# Intracellular crowding of macromolecules defines the mode and sequence of substrate uptake by *E. coli* and constrains its metabolic activity and growth

# Supporting Information, Part 1: Modeling framework

Q.K. Beg[1*], A. Vázquez[2*], J. Ernst[3], M.A. de Menezes[4], Z. Bar-Joseph[3], A.-L. Barabási[5] and Z.N. Oltvai[1]

[1]Department of Pathology, University of Pittsburgh, Pittsburgh, PA, 15261, USA

[2]The Simons Center for Systems Biology, Institute for Advanced Study, Princeton, NJ 08540, USA

[3]Machine Learning Department, Carnegie-Mellon University, Pittsburgh, PA, 15217, USA

[4]Instituto de Física, Universidade Federal Fluminense, Rio de Janeiro, 24210, Brazil.

[5]Department of Physics and Center for Complex Networks Research, University of Notre Dame, Notre Dame, IN 46556, USA

## Contents

## A. Mathematical framework

The Flux Balance analysis with Molecular Crowding modeling framework is implemented by solving the following optimization problem: maximize the growth rate subject to the constraints (1) and (4) (see Results in Main Text). The maximum growth rate corresponds to the biomass production rate, where biomass production is an auxiliary reaction (see Biomass reaction in the Supplementary Table S1a) containing as substrates the cellular components in their relative concentrations and as product the cell's biomass (Neidhardt *et al*, 1990).

We model the crowding coefficients $a_i$ as noise. The reported results were obtained assigning a random value to them from the gamma distribution

$$P(a) = \frac{\beta}{\langle a \rangle}\left(\frac{\beta}{\langle a \rangle}a\right)^{\beta-1}\exp\left(-\frac{\beta}{\langle a \rangle}a\right)$$

where $\beta>0$ and $<a>$ is the average crowding coefficient. There is no particular reason for this choice other than by changing $\beta$ we can explore different scenarios. For instance, for $\beta=1$ we obtain an exponential distribution, while for $\beta>>1$ we obtain a distribution that is almost concentrated around $a=<a>$. The results reported in Figures 1-3 of the paper were obtained using $\beta=3$ and running the simulations 1000 times to test the sensitivity of the results with respect to the specific $a_i$ values. Similar results are obtained using other $P(a)$ distributions (see Figs. S1-1 and S1-2).

The maximum growth rate $\mu$ for each carbon source was obtained assuming an unbound uptake rate for that carbon source and zero for all other carbon sources. The average crowding coefficient $<a>$ was fitted to obtain the minimum square deviation between the measured and model predicted growth rates, resulting in $<a>=0.0040\pm0.0005$ hour DW/mmol. However, the maximum growth rate on glucose and glycerol are more consistent with $<a>=0.0031\pm0.0001$ hour DW/mmol and $<a>=0.0053\pm0.0001$ hour DW/mmol, respectively.

To model the temporal order of substrate uptake we considered an initial concentration of 0.4 g/L for glucose, galactose, lactate, maltose-, glycerol, zero acetate concentration, and cell's dry weight $DW=0.00675$ g. The progression of the dry weight

and the external substrate concentrations were obtained from the integration of the differential equations

$$\frac{dDW(t)}{dt} = \mu(t) DW(t)$$

$$\frac{dC_m(t)}{dt} = \frac{\mu_m}{V} \sum_i S_{mi} f_i(t)\, DW(t)$$

where $m$ is restricted to external metabolites, $\mu_m$ is the molar mass of metabolite $m$ and $V$ is the working volume. The maximum growth rate $\mu(t)$ and the fluxes $f_i(t)$ are obtained by solving the FBAwMC model for the substrate concentration profile at time $t$. The value of $\langle a \rangle$ is smaller if glucose alone is consumed and larger if glycerol is consumed. Therefore, we solve three FBAwMC problems corresponding to the consumption of glucose alone ($\langle a \rangle$=0.0031 hour DW/mmol), consumption of all substrates except glycerol ($\langle a \rangle$=0.004 hour DW/mmol) and consumption of all substrates ($\langle a \rangle$=0.0053 hour DW/mmol), and selected the condition resulting in the maximum growth rate.

**B. Estimation of the crowding coefficients**

The crowding coefficients play a key role in our model. Therefore, it is important to estimate their values from experimental measurements. However, this is a challenging task, as it requires knowledge of the kinetic parameters associated with the *E. coli* enzymes, which are not available in all cases. Yet, we can obtain an estimate for the crowding coefficients by making use of the known turnover rate of several *E. coli* enzymes. The crowding coefficients (units of time×mass/mole) are defined by (see main text)

(S1) $$a_i = \frac{C v_i}{b_i}$$

where $C$ (units of mass/volume) is the density of the *E. coli* cytoplasm, $v_i$ (units of volume/mole) is the molar volume of enzyme $i$, and $b_i$ (units of 1/time) is the proportionality factor between the flux (units of mole/mass×time) and the enzyme concetration (units of mole/mass), i. e., $f_i = b_i E_i$. In general $b_i$ is decomposed into

(S2) $$b_i = x_i k_i$$

where $k_i$ (units of 1/time) is the enzyme turnover number, and $0 \le x_i \le 1$ (no units) is determined by the concentration of substrates, products and activators/inhibitors associated with the *i*th reaction. For example, if reaction *i* is a single substrate irreversible reaction and it is characterized by a Michaelis-Menten kinetics then

(S3) $$x_i = \frac{S}{S + K_i}$$

where $S$ is the substrate concentration and $K_i$ is the Michaelis-Menten or half-saturation constant. Thus, $x_i \approx 0$ at small substrate concentrations ($S << K_i$ ), while at saturation ($S >> K_i$ ) $x_i \approx 1$. Substituting (S2) into (S1) and assuming that enzymes are working near saturation ($x_i \approx 1$) we obtain

(S4) $$a_i \sim \frac{C v_i}{k_i}$$

where the symbol ~ indicates that this is an estimate and some variation is expected for enzymes that are not working at saturation. The magnitudes on the right hand side can be estimated using different sources of experimental data. The *E. coli* cytoplasmatic concentration is $C \approx 0.34$ g/ml (Zimmerman and Trach, 1991). The enzymes' molar volumes can be estimated from the equation

(S5) $$v_i^{(\text{molar})} = M_i v_i^{(\text{specific})}$$

where $M_i$ is the molar mass of enzyme *i* (units of mass/mole) and $v_i^{(\text{specific})}$ its specific volume (units of volume/mass). We estimate the specific volume from measurements of this magnitude for globular proteins (Lee et al, 1983). This data indicates that the specific volume has small variations among enzymes relative to the molar mass variations. Therefore, we take the same specific volume for all enzymes and equal to the average especific volume for the proteins reported in (Lee et al, 1983), resulting in $v_i^{(\text{specific})} \approx 0.73$ ml/g. Finally, we consider the molar mass and the turnover number of several *E. coli* enzymes reported in BRENDA (Schomburg et al, 2004). Using these data and equations (S4) and (S5) we computed $a_i$ for several *E. coli* enzymes (see Supplementary Table S1b), resulting in the distribution shown in Figure S1-3.

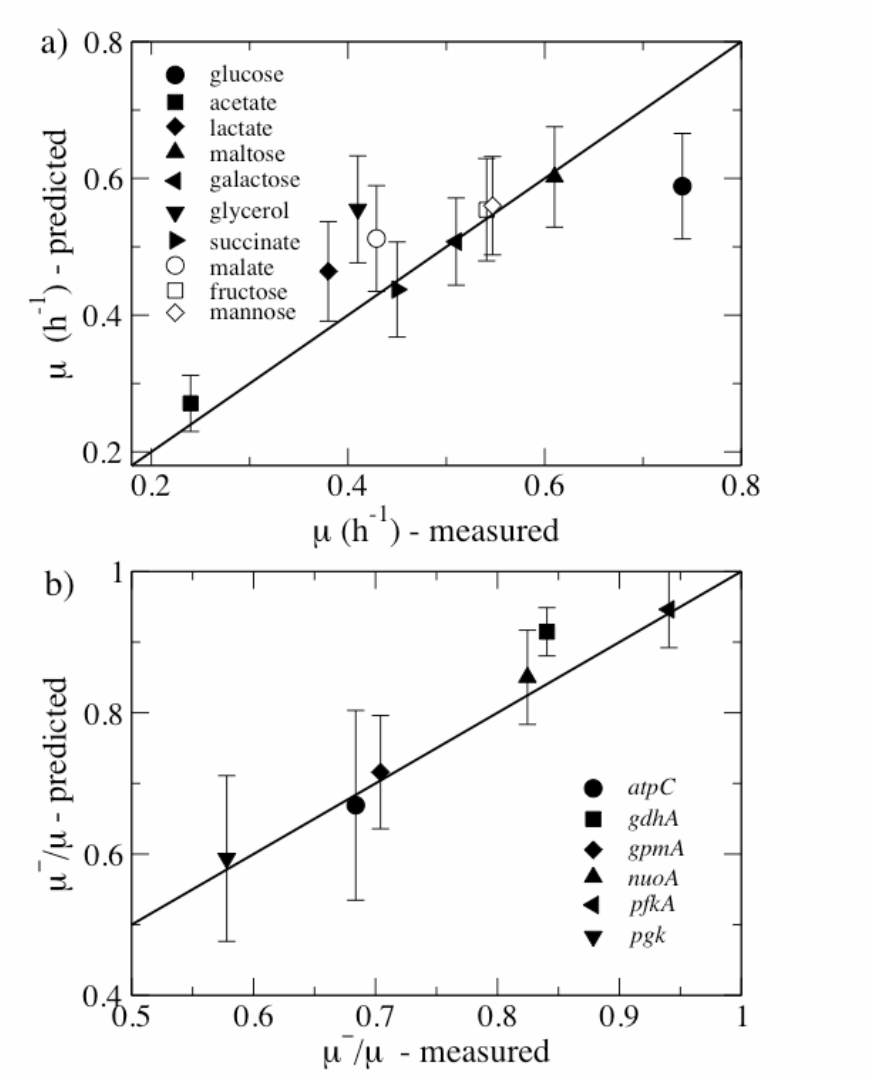


**Figure S1-1:** Predicted and measured maximum growth rate comparison

(a) Comparison between the predicted- (Y-axis) and measured (X-axis) growth rates $\mu$ of *E. coli* MG1655 grown in M9 minimal medium with different carbon sources. For a perfect match between experiments and theory the symbols should be on the black line. The symbols indicate the carbon substrate identified in the legend. The predicted growth rates were obtained using a uniform distribution for the crowding coefficients $a_i$ in the interval [0,2<*a*>] with <*a*>=0.0040 hour DW/mmol (see Methods). The error bars represent the standard deviation over 1000 sets of specific $a_i$ parameters. (b) Same plot for single gene deletion *E. coli* mutants growing in glucose, the deleted genes being indicated in the legend. The mutant growth rates $\mu^-$ are given relative to the predicted and measured maximum growth rate $\mu$ of wild type *E. coli* cells growing in glucose-limited medium.

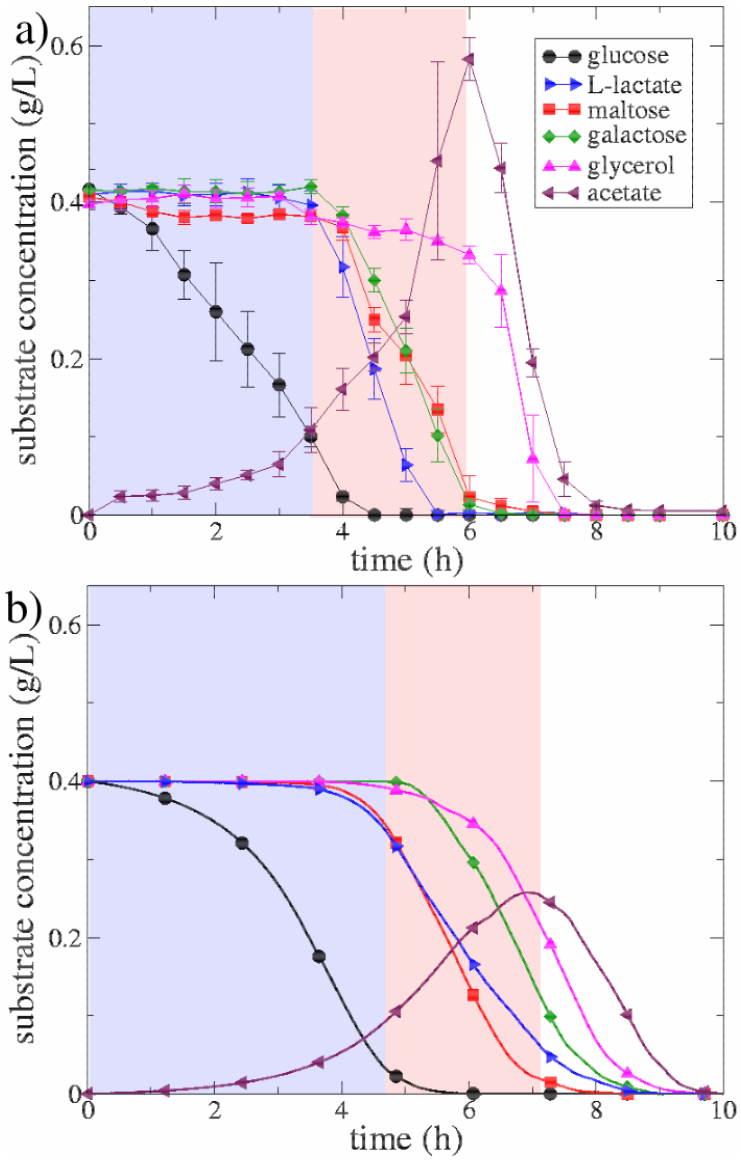


**Figure S1-2:** Predicted vs. measured hierarchy of substrate utilization. This figure is a reproduction of Figure 2b, after replacing the gamma distribution by a uniform distribution for the crowding coefficients. (a) The measured concentration of the indicated carbon sources in the growth medium. The experiments were performed in triplicate and averages and standard deviations are shown. Three substrate utilization phases, phase 1 (exclusive glucose), phase 2 (mixed substrate) and phase 3 (glycerol and acetate) are indicated in light blue, purple and while backgrounds, respectively. (b) Predicted substrate uptakes from the growth medium based on the FBAwMC model, using a uniform distribution for the crowding coefficients $a_i$ in the interval $[0,2\langle a\rangle]$ with $\langle a\rangle$=0.0040 hour DW/mmol (see Methods). The color coding for substrate uptake curves is identical in panels *a* and *b*.

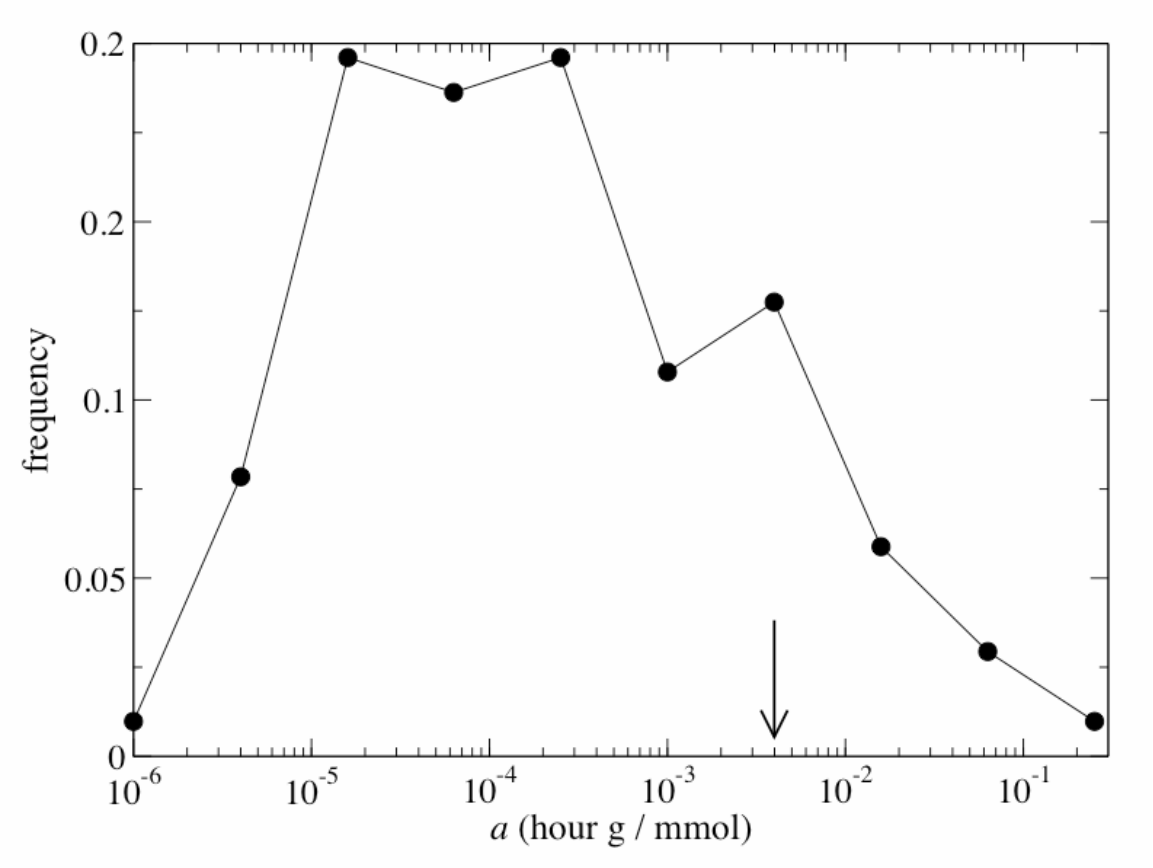


**Figure S1-3:** Distribution of the *a* coefficients as obtained from an independent estimate for about 100 *E. coli* enzymes. The arrow indicates the value of <*a*> obtained from the fit to the growth rate data.

## References


1. Neidhardt FC, Ingraham JL and Schaechter M (1990). Physiology of the bacterial cell: a molecular approach. Sinauer Associates, Sunderland, MA.
2. Lee B (1983). Calculation of volume fluctuation for globular protein models. PNAS 80; 622-626.
3. Schomburg et al (2004). BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 32; D431-433.
4. Zimmerman SB and Trach SO (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J Mol Biol 222; 599-620.