On the Explicit Use of Enzyme-Substrate Reactions in Metabolic Pathway Analysis

Angelo Lucia[1(✉)], Edward Thomas[1], and Peter A. DiMaggio[2]

[1] Department of Chemical Engineering,
University of Rhode Island, Kingston, RI 02881, USA
alucia@uri.edu
[2] Department of Chemical Engineering,
Imperial College London, London SW7 2AZ, UK

Abstract. Flux balance (or constraint-based) analysis has been the mainstay for understanding metabolic networks for many years. However, recently Lucia and DiMaggio [1] have argued that metabolic networks are more correctly modeled using game theory, specifically Nash Equilibrium, because it (1) captures the natural competition between enzymes, (2) includes rigorous chemical reaction equilibrium thermodynamics, (3) incorporates element mass balance constraints, and therefore charge balancing, in a natural way, and (4) allows regulatory constraints to be included as additional constraints.

The novel aspects of this work center on the explicit inclusion of enzyme-substrate reactions at the cellular length scale and molecular length scale protein docking information in metabolic network modeling. This multi-scale information offers the advantages of directly (1) computing cellular enzyme concentrations and activities, (2) incorporating genetic modification of enzymes, and (3) encoding the effects of age-related changes in enzymatic behavior (e.g., protein misfolding) within any pathway. Molecular length scale binding histograms are computed using protein-ligand docking and directly up-scaled to the cellular level. A small, proof-of-concept example from the Krebs cycle is presented to illustrate key ideas. Numerical results show that the proposed approach provides a wealth of quantitative enzyme information.

1 Introduction

While flux balance analysis (FBA) or constraint-based modelling (CBM) and its many variants have been used for metabolic pathway analysis for some time (see, for example, [2–12]), the Nash Equilibrium (NE) approach recently proposed by Lucia and DiMaggio [1] and extended in Lucia et al. [13] far outperforms all FBA and CBM methods because it is a first principles approach that incorporates rigorous chemical reaction equilibrium and elemental mass balances. As a result of its formulation, the NE approach has superior capabilities that naturally address (1) the competition among enzymes for resources in the metabolic pool, (2) substrate and co-factor charge balancing, and (3) regulatory controls.

G. Nicosia et al. (Eds.): MOD 2017, LNCS 10710, pp. 88–99, 2018.
https://doi.org/10.1007/978-3-319-72926-8_8

The NE approach is also predictive. While there are competing kinetic-based models of metabolic pathways, these methods generally require a large number of parameters that cannot be directly measured and must be determined through model regression. Thus kinetic approaches are correlative and not predictive, particularly for experimental conditions that deviate from the training data used to determine model parameters.

1.1 A Nash Equilibrium Approach to Metabolic Pathways

In this sub-section, only a brief summary of the Nash Equilibrium approach to modelling metabolic pathways is presented. The reader is referred to [1,13] for details. The key ideas behind the NE approach to metabolic pathway analysis are to (1) represent the network using first principles rigorous chemical reaction equilibrium and element mass balances and (2) view enzymes as players in a multi-player game, in which each enzyme minimizes the change in Gibbs free energy for the biochemical reaction it catalyzes subject to appropriate elemental mass balances (i.e., conservation of mass of carbon, hydrogen, oxygen, nitrogen, phosphorous and sulfur). This leads to the representation of any metabolic network as a set of N nonlinear programming sub-problems (NLPs), where the network objective function is defined by:

$$\frac{G(v)}{RT} = \sum_{j=1}^{N} \min \frac{G_j(v_j)}{RT} \tag{1}$$

where G is the Gibbs free energy, v denotes the vector of metabolic fluxes, R is the universal gas constant, T is absolute temperature and j denotes the j^{th} NLP sub-problem. The details of the Gibbs free energy and heat of formation data required in the NE formulation and a description of the cellular fluid model can be found in [1,13].

1.2 Element Mass Balances and Charge Balancing

The NE approach provides a natural way to ensure that atomic mass balances are satisfied within and across the metabolic network. In particular, appropriate element mass balances are included within each NLP sub-problem. Movement from one NLP sub-problem to the next automatically guarantees that element mass balances are satisfied since the outputs of one reaction are typically some or all of the inputs to the next reaction(s) in the network. For example, consider the first reaction in the Krebs (TCA) cycle given by:

$$\begin{aligned} C_4H_2O_5^{-2} \; + \; C_{23}H_{34}N_7O_{17}P_3S^{-4} \; + \; H_2O \\ \rightleftharpoons C_6H_5O_7^{-3} \; + \; C_{21}H_{32}N_7O_{16}P_3S^{-4} \; + \; H^+ \end{aligned} \tag{2}$$

in which oxaloacetate and acetyl-CoA combine with water to form citrate, coenzyme A, and hydrogen ions in the presence of the enzyme citrate synthase. The corresponding element balances for the reaction in Eq. 2 are the hydrogen,

nitrogen, oxygen, and carbon balances and are represented by the matrix-vector equation:

$$A_{ik,j}\ v_{k,j} = M_{i,j} \tag{3}$$

where the index i corresponds to the individual elements, index j denotes subproblem j (here $j = 1$ since the reaction under consideration is the first reaction in the network), index k corresponds to individual metabolites/cofactors, vector $v_{k,j} = (v_{1,1}, v_{2,1}, v_{3,1}, v_{4,1}, v_{5,1})^T$, vector $M_{i,j} = (H_1, N_1, O_1, C_1)^T$ and H_1, N_1, O_1, C_1 represent the molar amounts of hydrogen, nitrogen, oxygen, and carbon in that order for subproblem $j = 1$. The $k = 5$ independent fluxes (chemical species) in Eq. 3 are water, acetyl-CoA, co-enzyme A, oxaloacetate, and citrate respectively. H^+ is a dependent flux. Additionally, the full matrix $A_{ik,j}$ is

$$A_{ik,j=1} = \begin{pmatrix} 2 & 34 & 32 & 2 & 5 & 1 \\ 0 & 7 & 7 & 0 & 0 & 0 \\ 1 & 17 & 16 & 5 & 7 & 0 \\ 0 & 23 & 21 & 4 & 6 & 0 \end{pmatrix} \tag{4}$$

Note that while phosphorous and sulfur are also present, these elements, along with nitrogen, are fixed in the ratio N_7P_3S so only one of the element mass balances in the subset N, P, S is linearly independent and can be used as a constraint.

In the second reaction in the TCA cycle, the citrate from reaction 1 binds with the enzyme aconitase to form isocitrate. Considering only the overall metabolite reaction, we have

$$C_6H_5O_7^{-3} \rightleftharpoons iC_6H_5O_7^{-3} \tag{5}$$

Note that there is only a single independent element balance for this second reaction since it is simply an isomerization reaction.

The key points here are that:

1. If the element balances for the first reaction are satisfied, then the amount of citrate that is available for the second reaction preserves element mass balances.
2. Element balancing automatically accounts for correct charge balancing.

2 Explicitly Incorporating Enzyme-Substrate Reactions

To our knowledge there is no approach to metabolic network modeling and analysis that explicitly includes the binding and unbinding of substrates with enzymes and therefore no methodology capable of predicting enzyme concentrations and activities or their impact. The inclusion of enzyme-substrate binding/unbinding reactions opens up a wide range of possibilities that can provide important quantitative information such as:

1. The amount of a given enzyme needed to catalyze a given reaction.
2. The impact of changes in enzymatic activity on the steady-state behavior of metabolic networks.

Including enzyme-substrate reactions within the Nash Equilibrium formulation of a metabolic network is straightforward. However, it is well known that enzyme-substrate binding (also called protein docking) exhibits many minima and saddle points on the energy surface. To obtain this data, molecular length scale protein docking software (e.g., AutoDock [14], which was used exclusively in this work) can be used to create a look-up table of ranked enzyme-substrate binding energies from histograms and this information can be easily up-scaled to the cellular length scale for use in the NE calculations. In general, information for conformations (or docking solutions) with the lowest Gibbs free energy are up-scaled to and used at the cellular length scale - unless there are reasons for choosing a different conformation.

2.1 Enzyme-Substrate Reactions

General enzyme-substrate reactions can be described using the simple two-reaction sequence:

$$E + S \rightleftharpoons E - S \tag{6}$$

$$E - S \rightleftharpoons E + P \tag{7}$$

which represent, respectively, the binding of the enzyme, E, and substrate, S, to form a stable complex, denoted by $E - S$, followed by rearrangement, cleaving, or some other interaction and then subsequent unbinding to regenerate enzyme and produce product, P. Within the Nash Equilibrium framework, binding and unbinding are considered to reach chemical equilibrium.

2.2 An Example of Binding and Unbinding Reactions

Consider the simple example of the binding of citrate (shown in blue and red) with the iron sulfate complex (orange and yellow) of the enzyme aconitase as shown in Fig. 1. A sample of the output produced by AutoDock is given in Appendix A.1. Note that binding takes place at sites in the large binding pocket containing the iron sulfate complex, which is surrounded by α helix and β sheet portions of the enzyme. Moreover, this large pocket is the 'correct' binding site and the one that results in the production of isocitrate.

2.3 Multiple Minima from Protein Docking

It is well known that protein-ligand docking is a multi-minima problem, in which there are a large number of minima and saddle point solutions. For example, for the aconitase-citrate illustration, using just twenty-five (25) random starting points AutoDock located twenty-five different solutions or conformations in three separate binding pockets. Figure 2 shows three key solutions, which have corresponding Gibbs free energies of binding of -11.38, -6.72, and -6.22 kcal/mol respectively. Solution 1 (top left) is the global minimum and the one that leads to the conversion of citrate to isocitrate. Solutions 2 (top right) and 3 (bottom)

Fig. 1. Docking of citrate with aconitase (Protein Data Bank (PDB) ID: 1C96). (Color figure online)

represent conformations in which citrate binds to sites that are above the large binding pocket in the center of the enzyme and, as a result, do not convert citrate to isocitrate. These solutions can be ranked based on their respective Gibbs free energies and clustering information from the corresponding histogram (see Appendix A.2) can be used to define cluster efficiencies given by the rule

$$k_{ijm} = \frac{n_{ijm}}{\sum_{l=1}^{M} n_{ijl}}, \quad i = 1, ..., n_S;\ j = 1, ..., n_E;\ m = 1, ..., M \tag{8}$$

where i is a substrate index, n_S is the number of substrates, j denotes the enzyme index, n_E is the number of enzymes, m is the cluster index, and M is the total number of clusters. In the illustrative example, there is one substrate (citrate), one enzyme (aconitase) and three clusters, which gives

$$k_{111} = \frac{23}{25} = 0.92;\ k_{112} = \frac{1}{25} = 0.04;\ k_{113} = \frac{1}{25} = 0.04 \tag{9}$$

respectively for solutions 1, 2 and 3 shown in Fig. 2. Normally, the cluster with the highest efficiency is chosen unless there is reason to choose a different cluster efficiency. There are a few key points to note:

1. The total time for docking simulations is not prohibitive. For this small example, the total time to compute all twenty-five solutions was ∼10 min on a

laptop. This clearly indicates that generating a database of enzyme-substrate binding energies, clusters, and a cluster ranking is tractable - even for a large number of enzymes and substrates. Once this data is determined for a given enzyme-substrate pair, it never has to be computed again.

2. Our proposed approach for including molecular length scale enzyme-substrate information relies on the Protein Data Bank (PDB) and a tool for computing relative binding Gibbs free energies and cluster information (e.g., AutoDock).
3. In a more general sense, ranked enzyme-substrate efficiencies open up many possibilities, not the least of which is the capability to include behavior such as changes in enzyme activity due to genetic modifications, misfolding, ageing, and so on, provided of course the structural changes in the enzyme resulting from these modifications can be determined (e.g. via protein folding calculations, molecular dynamics, etc.).

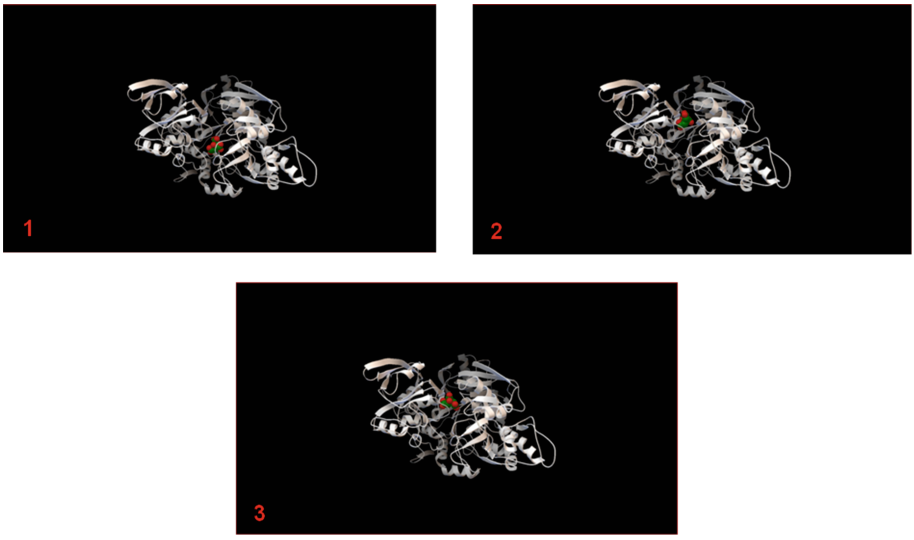

Fig. 2. Multiple binding solutions for aconitase-citrate docking.

2.4 A Multi-scale Methodology for Including Enzyme-Substrate Reactions

It is instructive to illustrate for the reader the way in which enzyme-substrate reactions are included in the NE framework for the purpose of determining enzyme activities and concentrations. Here again we use the example of citrate conversion to isocitrate to illustrate. The conversion of citrate to isocitrate can be treated as a two-reaction sequence, in which the first reaction is given by

$$C_6H_5O_7^{-3} \; + \; E \; \rightleftharpoons \; E - C_6H_5O_7^{-3} \tag{10}$$

where E = aconitase. The corresponding NLP sub-problem for this reaction is

$$\min \frac{G_1(v_1, v_2, v_3)}{RT} \tag{11}$$

subject to element balances

$$5v_1 + 0v_2 + 5v_3 = H \tag{12}$$

$$0v_1 + 1v_2 + 1v_3 = E \tag{13}$$

where H and E represent the amount of hydrogen and aconitase in the initial pool and the subscripts 1, 2 and 3 correspond to citrate, aconitase, and the aconitase-citrate complex, respectively. It is important for the reader to understand that enzymes are treated by assuming they undergo no change in mass. Only the substrates or metabolites undergo chemical change. As a result, element balancing of the enzyme is unnecessary, which avoids scaling and other complicating issues due to the typically large number of residues and corresponding molecular weight of enzymes.

The second reaction, Eq. 7, is the unbinding of isocitrate from aconitase and results in the NLP sub-problem given by

$$\min \frac{G_1(v_1, v_2, v_3)}{RT} \tag{14}$$

subject to the same set of mass balance constraints (i.e., Eqs. 12 and 13). The only difference here is that subscript 1 in Eqs. 12 and 13 now represents isocitrate. Note that charge balancing associated with the overall conversion of citrate to isocitrate remains unchanged in the presence of enzyme-substrate reactions.

2.5 Enzyme Activity

One way to get a measure of enzyme activity is to plot the rate of reaction as a function of substrate concentration. This leads to the simple expression for the rate of conversion of citrate to isocitrate (here for *E. coli*) given by

$$V_0 = \frac{v_P}{V} \tag{15}$$

where v_P represents the steady-state flux of product, V is the reaction volume of the appropriate compartment of the cell (e.g., cytosol in *E. coli*), and V_0 is the rate of catalysis. Biochemists usually express the rate of catalysis in terms of Michaelis-Menton kinetics using an equation of the form

$$V_0 = V_{max} \frac{[S]}{[S] + K_M} \tag{16}$$

where V_{max} is the maximum rate of catalysis, $[S]$ is the substrate concentration, and K_M is the Michaelis constant, which is defined as the substrate concentration that gives a reaction velocity equal to $V_{max}/2$.

Another important metric of enzyme activity is turnover number. The turnover number, k_{cat}, is the reaction rate constant associated with the conversion of enzyme-substrate complex to enzyme plus product (i.e., the rate constant associated with Eq. 7). The turnover number can be computed using the expression

$$k_{cat} = \frac{V_{max}}{[E]_T} = \frac{V_{max}}{[E] + [E - S]} \tag{17}$$

where $[E]_T$ is the total enzyme concentration.

3 Numerical Results

Numerical results for the inclusion of enzyme-substrate reactions within a Nash Equilibrium formulation are presented. To make the presentation clear, we focus on the citrate-aconitase-isocitrate example for *E. coli* from the previous sections (Eqs. 10–14) to provide proof-of-concept. Of specific interest is the quantitative determination of enzyme concentrations and activity metrics. All computations were performed on a Dell Inspiron laptop with the Lahey-Fujitsu LF95 compiler.

Table 1. Aconitase conversion of citrate to isocitrate at 25 °C.

Enzyme/Substrate	Flux (mmol/s)	Concentration (mM)
Citrate	0.066900	2.79040
Aconitase	0.106647	4.44828
Aconitase-citrate	0.022484	0.93782
Isocitrate	0.033813	1.41034

Table 1 shows the steady-state fluxes and concentrations for substrates, enzyme and enzyme-substrate complex for the conversion of citrate to isocitrate in the presence of aconitase for an initial pool of 0.1 mmol of citrate and 0.13 mmol of aconitase and temperature of 25 °C. Table 1 also shows that under the given conditions ∼33% of the citrate is converted to isocitrate and that ∼82% of the aconitase is regenerated.

Figure 3 shows the rate of isocitrate as a function of substrate concentration for 0.13 mM/s of aconitase and initial citrate concentrations ranging from 0.2 mM to 4 M at 25 °C and a cytosolic volume for *E. coli* of $V = 1\,\mu\text{m}^3$ (Fig. 1 in [15]).

Note that the reaction velocity increases as substrate concentration increases until the aconitase is saturated. At that point there are no more active sites available and the reaction velocity (enzyme activity) reaches a maximum rate of 0.091 M/s.

Figure 4, on the other hand, is an enlargement of Fig. 3 at low substrate concentration, which is necessary to graphically determine the Michaelis constant, K_M. From the value $V_{max}/2 = 0.0455$ M/s, the Michaelis constant is $K_M = 4.91$ mM. Using Eq. 17 and the saturated enzyme concentration of 5.386 mM, the predicted turnover number is 16.90 s^{-1}. Both metrics, $K_M = 4.91$ mM

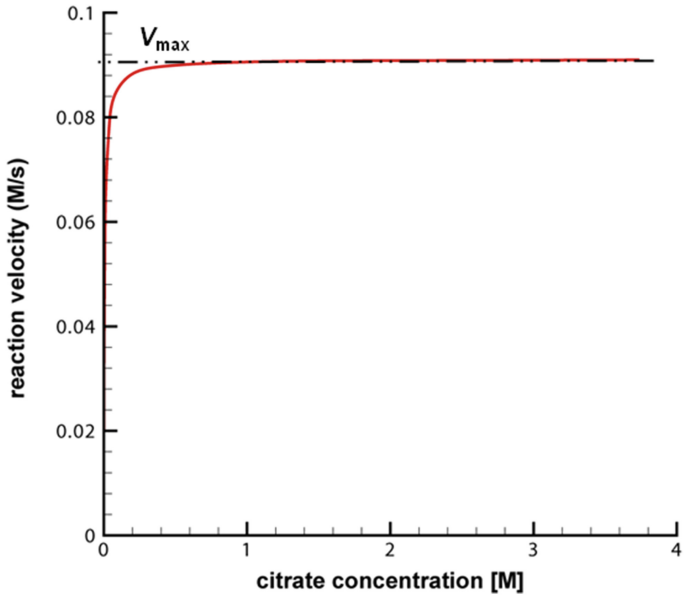

Fig. 3. Reaction velocity as a function of citrate concentration.

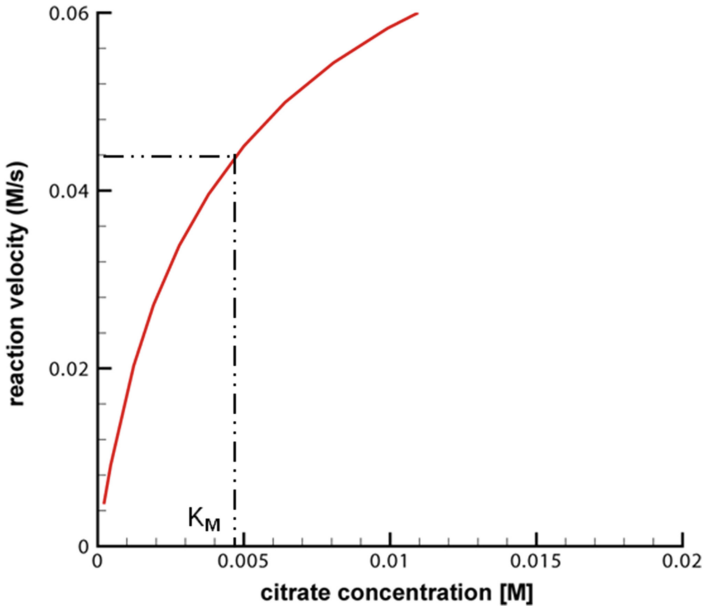

Fig. 4. Enlargement of V_0 as a function of low citrate concentration.

and $k_{cat} = 16.90\ \mathrm{s}^{-1}$ match published experimental data (i.e. $K_M = [1.16 - 11]$ mM in [16] and $k_{cat} = 13.5\ \mathrm{s}^{-1}$ in [17]) quite well.

Finally, from the histogram information, we use an enzyme efficiency of 92% and adjust the initial pool of active enzyme from 0.13 mmol to 0.1196 mmol. For the illustration in Table 1, this simply gives a slightly lower conversion of citrate to isocitrate (32 vs. 33%), slightly lower enzyme concentration (4.05 vs. 4.45 mM), and shifts the curve in Fig. 3 downward yielding a lower value of V_{max} (0.083 vs. 0.091 M/s) and a lower K_M (4.69 vs. 4.91 mM).

4 Conclusions

The inclusion of enzymatic reactions in a Nash Equilibrium framework for metabolic pathway analysis was presented. Results for a simple illustration of the conversion of citrate to isocitrate in the presence of aconitase clearly show that the proposed approach can be used to predict key metrics used to describe enzyme activity as well as enzyme and enzyme-substrate complex concentrations. The results presented in this work easily generalize to any enzymatic reaction and can be used for strain development via genetic modification, understanding epigenetics, therapeutics, and other biological tasks.

Appendix A.1

See Fig. 5.

```
LOWEST ENERGY DOCKED CONFORMATION from EACH CLUSTER
_____________________________________________________


Keeping original residue number (specified in the input PDBQ
file) for outputting.

MODEL        21
USER    Run = 21
USER    Cluster Rank = 1
USER    Number of conformations in this cluster = 23
USER
USER    RMSD from reference structure       = 0.503 A
USER
USER    Estimated Free Energy of Binding    =  -11.38 kcal/mol
[=(1)+(2)+(3)-(4)]
USER    Estimated Inhibition Constant, Ki   =    4.52 nM
(nanomolar)  [Temperature = 298.15 K]
USER
USER    (1) Final Intermolecular Energy     =  -12.88 kcal/mol
USER        vdW + Hbond + desolv Energy     =   -8.69 kcal/mol
USER        Electrostatic Energy            =   -4.18 kcal/mol
USER    (2) Final Total Internal Energy     =   +1.43 kcal/mol
USER    (3) Torsional Free Energy           =   +1.49 kcal/mol
USER    (4) Unbound System's Energy  [=(2)] =   +1.43 kcal/mol
USER
USER
USER
```

Fig. 5. Sample output from protein-ligand docking software.

Appendix A.2

See Fig. 6.

Clus -ter Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Clus	Histogram 5 10 15 20 25
1	-11.38	21	-10.27	23	########################
2	-6.72	13	-6.72	1	#
3	-6.22	6	-6.22	1	#

Fig. 6. Sample AutoDock histogram for protein-ligand docking.

References

1. Lucia, A., DiMaggio, P.A.: A Nash equilibrium approach to metabolic network analysis. In: Pardalos, P.M., Conca, P., Giuffrida, G., Nicosia, G. (eds.) MOD 2016. LNCS, vol. 10122, pp. 45–58. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-51469-7_4
2. Varma, A., Palsson, B.O.: Metabolic flux balancing: basic concepts, scientific and practical use. Nat. Biotechnol. **12**, 994–998 (1994)
3. Kauffman, K.J., Prakash, P., Edwards, J.S.: Advances in flux balance analysis. Curr. Opin. Biotechnol. **14**, 491–496 (2003)
4. Holzhutter, H.G.: The principles of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur. J. Biochem. **271**, 2905–2922 (2004)
5. Julius, A.A., Imielinski, M., Pappas, G.J.: Metabolic networks analysis using convex optimization. In: Proceedings of the 47th IEEE Conference on Decision and Control, p. 762 (2008)
6. Smallbone, K., Simeonidis, E.: Flux balance analysis: a geometric perspective. J. Theor. Biol. **258**, 311–315 (2009)
7. Murabito, E., Simeonidis, E., Smallbone, K., Swinton, J.: Capturing the essence of a metabolic network: a flux balance analysis approach. J. Theor. Biol. **260**(3), 445–452 (2009)
8. Lee, S., Phalakornkule, C., Domach, M.M., Grossmann, I.E.: Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. Comput. Chem. Eng. **24**, 711–716 (2000)
9. Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V.: Thermodynamic metabolic flux analysis. Biophys. J. **92**, 1792–1805 (2007)
10. Mahadevan, R., Edwards, J.S., Doyle, F.J.: Dynamic flux balance analysis in diauxic growth in Escherichia coli. Biophys. J. **83**, 1331–1340 (2002)
11. Patane, A., Santoro, A., Costanza, J., Nicosia, G.: Pareto optimal design for synthetic biology. IEEE Trans. Biomed. Circuits Syst. **9**(4), 555–571 (2015)
12. Angione, C., Costanza, J., Carapezza, G., Lio, P., Nicosia, G.: Multi-target analysis and design of mitochondrial metabolism. PLoS One **9**, 1–22 (2015)
13. Lucia, A., DiMaggio, P.A., Alonso-Martinez, D.: Metabolic pathway analysis using Nash equilibrium. J. Optim. (2017, in press)

14. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J.: Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. **16**, 2785–2791 (2009)
15. Milo, R.: What is the total number of protein molecules per cell volume? A call to re-think some published values. BioEssays **35**, 1050–1055 (2013)
16. Schomburg, I., Hofmann, O., Bänsch, C., Chang, A., Schomburg, D.: Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine. Gene Funct Dis. **3**, 109–118 (2000)
17. Villafranca, J.J., Mildvan, A.S.: The mechanism of aconitase action: I. Preparation, physical properties of the enzyme, and activation by iron (II). J. Biol. Chem. **246**, 772–779 (1971)