeling framework ORACLE (Optimization and Risk Analysis of Complex Living Entities) (Chakrabarti et al., 2013; Miskovic and Hatzimanikatis, 2011; Soh et al., 2012; Wang et al., 2004; Wang and Hatzimanikatis, 2006a; Wang and Hatzimanikatis, 2006b) to analyze possible enhancements of an *E. coli* strain engineered for production of BDO. ORACLE framework allowed us to integrate thermodynamics and available omics and kinetic data into a population of large-scale kinetic models. Many of these data have been acquired during the early stages of the strain development. The resulting kinetic models were then used to postulate metabolic engineering alternatives ensuring optimal performance with reduced byproduct secretion and fine-tuned redox balance and cofactor levels. More

precisely, we identified three modules that exert control over BDO/glucose yield and over the specific productivity of BDO: (*M1*) central glycolysis; (*M2*) the lower branch of TCA cycle; and (*M3*) the BDO production pathway (Fig. 1). Metabolic engineering strategies involving enzymes from the three modules differed in terms of their effects on cellular growth, byproduct formation, and intracellular states such as cytosolic redox. Our analysis revealed that selection of a strategy involving phosphoenolpyruvate carboxylase (PPC) and citrate synthase (CS) from module *M2* would result in the most significant improvement in BDO/glucose yield. In terms of process conditions, our models predicted that reduction of oxygen levels would lead to increase of BDO/glucose yield, which is consistent with the maximum theoretical yield analysis and with experimental studies.

2 Materials and Methods

2.1 Cultivation conditions

Replicate fermentations were performed with 1 L initial culture volume in 2-L Biostat B+ bioreactors (Sartorius Stedim Biotech) using modified M9 minimal medium (6.78 g/L Na_2HPO_4, 3.0 g/L KH_2PO_4, 1.5 g/L NH_4Cl, 1.0 g/L $(NH_4)_2SO_4$, 0.5 g/L NaCl, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 1 mL 1000x trace metals mixture (Teknova) supplemented with 10 g l^{-1} D-glucose and 0.2 mM IPTG to induce expression of plasmid-borne genes. The temperature was held at 35 °C, and the pH was held at 6.75 using 13.5% NH_4OH. The bioreactors were inoculated to initial ODs of 4.74 and 4.86 and immediately induced with 1.5 mM IPTG (0.5 M). Sparge and agitation were held constant at 0.4 sLPM and 800 rpm enabling a peak oxygen transfer rate of 60 mmol/kg/hr.

Concentrated glucose was fed in to maintain the glucose concentrations in the vessels between 10 and 30 g l^{-1}.

2.2 Bacterial Strain

Host Strain: *E. coli* strain 2731, the production strain for this study, is a derivative of strain 432 described by (Yim et al., 2011). Notable modifications include deletions in *adhE*, *ldhA*, *pflB*, *mdh*, and mutated versions of *gltA*, *lpdA,* and *arcA*. The *arcA* mutation, resulting in an 8 amino acid insertion in the ArcA protein (Iuchi and Lin, 1988; Silverman et al., 1991) renders the regulator largely inactive, and thus activates the oxidative TCA flux under low O2 conditions (Yim et al., 2011). The *gltA* and *lpdA* mutations are the same as described in (Yim et al., 2011). Strain 2731 contains chromosomally integrated copies of *sucD* (encoding CoA-dependant succinate semialdehyde dehydrogenase), *sucA* (encoding alpha-ketoglutarate decarboxylase), *4hbd* (encoding 4-hydroxybutyrate dehydrogenase), *cat2* (4-hydroxybutyryl CoA:acetyl-CoA transferase), and a GBL-hydrolyzing esterase. The sequences of heterologous genes are given in (Yim et al., 2011) and (Pharkya et al., 2014). The strain also contains deletions in *sad*, *gabD*, *aspA*, *poxB*, *ltaE*, *pckA*, *gltBD*, and *ndh*. The strain was transformed with the pZS*-13S-*ald-adh* plasmid as described below to enable conversion of 4-hydroxybutyryl-CoA into 4-hydroxybutyraldehyde and subsequently to 1,4-butanediol (BDO). More details regarding the construction of strain 2731 can be found in (Pharkya et al., 2014).

Plasmid: We used the base vector, pZS*13S, to express the aldehyde dehydrogenase (ALD-5B) and alcohol dehydrogenase genes (Barton et al., 2015) under control of the $P_{A1lacO-1}$ promoter. The vector backbone, pZS*13luc, was obtained from R. Lutz

(Expressys) and is based on the pZ Expression System (Lutz and Bujard, 1997). pZS*13luc contained the luciferase gene as a stuffer fragment. The luciferase stuffer fragment was replaced as described in (Yim et al., 2011). The plasmid contains the pSC101 origin of replication and an ampicillin resistance marker. The analytical tools and cloning methods were described in (Yim et al., 2011).

2.3 Core Reduced Stoichiometric Model

The core reduced stoichiometric model of *E. coli* was derived from the latest *E. coli* reconstruction iJO1366 (Orth et al., 2011)(Supplementary material S1 and S2). After adding the reactions and mass balances corresponding to the BDO production pathway, the resulting core model of recombinant BDO-producing *E. coli* strain is composed by 175 intracellular reactions, including the biomass reaction, and 106 metabolites distributed over the cytosol and the extracellular space (Fig. 1).

2.4 Construction of large-scale kinetic models

We used the well-established ORACLE (Optimization and Risk Analysis of Complex Living Entities) methodology that integrates in a consistent way thermodynamic and physicochemical constraints of the cellular metabolism along with diverse experimental data (fluxomics, metabolomics, transcriptomics, proteomics, and kinetics) into mathematical descriptions of the responses of the cellular metabolism to genetic and environmental perturbations (Chakrabarti et al., 2013; Miskovic and Hatzimanikatis, 2010; Soh et al., 2012; Wang et al., 2004; Wang and Hatzimanikatis, 2006a; Wang and Hatzimanikatis, 2006b). ORACLE allows us to build a population of

large-scale kinetic models that account for the uncertain and scarce information about the kinetic properties of enzymes. Uncertainty and lack of sufficient information in biological systems do not allow identifying a unique set of parameter values, and therefore many alternative sets of parameters can make the resulting kinetic models consistent with the available experimental information. ORACLE overcomes this limitation by "generating" a population of alternative kinetic models, which are consistent with the available information and which differ in the values of the estimated parameters. ORACLE is composed of several computational procedures that we have used in the following sequence (Fig. 2)

2.4.1 Computation of thermodynamically consistent flux profiles

The introduction of Second Law of Thermodynamics in the context of flux balance analysis (FBA) allows coupling the directionality of the fluxes and the levels of metabolite concentrations (Ataman and Hatzimanikatis, 2015; Henry et al., 2007; Kummel et al., 2006; Soh and Hatzimanikatis, 2010c; Soh and Hatzimanikatis, 2014). While we can eliminate infeasible loops in the steady-state metabolic models with no information about thermodynamics (Lewis et al., 2010; Schellenberger et al., 2011a), the Thermodynamics-based Flux Balance Analysis (TFA) additionally allows to eliminate thermodynamically infeasible flux directionalities and to integrate metabolomics data in the constraint-based analyses (Ataman and Hatzimanikatis, 2015; Henry et al., 2006; Henry et al., 2007; Soh and Hatzimanikatis, 2010b; Soh and Hatzimanikatis, 2014; Soh et al., 2012). TFA also allows to integrate information about other factors that can affect the metabolic responses by altering the standard change of Gibbs free energy of reactions such as pH, ionic strength and temperature. Using the core stoichiometric model of *E. coli* metabolism we incorporated the

thermodynamic constraints based on the available information on the Gibbs free energies of reactions (Alberty, 1994; Goldberg et al., 2004; Hadadi et al., 2015; Jankowski et al., 2008) and the available fluxomics and metabolomics data, and we performed TFA analysis to compute a thermodynamically feasible flux profile. Within this thermodynamically feasible flux profile each reaction is unidirectional, therefore its flux solution space is convex. We sampled this flux space, and we performed the Principal Component Analysis (PCA) on the obtained samples to find the representative steady-state flux vector that would characterize the studied physiology (Jolliffe, 2002; Soh et al., 2012).

2.4.2 Sampling of network consistent metabolite concentration levels

The space of metabolite concentrations that is coupled through the reaction directionalities with the computed thermodynamically feasible flux profile is convex. Moreover, the Gibbs free energy, ΔG, of reactions is a linear function of the natural logarithms of metabolite concentrations. This allowed us to use the Artificial-Centering Hit-and-Run sampler from the COBRA Toolbox for the sampling of metabolite concentration levels (Becker et al., 2007; Schellenberger et al., 2011b). The sampled levels of the metabolite concentrations were consistent with values of the Gibbs free energy, ΔG, and with the directionality of the reactions. Available estimates of metabolite concentration ranges acquired from experiments under similar physiological conditions were used as the bounds for the computational sampling of the metabolite levels (Soh et al., 2012). Using the values of the Gibbs free energy we computed the displacement of the enzymatic reactions from the thermodynamic equilibrium, which were consistent with the previously determined metabolite levels and flux profiles.

2.4.3 Estimation of reaction equilibrium displacements

For a uni-uni reaction with a substrate *S* and a product *P*, the displacement from the thermodynamic equilibrium is as follows:

$$\Gamma = \frac{1}{k_{eq}} \frac{P}{S}$$

where k_{eq} denotes the equilibrium constant defined as a ratio between the concentrations of the product and the substrate at the thermodynamic equilibrium. For a reaction which operates in the direction of a net production of the product $\Gamma < 1$, i.e. the Gibbs free energy difference is negative.

We computed the displacements of the reactions from thermodynamic equilibrium corresponding to the sampled metabolite concentrations (see Section 2.4.2), and we classified the reactions according to these displacements in the following sets: reactions that operate (i) strictly far from equilibrium (FE), i.e. with displacements $0 \leq \Gamma < 0.1$; (ii) with the middle displacement (MD), i.e. with displacements $0.1 \leq \Gamma \leq 0.9$; and (iii) strictly near equilibrium (NE), i.e. with displacements $0.9 < \Gamma \leq 1$. Some reactions could belong to more than one of these sets and we classified them as: (iv) FM, i.e. where $0 \leq \Gamma \leq 0.9$; (v) MN, i.e. where $0.1 \leq \Gamma \leq 1$; and (vi) FMN, i.e. these reactions could have any displacement from FE to NE ($0 \leq \Gamma \leq 1$).

2.4.4 Integration of kinetic information and parameterization of models

For each of enzymatic reactions within the metabolic network we assigned a kinetic mechanism using available information from the literature (Heinrich and Schuster, 1996; Reich and Sel'kov, 1981; Segel, 1975; Teusink et al., 2000) where this information was unavailable, we used generalized approximations of enzymatic mechanisms such as generalized reversible Hill (Hofmeyr and Cornish-Bowden, 1997), or convenience kinetics (Liebermeister and Klipp, 2006). The used kinetics

laws included reversible Michaelis-Menten kinetics, Uni-Bi, Bi-Uni, random Bi-Bi, ordered Bi-Bi, Bi-Ter, and Ter-Bi, etc. (Segel, 1975). Detailed information about the kinetic mechanisms used in these studies is available in the Supplementary Material S3.

We then integrated available values of kinetic parameters of enzymes from the literature and databases (Schomburg et al., 2013; Teusink et al., 2000; Wittig et al., 2012). For the enzymes with no prior or incomplete knowledge of K_m values, we sampled the space of K_m values through sampling of the degree of saturation of the enzyme active site (Wang et al., 2004), which is an alternative to sampling of the enzyme states and therefore to sampling of the K_m space (Miskovic and Hatzimanikatis, 2011). We found experimental information about 69 Michaelis constants (Schomburg et al., 2013; Wittig et al., 2012), for 37 out of 153 enzymatic reactions in the model (Supplementary material S5). Within these reactions, complete kinetic information (i.e., K_m values for every substrate and product) was available for 8 reactions (22%), whereas for 16 (43%) of the reactions K_m values for only one metabolite were found in the databases. Whenever a K_m value was available, we biased the distribution of the parameter samples toward this value such that the generated samples of kinetic constants were in the close proximity of the observed K_m.

2.4.5 Stability and consistency verification

We test the local stability of the steady state, and we reject samples for which their Jacobian matrix have positive eigenvalues (Andreozzi et al., 2015; Chakrabarti et al., 2013; Wang et al., 2004). This test is based on the assumption that the observable

flux profile is in a stable steady state at the studied time point. Furthermore, we consider pruning of the samples that are inconsistent with the experimentally measured responses of the metabolic network to gene perturbations and to changes in process conditions.

2.4.6 Computational analysis

We computed control coefficients based on the MCA framework to quantify the responses of the metabolic fluxes and intracellular metabolite concentrations to the change in (i) activities of each enzyme in the network, and (ii) the concentrations of extracellular metabolites (Wang et al., 2004; Wang and Hatzimanikatis, 2006b).

2.4.7 Data mining and identification of metabolic engineering strategies

We analyzed the populations of obtained control coefficients, and we postulated hypotheses about the expected responses of the studied metabolic network to changes in the system parameters. This analysis provided additional information about the couplings between fluxes within the network.

2.5 Control coefficients for multiple enzyme changes

Advancements in metabolic engineering and synthetic biology allowed manipulation of multiple genes simultaneously (Demeke et al., 2013), which emphasized the need to asses *in silico* the overall effect of these interventions on the states of the metabolic network such as product and by-product fluxes, substrate uptake, and redox potential. In order to quantify the responses of these states to alterations of the activities of multiple enzymes, we use the control coefficients for multiple enzyme changes for production of a certain compound as follows:

$$C_E^{V_{P_i}} = \sum_{e_j} \alpha_j C_{e_j}^{V_{P_i}} \quad (1)$$

where $C_{e_j}^{V_{P_i}}$ represents the flux control coefficient of the production rate of P_i with respect to an enzyme activity, e_j. The sum in Equation (1) is over all altered enzyme activities, E. The quantity a_j represents a relative factor of enzyme activity change e_j, which can be used to model different degrees of change of enzyme activity for all altered enzymes. For example, if e_k is increased by 50%, whereas for all other altered enzymes the enzyme activity is increased by 100%, then $a_k = \ln(0.5)$ and $a_j = \ln(2)$ for all other enzymes. In analogous manner, it is possible to define the control coefficient for multiple enzyme changes for the byproduct formation, carbon uptake, and other metabolic functions. The above mentioned control coefficients for multiple enzyme changes are a special case of the general formulation of control coefficients of metabolic functions with respect to a vector of parameters as proposed in (Hatzimanikatis et al., 1996).

2.6 Control coefficients for product selectivity

When there are several byproducts, we need to assess at the same time how metabolic engineering strategies affect the product formation, but also the overall carbon uptake and carbon loss. The product selectivity, i.e. the proportion of carbon flow toward the product P_i with respect to the carbon flow toward all byproducts, is defined as follows

$$S^i = \frac{n_{cp}^i V_{P_i}}{\sum_{j=1}^{n} n_{cp}^j V_{P_j}} \quad (2)$$

We use the general formalism presented in (Hatzimanikatis et al., 1996) to derive from Eq. (2) the product selectivity control coefficient, $C_q^{S_i}$:

$$C_q^{S^i} = \frac{d\ln(S^i)}{d\ln(q)} = C_q^{V_{P_i}} - \sum_{j=1}^{n} S^j C_q^{V_{P_j}} \quad (3)$$

where $C_q^{V_{P_i}}$ represents the flux control coefficient of the flux leading to product P_i with respect to a parameter q. The sum in the abovementioned definition (3) is over all byproducts. The values of this control coefficient will be high if a parameter q has a positive effect on the flux leading to product p_i, but a negative or negligible effect on the fluxes leading to other byproducts. This control coefficient allows us to assess the simultaneous impact of a change in a parameter, such as an enzyme activity, on production of several byproducts or on production of BDO. For example, for the selectivity of a group, G, of *m* products, S^G,

$$S^G = \frac{\sum_{k=1}^{m,m<n} n_{cp}^k V_{P_k}}{\sum_{j=1}^{n} n_{cp}^j V_{P_j}} = \sum_{k=1}^{m,m<n} S^k \quad (4)$$

the corresponding control coefficient is as follows:

$$C_q^{S^G} = \frac{1}{S^G} \sum_{k=1}^{m,m<n} S^k C_q^{V_{P_k}} - \sum_{j=1}^{n} S^j C_q^{V_{P_j}}. \quad (5)$$

2.7 List of enzymes in three identified modules with abbreviations

Module *M1*: glucose-6-phosphate isomerase (PGI), phosphofructokinase (PFK), fructose bisphosphate aldolase (FBA), phospho-glycerate mutase (PGM) and enolase (ENO). Module *M2*: aconitase, half-reaction A (ACONTa), NADP-dependent isocitrate dehydrogenase (ICDHyr), citrate synthase (CS) and phosphoenolpyruvate carboxylase (PPC). Module *M3*: 2-oxoglutarate carboxy-lyase (AKGD), semi-aldehyde dehydrogenase (SUCSALD), NAD-dependent 4-hydroxybutyrate dehydrogenase (4HBD), acetyl-CoA: 4-hydroxybutanoate CoA transferase (4HBCOAT), NADH-

dependent 4-hydroxybutyraldehyde dehydrogenase (4HBCOAR1), and NADPH-dependent 4-hydroxybutyraldehyde reductase (4HBDH2).

The complete list of enzymes and metabolites with the corresponding abbreviations is given in supplementary material S1.

3 Results and Discussion

3.1 Integration of quantitative physiology

The fermentation of BDO producing strain 2731 was performed as described in Materials and Methods. The data were analyzed and specific uptake and production rates were determined over the late growth phase, subsequently referred to as LG phase (i.e. at 24.5 hrs). Rates were determined by the best fit of the analytical and off-gas data within the constraints of carbon and electron balances (Table 1).

Table 1: The specific uptake and production rates higher than 0.002 mmol/gDW-hr at the LG phase.

Reaction	Uptake (mmol/gDW-hr)	Production (mmol/gDW-hr)
Glucose	2.77	-
Oxygen	3.28	-
CO_2	-	6.65
BDO	-	2.12
4HB	-	0.045
GBL	-	0.002
Glutamate	-	0.148
Acetate	0.076	-
Ethanol	-	0.234
Alanin	0.004	-
Lactate	-	0.018

Complete set of the data used for the study (i.e. measured exchange fluxes, extracellular concentrations in the culture medium and the intracellular metabolite

concentration measurements) to further constrain the solution space is provided in the supplementary material S4.

The stoichiometric model was configured with the following assumptions:

(i) The splitting ratio between glycolysis and pentose-phosphate pathway (PPP) was set to be 9:1 in accordance to the low PPP fluxes observed in this strain;

(ii) The model has three glucose transport reactions via the phosphoenolpyruvate-pyruvate phosphotransferase system in the periplasm (GLCptspp), simple diffusion (GLCt2pp), and ATP-dependent active transport (GLCabspp). We assumed that 95% of the glucose was uptaken through GLCptspp as it is the main glucose uptake system under batch conditions (Steinsiek and Bettenbrock, 2012);

(iii) Eleven knockouts were modeled by setting the corresponding fluxes to a minimal value of 10^{-4} mmol/gDW-h; while the gene knockouts could result in zero fluxes through the corresponding reactions, we choose to keep a minimal basal activity. While such basal activities do not impact significantly the main control analysis, it allows us to account for uncertain basal activity and describe potential reintroduction of the enzymes for future studies. The corresponding reactions were: pyruvate formate lyase (PFL), lactate dehydrogenase (LDH_D), malate dehydrogenase (MDH), cytochrome oxidase bo3 involving ubiquinol-8 with 4 protons (CYTBO3_4pp), NADPH-dependent glutamate synthase (GLUSy), Glycine Cleavage System (GLYCL), NADH dehydrogenase involving menaquinone-8 (NADH10), NADH dehydrogenase involving ubiquinone-8 (NADH5), NADH

dehydrogenase involving demethylmenaquinone-8 (NADH9), pyruvate oxidase (POX), and phosphoenolpyruvate carboxykinase (PPCK);

(iv) The minimum bound for the flux of ATP maintenance (ATPM) was set to 3.15 mmol/gDW-hr as it was assumed that the process operates in low-oxygen conditions (Feist et al., 2007);

(v) The reaction catalyzed by fructose 6-phosphate aldolase (F6PA) was assumed to operate in the condensation direction from glyceraldehyde 3-phosphate (g3p) to fructose 6-phosphate (f6p), as suggested in (Schurmann and Sprenger, 2001);

(vi) NADH-dependent flavin adenine dinucleotide reductase (FADRx) and NADPH-dependent flavin adenine dinucleotide reductase (FADRx2) are either simultaneously reducing flavin adenine dinucleotide (fad), or simultaneously oxidizing reduced flavin adenine dinucleotide (fadh2), in order to prevent flux loop cycling;

(vii) The 4-hydroxybutyraldehyde dehydrogenase favors NADPH cofactor (Burk et al., 2011), and therefore the flux through NADH-dependent BDO dehydrogenase (4HBDH1) was set to 10^{-4} mmol/gDW-h.

3.2 Physiology of BDO production: insights from TFA analysis

We used the stoichiometric model with integrated physiology data from Section 3.1 to perform the TFA analysis with the objective of maximizing biomass yield on glucose and we obtained the thermodynamically feasible steady state flux profile and the corresponding representative steady-state flux vector (see Section 2.4.1 and

supplementary material S6). The directionalities in the computed flux profile and the assumed directionalities in the genome-scale iJO1366 model (Orth et al., 2011) differed in three reactions: D-amino acid dehydrogenase (DAAD), NADPH-dependent flavodoxin reductase (FLDR2) and formate transport via proton symport (FORt2pp). Under given physiological conditions, we found that directionality of DAAD was coupled with directionality of FADRx2. Consequently, its directionality was determined by the assumption (vi) above. The network thermodynamics analysis suggested that, under the fermentation conditions, FLDR2 must operate in oxidative direction, which was in contrast to what had been suggested in (Orth et al., 2011). However, it was argued in the literature that Flavodoxin-NADP+ oxidoreductases are catalytically reversible reactions (Jenkins and Waterman, 1998) and therefore we used here the direction that is consistent with the network thermodynamics.

Flux Variability Analysis

We sampled the convex space of the computed flux profile and we analyzed the samples with respect to their variability (Fig. 3). As a measure of the variability of the flux solution space we used the coefficient of variation (CV), defined as ratio of the standard deviation to the mean of population of the analyzed quantity (Chakrabarti et al., 2013; Lequieu et al., 2011). Values of CV close to zero indicate a very tightly constrained flux. In contrast, large values of CV mark relatively weakly constrained flux.

In the studied case, flux variability results indicate: (i) the fluxes that are tightly constrained and therefore precisely determined by the stoichiometry and integrated

data; and (ii) the loosely constrained fluxes, i.e. the fluxes whose values in the reference steady-state flux vector have been estimated from the PCA analysis (see Section 2.4.1, Materials and Methods).

CV of the analyzed flux samples ranged from 0 to 1 implying a relatively well-constrained flux space (Fig. 3). Approximately 32% of the reactions were very tightly constrained (CV < 0.05), and they belonged to different pathways of the network (Fig. 3). For example, 40% of the reactions of the tricarboxylic acid cycle (TCA) were tightly constrained: aconitase: half reaction A (ACONTa), aconitase: half reaction B (ACONTb), NADP-dependent isocitrate dehydrogenase (ICDHyr), malate dehydrogenase (MDH) and citrate synthase (CS) (Fig. 3). In the Glycolysis/Gluconeogenesis pathway, seven of seventeen reactions were also tightly constrained: glyceraldehyde-3-phosphate dehydrogenase (GAPD), enolase (ENO), glucose-6-phosphate isomerase (PGI), phosphoglycerate kinase (PGK), phosphoglycerate mutase (PGM), and triose-phosphate isomerase (TPI).

Network thermodynamics and reaction displacements from thermodynamic equilibrium

We sampled metabolite concentrations that are thermodynamically consistent with the flux directionalities of the computed flux profile, and we compared the results obtained by the thermodynamics of individual reactions to the ones obtained from the network thermodynamics. Specifically, we found that considering only standard Gibbs free energies for 43 reactions (28%) could result in erroneous conclusions about the thermodynamically favorable reaction directionalities. For example, the standard Gibbs free energy corrected for pH and ionic strength of semi-aldehyde

dehydrogenase (SUCSALD) is +9.1 (kcal/mol) and this would imply that this reaction is unfavorable in the direction catalyzing the conversion of Succinyl-CoA to Succinate semialdehyde (Supplementary material S6). However, the concentration profiles of the participating metabolites made this the favorable direction with a Gibbs free energy ranging between -0.01 and -2.12 (kcal/mol).

We computed the displacements of the reactions from thermodynamic equilibrium (Materials and Methods, Section 2.4.3), and we observed that the displacement of a large majority of the reactions was either far from equilibrium, FE, (92 out of 152 reactions, i.e. 60%), or $0 \leq \Gamma < 0.9$, i.e. FM (27 out of 152 reactions, i.e. 18%) (Fig. 1). Out of remaining 34 reactions (22%), 24 reactions (16%) were $0.1 \leq \Gamma < 0.9$, i.e. MD, 4 reactions (3%) were $0.1 \leq \Gamma \leq 1$, i.e. MN, and only 5 reactions (3%) were near equilibrium, NE (Fig. 1). The detailed information about reaction displacements from thermodynamic equilibrium in the network is provided in the supplementary material S6.

When an enzyme of a reaction operates near thermodynamic equilibrium, a manipulation of its enzyme activity will not have a considerable impact on the state of the metabolic network (Heinrich and Schuster, 1996; Henry et al., 2007; Wang and Hatzimanikatis, 2006b). This implies that, only 9 reactions that were near equilibrium and MN might not have significant control over metabolic fluxes and metabolite concentrations in the network, while for the remaining 143 enzymes (94%) their contribution to the control over fluxes and metabolites will depend on their kinetics and saturation state.

3.3 Improving BDO production: insights from MCA

The primary aim of this study was to identify metabolic engineering targets for increasing BDO production and yield with respect to the glucose. We also investigated how the internal state of the cell characterized by NADH/NAD+ and NADPH/NADP+ ratio affect the choice of the metabolic engineering strategies. These ratios are important as many reactions in the network, particularly in the BDO production pathway, produce and consume redox potential and they also depend on the process conditions and physiology, such as oxygen availability and growth rates.

We used ORACLE (Chakrabarti et al., 2013; Miskovic and Hatzimanikatis, 2010; Miskovic et al., 2015; Soh et al., 2012; Wang et al., 2004; Wang and Hatzimanikatis, 2006a; Wang and Hatzimanikatis, 2006b) to perform a metabolic control analysis (MCA) around the representative steady-state flux vector at the LG phase (see Section 2.4.1). We generated a population of more than 370000 models, we rejected the ones that did not pass the stability test (see Section 2.4.5) and we obtained a final population of 238000 stable models, i.e., approximately 64% of the generated models were stable.

3.3.1 Strategies for improving BDO specific productivity and BDO/glucose yield

Three Modules of Enzymes

We identified three important modules, i.e. sets of enzymes, which had significant control over specific BDO productivity, BDO/glucose yield and other important cell outputs such as redox potential and process parameters such as oxygen availability. Module 1 (*M1*) contained enzymes from the central glycolysis pathway: PGI, PFK,

FBA, PGM and ENO. Module 2 (*M2*) was comprised of reactions from the TCA cycle: ACONTa, ICDHyr, CS and PPC, while Module 3 (*M3*) contained the reactions from the BDO product pathway: AKGD, SUCSALD, 4HBD, 4HBCOAT, 4HBCOAR1, and 4HBDH2 (Figure 2).

BDO Specific Productivity

Analysis of the control coefficients for BDO specific productivity (Materials and Methods) suggested several strategies for improving this output (Fig. 4, Panel B). For the enzymes in *M1*, we observed that two-fold (100%) increase in the enzyme activities of PFK or PGI would lead to 16-21% increase in BDO production. For the enzymes in *M2*, the computed flux control coefficients suggested that 100% increase of the activity of ACONTa, CS, ICDHyr and PPC would lead to approximately 2-6% of increase of specific BDO productivity (Fig. 4, panel B). Thus, ORACLE predicts that increase in BDO production could be achieved by increasing carbon flow from glycolysis towards the TCA and channeling it through the lower part of the TCA cycle. Increase in specific BDO productivity could also be achieved by increasing activity of the enzymes in the BDO producing pathway (*M3*): 4HBD, 4HBCOAT, 4HBDH2, AKGD and SUCSALD. The computed flux control coefficients suggested that 100% increase of the enzyme activity of 4HBDH2 would lead to 4% increase in specific BDO productivity (Fig. 4, panel B). Overall, it appears that for this engineered strain significant flux control has been shifted towards the upper part of the glycolysis and the lower branch of TCA cycle.

The large positive control coefficients of BDO specific productivity with respect to ATP synthase (ATPS4rpp), $C_{ATPS4rpp}^{14BDOt}$, and the large negative values of C_{ATPM}^{14BDOt} (Fig. 4, panel B) indicate that ATP availability has a large positive effect on BDO production. Specifically, 100% increase in ATPS4rpp and ATM would lead to 48.5% increase and 20% decrease, respectively, of the BDO specific productivity.

Yield

Even though all above-mentioned strategies suggested a change in BDO specific productivity, application of some of these strategies would result in mostly unchanged or even slightly reduced BDO/glucose yield (Fig. 4, panel A). Therefore, when engineering *E. coli* for improved BDO specific productivity one has to consider at the same time how alterations of enzymes affect the glucose uptake rate and the carbon loss through byproducts and CO_2.

For the enzymes in *M1*, a two-fold increase of activities of FBA, PFK and PGI would entail 0.4 to 6% reduction in the BDO/glucose yield. Indeed, a two-fold increase of activities of these enzymes would produce a higher positive effect on glucose uptake (7-21% increase) than on BDO specific productivity (6-19% increase) (Fig. 4, panels B and C). One possible cause for this could be that an increase in the overall flux through glucose uptake caused by increased activities of these enzymes results in overflow phenomena towards byproducts as it was also captured by the selectivity control coefficients (Fig. 5). This further demonstrates how kinetic modeling can be used to identify system responses and help us derive such testable hypotheses. In contrast, increase of ENO and PGM activities by 100% would increase BDO/glucose yield by less than 1%. As a matter of fact, the positive effects of 100% increase of

ENO and PGM on specific BDO productivity and glucose uptake would be comparable, i.e. this manipulation would entail 6.5-8.7% of increase of BDO specific productivity and 5.9-7.9% of increase in glucose uptake (Fig. 4, panels B and C).

A 100% increase in activities of the enzymes from *M2* had a consistently positive effect on BDO/glucose yield ranging from 2 to 12.5% of increase (Fig. 4 panel A). This effect is primarily due to a positive effect of these enzymes on the specific BDO productivity and almost no effect on the glucose uptake (Fig. 4, panels B and C). An exception is PPC whose 100% increase of activity would reduce the glucose uptake by 9.5% (Fig 4, panel C). This coupled with 3% increase in BDO specific productivity would eventually result in 12.5% increase in the BDO/glucose yield (Fig 4, panels A and B). These findings are in very good agreement with the experiments performed by Pharkya and coworkers where it was similarly found that increasing PPC and CS activity led to improved BDO specific productivity (Pharkya et al., 2014). Furthermore, as predicted by ORACLE, these interventions also reduced ethanol production by channeling more acetyl-CoA into the TCA cycle and lowering acetyl-CoA concentration (Pharkya et al., 2014). However, it must be noted that expression of these enzymes must be precisely tuned because acetyl-CoA is also a co-substrate in the BDO production pathway.

Increasing the activity of enzymes from M3 would consistently increase the BDO/glucose yield. These enzymes had almost no control over the glucose uptake and the positive impact of these enzymes on the BDO/glucose yield is due to their positive effect on the specific BDO productivity, i.e. we observed 0.8-5% of increase of this yield for a two-fold increase of activity of *M3* enzymes (Fig. 4, panels A-C).

Selectivity

Analysis of the fermentation data pointed out that at 24.5 hrs approximately 49% of carbon uptake was channeled into byproducts (Supplementary Figure 7). The vast majority of the carbon, i.e. 40%, was lost as CO_2. The remaining 9% of the lost carbon went to the other byproducts such as L-glutamate (4%), ethanol (3%) and biomass (1%**)**. We further observed that 4-hydroxybutyrate (4HB) and gamma-butyrolactone (GBL) were secreted from the cell at this time point. At the maximum theoretical yield, 2 moles of CO_2 are lost per 1 mole of glucose (33%). The amount of carbon lost as CO_2 would actually be higher when there is cell growth, because CO_2 is produced in biomass synthesis as well. However, the strain *operated* away from the maximum theoretical yield and therefore we observed excess CO_2 production, relatively to the theoretical minimum. 13C-flux analysis revealed that nearly all of excess CO_2 is either produced from the PPP or by diversion of the carbon flow at SuccCoA through SUCOAS which increased the flow around the TCA cycle (Yim et al., 2011). While the PPP provides NADPH (see Section 3.3.2), SUCOAS produces ATP via substrate-level phosphorylation in addition to ATP produced through ETC from the electrons produced via succinate dehydrogenase. Due to this extra ATP generated, less ATP must be produced via other means (e.g., from NADH via oxidative phosphorylation) when there is more flux through the complete TCA cycle. Thus the excess CO_2 may be directly related to the redox or energy needs of the cell.

We considered the impact of enzyme alterations in modules *M1*, *M2* and *M3* on the distribution of carbon towards BDO, CO_2 and other byproducts: glutamate, ethanol, pyruvate, lactate and biomass. To quantify this impact we computed control coefficients of selectivity of: (i) BDO production, C_q^{14BDOt} (see Eq. 3, Materials and Methods); these coefficients quantify the impact of enzyme alterations on the

proportion of the carbon flow toward BDO with respect to the carbon flow toward all byproducts (Materials and methods); (ii) CO_2 production, C_q^{CO2tpp}; these coefficients quantify the impact of enzyme alterations on the proportion of the carbon flow toward CO_2 with respect to the carbon flow toward all byproducts; and (iii) other byproducts, C_q^{SG} (see Eq. 5, Materials and Methods); these coefficients quantify the impact of enzyme alterations on the proportion of the carbon flow toward glutamate, ethanol, pyruvate, lactate and biomass with respect to the carbon flow toward all byproducts.

The performed sensitivity analysis indicated that an increase of activity of enzymes from *M1* would result in more carbon flow diverted from CO_2 toward glutamate, ethanol, pyruvate, lactate and biomass. Indeed, 100% increase of activity of enzymes from *M1* would entail 0.7-30.6% increase of the selectivity of the other byproducts, whereas the same alteration of these activities would result in 1.2-4.1% decrease of the selectivity of CO_2 (Figure 5; Supplementary Material S11 and Figure 12). Moreover, 100% increase of activity of FBA, PFK and PGI would decrease the selectivity of BDO by 0.1-2.2%, whereas 100% increase of activity of ENO and PGM would increase the selectivity of BDO by 0.8-1%.

An increase of activity of enzymes from *M2* would improve the selectivity of BDO and decrease the selectivity of other byproducts. As a matter of fact, 100% increase in activities of ACONTa, CS, ICDHyr or PPC would increase the selectivity of BDO by 2.2-11.3% and would decrease the selectivity of other byproducts by 2.4-78.5% (Fig. 5). Moreover, 100% increase in activities of ACONTa, and ICDHyr would entail 0.1-1.9% of decrease of the selectivity of CO_2, whereas 100% increase in activities of CS and PPC would increase of the selectivity of CO_2 by 0.2-2.5%. In this group of

enzymes PPC had the largest impact on the selectivities, i.e. a two-fold increase of PPC's activity would cause 11.3% and 2.5% increase in the selectivities of BDO and CO_2, respectively, and 78.5% decrease in the selectivity of other byproducts.

A metabolic engineering strategy for the improved BDO production involving enzymes from module *M3* would also consistently result in reduction of the carbon loss to other byproducts and in improvement of the BDO selectivity, whereas the selectivity of CO_2 would be mostly unchanged (Fig. 5). Specifically, 100% increase in the activities of AKGD, 4HBD, 4HBCOAT, 4HBCOAR1, 4HBCOAR2, SUCSALD or 4HBDH2 would result in reduction of the carbon loss to other byproducts by 4.8-20.7%, whereas the selectivity of BDO would be improved by 0.9-3.7% (Fig. 5).

Comparison of the metabolic engineering strategies involving alteration of the enzyme activities within *M1*, *M2* and *M3* suggested that strategies involving enzymes from *M2* and *M3* could be used for improvement of both the specific BDO productivity and BDO/glucose yield. In contrast, any strategy involving enzymes from *M1* would increase the specific BDO productivity but BDO\glucose yield would be reduced or unaltered due to increased carbon loss to other byproducts. Interestingly, the strategies involving *M3* enzymes are consistent with the experimental results. Of the improvements made between the strain studied in (Yim et al., 2011) and the strain used in current analysis, the most significant increases in BDO productivity and yield were made by improving the activity of the pathway enzymes, and optimizing the relative expression levels. These results and the analysis here suggest that the enzymes in *M1* could be the next targets for improved BDO productivity and yield.

Reaction displacements from thermodynamic equilibrium and control coefficients

As discussed in Section 3.2 the enzymes of reactions that operate near thermodynamic equilibrium have no control over the fluxes and metabolite concentrations in metabolic networks. However, we do not know in advance how far from equilibrium the operational state of an enzyme should be so that this enzyme has a control over the network. To answer this question we analyzed the correlation of the computed control coefficients for specific BDO productivity with the displacement from thermodynamic equilibrium of the corresponding enzymes. We observed that only the enzymes with the displacement greater than a limit value of 0.5 had control over specific BDO productivity (Fig. 6). In this group of enzymes, the ones closer to the limit of 0.5 (less than 0.47) belonged primarily to modules *M3* (4HBD, 4HBCOAR1 and 4HBCOAT) and *M1* (ENO and PGM)(Fig. 6), and indeed they had reduced control over the network fluxes.

3.3.2 Redox potential considerations

Cellular redox potential, i.e. NADH/NAD^+ and NADPH/NADP ratios, are important markers when considering strategies for metabolic engineering of an organism (Moreira dos Santos et al., 2003; Pitkanen et al., 2003; Thomas et al., 2007). In particular, increased redox potential is considered as an important factor in improving BDO production in (Zhuang et al., 2013).

We observed that most of perturbations in the metabolic network leading to increase in the NADH/NAD+ ratio lead to increase in the NADPH/NAPD+ ratio as well (Fig. 7). We hypothesized that this positive correlation could be due to strong activities of NAD transhydrogenase (NADTRHD) and NAD(P) transhydrogenase in

periplasm (THD2pp) as they balance conversion between NADH/NAD+ and NADPH/NAPD+ pairs. Increase of activities of *M1* enzymes would lead to increase in both the NADH/NAD+ ratio, and in the increase of the NADPH/NAPD+ ratio (Fig. 7). For example, 100% increase of enzyme activities of PFK, PGI, and PGM would result in increase by 25-50% of the NADH/NAD+ ratio, and in increase by 25-45% of the NADPH/NAPD+ ratio.

In contrast, 100% increase of activity of the enzymes from *M3* would lead to decrease of the redox potentials within the cell by 1-20%. The exceptions are 4HBCOAR2 and 4HBt, as 100% increase of activity of these enzymes would result in respectively 1% and 4% of increase of the NADH/NAD+ ratio (Supplementary Figure 9).

The enzymes from *M2* have a strong control over the redox potentials. The redox potentials would increase by: (i) ~15-17% for 100% increase of CS activity; (ii) ~20-45% for 100% decrease of PPC activity; and (iii) ~14-20% for 100% decrease of ICDHyr activity. Overall, these results correlate well with the role of these reactions in redox balancing. The reactions in *M1* and *M2* have a net production of NAD(P)H, whereas *M3* consumes this NAD(P)H to balance reducing equivalents. Increasing PPC activity could also lead to increased flux through the reductive TCA cycle, which would consume reducing equivalents; therefore it is negatively correlated with redox potential. Zhuang *et al.* predicted that increased activity of enzymes in the BDO production pathway would give higher BDO production, but it would be at the expense of the lower growth rate due to NADPH depletion (Zhuang et al., 2013). Increased consumption of NAD(P)H must be balanced with increased production elsewhere in metabolism. In the production strains it appears that the amount of

NADPH used in the product pathway by 4HBDH2 is equal to the amount produced by the ICDHyr upstream. However, additional NADPH is needed for growth, and in vitro analysis indicated some use of NADPH by the 4HBCOAR2 (Pharkya et al., 2014) . Any additional NAPDH required for growth and 4HB-CoA reductase should be produced by the PPP or by the excess flux in ICDHyr. The latter case could result in excess byproducts around the product pathway such as GBL, 4HB, and CO_2, whereas in the former case CO_2 is the main byproduct. An alternative source of NADPH is from NADH through NADTRHD (transhysdrogenase). However, native expression of the membrane-bound transhydrogenase was not sufficient to fill this gap. Indeed, deletion of *zwf* to channel all carbon through glycolysis resulted in reduction of BDO productivity, which could be restored by overexpression of *pntAB*, encoding the membrane-bound transhydrogenase (unpublished results). These observations suggest that PPP is an essential source for NADPH production for growth.

3.3.3 Oxygenation level considerations

The analyzed strains were cultivated in microaerobic conditions with measured oxygen uptake rate of 3.13 mmol/gDW-hr at the analyzed time point. Analysis of control coefficients for BDO production suggested that 50% decrease in oxygen uptake could lead to 2% increase in BDO/glucose yield (Fig. 4), and 160% decrease reduction in biomass (Supplementary Figure 8). This prediction is in a qualitative agreement with experiments performed with the current BDO production strains, where reduction of specific oxygen uptake rates by 50% led to approximately 5% improvement in BDO/glucose yield at comparable specific productivity. Stoichiometric BDO production is not possible under strict anaerobic conditions

because one mole NAD(P)H excess is created per BDO produced. This must be oxidized to generate ATP, which is needed for cell growth and maintenance. In addition, it has been demonstrated that simultaneous deletion of adhE, ldhA, and pflB renders the cell unable to grow anaerobically even though the proper distribution of fermentation products should enable redox balance and excess ATP generation (Mat-Jan et al., 1989; Stols and Donnelly, 1997). Further kinetic studies could explain how to fine-tune the redox potential and ATP generation.

4 Conclusions

We developed here a population of large-scale kinetic models of recombinant BDO producing *E. coli* that is consistent with the studied physiology. We used these models to identify potential targets for improved BDO specific productivity and BDO/glucose yield, and we found out that 20 key enzymes control these two outputs and they can be grouped in three modules. Interestingly, the performed flux variability analysis revealed that the fluxes of 13 out of these 20 key enzymes were tightly constrained. This is in consistency with the previous results and discussions that the genes with a tight flux range are subject to regulation (Bilu et al., 2006).

Consistent with earlier experimental studies, our models indicated that increasing activity of phosphoenolpyruvate carboxylase (PPC), citrate synthase (CS), and the enzymes in BDO production pathway would lead to most significant improvements in BDO production and yield and would also lead to reduction in carbon loss to other byproducts and reduction of the redox potential.

Though the analysis of computed control coefficients suggested that increasing flow of carbon into central glycolysis and away from PPP could increase the specific

productivity of BDO, it would also lead to increase in byproduct formation (more carbon loss) and increase in the intracellular redox potential. To avoid this, metabolic engineering strategies will require a fine-tuning of the carbon flows in the network, and we demonstrate here that kinetic models are excellent tools for performing such a task.

The present study points how we can combine multiple targets to fine-tune several objectives simultaneously. Future studies will benefit from using the models presented here and the multi-parametric, multi-objective optimization framework introduced by Hatzimanikatis et al. to improve the cellular performance (Hatzimanikatis et al., 1996).

Acknowledgements

S.A. was supported by the Swiss National Science Foundation. A.C. was supported through the BattleX RTD project, and K.C.S was supported through the MetaNetX RTD project both within SystemsX.ch, the Swiss Initiative for Systems Biology funded by the Swiss National Science Foundation. L.M and V.H. were supported by funding from the Ecole Polytechnique Fédérale de Lausanne (EPFL) and SystemsX.ch. We thank the Genomatica fermentation personnel for conducting the experiments.

Figures with captions

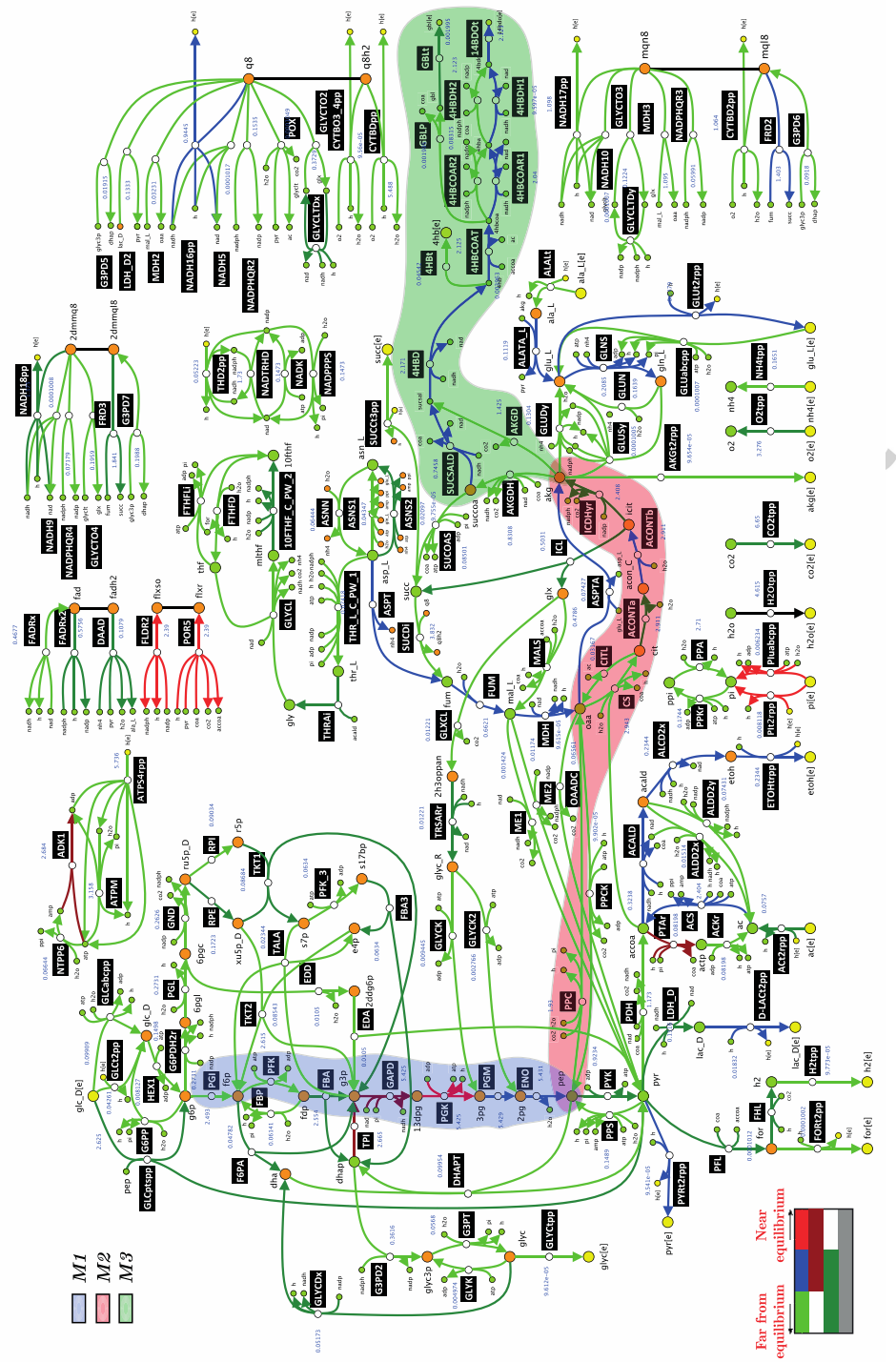

Figure 1: BDO producing *E. coli* metabolic network. Color-coding of the reactions denotes the distance from the thermodynamic equilibrium of their enzymes. The control over BDO production is centered around: Module 1 (*M1*) - focusing around PFK, FBA, GAPD, PGK and PGI; Module 2 (*M2*) - focusing around PPC, CS, ACONTa; Module 3 (*M3*) - focusing around AKGD, 4HBCOAT and 4HBDH2.

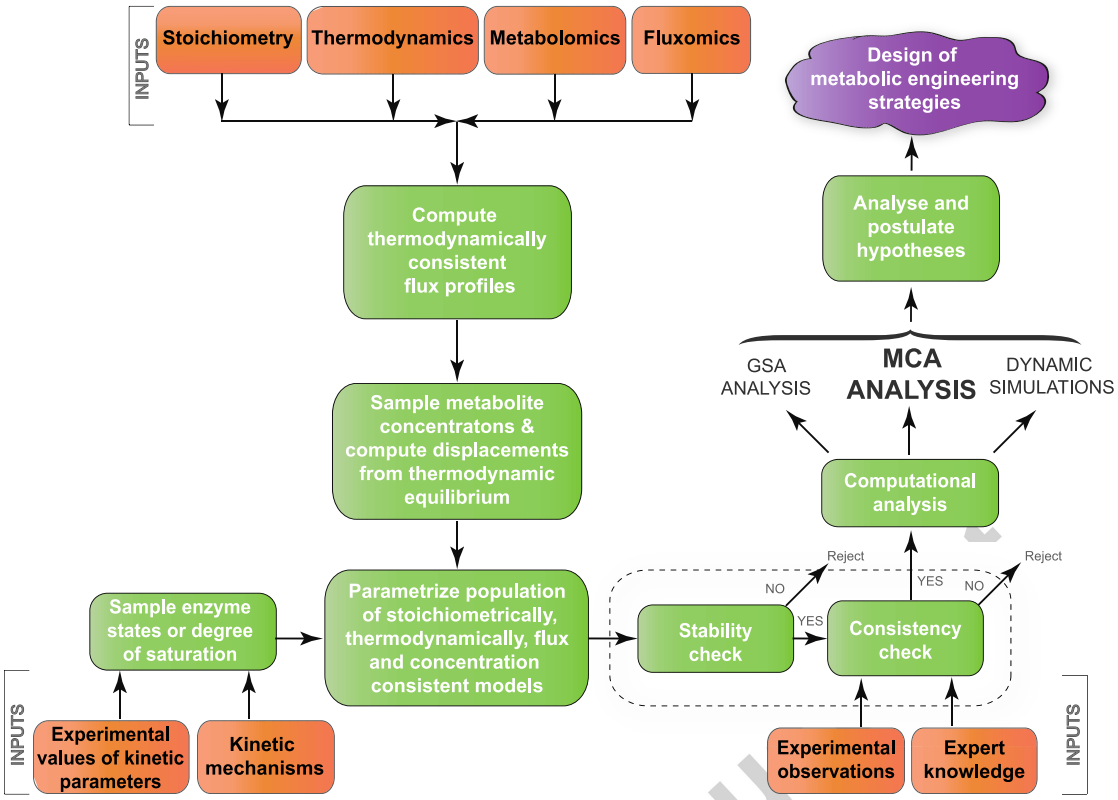

Figure 2: Workflow of the computational procedure for uncertainty analysis of metabolic networks within the ORACLE framework. The successive application of computational procedures integrates biological information from different levels and sources, thus refining kinetic models and providing guidance for metabolic engineering.

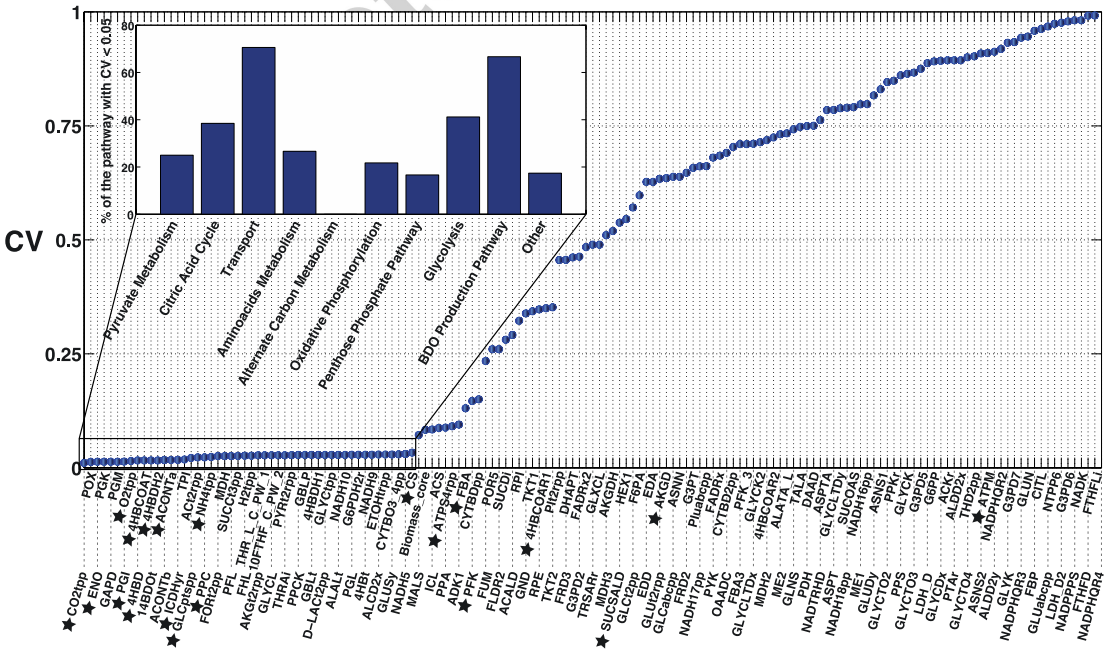

Figure 3: Coefficient of Variation (CV) of the flux estimates for the participating reactions. INSET – Percentage of the specific metabolic pathways that are tightly constrained (CV < 0.05) in the system. The reactions marked by stars correspond to the identified key enzymes.

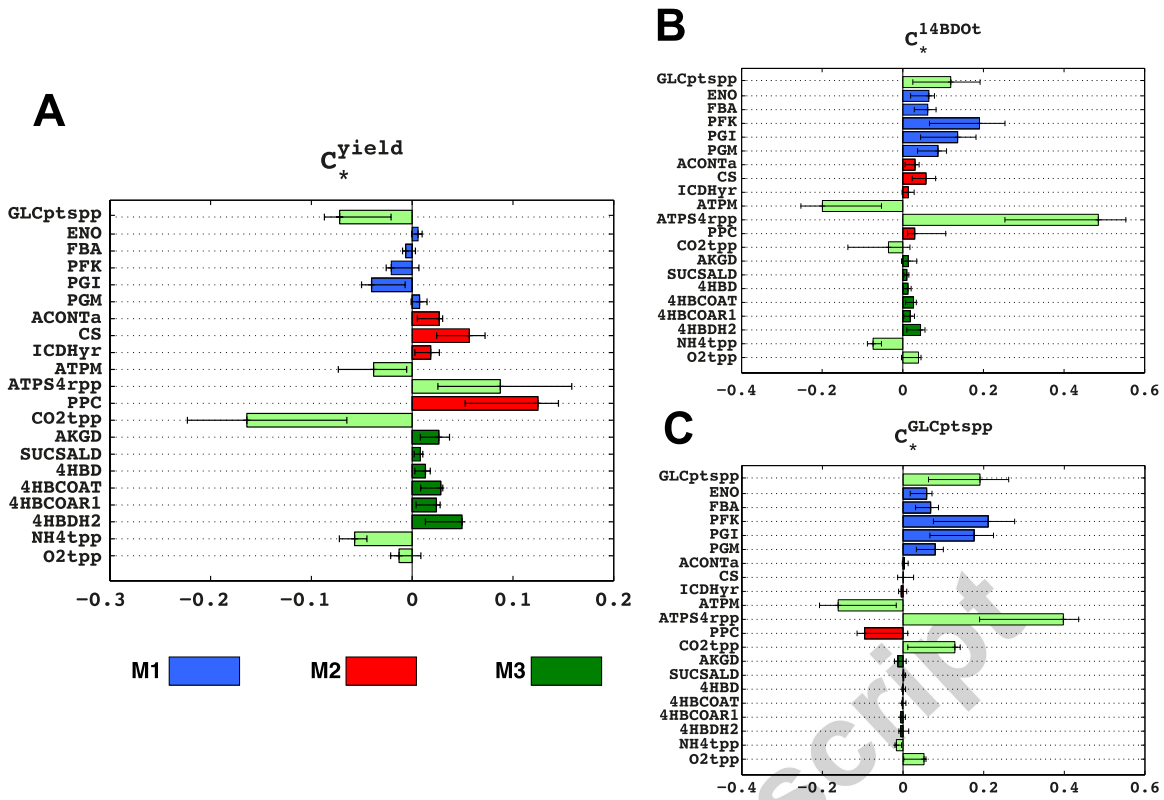

Figure 4: Control Coefficients for BDO/glucose yield (Panel A), BDO production, 14BDOt, (Panel B), and glucose uptake, GLCptspp (Panel C). Reactions from the three modules (*M1, M2* and *M3*) are highlighted with blue (*M1*), red (*M2*) and dark green (*M3*) fill.

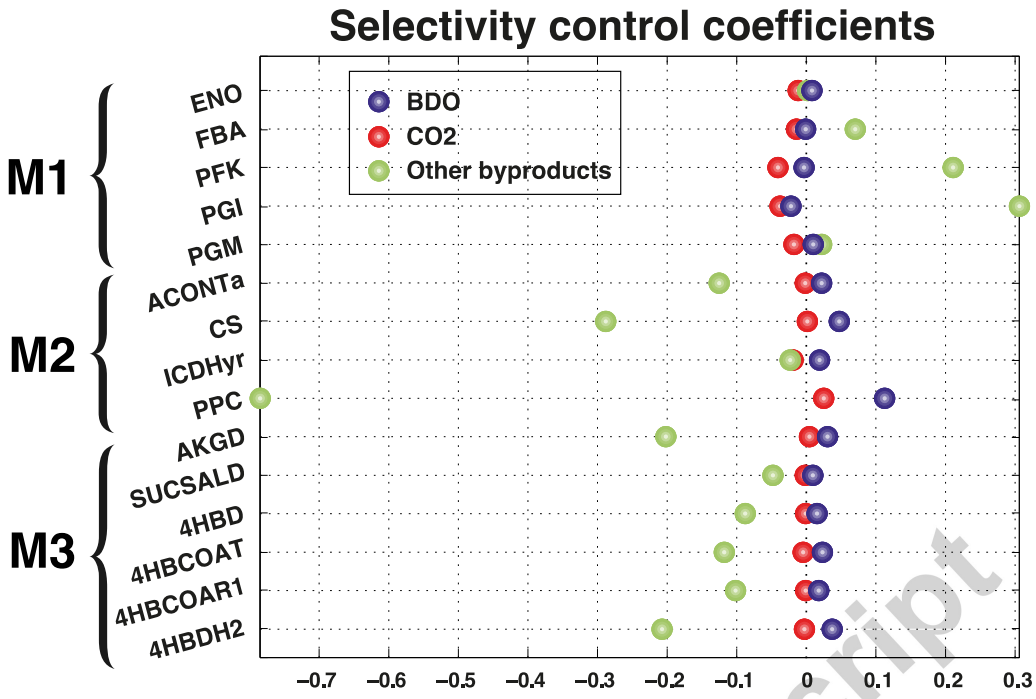

Figure 5: Mean selectivity control coefficients for: (i) BDO production, 14BDOt (blue dots); (ii) CO_2 production, CO2tpp (red dots); and (ii) production of other byproducts: L-glutamate, GLUabcpp and GLUt2rpp; ethanol, ETOHtrpp; lactate, D-LACTt2rpp; pyruvate, PYRt2rpp; and biomass, Biomass_core (green dots).

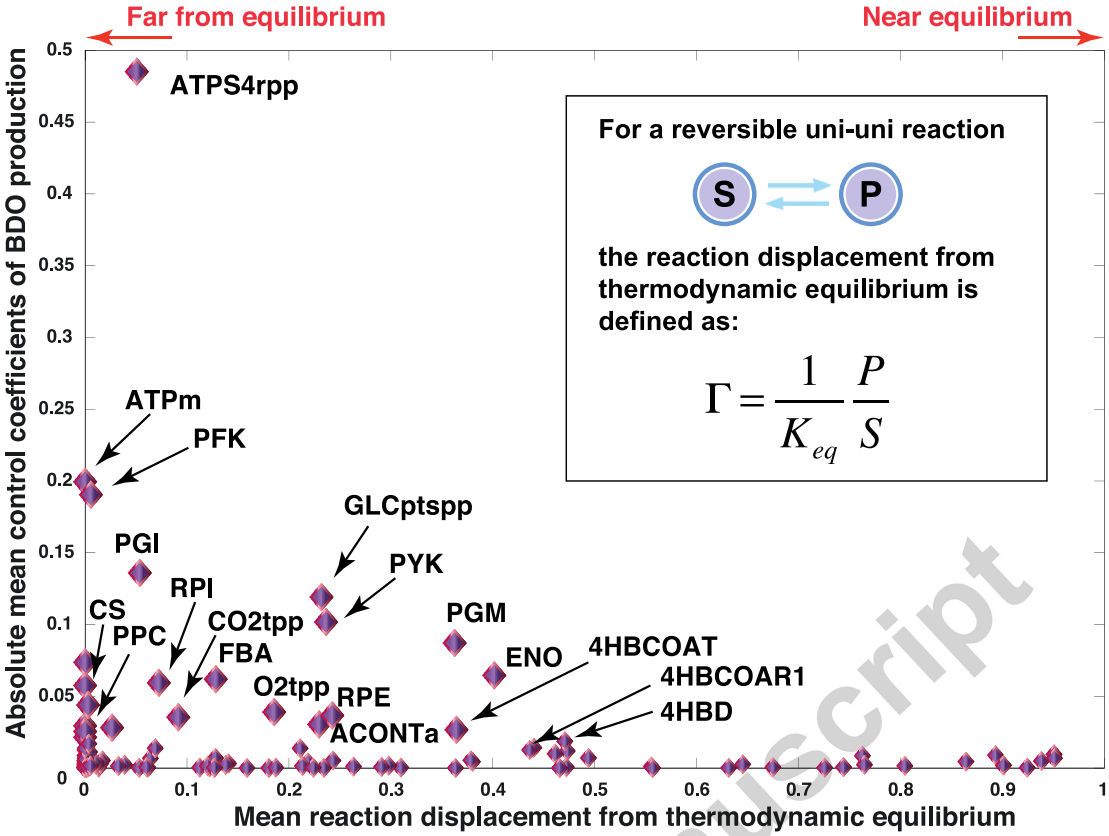

Figure 6: Correlation of the absolute mean control coefficients of specific BDO productivity with the mean displacement from thermodynamic equilibrium of the corresponding enzymes.

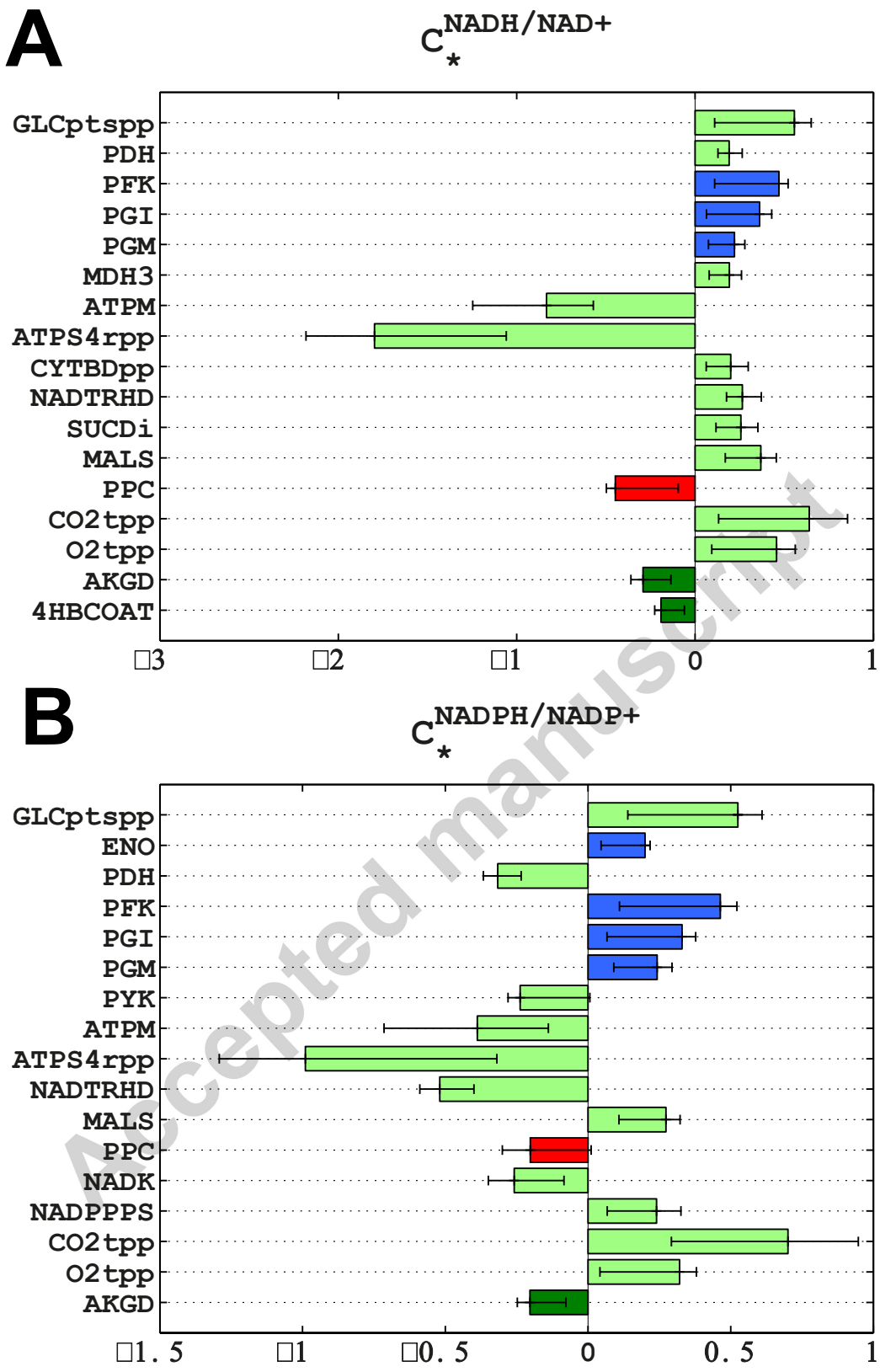

Figure 7: Control Coefficients for NADH/NAD+ ratio (redoxCyto) and NADPH/NAPD+ ratio (redoxPCyto). Reactions from the three modules (*M1, M2* and *M3*) are highlighted with blue (*M1*), red (*M2*) and dark green (*M3*) fill.

5 References

Alberty, R. A., 1994. Biochemical thermodynamics. Biochim Biophys Acta. 1207, 1-11.

Andreozzi, S., Miskovic, L., Hatzimanikatis, V., 2015. A Novel Approach to Characterization and Reduction of Uncertainty In the Kinetic Models of Genome-scale Metabolic Networks. Metabolic Engineering.

Ataman, M., Hatzimanikatis, V., 2015. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. Curr Opin Biotechnol. 36, 176-182.

Barton, N. R., Burgard, A. P., Burk, M. J., Crater, J. S., Osterhout, R. E., Pharkya, P., Steer, B. A., Sun, J., Trawick, J. D., Van Dien, S. J., Yang, T. H., Yim, H., 2015. An integrated biotechnology platform for developing sustainable chemical processes. J Ind Microbiol Biot. 42, 349-360.

Becker, S., Feist, A., Mo, M., Hannum, G., Palsson, B., Herrgard, M. J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nature Protocols. 2, 727-738.

Bilu, Y., Shlomi, T., Barkai, N., Ruppin, E., 2006. Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. Plos Comput Biol. 2, 932-938.

Borodina, I., Kildegaard, K. R., Jensen, N. B., Blicher, T. H., Maury, J., Sherstyk, S., Schneider, K., Lamosa, P., Herrgard, M. J., Rosenstand, I., Oberg, F., Forster, J., Nielsen, J., 2015. Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in Saccharomyces cerevisiae via beta-alanine. Metab Eng. 27, 57-64.

Burgard, A. P., Pharkya, P., Maranas, C. D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng. 84, 647-57.

Burk, M. J., Burgard, A. P., Osterhout, R. E., Sun, J., Microorganisms for the production of 1,4-butanediol. Google Patents, 2011.

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.

Chen, Y., Nielsen, J., 2013. Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. Current Opinion in Biotechnology. 24, 965-972.

Choi, S., Song, C. W., Shin, J. H., Lee, S. Y., 2015. Biorefineries for the production of top building block chemicals and their derivatives. Metabolic Engineering. 28, 223-239.

Demeke, M. M., Dietz, H., Li, Y. Y., Foulquie-Moreno, M. R., Mutturi, S., Deprez, S., Den Abt, T., Bonini, B. M., Liden, G., Dumortier, F., Verplaetse, A., Boles, E., Thevelein, J. M., 2013. Development of a D-xylose fermenting and inhibitor tolerant industrial Saccharomyces cerevisiae strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. Biotechnol Biofuels. 6.

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., Palsson, B. O., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology. 3, -.

Goldberg, R. N., Tewari, Y. B., Bhat, T. N., 2004. Thermodynamics of enzyme-catalyzed reactions - a database for quantitative biochemistry. Bioinformatics. 20, 2874-2877.

Hadadi, N., Ataman, M., Hatzimanikatis, V., Panayiotou, C., 2015. Molecular thermodynamics of metabolism: quantum thermochemical calculations for key metabolites. Physical Chemistry Chemical Physics. 17, 10438-10453.

Hatzimanikatis, V., Floudas, C. A., Bailey, J. E., 1996. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. Aiche Journal. 42, 1277-1292.

Hatzimanikatis, V., Li, C., Ionita, J. A., Henry, C., Jankowski, M. D., Broadbelt, L. J., 2005. Exploring the diversity of complex metabolic networks. Bioinformatics. 21, 1603-9.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems.

Henry, C., Jankowski, M. D., Broadbelt, L. J., Hatzimanikatis, V., 2006. Genome-scale thermodynamic analysis of Escherichia coli metabolism. Biophysical Journal. 90, 1453-61.

Henry, C. S., Broadbelt, L. J., Hatzimanikatis, V., 2007. Thermodynamics-based metabolic flux analysis. Biophysical Journal. 92, 1792-1805.

Hofmeyr , J., Cornish-Bowden, A., 1997. The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. Comp. Appl. Biosci. 13, 377-385.

Iuchi, S., Lin, E. C. C., 1988. Arca (Dye), a Global Regulatory Gene in Escherichia-Coli Mediating Repression of Enzymes in Aerobic Pathways. Proceedings of the National Academy of Sciences of the United States of America. 85, 1888-1892.

Jankowski, M., Henry, C., Broadbelt, L., Hatzimanikatis, V., 2008. Group Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks. Biophysical Journal. 95, 1487-1499.

Jenkins, C. M., Pikuleva, I., Waterman, M. R., 1998. Expression of eukaryotic cytochromes P450 in E. coli. Methods in molecular biology. 107, 181-93.

Jenkins, C. M., Waterman, M. R., 1998. NADPH-flavodoxin reductase and flavodoxin from Escherichia coli: characteristics as a soluble microsomal P450 reductase. Biochemistry. 37, 6106-13.

Jolliffe, I., 2002. Principal component analysis. Springer, New York.

Kim, I. K., Roldao, A., Siewers, V., Nielsen, J., 2012. A systems-level approach for metabolic engineering of yeast cell factories. FEMS Yeast Research. 12, 228-48.

Kummel, A., Panke, S., Heinemann, M., 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Molecular Systems Biology. 2.

Lee, S. Y., Mattanovich, D., Villaverde, A., 2012. Systems metabolic engineering, industrial biotechnology and microbial cell factories. Microbial Cell Factories. 11.

Leonard, E., Nielsen, D., Solomon, K., Prather, K. J., 2008. Engineering microbes with synthetic biology frameworks. Trends in Biotechnology. 26, 674-681.

Lequieu, J., Chakrabarti, A., Nayak, S., Varner, J. D., 2011. Computational Modeling and Analysis of Insulin Induced Eukaryotic Translation Initiation. Plos Comput Biol. 7.

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., Konig, R., Smith, R. D., Palsson, B. O., 2010. Omic data from evolved E-coli are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology. 6.

Lewis, N. E., Nagarajan, H., Palsson, B. O., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature Reviews Microbiology. 10, 291-305.

Liebermeister, W., Klipp, E., 2006. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. Theoretical Biology and Medical Modeling. 3.

Lutz, R., Bujard, H., 1997. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I-1-I-2 regulatory elements. Nucleic Acids Research. 25, 1203-1210.

Mat-Jan, F., Alam, K. Y., Clark, D. P., 1989. Mutants of Escherichia-Coli Deficient in the Fermentative Lactate-Dehydrogenase. Journal of Bacteriology. 171, 342-348.

Miskovic, L., Hatzimanikatis, V., 2010. Production of biofuels and biochemicals: in need of an ORACLE. Trends in Biotechnology. 28, 391-7.

Miskovic, L., Hatzimanikatis, V., 2011. Modeling of uncertainties in biochemical reactions. Biotechnology and Bioengineering. 108, 413-23.

Miskovic, L., Tokic, M., Fengos, G., Hatzimanikatis, V., 2015. Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes. Current Opinion in Biotechnology. 36, 1-8.

Moreira dos Santos, M., Thygesen, G., Kotter, P., Olsson, L., Nielsen, J., 2003. Aerobic physiology of redox-engineered Saccharomyces cerevisiae strains modified in the ammonium assimilation for increased NADPH availability. FEMS Yeast Research. 4, 59-68.

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. Molecular Systems Biology. 7, 535.

Pfleger, B. F., Gossing, M., Nielsen, J., 2015. Metabolic engineering strategies for microbial synthesis of oleochemicals. Metab Eng. 29, 1-11.

Pharkya, P., Burgard, A. P., Van Dien, S. J., Osterhout, R. E., Burk, M. J., Trawick, J. D., Kuckinskas, M. P., Steer, B., Microorganisms and methods for production of 4-hydroxybutyrate, 1,4-butanediol and related compounds US Patent 20,140,371,417, 2014.

Pitkanen, J. P., Aristidou, A., Salusjarvi, L., Ruohonen, L., Penttila, M., 2003. Metabolic flux analysis of xylose metabolism in recombinant Saccharomyces cerevisiae using continuous culture. Metabolic Engineering. 5, 16-31.

Reich, J. G., Sel'kov, E. i. E., 1981. Energy metabolism of the cell : a theoretical treatise. Academic Press, London ; New York.

Schellenberger, J., Lewis, N. E., Palsson, B. O., 2011a. Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models. Biophysical Journal. 100, 544-553.

Schellenberger, J., Richard, Q., Ronan, F., Ines, T., Hyduke, D., Bordbar, A., Feist, A., Bernhard, P., Daniel, Z., Jeffery, O., Joseph, K., Nathan, L., Sorena, R., 2011b. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols. 6, 1290-1307.

Schomburg, I., Chang, A., Placzek, S., Sohngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., Schomburg, D., 2013. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Research. 41, D764-72.

Schurmann, M., Sprenger, G. A., 2001. Fructose-6-phosphate aldolase is a novel class I aldolase from Escherichia coli and is related to a novel group of bacterial transaldolases. The Journal of Biological Chemistry. 276, 11055-61.

Segel, I. H., 1975. Enzyme Kinetics.

Silverman, P. M., Rother, S., Gaudin, H., 1991. Arc and Sfr functions of the Escherichia coli K-12 arcA gene product are genetically and physiologically separable. Journal of Bacteriology. 173, 5648-52.

Soh, K. C., Hatzimanikatis, V., 2010a. Dreams of Metabolism. Trends in Biotechnology. 28, 501-508.

Soh, K. C., Hatzimanikatis, V., 2010b. Network thermodynamics in the post-genomic era. Curr Opin Microbiol. 13, 350-7.

Soh, K. S., Hatzimanikatis, V., 2010c. Network thermodynamics in the post-genomic era. Current Opinion Microbiology. 13, 350-357.

Soh, K. S., Hatzimanikatis, V., 2014. Constraining the flux space using thermodynamics and integration of metabolomics data. Methods in Molecular Biology. 1191, 49-63.

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 Research. 12, 129-143.

Steinsiek, S., Bettenbrock, K., 2012. Glucose transport in Escherichia coli mutant strains with defects in sugar transport systems. Journal of Bacteriology. 194, 5897-908.

Stols, L., Donnelly, M. I., 1997. Production of succinic acid through overexpression of NAD(+)-dependent malic enzyme in an Escherichia coli mutant. Applied and Environmental Microbiology. 63, 2695-2701.

Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., van der Weijden, C. C., Schepper, M., Walsh, M. C., Bakker, B. M., van Dam, K., Westerhoff, H. V., Snoep, J. L., 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. European Journal of Biochemistry. 267, 5313-5329.

Thomas, R., Paredes, C. J., Mehrotra, S., Hatzimanikatis, V., Papoutsakis, E. T., 2007. A model-based optimization framework for the inference of regulatory interactions using time-course DNA microarray expression data. BMC Bioinformatics. 8, 228.

Wang, L., Birol, I., Hatzimanikatis, V., 2004. Metabolic Control Analysis under Uncertainty: Framework Development and Case Studies. Biophysical Journal. 87, 3750-3763.

Wang, L., Hatzimanikatis, V., 2006a. Metabolic engineering under uncertainty—II: Analysis of yeast metabolism. Metabolic Engineering. 8, 142-159.

Wang, L., Hatzimanikatis, V., 2006b. Metabolic engineering under uncertainty. I: Framework development. Metabolic Engineering. 8, 133-141.

Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E., Rojas, I., Muller, W., 2012. SABIO-RK-database for biochemical reaction kinetics. Nucleic Acids Research. 40, D790-D796.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., Van Dien, S., 2011. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat Chem Biol. 7, 445-52.

Zhuang, K., Yang, L., Cluett, W. R., Mahadevan, R., 2013. Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. DySScO strategy for strain design. Bmc Biotechnology. 13.

Highlights

- Population of large-scale kinetic models of recombinant *E. coli* constructed
- Metabolic engineering targets for the improvement of 1,4-butanediol production identified
- Computationally obtained targets consistent with the experimentally tested designs