# On the ordering of metabolite concentrations in *Escherichia coli*

Semidán Robaina Estévez *

*Institution*

**Abstract**

Metabolic networks at steady state operate under stoichiometric, kinetic and thermodynamic constraints. As a result, metabolite concentrations are not free to vary. It has been shown that certain metabolites in *Escherichia coli* show a natural ordering at steady state, *i.e.*, there are metabolite pairs for which one metabolite maintains a higher concentration level than the other under different steady states. Here, we provide an explanation to such order relations in a setting were *Escherichia coli*'s cells grow at steady state.

Several studies have shown that certain metabolites maintain an order relation in their concentrations under different steady states of *Escherichia coli* [Bennett et al., 2008, Bennett et al., 2009]. This order relation may be the result of diverse constraints operating at steady state, such as stoichiometric, thermodynamic and growth constraints. In the following, we will derive a theoretical explanation for this ordering of concentrations.

We describe the concentration dynamics of the biochemical network with the general system,

$$\frac{dx_i}{dt} = \sum_j n_{ij} v_j(x), \tag{1}$$

where, $n_{ij}$ represents the stoichiometric coefficient of metabolite $i$ in reaction $j$, with $n_{ij} < 0$ if it is a substrate of the reaction, $n_{ij} > 0$ if it is a product, $x$ the metabolite concentrations and $v(x)$ the metabolic fluxes, which are a function of the metabolite concentration — the exact form provided by the

selected kinetic law. All reactions in 1 are reversible [SRE], with the exception of the biomass production (pseudo)reaction, and the forward and backward direction are represented as two different reactions. Thus, we the net flux of a reaction $v_j = v_j^{for} - v_j^{back}$. We assume that cells are growing at steady state, thus the flux through the biomass reaction, $v_{bio} > \gamma v_{bio}^{max}$, with $\gamma \in [0, 1]$, a fraction of the theoretical maximum. [SRE]

SRE: Might as well include pre-defined irreversible reactions

SRE: I'll make a proper intro after I get results. Starting with the second law more appropriate

Thus far, we have determined that certain reactions are irreversible in an escenario where cells grow at steady state. Additionally, the second law of thermodynamics imposes that flux of free energy $g_j = \Delta^\circ G_j v_j < 0$ for a reaction to have non-zero flux [Kondepudi and Prigogine, 2014]. In our case, we have already determined the direction of the reaction with the linear programs in **??**, hence we only need the reaction Gibbs free energy

$$\Delta^\circ G_j < 0, \tag{2}$$

where,

$$\Delta^\circ G_j = \Delta^\circ G_{r(j)} + RT \sum_i n_{ij} \log x_i \tag{3}$$

in which $\Delta^\circ G_{r(j)} = \sum_i n_{ij} \Delta^\circ G_{f(i)}$. Further, the energies of formation $\Delta^\circ G_{f(i)}$ of the metabolites participating in the reaction can be estimated with the component contribution method [Noor et al., 2013].

Now, the logarithm is a monotonically increasing function, hence $\log x_p > \log x_q \implies x_p > x_q$. Furthermore, we can establish if $\log x_p > \log x_q$ with the following linear program,

$$\begin{aligned}
&z = \min_{\log x} \ \log x_p - \log x_q \\
&\text{s.t.} \\
&\Delta^\circ G_{r(j)} + RT \sum_i n_{ij} \log x_i < 0 \ \forall j \in \mathrm{R_I} \\
&\Delta^\circ G_r^\dagger - \epsilon \leq \Delta^\circ G_r \leq \Delta^\circ G_r^\dagger + \epsilon \\
&\log x_{min} \leq \log x \leq \log x_{max}
\end{aligned} \tag{4}$$

where $\mathrm{R_I}$ is the set of irreversible reactions computed with the linear programs in **??**, $\Delta^\circ G_r^\dagger$ the reaction Gibbs energy estimated by the component

*Corresponding autor: hello@semidanrobaina.com

contribution method with the corresponding estimation error, $\epsilon$. Finally, $x_{min}, x_{max}$ are metabolite concentration bounds — in this case $x_{min} = 10^{-8}$ to avoid numerical issues and $x_{max} = 0.1\,\mathrm{M}$ a reasonable upper bound considering the largest concentration obtained in experimental measurements [Bennett et al., 2008]. If $z \geq 0$ then we know that $\log x_p \geq \log x_q$ and thus $x_p \geq x_q$. Further, we have that $z = \log(x_p^*/x_q^*)$, hence $x_p^*/x_q^* = e^z$, where $x_i^*$ denote the concentration of metabolite $x_i$ at the optimum. Therefore, with this method, we can compute the minimum concentration ratio between pairs of metabolites that is required to make the system thermodynamically feasible, namely,

$$\frac{x_p}{x_q} \geq e^z. \tag{5}$$

Interestingly, this minimum ratio holds in any steady state of the system, and it is independent of reaction kinetics. A result that may provide a structural and thermodynamic explanation for the relationships observed in experimental measurements of metabolite concentrations. Note that the number of ordered pairs may increase if we have more reactions with available $\Delta^\circ G_r$. But, of course, if an order relation is established with a given set of constraints, augmenting the set of constraints can only add new order relations.

**NOTE**: I have discovered a 2013 paper [Tepper et al., 2013] in which authors implement a similar method. Interestingly, even though Noor is co-author, they do not consider the errors in $\Delta^\circ G_r$ like I did, also, they just establish concentration bounds of $x_{min} = 10^{-5}, x_{max} = 10^{-1}$ without further explanation. I do not include the constraints on minimizing total metabolite concentration and enzyme usage. Additionally, I'm optimizing the ratio, not absolute concentration values, which I think no one has done (check). In the former, two-step implementation, I got that the ratio $[H_p^+]/[H_c^+] \geq 18.2$ with $x_{min} = 10^{-7}$. This is an interesting result, it renders a minimum electric potential (due to pH difference) of about 72 mV which coincides with experimental observations. I obtained the previous result by using the Nerst equation for the proton transport $H_p^+ \rightleftharpoons H_c^+$:

$$\Delta^\circ G_r = zF\Delta\Psi + RT\log\frac{[H_c^+]}{[H_p^+]} \tag{6}$$

where $z = 1$ is the proton charge and $F = 96.5\,\mathrm{kJ.V^{-1}.mol^{-1}}$, the Faraday constant. For protons to transport back to the cytoplasm the system must

satisfy $\Delta^\circ G_r < 0$, hence, $-zF\Delta\Psi > RT \log [H_c^+]/[H_p^+]$ and so

$$\Delta\Psi > -\frac{RT}{zF} \log \frac{[H_c^+]}{[H_p^+]}, \tag{7}$$

substituting all the values, with $T = 310.15\,\mathrm{K}$, we obtain: $\Delta\Psi > 33.67\,\mathrm{mV}$. In addition, assuming that *Escherichia coli* cells have an average volume $V_{cell} = 1.3\,\mu\mathrm{m}^3$ then the minimum non-zero concentration is $7.3x10^{-8}\,\mathrm{M}$ which corresponds to 1 molecule of the compound per cell.

**SECOND NOTE** An alternative formulation of the problem is the following:

$$\begin{aligned}
z = & \min_{\substack{\log x \in \mathbb{R}^{\mathrm{m}}, \\ v \in \mathbb{R}^{\mathrm{n}}_{\geq 0}, \\ \Delta^\circ G_r \in \mathbb{R}^{\mathrm{k}}, \\ y \in \{0,1\}^k}} \log x_p - \log x_q \\
& \text{s.t.} \\
& Sv = 0 \\
& v_{bio} \geq \gamma v^*_{bio} \\
& \Delta^\circ G_{r(j)} + RT \sum_i n_{ij} \log x_i - (1 - y_j^{(+,-)}) M < 0 \; \forall j \in \mathrm{R}_{\Delta^\circ \mathrm{G_r}} \\
& \Delta^\circ G_r^\dagger - \epsilon \leq \Delta^\circ G_r \leq \Delta^\circ G_r^\dagger + \epsilon \\
& \log x_{min} \leq \log x \leq \log x_{max} \\
& v_{min} \leq v \leq v_{max} \\
& v_j^{(+,-)} \leq y_j^{(+,-)} v_{max(j)} \; \forall j \in \mathrm{R}_{\Delta^\circ \mathrm{G_r}} \\
& y^+ + y^- \leq 1
\end{aligned} \tag{8}$$

in which the two steps in **??** and 4 are merged together in a single MILP. Some reactions are considered irreversible, in which case only $y^+$ and $v^+$ exist. Further, some irreversible reactions have a default lower bound $v_{min(j)} > 0$, such as the biomass reaction, in which $v_{bio} \geq \gamma v^*_{bio}$, in this cases no binary variable is needed, since the reaction always carries flux in the feasible space. The advantage is that loopless fva is no longer needed, while the disadvantage is that the MILP is harder to solve than the previous LPs. Also, I need to split reversible reactions since the order of the reaction quotient and the value of the Gibbs energy is inverted! Actually, I don't see the advantage here... if a reaction is irreversible in the feasible space of **??** then it is also irreversible

in 6, it is the same feasible space with respect to flux values after all. No... because there could be two flux distributions in the feasible space in which one reaction is running in the opposite direction but in both cases render the same biomass production. We could use this flux distributions to find the minimum ratio of concentrations as well, but are eliminated in the first step since the reaction's direction is not consistent.

Actually, I think now that it's probably best to forget about finding fixed reactions and just do the MILP across all reactions, .i.e, all reactions with free energy data get a binary variable. This is because fixed reactions were found in a different optimization problem. Namely, the regular FBA LP problem. While we are imposing some reactions to carry non-zero flux in a larger MILP which includes other variables and constraints. Perhaps that's why I get infeasibilities. Actually, the infeasibilities were due to constraining reactions to carry non-zero flux when $y = 1$. Probably because I was using a generic lower bound rather than the actual lower bound of the feasible space. Also, it doesn't make sense to constraint the lower bound when $y = 1$: a thermodynamically feasible reaction doesn't have to carry flux, e.g., enzyme may not be expressed!

The previous method may take a lot of computational time. I'll try using a pre-processing step to select candidate concentration-ordered pairs, just like I did with the flux values. To this end, I'll have to implement my own sampling procedure. Namely, we first generate a random concentration vector within the allowed concentration ranges $x_{min} \leq x_{rand} \leq x_{max}$ and then find the closest feasible $x$ with the following MIQP:

$$
\begin{aligned}
\log x^* = & \underset{\substack{\log x \in \mathbb{R}^{\mathrm{m}}, \\ v \in \mathbb{R}^{\mathrm{n}}_{\geq 0}, \\ \Delta^\circ G_r \in \mathbb{R}^{\mathrm{k}}, \\ y \in \{0,1\}^k}}{\arg\min} \; || \log x_{rand} - \log x ||_2 \\
& \text{s.t.} \\
& Sv = 0 \\
& v_{bio} \geq \gamma v^*_{bio} \\
& \Delta^\circ G_{r(j)} + RT \sum_i n_{ij} \log x_i < 0 \; \forall j \in \mathrm{F}_{\Delta^\circ \mathrm{G_r}} \\
& \Delta^\circ G_{r(j)} + RT \sum_i n_{ij} \log x_i - (1 - y_j^{(+,-)}) M < 0 \; \forall j \in \mathrm{R}_{\Delta^\circ \mathrm{G_r}} \\
& \Delta^\circ G_r^\dagger - \epsilon \leq \Delta^\circ G_r \leq \Delta^\circ G_r^\dagger + \epsilon \\
& \sum_i x_i \leq \; 300\,\mathrm{mM} \\
& \log x_{min} \leq \log x \leq \log x_{max} \\
& v_{min} \leq v \leq v_{max} \\
& y_j^{(+,-)} v_{min(j)} \leq v_j^{(+,-)} \leq y_j^{(+,-)} v_{max(j)} \; \forall j \in \mathrm{R}_{\Delta^\circ \mathrm{G_r}} \\
& y^+ + y^- \leq 1
\end{aligned}
\tag{9}
$$

Alternative, we can use the $\ell_1$ norm instead of the $\ell_2$ norm and solve a MILP instead (doable in cvxopt) directly. To this end, we need to transform the objective and add two intermediate variables $q^+, q^-$:

$$
\begin{aligned}
\log x^* = & \underset{\substack{\log x \in \mathbb{R}^{\mathrm{m}}, \\ v \in \mathbb{R}^{\mathrm{n}}_{\geq 0}, \\ \Delta^\circ G_r \in \mathbb{R}^{\mathrm{k}}, \\ y \in \{0,1\}^k}}{\arg\min} \; \sum_{i=1}^{m} (q_i^+ + q_i^-) \\
& \text{s.t.} \\
& q^+ - q^- = \log x_{rand} - \log x \\
& q^+, q^- \geq 0 \\
& \ldots
\end{aligned}
\tag{10}
$$

We just need to repeat this process n times to generate a sample. Then eliminate all metabolite pairs with inconsistent order relation in the sample.

# References

[Bennett et al., 2009] Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology*, 5(8):593–599.

[Bennett et al., 2008] Bennett, B. D., Yuan, J., Kimball, E. H., and Rabinowitz, J. D. (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols*, 3(8):1299–1311.

[Kondepudi and Prigogine, 2014] Kondepudi, D. and Prigogine, I. (2014). *Modern Thermodynamics: From Heat Engines to Dissipative Structures: Second Edition*. Wiley.

[Noor et al., 2013] Noor, E., Haraldsdóttir, H. S., Milo, R., and Fleming, R. M. T. (2013). Consistent Estimation of Gibbs Energy Using Component Contributions. *PLoS Computational Biology*, 9(7):e1003098.

[Tepper et al., 2013] Tepper, N., Noor, E., Amador-Noguez, D., Haraldsdóttir, H. S., Milo, R., Rabinowitz, J., Liebermeister, W., and Shlomi, T. (2013). Steady-State Metabolite Concentrations Reflect a Balance between Maximizing Enzyme Efficiency and Minimizing Total Metabolite Load. *PLoS ONE*.