Constraining Mechanism Based Simulations to Identify Ensembles of Parametrizations to Characterize Metabolic Features

Riccardo Colombo[1,2(✉)], Chiara Damiani[1,2], Giancarlo Mauri[1,2], and Dario Pescini[1,3]

[1] SYSBIO.IT Centre of Systems Biology, Piazza Della Scienza 2, 20126 Milano, Italy
[2] Department of Informatics, Systems and Communication, University of Milano-Bicocca, Viale Sarca 336, 20125 Milano, Italy
riccardo.colombo@disco.unimib.it, {chiara.damiani,giancarlo.mauri}@unimib.it
[3] Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via Bicocca Degli Arcimboldi 8, 20126 Milano, Italy
dario.pescini@unimib.it

Abstract. Constraint-based approaches have been proven useful to determine steady state fluxes in metabolic models, however they are not able to determine metabolite concentrations and they imply the assumption that a biological process is optimized towards a given function. In this work we define a computational strategy exploiting mechanism based simulations as a framework to determine, through a filtering procedure, ensembles of kinetic constants and steady state metabolic concentrations that are in agreement with one or more metabolic phenotypes, avoiding at the same time the need of assuming an optimization mechanism. To test our procedure we exploited a model of yeast metabolism and we filtered trajectories accordingly to a loose definition of the Crabtree phenotype.

Keywords: Systems biology · Mechanistic simulations · Steady state · ODEs · Fluxes · Ensembles · Kinetic parameters · Metabolism

1 Scientific Background

It is nowadays evident that biological processes must be described in terms of complex networks of non linear interactions involving several entities (genes, transcripts, proteins, metabolites) giving rise to emergent behaviors. This awareness, coupled with the fact that biological complex systems can be effectively analyzed only by means of mathematical modeling and simulations, gave rise during the last two decades to Systems Biology, a new discipline integrating computational modeling and "wet" experimental approaches [3]. In particular the study of metabolism has widely took advantage of Systems Biology approaches

A. Bracciali et al. (Eds.): CIBB 2016, LNBI 10477, pp. 107–117, 2017.
DOI: 10.1007/978-3-319-67834-4_9

usually describing metabolic networks as hypergraphs in which nodes represent metabolites and edges indicate reactions [22].

Simultaneously to the development of computational techniques, progresses in high throughput technologies opened the "omics data" era characterized by a thrive of genome-scale metabolic reconstructions tailored on different cell types (unicellular organisms [2], healthy and diseased tissues [9]). However, technological limitations do not allow yet to investigate these genome-scale models by means of mechanism-based approaches (i.e., by simulating their temporal dynamics) [5]. For this reason, these models are currently studied through constraint-based approaches [11] exploiting information on metabolic network structure and assuming a pseudo-steady state for internal metabolites, thereby disregarding temporal evolution of the system: focusing on the metabolic steady state has been proven to be a valid assumption due to experimental evidences showing that *in vivo* metabolism reaches the steady state in few seconds [21].

The stoichiometric information retrievable from the structure of the metabolic network is the core of constraint-based modeling, indeed the stoichiometric matrix associated to a metabolic network mathematically defines changes in metabolites quantities when reactions are applied. The imposition of mass balance and of additional constraints (e.g., irreversibilities or boundaries on fluxes) allows to determine a feasible solution space containing flux distributions (i.e., flux values for each reaction in the model) reachable by the system and representing different functional states. Lastly, under the assumption that the metabolic system is optimized towards a given goal, optimization methods as flux balance analysis (FBA) [17], can be used to determine an optimal flux distribution that maximizes or minimizes a given metabolic task determined by an objective function (OF).

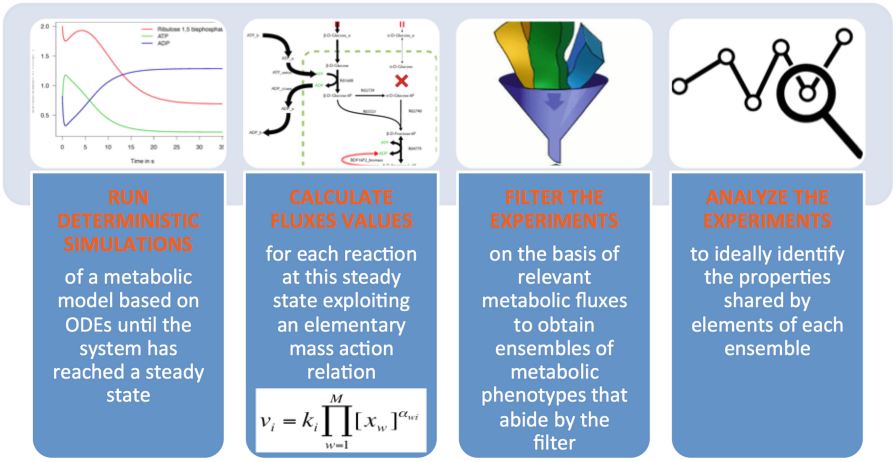


Fig. 1. Schematic workflow illustrating the four main phases of the computational procedure.

Recent studies [10] showed how the selection of the appropriate OF is essential when performing FBA investigations. This is due to the fact that often it is not possible to determine the formulation of the OF, but also to the fact that it is not possible to determine if the system is found in a sub-optimal state.

In a previous work, using an extension of FBA that we named Ensemble Evolutionary FBA (eeFBA) [6], we analyzed the capability of a cell to pursue alternative metabolic behaviors by altering its fluxes, or in other words, we determined which flux distributions are able to give rise to a specific metabolic behavior. To perform the analyses we generated a set of random OFs to be optimized by means of linear programming. We then filtered them accordingly to the definition of different metabolic phenotypes in order to obtain distinct ensembles of solutions that comply with defined phenotypes.

However, both with FBA and eeFBA it is not possible to determine the extent of metabolic concentrations when the system is at steady state due to the lack of information on kinetic constants. In the present paper we propose a novel strategy, where the ensembles of alternative phenotypes are still populated according to fluxes properties but extracting steady states from mechanism based simulations parametrized using initial concentrations retrieved from the literature and randomly sampled kinetic constants.

By doing this, we are able to infer ensembles of metabolite concentrations at steady state that are in accordance with a given metabolic phenotype independently from the definition of an appropriate OF and from the assumption that the cell is optimizing towards a specific objective.

2 Materials and Methods

With the procedure here devised and illustrated in Fig. 1, we setup several "experiments" that we define as follows: for each random parametrization we execute a number ($\mathcal{N}$) of simulations using, for each of them, a different but constant concentration of nutrient (e.g., glucose).

To perform the procedure, we firstly run deterministic simulations of a metabolic model based on ODEs and exploiting the LSODA solver until the system has reached a steady state. We then calculate fluxes values v_i for each reaction i at this steady state exploiting an elementary mass action relation:

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

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. At this stage, we filter the experiments on the basis of relevant metabolic fluxes to obtain ensembles of metabolic phenotypes that abide by the filter. In particular, in this work, to filter the experiments we used the same phenotype definition already published in [6]. The last step of the procedure is to analyze the experiments to ideally identify the properties shared by elements of each ensemble such as the presence of putative subphenotypes.

To test the procedure herein developed we defined the metabolic phenotype expressed by the Crabtree effect [8], a well known biological phenomenon taking place in some yeasts like *Saccharomyces cerevisiae*, implying a production of ethanol by fermentation when high concentrations of glucose are available in the extracellular environment preferring fermentation—regardless the availability of oxygen—with respect to the more energetically efficient oxidative phosphorilation (OXPHOS). In our test case we call "Crabtree-positive" the phenotype exhibiting the enhanced fermentation. On the same line, we call "Crabtree-negative" the phenotype of those yeasts (like *Kluyveromyces*) not showing the peculiar experimental characteristics.

To evaluate the effectiveness of the procedure in discriminating the two phenotypes and in selecting corresponding ensembles of kinetic constants and steady state metabolic concentrations, we used a simplified model of yeast metabolism (illustrated in Fig. 2) developed in [6] and designed to take into account only those pathways (metabolites and reactions) involved in the emergence of the Crabtree effect (CE).

To determine the initial concentrations of metabolites involved in the yeast model, we mined the literature and we set them accordingly to the average values illustrated in Smallbone et al. [20], and Canelas et al. [4]. From the *in vivo* experimental point of view this effect can be observed as the concomitant presence of alcoholic aerobic fermentation and reduction of OXPHOS rate when the glucose uptake from the medium progressively increases (e.g. by means of incremental glucose pulses added to yeast medium culture). The given biological definition of the CE however must be mathematically translated in order to formally and unequivocally determine metabolic response constraints defining the Crabtree-positive (C$\oplus$) and Crabtree-negative (C$\ominus$) phenotypes.

To this end we evaluated fluxes that in the model are proxies for OXPHOS—the sum of fluxes (v_o) for the two reactions summing up respiration, illustrated with A and B in Fig. 2— and alcoholic fermentation—ethanol secretion flux (v_e) represented with C in Fig. 2—traditionally defining CE. Furthermore, due to experimental observations [18] of marked differences between the two yeast phenotypes with regard to glucose uptake kinetics, we consider v_o and v_e as a function of glucose uptake v_g represented by series of glucose uptake concentrations maintained "in feed" and defined by the expression $\{v_g^i \mid \forall\, i < j \quad v_g^i < v_g^j\}_{i,j=1,\ldots,L}$ representing the set of constant glucose concentrations at which each simulation is run.

To formally define C$\oplus$ we start from the observation that under this phenotype, the ratio of alcoholic fermentation over respiration increases proportionally to the glucose uptake, implying that at the maximum "in feed" glucose concentration evaluated, the ethanol secretion flux must have higher flux values with respect to the respiratory flux. In addition for the C$\oplus$ case, we imposed that respiration and ethanol secretion should have at least one value different from zero. Formally these constraints relative to the C$\oplus$ phenotype are summarized by logical expressions shown in Eq. 2.

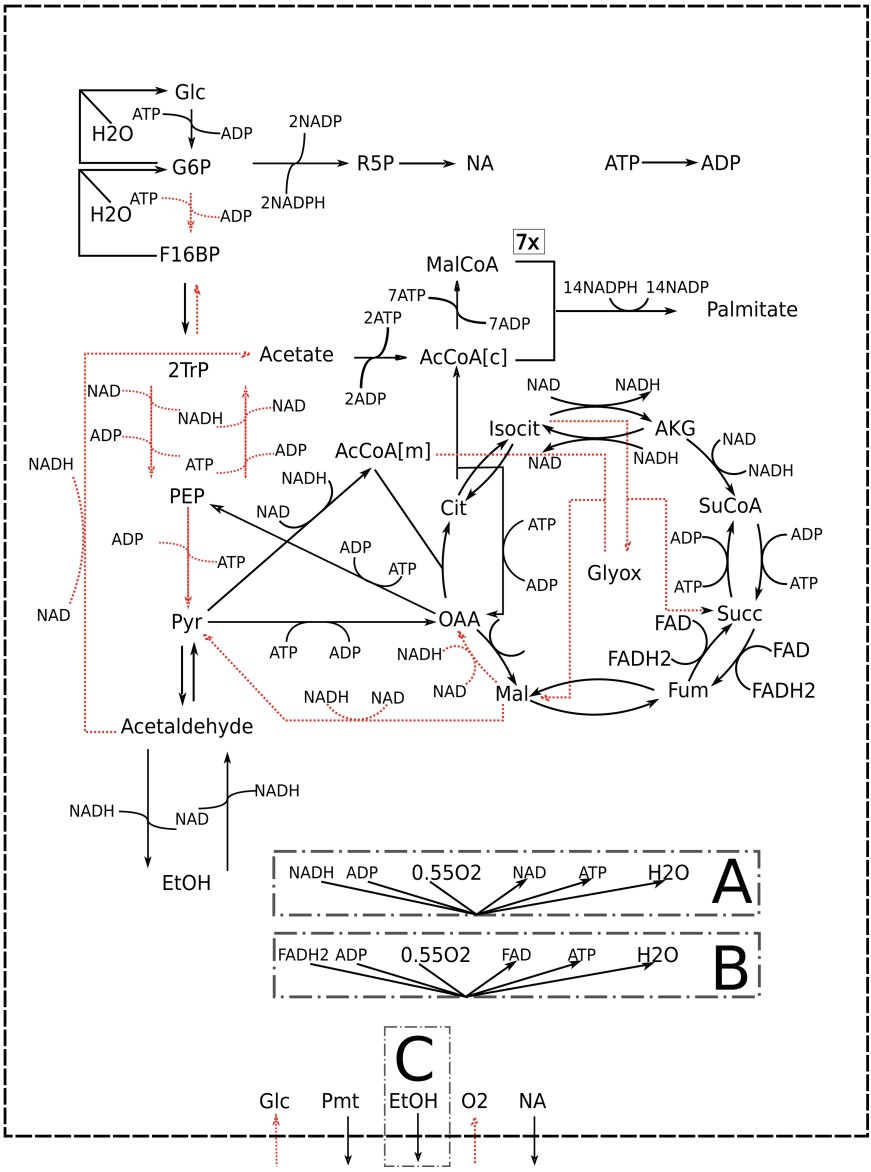


Fig. 2. Diagram of the yeast metabolism core model. The model consists in 48 reactions and 34 metabolites representing the main metabolic pathways. Only glucose is considered as carbon source. The directionality of reactions has been imposed according to literature. A and B indicate reactions modeling respiration, while C labeled reaction indicates the proxy for fermentation. Black solid arrows indicate reactions that significantly differ between C⊕ and C⊖ accordingly to the Kolmogorov-Smirnov test described in Sect. 3, red dashed arrows are not significant reactions accordingly to the same test. (Color figure online)

$$\left(\sum_{l=1}^{L} v_e(v_g^l) > 0\right) \wedge \left(\sum_{l=1}^{L} v_o(v_g^l) > 0\right) \wedge \left(v_e(v_g^L) - v_o(v_g^L) > 0\right) \tag{2}$$

Once we filtered the experiments to populate the C$\oplus$ ensemble, on the same line we filtered the remaining experiments by means of a logical expression formulated for C$\ominus$ in which we do not impose specific fermentation levels, but instead we verify that respiration flux does not overtake ethanol secretion as a function of glucose concentrations, implying also that respiration must increase as a function of "in feed" glucose concentrations. In Eq. 3 we define the expression for the C$\ominus$ phenotype as follows:

$$\left(v_o(v_g^1) - v_o(v_g^L) < 0\right) \tag{3}$$

In this paper, we focus on the relationship between fermentation and respiration at the extremes of the considered interval of glucose uptake (rather than at intermediate levels) to examine a wider extent of emergent behaviors that are able to satisfy Eqs. 2 and 3.

In order to simulate the dynamics of the model until it reaches the steady state, we use a set of ordinary differential equations (ODEs) defined assuming a mass action kinetic. To numerically solve the ODEs system, we use the efficient software library LSODA (Livermore solver for ODEs with automatic method) [19]; in particular in this paper we used the LSODA version implemented in SciPy [12] (a collection of scientific packages for Python). To gain more flexibility we uncoupled data production (simulations) and their analysis. In order to manage a wide amount of data, to store them and to perform queries we setup a database exploiting PyTables [1], a package for managing hierarchical datasets designed to efficiently and easily cope with extremely large amounts of data.

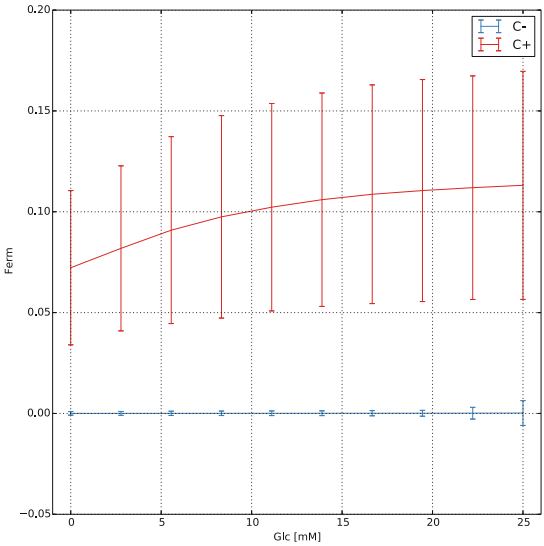
PyTables is built on top of the HDF5 library, using the Python language and the NumPy package.

To obtain the ensembles of metabolic phenotypes that sustain C$\oplus$ and C$\ominus$, we firstly performed several "experiments" randomly defining, for each of them, the set of kinetic constants and performing a simulation for every level of nutrient (glucose). Once we obtained the experimental data set we populated the ensembles implementing the Boolean filter defined in Expression Eq. 2 for the C$\oplus$ ensemble and Expression Eq. 3 for the C$\ominus$ ensemble.

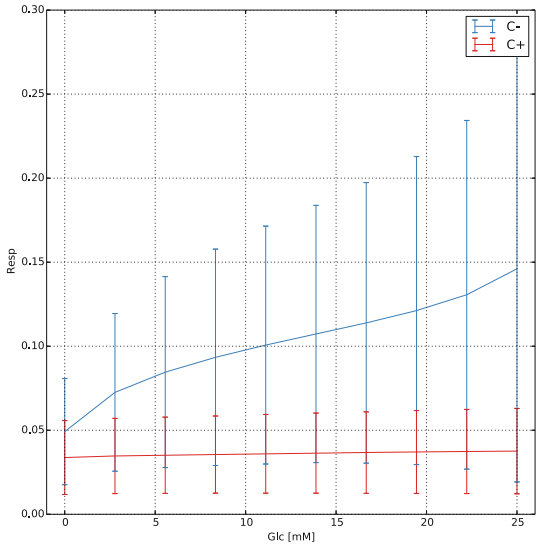
3 Results

To test the procedure on the simplified yeast model, we tossed multiple different random sets of kinetic constants, for each of them we performed 10 different simulations evenly sampling the glucose interval $[0, 25]$ mMol, keeping the concentration constant throughout the simulation (i.e. glucose is "in feed") time of 50 s (defined accordingly to [21]).

After the simulation we checked that the system reached the steady state: we calculated the standard deviation (σ) for every species in the system during



(a) Fermentation flux



(b) Respiration flux

Fig. 3. Lines represent average flux values for C⊕ (red) and C⊖ ensembles at the variation of the glucose level, associated error bars indicate the $\pm\sigma$ values. (Color figure online)

the last 10% of the simulation time, subsequently we summed the σs and we divided the value for the number of species not "in feed". If the value was less than 1% we considered the system at steady state and we retained the random parametrization, otherwise we discharged it. We iterated the procedure to obtain 10^4 random sets of kinetic constants, discarding a total of 23199 parametrizations. The total computational time to produce the data set has been 5.5 h to run ODEs simulations on a MacBookPro (CPU 2.6 GHz Intel Core i7, RAM 16 GB) and producing 268 Mb of data. After filtering the data set we obtained an ensemble of 7901 C$\ominus$ solutions and ensemble of 29 C$\oplus$ solutions.

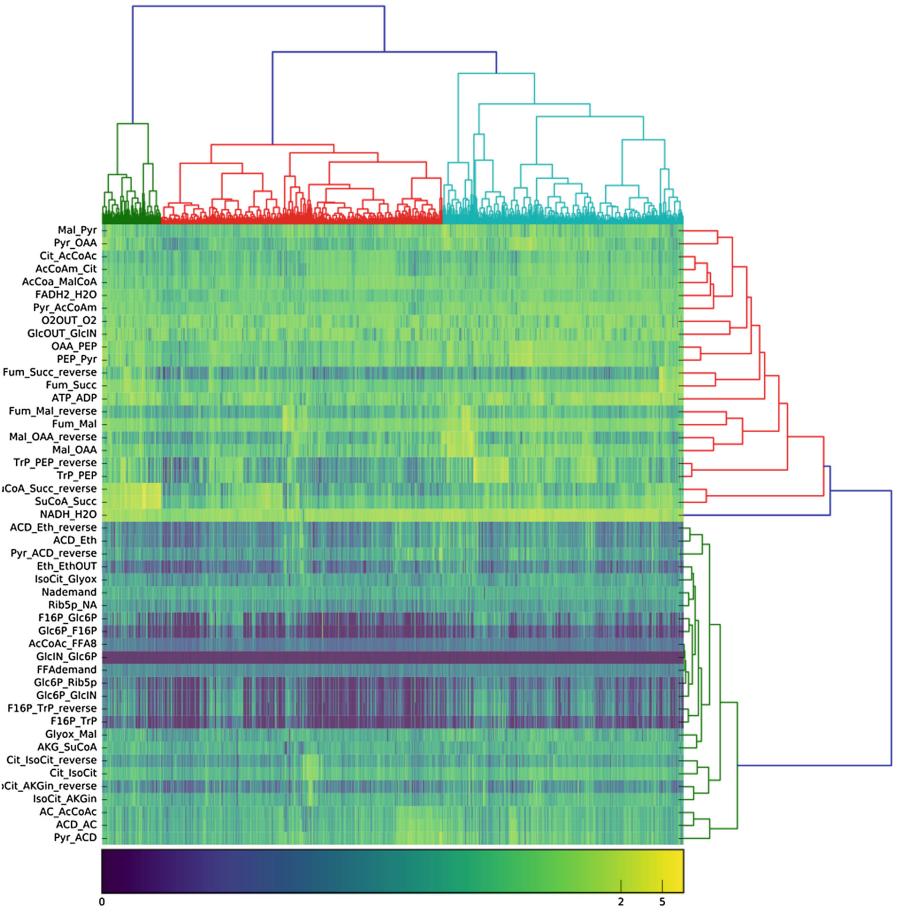


Fig. 4. Hierarchical clustering performed on both reactions (rows) and solutions (columns). The associated heatmap illustrates flux values for every reaction and for every solution at steady state. Reactions are indicated as substrate_product, with substrate and product being one among the reaction substrates and products respectively. Reverse reactions are considered separately and are indicated with the suffix "reverse". (Color figure online)

Results indicate that less than 0.3% of the total random parametrizations led to a C⊕ phenotype, while the 79% of the cases were assigned to the C⊖ ensemble.

Moreover, the 21% of the random parametrizations were not assignable to one of the two ensembles. Comparing these results with those previously published with the eeFBA approach [6] it is possible to notice that in spite of having different proportions (2% against 0.3% for C⊕ and 11% against 79% for C⊖) there still a higher probability to observe the metabolic response typical of the Crabtree-negative yeasts than of observing the Crabtree effect. The most marked difference with the eeFBA approach can be identified when comparing solutions that were not assigned to any ensemble (87% against 21%).

To improve the soundness of the biological readout, we analyzed the resulting ensembles by means of a hierarchical clustering performed on the global data set computing euclidean distances on the flux values matrix (i.e. the flux distribution in the different solutions). In particular we clustered both reactions (rows in Fig. 4) and solutions (columns in Fig. 4), overall we stress the fact that C⊕ and C⊖ solutions form two main separate clusters (C⊕ : small green cluster on columns, C⊖ : red and light blue clusters on columns) and, for what concerns reactions, on rows three main clusters are clearly evident (red, green and blue); however there is no strict correspondence between reaction clustering and biochemical pathways (e.g., reactions of both TCA cycle and glycolysis are split among two different clusters).

To further characterize the ensembles identifying those fluxes significantly different between the two ensembles we exploited a Kolmogorov-Smirnov test [13], a non-parametric hypothesis test procedure able to discriminate if two samples derive from the same distribution without investigating the actual shape of the distributions. The statistical test has been performed using the flux values of each reaction at steady state for all the 10 non-null levels of glucose.

As a result we obtained 33 reaction fluxes (over a total of 48) that are significantly different for at least 9 out of 10 levels of glucose (indicated with black solid arrows in Fig. 2), setting a p-value threshold of 0.05. Obviously among these are the reactions used to discriminate between the C⊖ and C⊕ phenotypes.

4 Conclusion

Constraint-based models have been effectively used to study metabolic fluxes at steady state, however they are not able to provide information on the temporal evolution of the system during the transient phase previous to the steady state. Moreover with constraint-based methods it is not possible to infer the metabolic concentrations at steady state due to the fact that there is no information about kinetic constants (a metabolic flux is determined using the Eq. 1).

The technique developed in the present work together with the current calculating capacity, enables to overcome this limitations by means of mechanism-based simulations (parametrized used random kinetic constants and initial molecular concentrations retrieved in literature), calculation of metabolic fluxes at

steady state and selection of those solutions (sets of kinetic constants and molecular concentrations at steady state) that are in agreement with phenotype definitions. Results shown in Fig. 3, illustrates that by a simple filtering of pivotal fluxes for respiration and fermentation at the boundary levels of nutrient uptake, the devised method is able to discriminate between C⊕ and C⊖ metabolic phenotypes. Indeed, as it happens *in vivo*, fermentation flux is higher in simulated C⊕ solutions with respect to C⊖ for every level of glucose, while respiration shows an opposite behavior (the average respiration flux in C⊖ ensemble is higher than in C⊕). Moreover, the hierarchical clustering illustrating a separation between C⊕ and C⊖ solutions as well as the Kolmogorov-Smirnov test identifying as statistically different pivotal reactions for the identification of the two phenotypes, provide a further support for the obtained results.

In conclusion, in this work we provided a proof of concept for a computational framework able to discriminate between different metabolic phenotypes in order to retrieve ensembles of putative steady state metabolic concentrations and kinetic constants without the need of assuming that the cell is optimized towards a specific behavior.

Currently we are exploiting the computational framework here described to investigate the linking between alterations in metabolic fluxes and shifts in metabolite levels. Briefly, preliminary results suggest that metabolite levels exhibit a poor correlation with variations in flux values of reactions directly involving them. At the same time, results show a stronger linkage with variations in fluxes that are distantly located in the network [7].

In the next future, we plan to expand the set of sampled random kinetic constants and to implement a more efficient strategy to determine the metabolic steady state (e.g. exploiting the NLEQ2 algorithm [16]). Lastly, we will consider the feasibility of using parallel and high performance computing techniques [14,15] in order to speed-up simulations.

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

References

1. Alted, F., Vilata, I., et al.: PyTables: hierarchical datasets in python (2002). http://www.pytables.org
2. Aung, H.W., Henry, S.A., Walker, L.P.: Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. Ind. Biotechnol. **9**(4), 215–228 (2013)
3. Alberghina, L., Westerhoff, H.V. (eds.): Systems Biology: Definitions and Perspectives. Topics in current genetics, 13th edn. Springer, Heidelberg (2005)
4. Canelas, A.B., van Gulik, V.M., Heijnen, J.J.: Determination of the cytosolic free NAD/NADH ratio in saccharomyces cerevisiae under steady-state and highly dynamic conditions. Biotechnol. Bioeng. **100**(4), 734–743 (2008)

5. Cazzaniga, P., Damiani, C., Besozzi, D., Colombo, R., Nobile, M.S., Gaglio, D., Pescini, D., Molinari, S., Mauri, G., Alberghina, L., et al.: Computational strategies for a system-level understanding of metabolism. Metabolites **4**(4), 1034–1087 (2014)
6. Damiani, C., Pescini, D., Colombo, R., Molinari, S., Alberghina, L., Vanoni, M., Mauri, G.: An ensemble evolutionary constraint-based approach to understand the emergence of metabolic phenotypes. Nat. Comput. **13**(3), 321–331 (2014)
7. Damiani, C., Colombo, R., Di Filippo, M., Pescini, D., Mauri, G.: Linking alterations in metabolic fluxes with shifts in metabolite levels by means of kinetic modeling. In: Rossi, F., Piotto, S., Concilio, S. (eds.) WIVACE 2016. CCIS, vol. 708, pp. 138–148. Springer, Cham (2017). doi:10.1007/978-3-319-57711-1_12
8. De Deken, R.H.: The crabtree effect: a regulatory system in yeast. J. Gen. Microbiol. **44**(2), 149–156 (1966)
9. Di Filippo, M.: Zooming-in on cancer metabolic rewiring with tissue specic constraint-based models. Comput. Biol. Chem. **62**, 60–69 (2016)
10. Feist, A., Palsson, B.: The biomass objective function. Curr. Opin. Microbiol. **13**(3), 344–349 (2010)
11. Gianchandani, E.P., Chavali, A.K., Papin, J.A.: The application of flux balance analysis in systems biology. Wiley Interdisc. Rev. Syst. Biol. Med. **2**(3), 372–382 (2010)
12. Jones, E., Oliphant, E., Peterson, P., et al.: SciPy: open source scientific tools for python (2001). http://www.scipy.org/
13. MacFarland, T.W.W., Yates, J.M.M.: Introduction to Nonparametric Statistics for the Biological Sciences Using R. Springer, Cham (2016)
14. Nobile, M.S., Besozzi, D., Cazzaniga, P., et al.: GPU-accelerated simulations of mass-action kinetics models with cupSODA. J Supercomput **69**(1), 17–24 (2014)
15. Nobile, M.S., Cazzaniga, P., Besozzi, D., et al.: CuTtauLeaping: a GPU-powered tau-leaping stochastic simulator for massive parallel analyses of biological systems. PLoS ONE **9**(3), e91963 (2014)
16. Olivier, B.G., Rohwer, J.M., Hofmeyr, J.H.S.: Modelling cellular systems with PySCeS. Bioinformatics **21**(4), 560–561 (2005)
17. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? Nat. Biotechnol. **28**(3), 245–248 (2010)
18. Papini, M., Nookaew, I., Uhlén, M., Nielsen, J.: Scheffersomyces stipitis: a comparative systems biology study with the crabtree positive yeast saccharomyces cerevisiae. Microb. Cell Fact. **11**, 136 (2012)
19. Petzold, L.: Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. SIAM J. Sci. Stat. Comp. **1**(4), 136–148 (1983)
20. Smallbone, K., et al.: A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. FEBS Lett. **587**(17), 2832–2841 (2013)
21. Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., Reuss, M.: In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: I experimental observations. Biotechnol. Bioeng. **55**(2), 305–316 (1997)
22. Zhao, J., Yu, H., Luo, J., Cao, Z., Li, Y.: Complex networks theory for analyzing metabolic networks. Chin. Sci. Bull. **51**(13), 1529–1537 (2006)