

Exploring the Concentration Space of Genome Scale Metabolic Networks

Jan Schellenberger^{1,*}, Ellen A. Tsai², Bernhard Ø. Palsson²

1. Bioinformatics Department, University of California San Diego, La Jolla, CA, USA

2. Bioengineering Department, University of California San Diego, La Jolla, CA, USA

*E-mail: jschelle@ucsd.edu

Genome scale metabolic reconstructions

Genome scale metabolic network reconstructions have been used extensively to study the properties of whole cell metabolism. Because of the size of these networks, fully kinetic models are not feasible and other methods, such as constraint based modeling, are required [1]. By imposing a steady state assumption on the system it is possible to calculate the range of flux distributions through a network. Recently there have been efforts [2, 3] to reconcile these flux distributions with metabolic concentration data which is becoming available from various metabolomics projects [4].

Defining the concentration space

In order for a reaction to proceed in a forward direction, $\Delta G_r < 0$. The Gibbs free energy (ΔG_r) of a reaction is composed of the energy of formation (ΔG_0), which is assumed constant and known, and a term related to the concentrations of the participating metabolites.

$$\Delta G_r = \Delta G_0 + RT \sum S_{i,j} \ln[c_i] \quad (1)$$

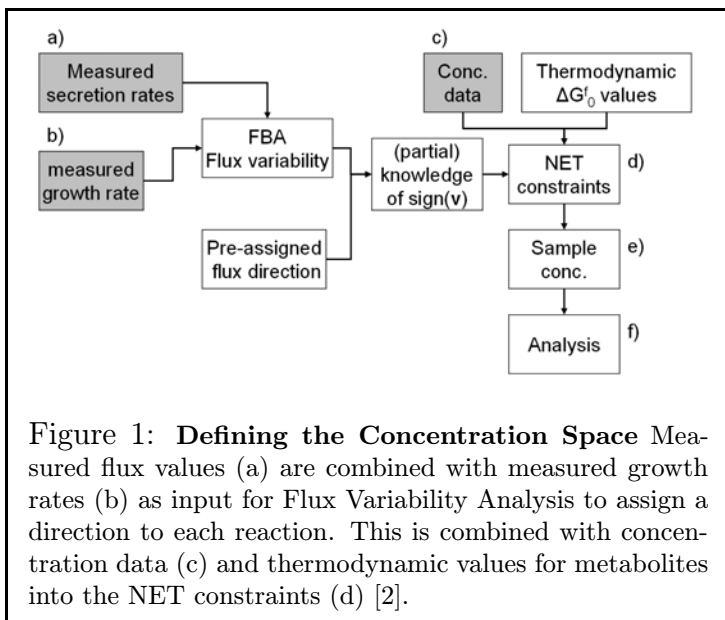


Figure 1: **Defining the Concentration Space** Measured flux values (a) are combined with measured growth rates (b) as input for Flux Variability Analysis to assign a direction to each reaction. This is combined with concentration data (c) and thermodynamic values for metabolites into the NET constraints (d) [2].

By assigning a sign to ΔG_r for all or some of the reactions, constraints are placed on concentrations of many metabolites creating a space of feasible solutions.

Figure 1 outlines the procedure for defining and characterizing the concentration space. Flux variability is performed on the network to restrict the direction of as many reactions as possible. Equation 1 is applied to every reaction creating a set of linear constraints on the concentrations. To study the resulting concentration space, an Artificially Centered Hit and Run (ACHR) algorithm [5] is used to draw uniformly distributed concentration which are consistent with observed and calculated flux states.

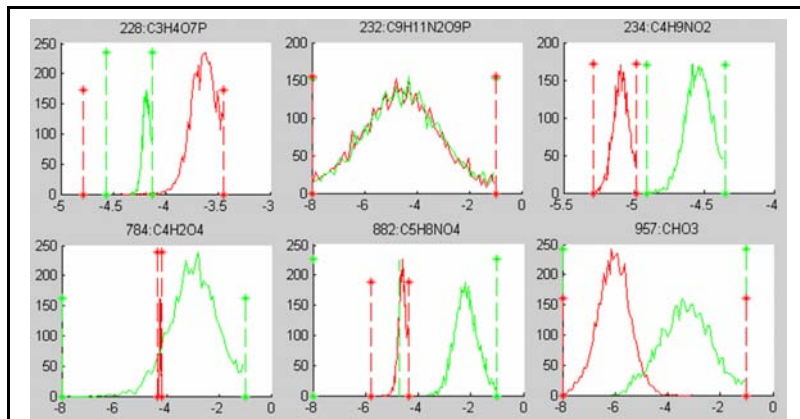


Figure 2: Sampling the Concentration Space to Compare two strains Distributions of 5000 sampled points between two strains of glycerol evolved PGI knockouts of *E. coli* (green and red). The metabolites shown are: 3-Phospho-D-glycerate (3pg), 3ump (extracellular), 4-Aminobutanoate (4abut), fumarate, D-glutamate, and bicarbonate (hco3). The star ‘*’ bound indicates the concentrations which were imposed on the system based on measured values or arbitrarily large physiological bounds ($10^{-8}M$ to $10^{-1}M$). The dashed vertical line indicates the feasible bounds on the concentrations due to thermodynamic network constraints.

Results

Figure 2 shows the sampled concentration spaces from two strains of Δ PGI *E. coli* using the *E. coli* iAF1260 metabolic reconstruction [6]. The distributions of each metabolite indicate the different probabilities within the space.

1. **Additional constraints** While only constraining 74 measured metabolite concentrations it is possible to obtain constraints on an additional 32 compounds. This results from network properties, namely that all reactions must satisfy $\Delta G_r < 0$ and shared metabolites often have a narrower range of feasible concentrations. Additionally, the distributions of the concentrations indicate that certain concentrations are highly improbable and allow for quantification of differences between two conditions.
2. **Correlations between metabolites** (not shown): By performing a correlation analysis between different metabolites, it is possible to find which concentrations

vary together. Both positive and negative correlations were found indicating the presence of ‘metabolic pools’ analogous to reaction co-sets [7].

Conclusion

With the availability of thermodynamic constants (ΔG_0) for most metabolites and the emerging availability of high throughput concentration measurements, there is a need to develop tools for analysis of these data. After having been successfully used to analyze the flux space and kinetic space [8], Monte Carlo sampling has now been shown to describe the complex constraints imposed on the concentration space as well.

References

- [1] **Price ND, Reed JL, Palsson BØ** (2004) *Genome-scale models of microbial cells: evaluating the consequences of constraints*. Nat Rev Microbiol **2**:886–897.
- [2] **Kümmel A, Panke S, Heinemann M** (2006) *Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data*. Mol Syst Biol **2**:1744–4292.
- [3] **Henry CS, Broadbelt LJ, Hatzimanikatis V** (2007) *Thermodynamics-based metabolic flux analysis*. Biophys J **92**:1792–1805.
- [4] **Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, et al.** (2007) *Multiple high-throughput analyses monitor the response of e. coli to perturbations*. Science **316**:593–7.
- [5] **Kaufman DE, Smith RL** (1998) *Direction choice for accelerated convergence in hit-and-run sampling*. Operations Research **46**:84–95.
- [6] **Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce A, et al.** (2007) *A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information*. Molecular Systems Biology **3**:121.
- [7] **Papin JA, Stelling J, Price ND, Klamt S, Schuster S, et al.** (2004) *Genome-scale in silico models of e. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states*. Genome Research **14**:1797–1805.
- [8] **Famili I, Mahadevan R, Palsson BØ** (2005) *k-cone analysis: determining all candidate values for kinetic parameters on a network scale*. Biophys J **88**:1616–1625.