Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-23 (2022) 109–114

Using parameter sampling to explore key properties of *E. coli's* anaerobic metabolism*

Giulia Slaviero * **Ralf Steuer** ** **Steffen Klamt** ***

* *Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany (e-mail: slaviero@mpi-magdeburg.mpg.de)*
** *Institute for Biology, Humboldt-University of Berlin, Berlin, Germany (e-mail: steuer.ralf@gmail.com)*
*** *Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany (e-mail: klamt@mpi-magdeburg.mpg.de)*

Abstract: The high uncertainty and incomplete knowledge of kinetic parameters is a major challenge for using kinetic metabolic models. Literature values are scarce and fitting procedures require large experimental data sets and complex computations. In this study we show that Monte-Carlo sampling of kinetic parameters allows the identification of system's properties such as stability and flux control patterns. Applying this computationally simple method to the anaerobic central metabolism of *E. coli* we determine that only few network parameters are directly correlated with stability. We show that low enzyme concentrations often correspond to positive eigenvalues, suggesting that enzyme-optimized pathways lack in flexibility. Analysis of the distribution of flux control coefficients reveals that the highest control on the network rates is exerted by reactions utilizing ATP. Finally, a comparison with experimental evidence confirms this method's potential to qualitatively analyze networks for which no information on parameters are available.

Keywords: Metabolic engineering, parameter identification, kinetic modelling, stability, probabilistic parameter sampling, Monte-Carlo sampling.

1. INTRODUCTION

The main issue when dealing with kinetic models of metabolism is the high number and uncertainty of parameters and the extensive amount of experimental data required to fit them. Throughout the years many parameter estimation techniques have been established, but still the highly non-linear nature of kinetic models does not guarantee univocal numerical identifiability of all parameters (i.e. local minima in optimization routines, biases in experimental data, etc). This issue can be overcome using stochastic methods to exploit network properties. One possibility is to employ a Monte-Carlo procedure to sample model variables, as first applied by Wang et al. (2004) and Steuer et al. (2006), and subsequently adapted by Murabito et al. (2011), to randomly sample kinetic constants. This latter method provides a probabilistic approach to analyze stability properties and control patterns of the network, knowing only the concentrations and fluxes of a reference steady state. Additionally, it does not require any expensive calculations, and can therefore be applied to very complex and large networks. In this study, starting from the network structure of a known *E. coli* kinetic model, and without any prior knowledge on the kinetic constants, we (i) identify sources of stability and instability for the system, (ii) analyze the control patterns in the network, (iii) compare it with literature data and (iv) show the importance of the regulatory structure.

2. SAMPLING PROCEDURE

The Monte-Carlo sampling procedure used here is based on the method of Murabito et al. (2014) and can be summarized in four main steps.

1) *Defining system structure.* To begin, network stoichiometry and rate laws are defined. Note that if no information is available on the reaction mechanism, also simplified or standardised rate equations can be used e.g. convenience kinetics (Liebermeister et al., 2010). Next, steady state fluxes $\mathbf{v}^0$, concentrations $\mathbf{S}^0$, thermodynamic constants $\mathbf{k}_{eq}$ and turn over constants $\mathbf{k}_{cat}$ are retrieved. These values can be directly measured, computed or found in the literature.

2) *Parameters sampling.* Random sampling is performed inside the sampling intervals. For parameter $k_{i,j}$ participating in reaction j and being related to metabolite i the sampling space in this study is expressed as

$$k_{i,j} \in [b_L \cdot S_i^0, b_U \cdot S_i^0], \tag{1}$$

with relative bounds b_L and b_U.

3) *Anchoring sampled instances to the given steady state.* Enzyme concentrations are computed (Eq. (4)) to ensure

* This work was founded by the European Research Council (721176)

10.1016/j.ifacol.2023.01.024

consistency of parameter sets with the observed reference steady state $\mathbf{v}^0$ and $\mathbf{S}^0$.

For a generic reaction ($\sum R_j \leftrightarrow \sum P_j$), characterized by enzyme concentration E and mass ratio Γ,

$$\nu = E_i \cdot k_{cat,i} \cdot \frac{\prod \frac{R_j}{k_{i,j}} \cdot (1 - \frac{\Gamma}{k_{eq,i}})}{1 + \prod \frac{R_j}{k_{i,j}} + \prod \frac{P_b}{k_{i,j}}} \tag{2}$$

$$= E_i \cdot k_{cat,i} \cdot f(\mathbf{S}, \mathbf{k}) \tag{3}$$

hence

$$E_i = \frac{\nu_i}{k_{cat,i} \cdot f(\mathbf{S}, \mathbf{k})} \tag{4}$$

4) *Evaluating stability and further system properties.* The stability of the sampled sets is evaluated. Stability is ensured if all real parts of the eigenvalues of the Jacobian matrix (Eq. (5)) are negative. Further information on network dependencies can be gathered by computing concentration (Eq. (6)) and flux control coefficients (Eq. (7)).

$$\mathbf{J}' = \mathbf{N}_R \cdot \frac{\partial \mathbf{v}}{\partial \mathbf{S}}\Big|_{\mathbf{S}^0} \cdot \mathbf{L} \tag{5}$$

$$\mathbf{C}^S = -(\mathbf{D}^S)^{-1} \cdot \mathbf{L} \cdot \mathbf{J}'^{-1} \cdot \mathbf{N}_R \cdot \mathbf{D}^J \tag{6}$$

$$\mathbf{C}^J = 1 + (\mathbf{D}^J)^{-1} \cdot \frac{\partial \mathbf{v}}{\partial \mathbf{S}}\Big|_{\mathbf{S}^0} \cdot \mathbf{D}^S \cdot \mathbf{C}^S \tag{7}$$

Here $\mathbf{N}_R$ and $\mathbf{L}$ are respectively the reduced stoichiometric and link matrices, respectively. $\mathbf{D}^S$ and $\mathbf{D}^J$ are diagonal matrices with the steady state concentrations and fluxes values on the diagonal. For more information on metabolic control analysis see Hofmeyr (2000). For a more extensive and complete description of the method see Murabito et al. (2011).

3. RESULTS

The aforementioned sampling and analysis procedure has been applied to a kinetic model of the anaerobic central metabolism of *E coli*. The network structure (Fig. 1) as well as rate equations and steady state reactions and concentrations vectors have been retrieved from the model used in Boecker et al. (2021) (PTS and PPC kinetic laws have been slightly modified). All Michaelis-Menten as well as inhibition and activation constants have been sampled one order of magnitude around the concentration value of the related metabolic compound (Eq. 1). The sampling procedure has been performed for $2 \cdot 10^4$ iterations using *Matlab R2021a*.

3.1 Stability analysis

Of all sampled parameter sets a very high percentage (85%) results in stable models (all eigenvalues have negative real parts). Even increasing the sampling interval to two orders of magnitude above and below the reference concentration, the stable sets remain almost 70%. This broad network stability can be investigated in more detail by analyzing the ensemble of Jacobian matrices.

Firstly, we observed that throughout all sampled parameter sets, diagonal elements of the Jacobian matrix are always negative. Ivanov et al. (2016) demonstrated that such

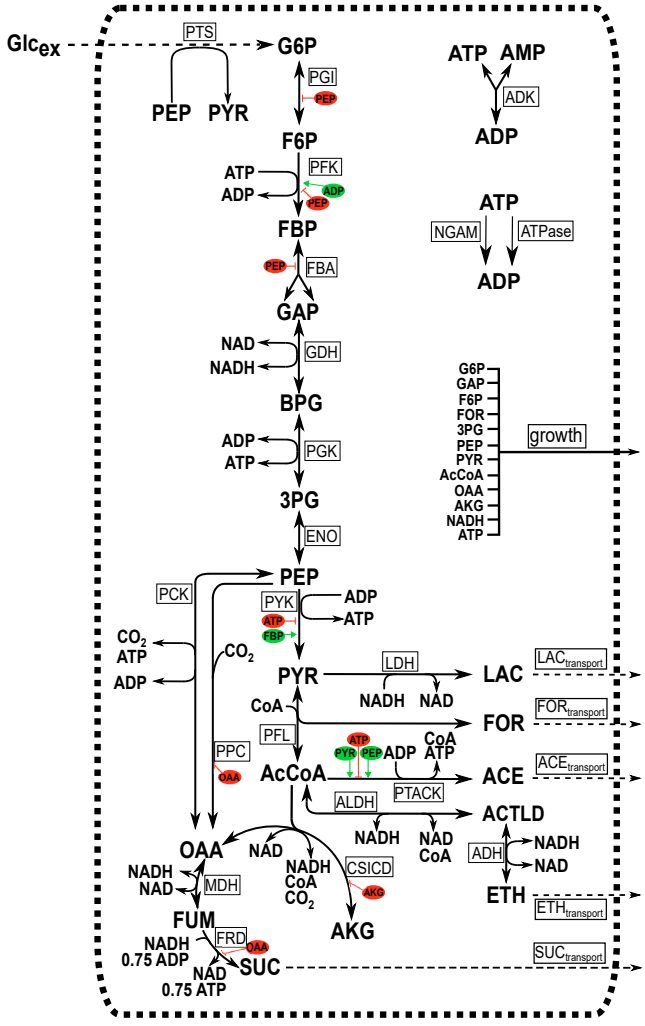

Fig. 1. Scheme of the *E. coli* anaerobic metabolism used in this study. Metabolites are shown in black, red and green circles represent inhibitors and activators respectively, reaction names are set in boxes. For the complete description of the model structure see Boecker et al. (2021). Abbreviation of reaction names: PTS: phoshotransferase system; PGI: glucose-6-phosphate isomerase; PFK: phosphofructokinase; FBA: fructose-bisphosphate aldolase (the associated reaction was lumped with the reaction of the triose-phosphate isomerase (TPI) thus yielding two molecules of GAP); GHD: glyceralde- hyde-3-phosphate dehydrogenase; PGK: phosphoglycerate kinase; ENO: enolase (the reaction of this enzyme was lumped with the reaction of the phosphoglycerate mutase); PYK: pyruvate kinase; PFL: pyruvate formate lyase; LDH: lactate dehydrogenase; PTACK: lumped reaction of acetate kinase and phosphate acetyltransferase; ALDH: acetaldehyde-CoA dehydrogenase; ADH: alcohol dehydrogenase; PCK: phosphoenolpyruvate carboxykinase; PPC: phosphoenolpyruvate carboxylase; CSICD: lumped reaction of citrate synthase, aconitate hydratase A, aconitate hydratase B and isocitrate dehydrogenase; MDH: lumped reaction of malate dehydrogenase and fumarase; FRD: lumped reaction of fumarate reductase and of other reactions involved in fumarate reduction.

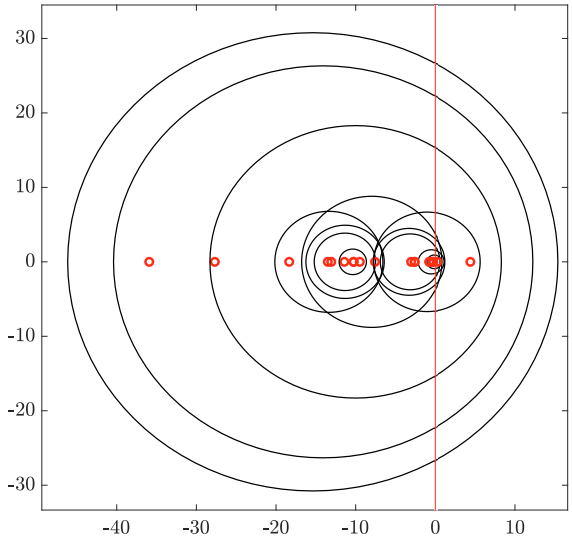

Fig. 2. Details of the Gershgoring circles for an unstable parameter set. Each diagonal element of the Jacobian matrix is a circle center with a radius given by the sum of the absolute values of the off-diagonal elements in the respective row. Red points denote the real part of the eigenvalues, the red line splits the positive from the negative abscissa.

a peculiarity is always granted for networks characterized by monotonic reactions without regulations. Furthermore, Du et al. (2017) showed that for mass action kinetics negative diagonals favour stability. These considerations together with the observations made all over random-chosen kinetic sets, suggest an intrinsic tendency towards stability also of networks characterized by non-monotonic regulated reactions.

Secondly, the eigenvalues position (hence sign) can be explained testing the rows of the Jacobian matrix for diagonal dominance and applying the Gershgoring circle theorem (Flach and Schnell, 2010). Diagonal dominance is defined when the absolute value of the sum of the off-diagonal elements is smaller than the absolute value of the diagonal element itself ($|\mathbf{J}_{ii}| > \sum_{k\neq i} |\mathbf{J}_{ik}|$). Similarly, each row of the Jacobian matrix forms a Gershgoring circle, with the diagonal element $\mathbf{J}_{ii}$ being the center and the absolute value of the off-diagonal sum $\sum_{k\neq i} |\mathbf{J}_{ik}|$ being the radius. The theorem states that all eigenvalues fall inside a Gershgorin circle, eventhough not all Gershgorin circles will contain an eigenvalue. Therefore, if a row is strongly diagonally dominant, with a negative diagonal value, it constraints the eigenvalue to be negative, and therefore stable (Du et al., 2017). On the other hand in absence of diagonal dominance the radius of the circle gets larger and reaches possibly positive values of the x-axis, increasing the probability of the eigenvalues to be positive (see Fig. 2).

In all sampled sets (even without regulation) only 6 rows of the Jacobian matrix show diagonal dominance, particularly the ones corresponding to GAP, SUC, FOR, LAC, ACE and ETH. Interestingly, with exception of GAP, diagonal dominance is related only to fermentation products. These metabolites are situated in the network periphery and are characterized just by a production and a consumption rate. Being the Jacobian matrix directly related to the matrix of flux derivatives ($d\mathbf{v}/d\mathbf{S}$), follows that by an enhanced model complexity (i.e. increased stoichiometric matrix density) also the number of the off-diagonal elements increases and thus also their absolute sum. Consequently, with exception of some species, the elements causing instability are spread across the system.

3.2 Correlation between parameters and stability.

As a next step, the effect of single parameter variations on stability has been studied. For its assessment, Spearman correlation has been used, and parameters ranked according to the absolute value of their correlation coefficient. As a result, parameters of the PFK, PTS and PYK reactions show the highest degree of correlation with the maximum real part of the eigenvalues (see Fig. 3). As expected, the majority of those parameters are part of glycolysis. Glycolysis is known to be tightly controlled by feed-back and feed-forward allosteric regulations that are required to readily adjust metabolism in case of disturbances (Mulukutla et al., 2014). Repeating the same analysis in the model without allosteric regulations shows a shift in the correlation hierarchy. In this case parameters related to PTACK, CSICD and FBA show the highest ranking. It is more likely that in the absence of control loops the metabolism is governed by equilibrium reactions (FBA) or by the flux-demand of the end-reactions (CSICD and PTACK). Interestingly, even without allosteric control, the parameters with the highest absolute correlation coefficients remain the ones participating in reactions that would otherwise be regulated, a phenomena also observed in (Grimbs et al., 2007). Such a result could help to explain why evolution selected regulation mechanisms on some reactions but not on others.

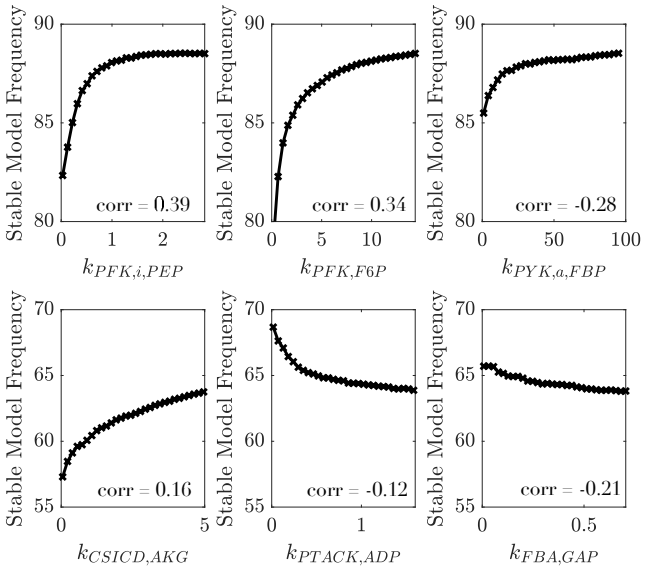

Fig. 3. Relationship between kinetic parameters and the relative frequency in which the corresponding models are stable. Shown are three parameters with the highest correlation coefficients, for the scenario with regulation (upper row), and without regulation (lower row). The respective value of the correlation coefficient is shown in each subplot.

3.3 Enzyme concentration and relaxation time

Next, starting from Eq. (4) the total enzyme amount required for each parameter set is calculated as $\sum E_i$. As shown in Fig. 4, it emerges that high enzyme concentrations are usually related to stable models. More precisely, the third quantile for the stable parameter sets is almost 30% higher than the one of the unstable ones. As an explanation we hypothesize that, when the enzymes are very low, or, equivalently, if the enzymes operate close to saturation, not much flexibility is left to the network. Hence in case of perturbations, very long relaxation times are required, pushing the system towards instability.

3.4 Distribution of flux control coefficients

The reliability of predictions made by random parameter sampling are here explored using the fitted quantitative model of Boecker et al. (2021). Computing the flux control coefficients according to Eq. (7) provides the probabilistic flux control distributions shown in Fig. 5*a*. Columns represent the controlling enzyme, rows represent the affected reactions. Red lines indicate the value of the FCCs obtained using the quantitative model of Boecker et al. (2021).

First of all, the width of the distribution is analyzed. Broad distributions imply reactions very sensitive to parameter changes, vice versa, narrow distributions imply that control coefficients are rather insensitive to parameter variations. Typically, reactions close to equilibrium have narrow profiles centered around zero (Steuer and Junker, 2009), here confirmed by PGI and ENO. In general, in this study, distributions are not very broad, hinting to a model that is robust with respect to parameter variations. Exception is made for the very broad control distributions exerted by several reactions on the rows of MDH and FRD. These reactions of the succinate branch on the one hand benefit from increased metabolic fluxes, on the other hand depend on PEP as substrate. In any case, such spread profiles indicate a great sensitivity to parameter variations and suggest that kinetic rate equations and regulations should be further investigated.

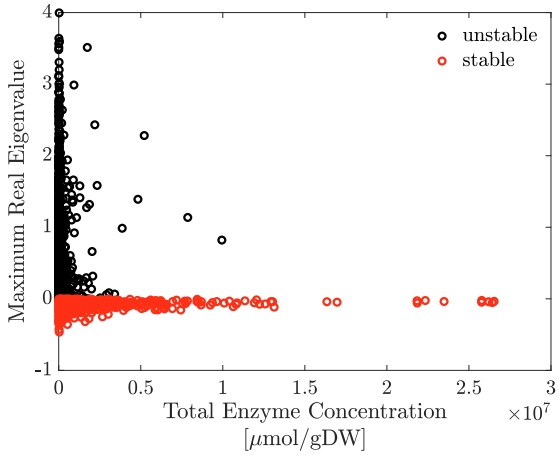

Fig. 4. Relationship between the maximal real part of the eigenvalues and the total enzyme concentration for each sampled parameter set.

Next, the analysis of the probabilistic distributions provides information on the control pattern in the network, and is in good agreement with experimental evidences and intuitive expectations. As an example, growth rate and glucose uptake rate have a positive control on almost all fluxes as reported also in Millard et al. (2017). Similarly, PFK is one of the reactions with the largest control spectrum and has a positive average value of the control coefficient on glycolysis ($c^{glycolysis}_{PFK,average} = 0.23$) supported by the measurement of Emmerling et al. (1999). In the same study it was reported that overexpression of the PFK increases lactate production with decreased ethanol synthesis, confirmed by the higher control coefficient for LDH ($c^{LDH}_{PFK,average} = 0.38$, $c^{ADH}_{PFK,average} = 0.16$). Also the reaction producing acetate shows a predominantly positive control on metabolism, pulling a higher flux through glycolysis and matching the experimental measurements of Schütze et al. (2020). Less intuitive is the negative control that PYK exerts on the glucose uptake rate. In fact, its increase diminishes the concentration of PEP, confirmed by a negative average concentration control coefficient $c^{PEP}_{PYK,average} = -0.22$. Such a reduction, even if hampering allosteric inhibitions through the glycolysis, considerably depletes substrate availability for PTS, thus decreasing its rate.

Interestingly, even if more than 35% of the fitted model's parameters do not fall into the intervals used for the sampling in this study, its FCCs (red lines in Fig. 5*a*-*b*) fall always within the probabilistic distribution, and in many cases match the distribution center. This fact confirms that the network structure has a higher impact on the system physiology than the particular choice of parameter values. Such an observation is confirmed computing the FCCs for the scenario in which regulatory interactions are removed (Fig. 5*b*). In this case almost all distributions are, on average, broader, especially the ones for GROWTH, PFL and PTACK. This denotes an enhanced sensitivity to parameter change, hence a loss in model robustness. A clearer demonstration on how the lack of regulation degrades prediction capability is obtained analyzing the change in FCCs distribution's signs (Fig. 5*c*-*d*). For instance, in the regulated network, the NGAM reaction accounting for ATP hydrolysis for non-growth associated maintenance, has over 80% of positive flux control coefficients. Indeed, a forced decrease in ATP levels was shown in (Boecker et al., 2021) to activate the PFK reaction and induce a higher glycolytic flux. Contrarily, without regulations NGAM FCCs are mainly negative, differing from experimental evidences. Also GDH and PGK, reactions proven to have a negative control on the glucose uptake rate (Jian et al., 2017), get in the unregulated scenario a positive effect, probably because of the absence of the inhibition action of PEP on the upper glycolysis, and the possibility of its accumulation.

4. CONCLUSION

In this study we have shown how it is possible to analyze key properties of *E. coli* anaerobic metabolism using a Monte Carlo sampling technique without any prior knowledge of kinetic parameters values. The evaluation of the ensemble of Jacobian matrices delivers insights into the

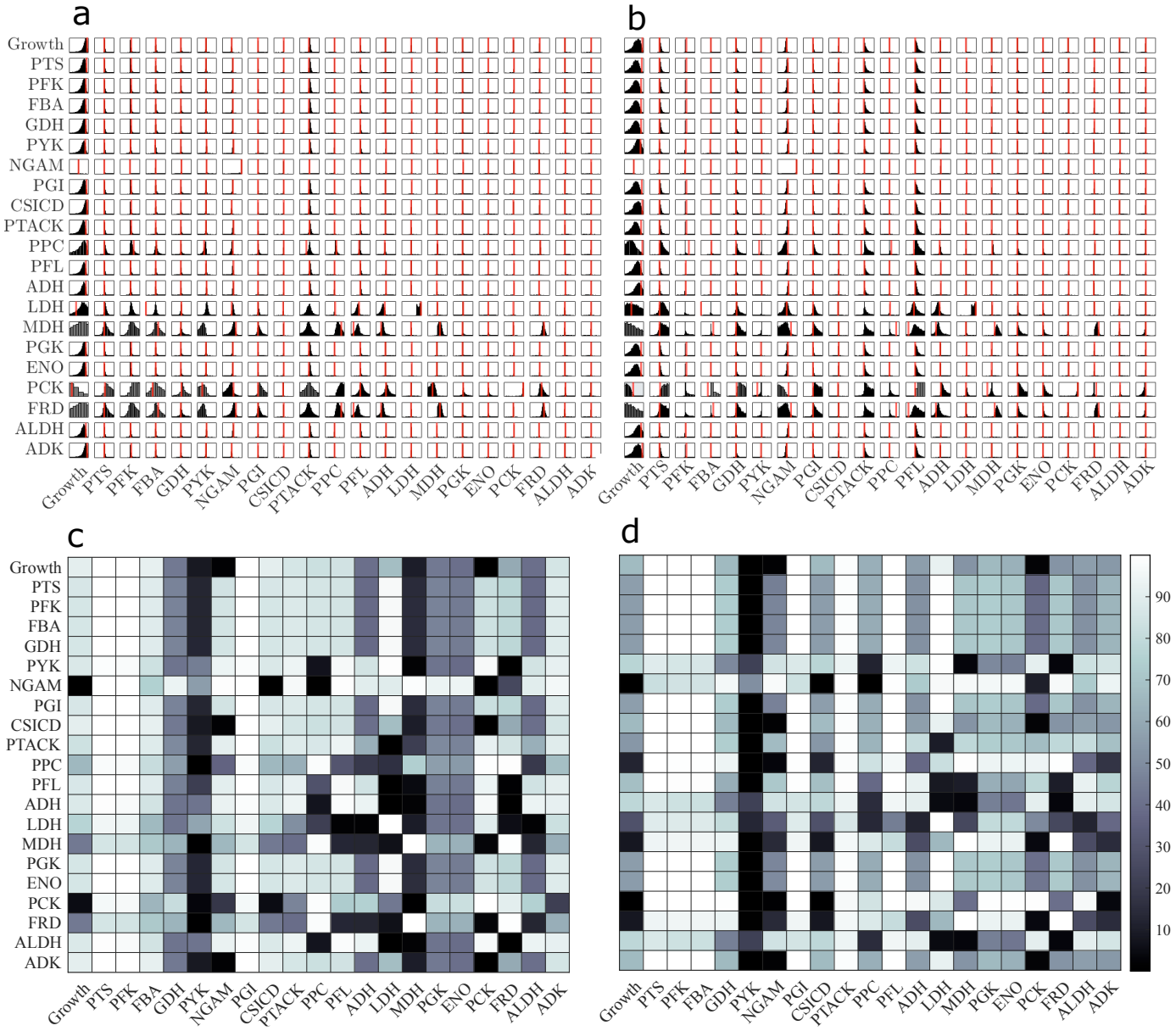

Fig. 5. $(a-b)$Probabilistic distributions of scaled flux control coefficients, for the model with (a) and without regulation (b). The red lines indicate the FCCs computed with the model of Boecker et al. (2021), with modifications described in the main text (Section 3). Columns indicate the controlling enzyme (cause) and rows the controlled reaction (effect). The x-asis spans between [-1 1].
$(c-d)$ Sign distributions of the flux control coefficients for the model with (c) and without regulation (d). Dark colors indicate mainly negative distributions, while light color indicates that almost all FCCs lie on the positive abscissa.

stability properties of the network. In particular, analyzing for diagonal dominance we could identify metabolites, which are not linked with instability. We concluded that parameters related to glycolysis and regulated reactions have the highest influence on the model eigenvalues. Furthermore, we could note that low total enzyme concentrations (i.e. enzymes operating close to saturation) often correspond to unstable responses, while unsaturated pathways are more frequently stable. We identified and validated, using literature data, important control properties of the network. Moreover, we showed that the flux control coefficients calculated with the fitted model (in which half of the parameters are taken from the literature) fall in the stochastic FCCs distributions, often matching the center. Finally, we observed that a deletion of allosteric regulations causes an increased number of unstable parameter sets and in a degradation of the accuracy of the flux control predictions.

REFERENCES

Boecker, S., Slaviero, G., Schramm, T., Szymanski, W., Steuer, R., Link, H., and Klamt, S. (2021). Deciphering the physiological response of *Escherichia coli* under high atp demand. *Molecular Systems Biology*, 17(12), e10504. doi:https://doi.org/10.15252/msb.202110504.

Du, B., Zielinski, D.C., and Palsson, B.O. (2017). Topological and kinetic determinants of the modal matrices of dynamic models of metabolism. *PLOS ONE*, 12(12), 1–21. doi:10.1371/journal.pone.0189880.

Emmerling, M., Bailey, J.E., and Sauer, U. (1999). Glucose catabolism of *Escherichia coli* strains with increased activity and altered regulation of key glycolytic enzymes. *Metabolic Engineering*, 1(2), 117–127. doi: https://doi.org/10.1006/mben.1998.0109.

Flach, E.H. and Schnell, S. (2010). Stability of open pathways. *Mathematical Biosciences*, 228(2), 147–152. doi:https://doi.org/10.1016/j.mbs.2010.09.002.

Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H.G., and Steuer, R. (2007). The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. *Molecular Systems Biology*, 3(1), 146. doi:https://doi.org/10.1038/msb4100186.

Hofmeyr, J.H.S. (2000). Metabolic control analysis in a nutshell. *Proceedings of the International Conference on Systems Biology,Pasadena, California*, 291 – 300.

Ivanov, O., van der Schaft, A., and Weissing, F.J. (2016). Steady states and stability in metabolic networks without regulation. *Journal of Theoretical Biology*, 401, 78–93. doi:https://doi.org/10.1016/j.jtbi.2016.02.031.

Jian, X., Li, N., Chen, Q., and Hua, Q. (2017). Model-guided identification of novel gene amplification targets for improving succinate production in *Escherichia coli* nzn111. *Integr. Biol.*, 9. doi:10.1039/C7IB00077D.

Liebermeister, W., Uhlendorf, J., and Klipp, E. (2010). Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics*, 26(12), 1528–1534. doi:10.1093/bioinformatics/btq141.

Millard, P., Smallbone, K., and Mendes, P. (2017). Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*. *PLOS Computational Biology*, 13(2), 1–24. doi: 10.1371/journal.pcbi.1005396.

Mulukutla, B.C., Yongky, A., Daoutidis, P., and Hu, W.S. (2014). Bistability in glycolysis pathway as a physiological switch in energy metabolism. *PLOS ONE*, 9(6), 1–12. doi:10.1371/journal.pone.0098756.

Murabito, E., Smallbone, K., Swinton, J., Westerhoff, H.V., and Steuer, R. (2011). A probabilistic approach to identify putative drug targets in biochemical networks. *Journal of The Royal Society Interface*, 8(59), 880–895. doi:10.1098/rsif.2010.0540.

Murabito, E., Verma, M., Bekker, M., Bellomo, D., Westerhoff, H.V., Teusink, B., and Steuer, R. (2014). Montecarlo modeling of the central carbon metabolism of *Lactococcus lactis*: Insights into metabolic regulation. *PLOS ONE*, 9(9), 1–15. doi:10.1371/journal.pone.0106453.

Schütze, A., Benndorf, D., Püttker, S., Kohrs, F., and Bettenbrock, K. (2020). The impact of *ackA*, *pta*, and *ackA-pta* mutations on growth, gene expression and protein acetylation in *Escherichia coli* k-12. *Frontiers in Microbiology*, 11. doi:10.3389/fmicb.2020.00233.

Steuer, R., Gross, T., Selbig, J., and Blasius, B. (2006). Structural kinetic modeling of metabolic networks. *Proceedings of the National Academy of Sciences*, 103(32), 11868–11873. doi:10.1073/pnas.0600013103.

Steuer, R. and Junker, B.H. (2009). Computational models of metabolism: Stability and regulation in metabolic networks. 105–251. doi: https://doi.org/10.1002/9780470475935.ch3.

Wang, L., Birol, I., and Hatzimanikatis, V. (2004). Metabolic control analysis under uncertainty: Framework development and case studies. *Biophysical Journal*, 87(6), 3750–3763. doi: https://doi.org/10.1529/biophysj.104.048090.