RESEARCH

Emerging ensembles of kinetic parameters to identify experimentally observed phenotypes

Riccardo Colombo[1,2*†], Chiara Damiani[1,2], David GIlbert[4], Monika Heiner[5], Giancarlo Mauri[1,2] and Dario Pescini[3,2†]

Abstract

Determine the value of kinetic constants for a metabolic system in the exact physiological conditions is an extremely hard task. However, this kind of information is of pivotal relevance to effectively simulate a complex biological phenomenon such as metabolism.

To overcome this issue, we propose to investigate emerging properties of ensembles of sets of kinetic constants leading to a biological readout observed in different experimental conditions. To this aim, we exploit information retrievable from constraint-based analyses (i.e. metabolic flux distributions at steady state) with the goal to generate putative values for kinetic constants exploiting the mass action law. The sets retrieved from the previous step will be used to parametrize a mechanistic model whose simulation will be performed to reconstruct the dynamics of the system (till the reaching of the metabolic steady state) for each experimental condition. Every parametrization that is in accordance with experimental data is collected in an ensemble whose features are analyzed to determine the emergence of properties of a phenotype.

In a previous work, we fruitfully tested the devised procedure using a toy model of *S. cerevisiae*. In this work we are extending the approach to identify kinetic parameters for a more complex metabolic system, analyzing five different experimental conditions associated to the ECC2comp model recently published by Hädicke and collaborators.

Keywords: ensembles; fluxes; kinetic parameters; mechanistic simulations; metabolism; ODEs; steady state; Systems Biology

Background

Systems Biology approaches in metabolic modeling

Advances in the understanding of biological processes revealed that living organisms must be analyzed keeping into account the complex network of interactions among different entities such as genes, transcripts, proteins and metabolites in order to decipher emergent behaviors and regulatory processes. It is today evident that the complexity of biological systems can be thoroughly investigated only exploiting mathematical modeling and simulation. To deal with this need, at the beginning of the present century, a new discipline, called Systems Biology, started to propose an integrated approach encompassing computational modeling and traditional "*wet*" experiments [3] to unravel the complexity of biological systems. In the context of Systems Biology, the study of metabolic processes (where usually metabolic networks are represented as hypergraphs in which nodes represent metabolites and edges indicate reactions [34]) has seen great interest especially due to the fruitful applications in metabolic engineering [29].

Omics data in metabolic modeling

Increased computational capabilities and technological innovations opened the so called "omics revolution", a new era characterized by the availability of an unprecedented amount of data. These high throughput information allowed the generation of more and more detailed genome-scale metabolic reconstructions, defined *ad hoc* for different cell types (unicellular organisms [2], healthy and diseased tissues in mammalian [11]). Besides this, there are still technological hindrances preventing mechanistic simulation of genome-scale metabolic model to be applied: currently, simulated temporal dynamics of metabolic concentrations

*Correspondence: riccardo.colombo@disco.unimib.it
[1]Department of Informatics, Systems and Communication, University of Milan - Bicocca, Viale Sarca, 336, 20126 Milan, Italy
[2]SYSBIO - Centre of Systems Biology, Piazza della Scienza, 2, 20126 Milan, Italy
Full list of author information is available at the end of the article
†Equal contributor

are available only for small models due to shortage of parameters and computational costs [5].

Constraint-based methods

The points above raised, determine the current strategy in metabolic modeling, namely the exploitation of the so called constraint-based approaches [15], a modeling framework making use of information on the structure of the metabolic network and assuming that internal metabolites reach a (pseudo) steady-state concentration. Even if these approaches neglect the temporal evolution of the system, they can be considered a valid framework to describe metabolism because of experimental studies pointing out that *in vivo* metabolism reaches the steady state in few seconds [33]. The core of constraint-based modeling is the stoichiometric information: it can be retrieved from the structure of the metabolic network and defines the stoichiometric matrix, i.e. the table illustrating changes in metabolites quantities due to the firing of reactions.

Moreover constraint-based approaches define, the mathematical space containing flux distributions (i.e. flux values for each reaction in the model) that can be reached by the system and representing different functional states. This "feasible solution space", can be determined by imposing additional constraints such as the mass balance and setting boundaries on fluxes (e.g. to determine their reversibility). Once the stoichiometric matrix and the boundaries are defined, it is possible to assume that the system is optimal toward a given task (an Objective Function – OF – to be maximized or minimized), and subsequently an optimal flux distributions can be calculated by means of optimization techniques such as flux balance analysis (FBA) [25].

Choosing an appropriate formulation of the OF is of paramount importance when conducting FBA, however often its exact formulation is not definable, and because it is reasonable to assume that the system is found in a sub-optimal state (see [14] for further information).

Ensemble FBA

To analyze the potentiality of a cell to explore alternative metabolic behaviors by altering its fluxes, we defined the Ensemble Evolutionary FBA (eeFBA) [8] an extension of FBA with the goal to investigate putative flux distributions that can give rise to a specific metabolic behavior. With eeFBA, analyses are performed generating a set of random OFs that is subsequently optimized by means of linear programming. Following this, computed flux distributions are filtered on the basis of one or more metabolic phenotypes definitions, retrieving ensembles of solutions that are in agreement with the defined phenotypes.

Retrieving kinetic parameters form a mechanism-based ensemble approach

Despite this extension of the traditional constraint-based approaches, due to lack of information on kinetic constants, either with FBA or eeFBA it is not possible to determine metabolic concentrations at steady state. In a recent paper [6], we proposed a strategy, where ensembles of phenotypes are populated according to fluxes properties, however steady states are there retrieved from mechanism based simulations whose parameters are determined using initial concentrations from the literature (whenever possible) and kinetic constants randomly sampled.

With the above described procedure we have been able to determine steady state metabolic concentrations that satisfy the definition of a given metabolic phenotype. It is worth to underline that, this readout is obtained without defining an OF and avoiding to assume that the cell is performing an optimization towards a certain objective.

Moreover, under the assumption that kinetic constants can assume different values under various environmental conditions due to enzymatic regulation, with the same procedure it has also been possible to associate, to each phenotype (and for every reaction in the model), a set of rate constants. These kinetic constants retrieved for each phenotype are parametrizations of a mechanistic model that, when dynamically simulated, is able to generate time courses in agreement with the phenotype definition. Strikingly, our method can be used to predict ensembles of rate constants that are in agreement with a given condition of interest only by providing its definition (metabolic phenotype) and a flux distribution for the same condition, obtained by means of FBA.

EColiCore2: a case study

In [6], we applied our procedure exploiting a toy metabolic model of *S. cerevisie* metabolism and filtering trajectories accordingly to a definition of the Crabtree phenotype. In the present paper, we aim at investigating a more realistic metabolic reconstruction focusing on *Escherichia coli*, the prokaryotic model organism for which a number of core models have been built in a bottom-up fashion and are currently retrievable from literature. A notable example, due to its wide exploitation, is the *E. coli* core model illustrated in Orth et al. [26]. However there is a relative scarcity of genome scale reconstruction for this bacteria. Belonging to this second category is iAF1260 [13], from which it was manually derived the EColiCore1 reconstruction, a model that has mainly testing and training purposes but is not completely consistent with the genome wide model from which it is derived.

Starting from the genome wide model from iJO1366 [27], Hädicke et al. in [18] aimed at reconstructing a metabolic model of the central metabolism of *E. coli* called EColiCore2. This model, built with the final goal to establish a reference core model for *E. coli* constraint-based analyses, has been derived reducing redundancies in biosynthetic routes and maintaining the degrees of freedom in the central metabolism, moreover this core model is completely consistent with its genome-wide counterpart. One key aspect of EColiCore2 is its ability to reproduce pivotal aspects of iJO1366, achieving a notable complexity reduction without losing its ability to depict emerging behaviors of *E. coli* metabolism.

Methods

The procedure defined in [6] and schematically represented in Figure 2, has been used here to setup the "experiments" hereafter illustrated: for every tossed random parametrization (i.e. a set of random kinetic constants, one for each mass action reaction in the model, whose value has been uniformly sampled in the interval [0,100]) we perform a number ($\mathcal{N}$) of simulations using, for each of them, constant concentrations of nutrients (e.g., glucose) and ions. Every different combination of nutrients and ions is here called metabolic phenotype. Every set of random kinetic constants is exploited to produce simulations using every metabolic phenotype defined in the study.

Run Deterministic simulations

The procedure is executed running mechanism-based simulations exploiting the deterministic framework provided by ODEs and using the LSODA solver to simulate the metabolic model until the achievement of a steady state for internal metabolites concentrations (A in Figure 2).

Calculate flux values

Afterwards, fluxes values v_i are calculated for every reaction (when the dynamic reaches the steady state) by means of the relation expressed by the mass action equation illustrated in Equation 1.

$$v_i = k_i \prod_{w=1}^{M} [\chi_w]^{\alpha_{wi}} \tag{1}$$

where k_i is the rate constant of reaction i, $[\chi_w]$ is the concentration of species w and α_{wi} the stoichiometric coefficient with which species w participate to reaction i (B in Figure 2).

Filter the experiments

Once flux values have been obtained, experiments are filtered exploiting key metabolic fluxes in order to populate ensembles of metabolic phenotypes that are in agreement with the filter definition (C in Figure 2). In particular, in this work, to filter the experiments we defined 5 different phenotypes based on FBA simulations presented in [18].

Analyze the experiments

Finally (D in Figure 2), it is possible to analyze the experiments identifying properties shared by elements of each ensemble (e.g. investigate the presence of putative subphenotypes or evaluate which reactions exhibit kinetic constants whose value is different from the average).

Phenotypes selection

To test the procedure herein described we defined 5 different metabolic phenotypes ("protected phenotypes" in [18]) built on the basis of the nutrient supplied and the oxygenation state (Table 1): exp1 - aerobic growth on glucose, exp2 - anaerobic growth on glucose, exp3 - aerobic growth on acetate, exp4 - aerobic growth on succinate, exp5 - aerobic growth on glycerol.

To evaluate the effectiveness of the procedure in discriminating the 5 phenotypes and in selecting corresponding ensembles of kinetic constants and steady state metabolic concentrations, we used the simplified model of *E. coli* named ECC2comp presented in [18] and illustrated in Figure 1. To generate ECC2comp, authors exploited NetworkReducer [12] an algorithm able to automatically compress metabolic models lumping linear chains of reactions in a single cumulative equation and removing elements (metabolites and reactions) that are non essential to represent key metabolic functions defined "protected functions". An adapted version of the compressed "core" model is composed by 114 irreversible reactions and 93 metabolites (of which 60 are internal metabolites, while 33 are external metabolites). The final core model used in this study is provided in Additional File 1.

To determine the initial concentrations of metabolites involved in the *E. coli* model, we mined the literature and we set them accordingly to the average values illustrated in the *E. coli* Metabolome Database (ECMDB) [16], an expertly curated database containing extensive metabolomic data and metabolic pathway diagrams about *Escherichia coli* (strain K12, MG1655). The ECMDB contains 3755 entries for metabolites and small molecules manually compiled including identification, taxonomy, concentrations, spectra, physical and biological properties. Information are derived from "original" data and from metabolic reconstructions, scientific articles, textbooks and other

electronic databases. For metabolites not having a concentration in ECMDB, we used the average value calculated over the retrieved values. The set of metabolic concentrations for each phenotype is provided in Additional File 2.

Phenotypes defined in this section need to be translated using a mathematical formalism in order to unequivocally establish metabolic response constraints characterizing the phenotypes. To this end we evaluated fluxes that in the ECC2comp model are proxies for the 5 phenotypes listed in Table 1.

Populating the ensembles

To perform the procedure illustrated in this section, we implemented a set of scripts in plain vanilla Python available on GitHub (see Additional File 3). Dynamic simulations of the *E. coli* "core" metabolic model (step A in Figure 2) have been performed till the reaching of the steady state exploiting a set of ordinary differential equations (ODEs) determined under the mass action kinetic assumption. The numerical integration of the ODEs system has been realized exploiting the software library LSODA (Livermore solver for ODEs with automatic method) [30] efficiently implemented in SciPy [19].

Due to large volume of data produced with simulations (stored on GitHub, see Additional File 3), we decided to separate data generation and analysis phases. An efficient way to organize and access simulation outputs, is to store them in a database. In particular we here exploited PyTables [1], a package for managing hierarchical datasets designed to efficiently and easily cope with extremely large amounts of data. PyTables makes use of the NumPy package and of the HDF5 library under the Python language.

Ensembles of kinetic constants sustaining the 5 different metabolic phenotypes have been populated by performing a large number of "experiments" conducted first randomly defining, for each of them, the set of kinetic constants and then executing a simulation for each given experimental condition (i.e., setting nutrients availability – glucose, acetate, succinate, glycerol – and oxygenation state). To populate the ensembles of kinetic constants, we filtered the experimental data set implementing conditions on the basis of reactions illustrated on columns in Table 2. Here, values indicate flux calculations obtained by simulating the ECC2comp model under the 5 different experimental conditions (see Table 1) with FBA.

In particular, to build filters we evaluated only ECC2comp reactions: (A) having non zero flux value in just one of the experimental conditions (Table 2, in bold), (B) having non zero flux value for reactions defining the experimental condition (i.e., nutrients and oxygenation state) (Table 2, in italic).

An experimental set of kinetic constants is assigned to a given ensemble (metabolic phenotype) only if, for a reaction, starting from mechanistic simulations, non zero fluxes can be calculated evaluating the last time point (assumed to be at steady state, see section Results for the determination of the steady state) and present in the filtering constraint for the same reaction. Formally these constraints relative to the phenotypes are summarized by logical expressions shown in Equations 2 to 6 where v_i represent the metabolic flux through the i reaction.

$$\textbf{exp1}: \left(v_{G6PDH2r} > 0\right) \wedge \left(v_{O2Up} > 0\right) \wedge \left(v_{GND} > 0\right) \wedge \left(v_{PGL} > 0\right) \wedge \left(v_{GLCptspp} > 0\right) \tag{2}$$

$$\textbf{exp2}: \left(v_{AcEx} > 0\right) \wedge \left(v_{ALCD2x} > 0\right) \wedge \left(v_{EthEx} > 0\right) \wedge \left(v_{GLCptspp} > 0\right) \wedge \left(v_{O2Up} = 0\right) \tag{3}$$

$$\textbf{exp3}: \left(v_{O2Up} > 0\right) \wedge \left(v_{MALS} > 0\right) \wedge \left(v_{ICL} > 0\right) \wedge \left(v_{AcUp} > 0\right) \tag{4}$$

$$\textbf{exp4}: \left(v_{O2Up} > 0\right) \wedge \left(v_{SUCCt2_{2}pp} > 0\right) \wedge \left(v_{ME2} > 0\right) \tag{5}$$

$$\textbf{exp5}: \left(v_{O2Up} > 0\right) \wedge \left(v_{GLYK} > 0\right) \wedge \left(v_{F6PA} > 0\right) \wedge \left(v_{GLYCDx} > 0\right) \tag{6}$$

Results

To test the procedure on the simplified *E. coli* model, we tossed multiple different random sets of kinetic constants, keeping the concentration of ions and exchanged species constant throughout the simulation (i.e., ac_ex, ca2_ex, cl_ex, co2_ex cobalt2_ex, cu2_ex, fe2_ex, fe3_ex, for_ex, glc_DASH_D_c, glc_DASH_D_ex, glc_DASH_D_p, h_ex, h2_ex, h2o_ex, k_ex, mg2_ex, mn2_ex, mobd_ex, MTHTHF_ex, nh4_ex, ni2_ex, o2_ex, pi_ex, so4_ex, succ_ex and zn2_ex are "in feed") time of 100 seconds, defined accordingly to [33] in order to allow the reaching of the metabolic steady state after a perturbation (e.g. a pulse of nutrient).

Every simulation is considered at steady state if every species standard's deviation on concentrations (σ) computed over the last 10% of the simulation time and summed up (normalizing over the number of species "not in feed"), is smaller than 0.1%. In a positive case, the random parametrization is retained, otherwise is dropped. To obtain a dataset of 10^4 random sets of kinetic constants, we performed a total of 11520 samplings, thereby discarding the 13.2% of performed simulations. The total computational time to produce the data set has been 1d 2h 20min to run ODEs simulations on a workstation (8 x CPU 3.8 GHz Intel Core i7, RAM 32 GB) and producing 20.3 GB of data.

The input of the filtering procedure has been a dataset composed of $5 \cdot 10^4$ simulations, i.e. 10^4 random sets of kinetic constants tested over 5 experimental conditions (the 5 protected phenotypes). Imposing a threshold considering fluxes having value less than 10^{-10} to be 0, from the dataset of $5 \cdot 10^4$ solutions 15267 has been assigned to exp1, 101 to exp2, 19616 to exp3, 22719 to exp4 and 16033 to exp5 (lower part of Figure 3 where we explore the cardinality of solutions, i.e. the number of random parametrizations that are assigned to one or more phenotypes at the same time). From Figure 3 we have been able to retrieve kinetic constants common to all the anaerobic phenotypes (1345 in Figure 3) but not for the anaerobic phenotype (exp2 – 2 in Figure 3). This reflects the consistent metabolic differences that can be pointed out *in vivo* between aerobic and anaerobic conditions.

In connection to this issue, we can notice that combinations exp2 - exp3 (23 in Figure 3) and exp2 - exp5 (25 in Figure 3) are empty sets due to the fact that in phenotype exp2 (anaerobic) reactions sustaining respiration are blocked (e.g. in TCA cycle the flux for reaction CS, leading to citrate is almost zero – see Figure 4) while in exp3 and exp5 (aerobic conditions) the same reactions are active.

To compare flux values at steady state for each reaction in the system before and after the filtering, we draw the heatmap of Figure 4 where rows list reactions, columns list set of dynamics ensembles associated to each metabolic phenotype and the color represents the median value of that ensemble for that reaction at steady state (range $[1 \cdot 10^{-13}, 1 \cdot 10^{1}]$). We made two distinct association of the dynamics to the phenotypes, the columns labeled as sC# have the dynamics assigned according to their initial condition, whereas columns labeled as fC#, the filtered ones, are populated with the dynamics that satisfies phenotypes constraints at steady state, disregarding their initial condition. Overall, it is possible to notice that flux values in sC# and fC# for a given phenotype exhibit almost always a comparable flux value, there with only few exceptions to this behavior (e.g. reactions: O2Up_reverse less active in the fC2; h2Ex, PGM, PGM_reverse, PGK less active in fC3, SUCCt2_2pp more active in fC3; GlcUp and GLCt2pp more active in fC4). Moreover, comparing the different phenotypes, it is possible to notice that exp5 (sC and fC5) has fluxes values dissimilar to the other 4 phenotypes.

To better characterize the ensembles we also plotted the median and the standard deviation for kinetic constants values retrieved for each ensemble after the filtering. Results illustrated in Figure 5 show that there are little but non negligible differences in the median of kinetic constant values for all the reactions of the model (e.g. exp1 has different median values for h2o_Ex_reverse, F6PA_reverse, PGL, PGI, GND, h2o_Ex; exp5 has constant associated to ATPM greater than the average). Furthermore, supporting findings emerged analyzing cardinalities (Figure 3), median values for the group of four aerobic phenotypes are very similar, while the the medians for anaerobic phenotype are different form the previous group.

Discussion

The analysis of average concentrations and relatives standard deviations for molecular species during time courses shed light on some relevant issues hereafter discussed.

Overall, we underline that standard deviation values (σ) are small and few parametrizations (only 13% of the total) are discarded, suggesting that for the sampled interval $[0, 1 \cdot 10^2]$ metabolism is robust towards kinetic constant variation. This parameter insensitivity has been further investigated in [17] where authors showed that many models in Systems Biology exhibit a "sloppy" spectrum of parameter sensitivities, concluding that besides the mere estimation of the parameter value, the community should focus on analyze models in a predictive fashion.

Concerning biomass (Figure 6) it is possible to notice that the concentration of this faux species is increasing

over time for all the metabolic phenotypes. Interestingly, when we tested a further experimental condition (exp0 – not used as metabolic phenotype) representing an enriched growth media (i.e., when all the nutrients are simultaneously available), this turned out not to be the condition leading to the maximal level of biomass (this is for instance the aerobic growth on succinate, exp4 – purple line).

Connected with the biomass accumulation, time courses illustrate that key metabolic pathways are active during simulation, indeed the *E. coli* is performing for example both alcoholic fermentation (Figure 7, top) and TCA cycle considered here a proxy for respiration (Figure 7, bottom).

Data supporting the actual activation of biochemical pathways in the model are also the presence of steady states for cofactors such as NAD/NADH and NADP/NADP which appear to be dynamically interconverted (Figure 8) indicating that metabolic pathways are maintaining the system energetically active and able to generate biomass.

Focusing on the set of kinetic constants assigned to the different metabolic phenotypes, the procedure illustrated in the present paper led to the population of all the 5 phenotypes and to the identification of a subset of kinetic constants assignable to the four aerobic conditions. Unfortunately, there is no single "universal" parametrization assignable to all the 5 phenotypes. This fact could be determined by different causes such as an under sampling of random kinetic constants, a too narrow sampling interval (2 orders of magnitude here), or an excessively relaxed filtering condition not allowing a complete discrimination among the phenotypes.

Furthermore, the evaluation of median and standard deviation for the kinetic constants belonging to the 5 ensembles suggest that there are only few reactions that has to be finely tuned in order to direct the system towards a specific metabolic phenotype, a fact that suggests once more that metabolism is a system particularly robust towards perturbations. In this case a global sensitivity analysis would help to more specifically investigate the issue of robustness.

Conclusions

Constraint-based models have been successfully implemented to study metabolic fluxes at steady state, nevertheless, information about the temporal evolution of the system during the transient phase preceding the steady state can not be derived from them. In addition, the metabolic concentrations at steady state can not be deduced from constraint-based methods since there is no information about kinetic constants. Computational approaches developed in [6] and exploited in the present work are an improvement designed to override limitations by exploiting mechanism-based simulations. Here, initial conditions are partially retrieved from literature (molecular concentrations) and kinetic constants are randomly determined. Figure 3 sums up the readout of the procedure: through a filtering procedure based on a loose definition of the 5 experimental conditions (metabolic phenotypes) involving some key reactions, the developed method is able to assign random sets of kinetic constants to one or more metabolic phenotypes.

With the present contribution we aimed at improving and testing a computational framework capable to retrieve ensembles of kinetic constants that can be associated to different metabolic phenotypes. It is worth to be underlined that, in contrast with constraint-base approaches, our method is not assuming that metabolism is optimized to perform a specific task. We underline that the methodology here used can be exploited to retrieve ensembles of kinetic constants for any metabolic phenotype providing only its formal definition (in terms of nutrients supplied and oxygenation state together with an estimation of initial concentrations for modeled species) and a flux distribution obtained by means of a constraint-based simulation (for which no kinetic parameters are needed).

For what concerns perspectives, we plan to better characterize the metabolic steady state by exploiting more efficient strategies to calculate if the system reaches a stationary condition. Among these strategies a promising approach is the use of the NLEQ2 nonlinear root-finding algorithm [24]. Moreover, we are considering to significantly expand the sampled set of kinetic constants through a significant speed-up of simulations achieved by means of high performance and parallel computing applied to Systems Biology modeling problems [21][22].

Competing interests

The authors declare that they have no competing interests.

Author's contributions

RC, DP, DG, MH conceived and designed the study, DG, MH acquired data DP and RC performed the analysis and interpretation of data, RC drafted the manuscript, CD and GM provided critical revision and suggestions.

Acknowledgements

This work has been supported by SYSBIO Centre of Systems Biology, through the MIUR grant SysBioNet—Italian Roadmap for ESFRI Research Infrastructures.

Author details

[1]Department of Informatics, Systems and Communication, University of Milan - Bicocca, Viale Sarca, 336, 20126 Milan, Italy. [2]SYSBIO - Centre of Systems Biology, Piazza della Scienza, 2, 20126 Milan, Italy. [3]Department of Statistics and Quantitative Methods, University of Milan - Bicocca, Via Bicocca degli Arcimboldi, 8, 20126 Milan, Italy. [4]College of Engineering, Design and Physical Sciences, Brunel University, Middlesex, UB8 3PH London, Uxbridg, UK. [5]Computer Science Department, Brandenburg University of Technology Cottbus-Senftenberg, Walther-Pauer-Str. 2, D-03046 Cottbus, Germany.

References

1. F. Alted, I. Vilata et al. PyTables: Hierarchical Datasets in Python. *http://www.pytables.org*, 2002-
2. H. W. Aung, S. A. Henry, L. P. Walker. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Industrial Biotechnology*, 9(4):215–228,2013.
3. L. Alberghina and H. V. Westerhoff, editors. *Systems Biology. Definitions and Perspectives*, volume 13 of *Topics in Current Genetics*. Springer, 2005.
4. A. B. Canelas, V. M. van Gulik, J J Heijnen. Determination of the cytosolic free NAD/NADH ratio in *Saccharomyces cerevisiae* under steady-state and highly dynamic conditions. *Biotechnology and bioengineering*, 100(4):734–743, 2008.
5. P. Cazzaniga, C. Damiani, D. Besozzi, R. Colombo, M. S. Nobile, D. Gaglio, D. Pescini, S. Molinari, G. Mauri, L. Alberghina, et al. Computational strategies for a system-level understanding of metabolism. *Metabolites*,4(4):1034–1087, 2014.
6. R. Colombo, C. Damiani, G. Mauri, D. Pescini. Constraining mechanism based simulations to identify ensembles of parametrizations to characterize metabolic features. To appear in *Lecture Notes in Bioinformatics (LNBI)*, 2017
7. S. Cortassa, M. A. Aon, B. OÕRourke, R. Jacques, H.-J. Tseng, E. Marbán, and R. L. Winslow. A computational model integrating electrophysiology, contraction, and mitochondrial bioenergetics in the ventricular myocyte. *Biophysical Journal*, 91(4):1564–1589, 2006.
8. C. Damiani, D. Pescini, R. Colombo, S. Molinari, L. Alberghina, M. Vanoni, G. Mauri. An ensemble evolutionary constraint-based approach to understand the emergence of metabolic phenotypes. *Natural Computing*, 13(3):321–331, 2014.
9. C. Damiani, R. Colombo, M. Di Filippo, D. Pescini, G. Mauri. Linking alterations in metabolic fluxes with shifts in metabolite levels by means of kinetic modeling. Accepted to *Communications in Computer and Information Science*, 2017.
10. R. H. De Deken. The Crabtree effect: a regulatory system in yeast. *Journal of General Microbiology*, 44(2):149–56, 1966.
11. M. Di Filippo, R. Colombo, C. Damiani, D. Pescini, D. Gaglio, M. Vanoni, L. Alberghina, G. Mauri. Zooming-in on cancer metabolic rewiring with tissue specic constraint-based models. *Computational Biology and Chemistry*, 62:60–69, 2016.
12. P. Erdrich, R. Steuer, S. Klamt. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC systems biology*, 9(1):48, 2015.
13. A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, B. Ø. Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information *Molecular systems biology*, 3(1):121, 2007.
14. A. Feist and B. Palsson. The biomass objective function. *Current Opinion in Microbiology*, 13(3):344–349, 2010.
15. E. P. Gianchandani, A. K. Chavali, and J. A. Papin. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(3):372–382, 2010.
16. A. C. Guo, T. Jewison, M. Wilson, Y. Liu, C. Knox, Y. Djoumbou, P. Lo, R. Mandal, R. Krishnamurthy, D. S. Wishart. ECMDB: the E. coli Metabolome Database. *Nucleic acids research*, 41(D1):D625–D630, 2013.
17. R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*,
18. O. Hädicke and S. Klamt. EColiCore2: a reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model. *Scientific Reports*, 7:39647, 2017.
19. E. Jones, E. Oliphant, P. Peterson, et al. SciPy: Open Source Scientific Tools for Python. *http://www.scipy.org/*, 2001-
20. T. W. MacFarland and J. M. Yates. Introduction to nonparametric statistics for the biological sciences using R. Springer, 2016.
21. M. S. Nobile, D. Besozzi, P. Cazzaniga, et al. GPU-accelerated simulations of mass-action kinetics models with cupSODA. *J Supercomput*, 69(1):17–24, 2014.
22. M. S. Nobile, P. Cazzaniga, D. Besozzi, et al. cuTauLeaping: A GPU-powered tau-leaping stochastic simulator for massive parallel analyses of biological systems. *PLoS ONE*, 9(3):e91963, 2014.
23. J. S. Oliveira, C. G. Bailey, J. B. Jones-Oliveira, D. A. Dixon, D. W. Gull, M. L. Chandler. A computational model for the identification of biochemical pathways in the Krebs cycle. *Journal of Computational Biology*, 10(1):57–82, 2003.
24. B. G. Olivier, J. M. Rohwer, J. H. S. Hofmeyr. Modelling cellular systems with PySCeS. *Bioinformatics*, 21(4):560–561, 2005.
25. J. D. Orth, I. Thiele, B. Ø. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–8, 2010.
26. J. D. Orth, R. M. T. Ronan , B. Ø. Palsson. Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal plus*, 4:1, 2010.
27. J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist, B. Ø. Palsson. A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011, Molecular systems biology, 7(1):535, 2011.
28. M. Papini, I. Nookaew, M. Uhlén, J. Nielsen. *Scheffersomyces stipitis*: a comparative systems biology study with the Crabtree positive yeast *Saccharomyces cerevisiae*. *Microbial Cell Factories*, 11:136, 2012.
29. J. M. Park, T. Y. Kim, S. Y. Lee Constraints-based genome-scale metabolic simulation for systems metabolic engineering, *Biotechnology advances*, 27(6):979–988, 2009.
30. L. Petzold. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations *SIAM J Sci Stat Comp*,1(4):136–148, 1983.
31. K. Smallbone, H. L. Messiha, K. M. Carroll, C. L. Winder, N. Malys, W. B. Dun, E. Murabito, N. Swainston, J. O. Dada, F. Khan, et al. A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes, *FEBS letters*, 587(17):2832–2841, 2013.
32. B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, et al. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry*, 267(17):5313–5329, 2000.
33. U. Theobald, W. Mailinger, M. Baltes, M. Rizzi, M. Reuss. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. *Biotechnology and bioengineering*, 55(2):305–316, 1997.
34. J. Zhao, H. Yu, J. Luo, Z. Cao, Y. Li. Complex networks theory for analyzing metabolic networks *Chinese Science Bulletin*, 51(13):1529––1537, 2006.

Tables

Table 1 Protected phenotypes. Phenotypes and maximal growth rate in the core model ECC2 obtained with FBA.

ID	Description	Reached μ (ECC2)
exp1	aerobic growth on glucose	0.982
exp2	anaerobic growth on glucose	0.289
exp3	aerobic growth on acetate	0.244
exp4	aerobic growth on succinate	0.492
exp5	aerobic growth on glycerol	0.563

Additional Files

Additional file 1 — ECC2C.xml
SBML file for the ECC2comp model of *E. coli* used for the analysis.

Additional file 2 — X0etc.xlsx
In tab "conc" are listed initial concentrations of metabolites for the 5 different phenotypes. In tab "FeedNoFilt" are listed metabolites provided "in feed" and metabolites not evaluated to verify the steady state.

Additional file 3 — Github
The generated dataset (ECcoliExpsParam_10_Filter_0.001.h5) and python scripts implemented for this study are deposited on a GitHub repository at http://github.com/riccardocolombo/kineticensemble

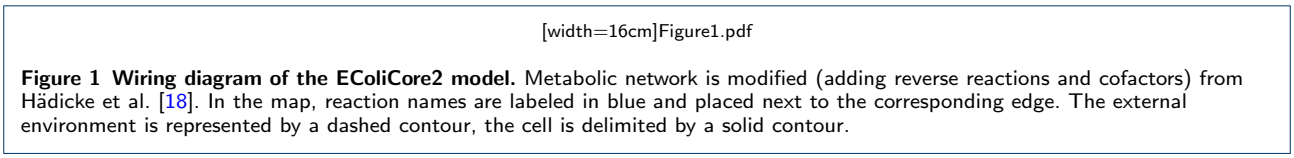

Figure 1 Wiring diagram of the EColiCore2 model. Metabolic network is modified (adding reverse reactions and cofactors) from Hädicke et al. [18]. In the map, reaction names are labeled in blue and placed next to the corresponding edge. The external environment is represented by a dashed contour, the cell is delimited by a solid contour.

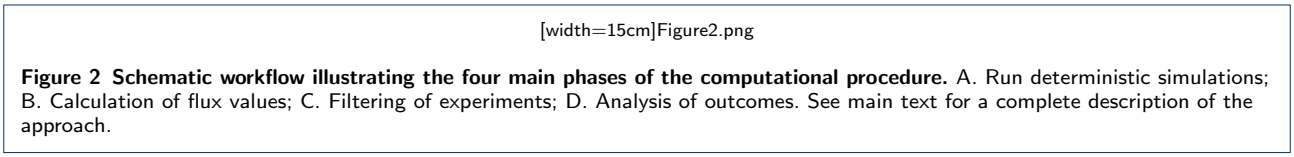

Figure 2 Schematic workflow illustrating the four main phases of the computational procedure. A. Run deterministic simulations; B. Calculation of flux values; C. Filtering of experiments; D. Analysis of outcomes. See main text for a complete description of the approach.

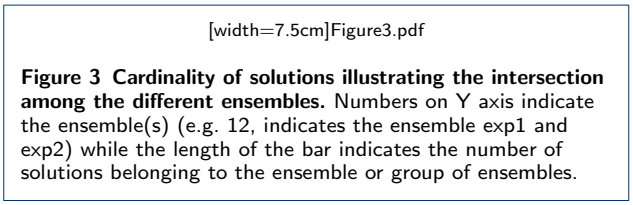

Figure 3 Cardinality of solutions illustrating the intersection among the different ensembles. Numbers on Y axis indicate the ensemble(s) (e.g. 12, indicates the ensemble exp1 and exp2) while the length of the bar indicates the number of solutions belonging to the ensemble or group of ensembles.

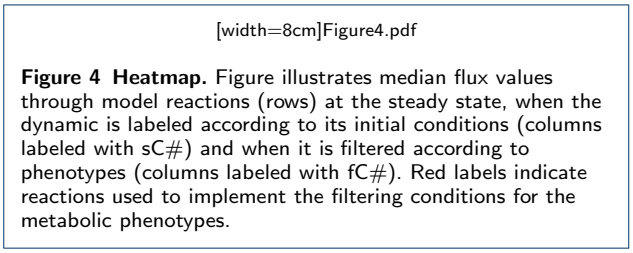

Figure 4 Heatmap. Figure illustrates median flux values through model reactions (rows) at the steady state, when the dynamic is labeled according to its initial conditions (columns labeled with sC#) and when it is filtered according to phenotypes (columns labeled with fC#). Red labels indicate reactions used to implement the filtering conditions for the metabolic phenotypes.

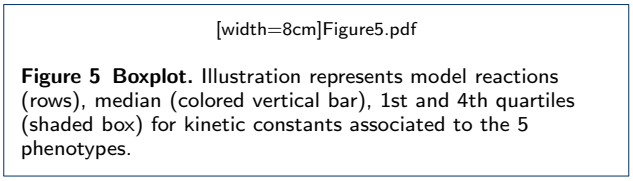

Figure 5 Boxplot. Illustration represents model reactions (rows), median (colored vertical bar), 1st and 4th quartiles (shaded box) for kinetic constants associated to the 5 phenotypes.

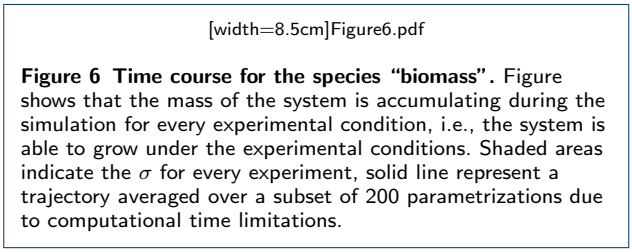

Figure 6 Time course for the species "biomass". Figure shows that the mass of the system is accumulating during the simulation for every experimental condition, i.e., the system is able to grow under the experimental conditions. Shaded areas indicate the σ for every experiment, solid line represent a trajectory averaged over a subset of 200 parametrizations due to computational time limitations.

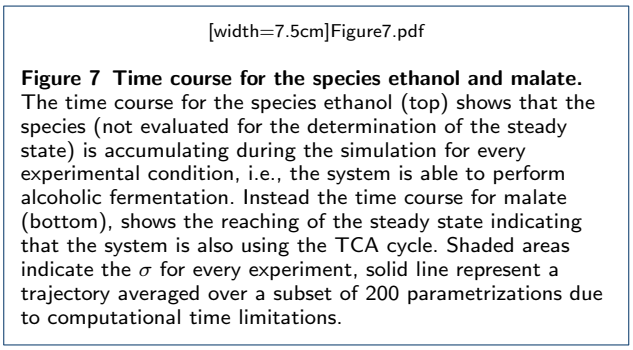

Figure 7 Time course for the species ethanol and malate. The time course for the species ethanol (top) shows that the species (not evaluated for the determination of the steady state) is accumulating during the simulation for every experimental condition, i.e., the system is able to perform alcoholic fermentation. Instead the time course for malate (bottom), shows the reaching of the steady state indicating that the system is also using the TCA cycle. Shaded areas indicate the σ for every experiment, solid line represent a trajectory averaged over a subset of 200 parametrizations due to computational time limitations.

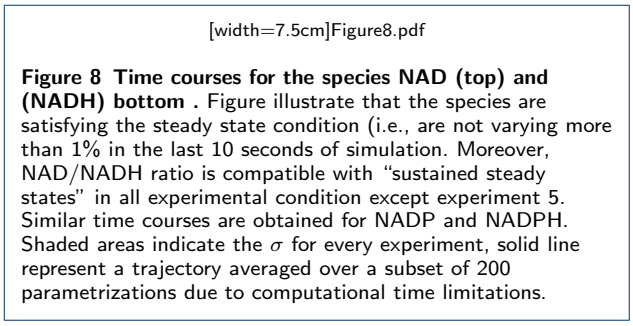

Figure 8 Time courses for the species NAD (top) and (NADH) bottom . Figure illustrate that the species are satisfying the steady state condition (i.e., are not varying more than 1% in the last 10 seconds of simulation. Moreover, NAD/NADH ratio is compatible with "sustained steady states" in all experimental condition except experiment 5. Similar time courses are obtained for NADP and NADPH. Shaded areas indicate the σ for every experiment, solid line represent a trajectory averaged over a subset of 200 parametrizations due to computational time limitations.

Table 2 Flux values used to set up filters in order to populate the 5 ensembles of kinetic constants corresponding to experimental conditions. Target fluxes have been calculated by means of FBA experiments. Reactions having non zero value in only one experimental condition are in bold; reactions defining the experimental condition (i.e., specific nutrients and oxygenation state) are in italic.

	exp1	exp2	exp3	exp4	exp5
G6PDH2r	**4.142**	0.000	0.000	0.000	0.000
O2Up	*17.587*	0.000	*9.451*	*13.735*	*10.699*
GlycUp	0.000	0.000	0.000	0.000	*10.000*
MALS	0.000	0.000	**2.627**	0.000	0.000
AcEx	0.000	**7.835**	0.000	0.000	0.000
GND	**4.142**	0.000	0.000	0.000	0.000
SUCCt2_2pp	0.000	0.000	0.000	*10.000*	0.000
PGL	**4.142**	0.000	0.000	0.000	0.000
F6PA	0.000	0.000	0.000	0.000	**4.586**
SuccUp	0.000	0.000	0.000	*10.000*	0.000
GLYCDx	0.000	0.000	0.000	0.000	**4.586**
ALCD2x	0.000	**7.806**	0.000	0.000	0.000
ICL	0.000	0.000	**2.627**	0.000	0.000
EthEx	0.000	**7.806**	0.000	0.000	0.000
GLCptspp	*10.000*	*10.000*	0.000	0.000	0.000
AcUp	0.000	0.000	**10.000**	0.000	0.000
GlcUp	*10.000*	*10.000*	0.000	0.000	0.000
ME2	0.000	0.000	0.000	**3.488**	0.000
GLYK	0.000	0.000	0.000	0.000	**5.414**