Genome-scale estimate of the metabolic turnover of \textit{E. Coli} from the energy balance analysis

D De Martino

Institute of Science and Technology Austria (IST Austria), Am Campus 1, Klosterneuburg A-3400, Austria
Center for life nanoscience, Istituto Italiano di Tecnologia, CLNS-IIT, Viale Regina Elena 291, I-00161, Rome, Italy
E-mail: ddemarti@ist.ac.at

Keywords: metabolic networks, metabolic turnover, energy balance analysis
Supplementary material for this article is available online

Abstract

In this article the notion of metabolic turnover is revisited in the light of recent results of out-of-equilibrium thermodynamics. By means of Monte Carlo methods we perform an exact sampling of the enzymatic fluxes in a genome scale metabolic network of \textit{E. Coli} in stationary growth conditions from which we infer the metabolites turnover times. However the latter are inferred from \textit{net} fluxes, and we argue that this approximation is not valid for enzymes working nearby thermodynamic equilibrium. We recalculate turnover times from \textit{total} fluxes by performing an energy balance analysis of the network and recurring to the fluctuation theorem. We find in many cases values one of order of magnitude lower, implying a faster picture of intermediate metabolism.

1. Introduction

The recent wealth of data coming from genome sequencing in biology \cite{1} is eager for unifying schemes, interpretations and insights that could come from physics. In particular metabolism, the ubiquitous and highly conserved enzymatic network devoted to free energy transduction in every cell, has been the subject of structural reconstructions at the scale of the whole genome \cite{2}. Metabolism has deep physical roots and thus a long standing tradition of physical modeling efforts \cite{3}. A current challenge faced by physics thus concerns the extension of such efforts in large scale models. On one hand we lack detailed information on many parameters, on the other, even simple models with minimal assumptions lead to difficult computational issues. Much attention has been paid to the structural properties of metabolic networks \cite{4}, but on the other hand metabolism is inherently dynamical and a fundamental inherently physical question regard the assessment of its typical timescales. In this article we will analyze metabolites turnover times at the scale of the whole genome in a metabolic network model of \textit{E. Coli}. We will point out that such analysis requires to integrate thermodynamic information and thus the evaluation of an energy balance analysis of the network \cite{5,6}.

If we model a metabolic system in terms of the dynamics of the concentration levels, assume well-mixing (no space) and neglect noise (continuum limit), we still have a very large non-linear dynamical system whose parameters are not currently known in their entirety to the scale of the whole genome. For a chemical reaction network in which \(M\) metabolites participate in \(N\) reactions with the stoichiometry encoded in a matrix \(S = \{S_{ij}\}\), the concentrations \(c_i\) change in time according to mass-balance equations

\[
\dot{c} = S \cdot v
\]  

(1)

where \(v_i\) is the flux of the reaction \(i\) that is in turn a function of the concentration levels \(v_i(c)\). Upon considering a steady state (homeostasis), i.e. a flux configuration satisfying

\[
S \cdot v = 0
\]  

(2)

we could in principle rigorously determine the timescales by performing a linear stability analysis of these steady states, i.e. upon linearizing the laws \(v_i(c)\) and finding the spectrum of the resulting matrix. Such calculation requires knowledge of the elasticity coefficients \cite{7} \(\frac{\partial v_i}{\partial c_j}\) and in turn of the reaction laws with their parameters, that is not the case in large scale models. A widely employed approximation, if at least fluxes and concentrations are known, is to consider the metabolites turnover times \cite{8,9} \(\tau\), i.e. the ratio between the
concentration of a given metabolite $c$ and the flux of production $P$ (or equivalently destruction $D$), that is the same in the steady state, i.e.

$$
\dot{c} = P - D = 0, \quad \tau = \frac{c}{P}.
$$

The turnover time is a measure of the average life-time of the molecules before undergoing conversion and of the time required to reach the stationary state [10]. However, we point out that, especially nearby thermodynamic equilibrium, net fluxes result from the difference between forward and backward contributions $\nu = \nu^+ - \nu^-$. This would imply that production and destruction fluxes split as well and the resulting turnover time could be lower [10]:

$$
\dot{c} = (P^+ + D^-) - (D^+ + P^-) = 0, \quad \tau = \frac{c}{P^+ + D^-}.
$$

It should be understood that these contributions directly affect even the calculation of relaxation times from a more rigorous linear stability analysis. Now, if we have information about the net flux $\nu$ and the free energy $\Delta G_i$ we can estimate the backward and forward contribution $\nu^+, \nu^-$ from a simplified form [11] of one of the main result in out-of-equilibrium thermodynamics, the fluctuation theorem [12–14]:

$$
\nu^+ \nu^- = e^{-\Delta G_i / RT}.
$$

It has been shown that this simplified form is valid for the mass action law and the reversible Michaelis–Menten kinetics, and it has been proposed to be valid in general [11]. For instance, consider a metabolite in the red cell, glucose-6-phosphate, produced in glycolysis [15] by an irreversible reaction, hexokinase ($\Delta G_1 \approx -29$ kJ mol$^{-1}$) and consumed by a reversible one, phosphoglucoisomerase ($\Delta G_2 \approx -2.9$ kJ mol$^{-1}$). We can calculate at $RT = 2.5$ kJ mol$^{-1}$ that the turnover time $\tau$ from net fluxes overestimates the one $\tau$ from total fluxes by a factor

$$
\frac{\tau_0 - \tau}{\tau} \approx \frac{1}{e^{\frac{29}{2.5}} - 1} \approx 45%.
$$

In the following we will show the results of a sampling of the steady state fluxes in a genome scale metabolic network model for the bacterium E. Coli under fixed average stationary growth conditions. Then, the results of an energy balance analysis of the network will be presented in order to estimate total fluxes from the simplified form of the aforementioned fluctuation theorem. We will calculate metabolites turnover times from net fluxes, correct them from estimate of total fluxes and show that in the latter case they can be much lower. We will finally draw out some conclusions, for instance regarding how this new faster picture of the intermediate metabolism affects the well-mixing hypothesis.

2. Results and discussion

We consider the steady state fluxes of the metabolic network model of E. Coli iJR904 [16] in a glucose limited minimal medium under fixed average stationary growth conditions (see section 3.1).

In constraints-based modeling, apart from mass balance constraints, fluxes are bounded in certain ranges $\nu_i \in [\nu_i^{\text{min}}, \nu_i^{\text{max}}]$ that take into account thermodynamic irreversibility, kinetic limits and physiological constraints. The set of constraints

$$
S \cdot \mathbf{v} = 0, \quad \nu_i \in [\nu_i^{\text{min}}, \nu_i^{\text{max}}]
$$

defines a convex closed set in the space of reaction fluxes: a polytope from which feasible steady states can be efficiently inferred with Monte Carlo methods [17] (see section 3.2.1). In order to fix average stationary growth conditions, fluxes have to be sampled from max entropy Boltzmann distributions.

Once we have the flux distributions, we can single out for each metabolite the net production flux, that is the sum of positive definite terms (and that it is equal to the net consumption flux under our steady state assumption), if we have information about the concentration levels we can thus calculate the turnover times:

$$
\dot{c}_i = P_i - D_i = 0, \quad P_i = \sum_i \theta(S_{ij} \nu_j)S_{ij} \nu_i, \quad D_i = - \sum_i \theta(-S_{ij} \nu_j)S_{ij} \nu_i, \quad \tau_i = \frac{c_i}{P_i},
$$

where $\theta$ is the Heaviside step function.

We show in figure 1 in ordered fashion the average of metabolites production fluxes (log scale, red plus). They span 8 orders of magnitude ranging from $10^3$ to $10^{-5}$ mmol gDW$^{-1}$ h$^{-1}$, thus giving an highly heterogeneous picture of intermediate metabolism. This is consistent with the time-hierarchy hypothesis, by which typical metabolic timescales should be highly heterogeneous in order to suppress instabilities [18–20]. As we mentioned in the introduction we can go a step beyond and calculate forward and backward contribution of reaction fluxes by using the fluctuation theorem from the knowledge of the reactions free energies. These can be estimated by performing an energy balance analysis of the network, that is performed along the lines described in the section 3.2.2.

Once free energies have been retrieved, the backward and forward contributions to fluxes can be calculated from the fluctuation theorem:

$$
\nu_i^+ = \frac{\nu_i}{1 - e^{\Delta G_i / RT}}, \quad \nu_i^- = \frac{\nu_i}{e^{-\Delta G_i / RT} - 1}.
$$
The total fluxes are reported in figure 1 (green crosses) in log scale comparing them with net values. We can see that for many metabolites the corrected production flux can be one order of magnitude higher when it is processed by enzymes working nearby thermal equilibrium, that would imply faster relaxation times of the system, for example in its response to perturbations. In particular we highlight that 2-phosphoglycerate has the biggest change, i.e. the total production flux is 32 times the net for comparison the total production flux of ATP is \(\approx 8\) times bigger than the net.

We illustrate this correction at work for ATP in figure 2 where we show the distribution of the turnover times inferred from net and total fluxes assuming a concentration of \(9.6\) mM [21]; we pass from a mean of 1.8 s (consistent with previous estimates reported in databases [22]) to 0.24 s upon using the correct value from total fluxes. In figure 3 we plot in log scale turnover times from net and total fluxes for several metabolites for which we have used the concentration levels measured in [21] We present in table 1 for comparison the turnover estimates from net and total fluxes of some key metabolites ruling the energy charge (AMP, ADP, ATP) and reductive power (NAD, NADH, NADP, NADPH) of the metabolic system. We report in an excel sheet in the supplementary material the full data on net and total production fluxes as well as turnover times for the metabolites of the network being analyzed. Finally we highlight that a good lesson of the kind of insights that the fluctuation theorem can give for the turnover estimate comes from the analysis of pyruvate turnover. On one hand, the larger (net) flux of pyruvate production and consumption (around \(20\) mmol GDW\(^{-1}\)h\(^{-1}\)) is carried respectively by GLCpts (glucose transport via PEP:PYR phospho-transferase system) and pyruvate dehydrogenase, being both strongly out of equilibrium (\(\Delta G \approx 20\) kcal mol\(^{-1}\)). On the other hand, pyruvate is engaged in reversible reactions, like alanine transaminase (ALALT_L), that albeit carrying quite low net flux (around \(1.7\) mmol GDW\(^{-1}\)h\(^{-1}\) in absolute value), it turns out to give a consistent contribution to the pyruvate turnover given its free energy value nearby equilibrium (\(\Delta G \approx 0.3\) kcal mol\(^{-1}\)). In figure 4 we show net, backward and forward flux distributions of ALALT_L (x-axis is in log scale).

3. Materials and methods

3.1. Data

The network employed is the E. Coli iJR904 [16] genome scale reconstruction. This network consists of \(N = 1075\) reactions among \(M = 761\) metabolites. Upon considering a glucose limited minimal medium we are left, after removing blocked reactions and leaf metabolites with \(N = 667\) reactions among \(M = 450\) metabolites. The resulting polytope has dimension \(D = 233\). We set a maximal glucose uptake of \(u_g = 12\) mmol GDW\(^{-1}\)h\(^{-1}\) and fix a minimal ATP maintenance of \(v_{ATP} = 7.6\) mmol GDW\(^{-1}\)h\(^{-1}\). The rest of the bounds are essentially the reversibility assignments of the model iJR904. In order to obtain realistic flux distributions we fix the average growth rate at \(\langle \lambda \rangle = 0.65\) h\(^{-1}\) through max entropy Boltzmann sampling. In regard to the data for the energy balance analysis we recur to the free energy formation estimates coming from a genome scale application of the group contribution methods from [23, 24]. In order to calculate turnover times we have recurred to the measure of absolute concentrations reported in [21] when available. The values of chemical potentials enter as an initial prior for the relaxation algorithm in order to solve Gibbs inequalities for reaction fluxes and perform the energy balance analysis (see subsection of methods). When metabolites concentration are not available from [21], they have been set to a initial

![Figure 1. Net and total metabolites production fluxes and from sampling steady states of the metabolic network model of E. Coli iJR904 in a minimal glucose medium under fixed average stationary growth conditions.](image-url)
value of $10^{-5}$ M. Data are integrated using the standard formula widespread for equilibrium diluted systems $g = g_0 + RT \log(c)$. Given the error on chemical potentials from data, the prior has been extracted from a Gaussian distribution. Values are referred to standard biochemical conditions for temperature and pH. In order to implement the latter conditions the chemical potential of water and hydrogen ion are kept fixed during the relaxation dynamics.

### 3.2. Computational methods

#### 3.2.1. Sampling of steady states

A uniform sampling of a convex polytope in high dimension is usually performed with Markov chain Monte Carlo methods, since an exact enumeration of the vertex would be infeasible due to their exponential number and static rejection methods [25] would suffer as well from high-dimensionality issues. We have recurred to well known hit and run markov chain (HR) that is defined as follows. Given a $D$-dimensional convex set $P$, from which one wants to sample from, and a point inside $x_t \in P$

(i) Choose a uniformly distributed direction $\theta_t$, that is, a point extracted from the uniform distribution on the $D$-dimensional unit sphere. This can be done with the Marsaglia method, i.e. by generating $D$ independent gaussian random
variables $\theta_i$ with zero mean and unit variance, and then normalizing the vector to unit length;

(ii) Extract $\lambda$ uniformly from the interval $[\lambda_{\text{min}}, \lambda_{\text{max}}]$, where $\lambda_{\text{min}}$ ($\lambda_{\text{max}}$) is the minimum (maximum) value of $\lambda$ such that $x_i + \lambda \theta_i \in P$;

(iii) Compute the point $x_{t+1} = x_t + \lambda \theta_i$, increment $t$ by one and start again. This algorithm has been investigated in much detail in mathematical literature, showing nice convergence properties: it is guaranteed by detailed balance to converge to the uniform distribution with a mixing time that scales as [27]

$$
\tau = O\left(D^2 \left(\frac{R}{r}\right)^3\right)
$$

(10)

where $\tau$ and $R$ are the radius of respectively the biggest (smallest) inscribed (inscribing) sphere. The factor $R/r$ (‘sandwiching ratio’) can lead to ill-conditioning issues, since the timescales of metabolic fluxes are typically very heterogeneous. Such factor can be reduced to a polynomial of the space dimension with a polynomial time algorithm that finds a rounding ellipsoid [28]. We have rounded the polytope with an ellipsoid founded with the flux variability analysis [29], that has been shown to reduce the sandwiching ratio to values that allow an efficient sampling [30].

Finally, in order to fix the average growth rate we sample fluxes from a max entropy Boltzmann distribution, i.e. we extract them from

$$
P(\nu) \propto e^{\beta \lambda(\nu)}.
$$

(11)

This distribution is the less unbiased distribution that fixes the average growth rate, interpolating between the uniform sampling ($\beta = 0$) and the flux balance analysis max growth solution ($\beta \to \infty$). In order to fix $\langle \lambda \rangle = 0.65 \text{ h}^{-1}$ we have to fix the lagrange multiplier $\beta = 400$.

3.2.2. Energy balance analysis
Reactions free energies should have the opposite sign of the flux reaction directions ($\text{sign}(\nu) \Delta G_i \leq 0$) and upon decomposing them in terms of the chemical potentials, these Gibbs inequalities become a system of linear inequalities for the chemical potentials:

$$
\xi_{ij} = -\text{sign}(\nu_i)S_{ij},
$$

$$
\sum_{\mu} \xi_{ij} \delta_{ij} > 0 \quad \forall i.
$$

(12)

A solution to this system can be found by relaxational algorithms [31, 32] starting from an experimental prior (see section 3.1) By means of the Gordan theorem such a system has solutions if and only if the dual system:

$$
\sum_i \xi_{ij} k_i = 0 \quad \forall \mu,
$$

$$
k_i \geq 0, \quad k \neq 0
$$

(13)

has no solutions. The solutions of the dual system are the so called unfeasible cycles [33, 34] that for this network have been enumerated in [33].

4. Conclusions
In this work we have revisited the notion of metabolic turnover in the light of recent results in out-of-equilibrium thermodynamics. The net flux of a reaction working nearby thermodynamic equilibrium results from a contribution of backward and forward fluxes, whose value can be inferred from the fluctuation theorem upon knowledge of the free energy. The higher resulting total fluxes lead to effectively faster relaxation times in metabolic systems, a notion that can be captured at an approximated level by the computation of metabolites turnover times. We performed a sampling of the steady states of an E. Coli genome scale constraint based metabolic network model under stationary fixed average growth conditions, we performed an energy balance analysis of the network and we have estimated total production fluxes from the fluctuation theorem. We have shown that metabolites turnover times estimated in this way can be as far as one order of magnitude lower than the ones inferred from net fluxes. Such reduction of turnover times could lead in principle...
to values that are below typical metabolites diffusion times and thus it could affect the well-mixing hypothesis, that is at the core of our approach and constraint-based modelling in general. An approximate estimate in which we consider a diffusion constant of $D \approx 200 \, \mu m^2 s^{-1}$, and that the diameter of E. Coli is $d \approx 1 \mu m$, lead to an order of magnitude estimate $t_d \approx \frac{d^2}{6D} \approx 10^{-3} \, s$. The turnover time of all the metabolites we have calculated is above this threshold apart from NADP and adenosine. Such analysis could be performed in principle for any metabolic network upon knowledge of the free energies landscape. In particular, we have chosen to analyse typical steady states in a glucose limited minimal medium, but any experimental measurement could be used to further constraint the space and analyse the turnover for any particular experimental condition. Finally, a more rigorous estimate of the true relaxation times of a metabolic system would require genome-scale insights on enzymatic kinetic laws, including allosteric regulations, an aspect that would require further investigations.

Acknowledgments

The author thanks A De Martino for interesting discussions. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. [291734].

References

[1] Marx V 2013 Biology: the big challenges of big data. Nature 498 255–60
[2] Palsson B O 2015 Systems Biology (Cambridge: Cambridge University Press)
[3] Heinrich R and Schuster S 1996 The Regulation of Cellular Systems (New York: Chapman and Hall)
[4] Braakman R and Smith E 2013 The compositional and evolutionary logic of metabolism Phys. Biol. 10 011001
[5] Beard D, Liang S and Qian H 2002 Energy balance for analysis of complex metabolic networks Biophys. J. 83 79
[6] Beard D, Babson E, Curtis E and Qian H 2004 Thermodynamic constraints for biochemical networks J. Theor. Biol. 228 327
[7] Fell D and Cornish-Bowden A 1997 Understanding the Control of Metabolism vol 2 (London: Portland)
[8] Easterby JS 1973 Coupled enzyme assays: a general expression for the transient Biochim. Biophys. Acta (BBA)—Enzymol. 293 552–8
[9] Easterby JS 1981 A generalized theory of the transition time for sequential enzyme reactions Biochem. J. 199 155–61
[10] Reich J G and Selkov E E 1981 Energy Metabolism of the Cell. A Theoretical Treatise (London: Academic)
[11] Beard DA and Qian H 2007 Relationship between thermodynamic driving force and one-way fluxes in reversible processes PLoS ONE 2 e144
[12] Gallavotti G and Cohen E G D 1995 Dynamical ensembles in nonequilibrium statistical mechanics Phys. Rev. Lett. 74 2694
[13] Gaspard P 2004 Fluctuation theorem for nonequilibrium reactions J. Chem. Phys. 120 8998–905
[14] Schmiedl T and Seifert U 2007 Stochastic thermodynamics of chemical reaction networks J. Chem. Phys. 126 044101
[15] Berg JM, Tymoczko JL and Stryer L 2006 Glycolysis and gluconeogenesis Biochemistry (London: WH Freeman)
[16] Reed J L, Vo T D, Schilling C H, Palsson B O et al 2003 An expanded genome-scale model of Escherichia coli k-12 (iJR904 gsm/ gpr) Genome Biol. 4 R54
[17] Wiback S, Famili I, Greenberg H and Palsson B-O 2004 Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space J. Theor. Biol. 228 437
[18] Park DJ M 1974 The hierarchical structure of metabolic networks and the construction of efficient metabolic simulators J. Theor. Biol. 46 51–74
[19] Reich J G and Selkov E E 1975 Time hierarchy, equilibrium and non-equilibrium in metabolic systems Biosystems 7 39–50
[20] Heinrich R, Rapoport S M and Rapoport T A 1978 Metabolic regulation and mathematical models Prog. Biophys. Mol. Biol. 32 1–82
[21] Bennett B D, Kimball E H, Gao M, Osterhout R, Dien S J V and Rabinowitz J D 2009 Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli Nat. Chem. Biol. 5 393–9
[22] Milo R, Jorgensen P, Moran U, Weber G and Springer M 2010 Biomarkers the database of numbers in molecular and cell biology Nucl. Acids Res. 38 D750–3
[23] Jankowski M, Henry C, Broadbent L and Hatzimanikatis V 2008 Group contribution method for thermodynamic analysis of complex metabolic networks Biophys. J. 95 1487
[24] Henry C, Jankowski M, Broadbent L and Hatzimanikatis V 2006 Genome-scale thermodynamic analysis of Escherichia coli metabolism Biophys. J. 90 1453–61
[25] Price ND, Schellenberger J and Palsson B O 2004 Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies Biophys. J. 87 2172–86
[26] Turcin V 1971 On the computation of multidimensional integrals by the Monte Carlo method Theory Probab. Appl. 16 720–4
[27] Lovász L 1999 Hit-and-run mixes fast Math Program 86 443–61
[28] Lovász L 1987 An Algorithmic Theory of Numbers, Graphs and Convexity vol 50 (Philadelphia: SIAM)
[29] Mahadevan R and Schilling C H 2003 The effects of alternate optimal solutions in constraint-based genome-scale models Metab. Eng. 5 264
[30] De Martino D, Mori M and Parisi V 2013 Uniform sampling of steady states in metabolic networks: Heterogeneous scales and rounding PLoS ONE 10 e0122670
[31] Krauth W and Mezard M 1987 Learning algorithms with optimal stability in neural networks J. Phys. A: Math. Gen. 20 L745
[32] Motzkin TS and Schoenberg J J 1954 The relaxation method for linear inequalities Can. J. Math. 6 393–404
[33] Martino D De 2013 Thermodynamics of biochemical networks and duality theorems Phys. Rev. E 87 053108
[34] Müller A 2012 Thermodynamic constraints in metabolic networks MSc Thesis Freie Universitt Berlin, Fachbereich Mathematik und Informatik