Parameter estimation for iJO1366-based E. coli RBA model

May 2019

Contents

1 Introduction

Here we detail the procedure of parameterizing a Resource Balance Analysis (RBA)(3) model of model bacterium *Escherichia coli* based on the genome scale metabolic reconstruction *iJO1366* (10). We first explain how one can compute parameters related to compartment densities, those being the cytosolic density and total protein concentration, both in $[\frac{mmol.AA}{gCDW}]$. These parameters enable RBA models to give meaningful quantitative predictions of appropriate scale and unit. We then proceed to explain how one can compute the efficiencies of molecular machines, such as ribosomes, chaperones, secretion apparatus and enzymes. We shortly explain the choice of some default values in the model. Finally, we elaborate an additional parameter modification of the model - the efficiency of the respiratory chain.

2 Compartment densities

Compartment density constraints are a part of the RBA cell model formulation (3; 4). They mathematically describe macromolecular occupancy of different compartiments, and are for the sake of simplicity expressed in terms of an average-sized amino acid. Since size of an aminoacid is not trivial to determine, especially in the bound-polymeric state, we use molecular weight as its proxy.

Average amino acid molecular weight $\bar{MW}_{AA}$ will depend on the percentage of each amino acid in the total amino acid pool. The amino acid composition data was taken from (9), and the weighted average amino acid molecular weight (without water) is computed to be $\bar{MW}_{AA} = 108.3\frac{g}{mol}$.

Average rRNA molecular weight $\bar{MW}_{RNA}$ is computed from the composition of the ribosome, and is $\bar{MW}_{RNA} = 340\frac{g}{mol}$.

Ribosome conversion factor $d_{r/AA}$ is a constant describing the molecular weight of all the nucleic acids in one ribosome, expressed in terms of number of average amino acids. Total number of nucleic acids in an *E. coli* is $N_{NA/r} = 4593$. This parameters is computed to be $d_{r/AA} = N_{NA/r} \times \frac{\bar{MW}_{RNA}}{\bar{MW}_{AA}} = 4593 \times 3.14 = 14422$. We use the same conversion as introduced by (8).

2.1 Cytosolic density

Cytosolic density, or bouyant cell density, is a constant independent of the growth rate for fast growing bacteria (6; 5), representing portion of space taken up by all cytosolic molecules. RBA formulates this mathematically as a constraint, and, as in (8), it uses this constant to represent the space taken up by all the cytosolic enzymes and ribosomes, thus taking into account the protein and rRNA content. To compute the cytosolic density, we consolidated three relevant datasets (1; 8; 9) and have thus chosen a growth rate which is present in all three datasets, that of $1[h^{-1}]$. Cytosolic density can be expressed as a sum of portion of protein that is assigned to cytosol and of the RNA

content of ribosomes converted to average amino acid:

$$D_{cyt} = p_{cyt}(\mu)P_{tot}(\mu) + d_{r/AA}R(\mu) \qquad [\frac{mmol.AA}{gCDW}] \tag{1}$$

Fraction of protein assigned to cytosol $p_{cyt}(\mu)$ is estimated using the package `RBApy.estim` from proteomics data (12) as a linear function of the growth rate. The process of estimation is described in the supplementary text S6, to be found here (link).

$$p_{cyt} = 0.04\mu + 0.73 \tag{2}$$

What remains to be computed are the total amino acid and ribosome concentration. To compute the total amino acid concentration in $[\frac{mmol}{gCDW}]$, we use the comprehensive dataset of (1), and express it as:

$$C_{AA/CDW} = N_{AA/CDW}[\frac{\#AA}{\mu gCDW}] \times \frac{1}{N_a}[mol] \times 10^6 \times 10^3 \qquad [\frac{mmol.AA}{gCDW}] \tag{3}$$

N_a is the Avogadro constant, and $N_{AA/CDW}$ is the number of amino acids in a μg of cell dry weight. For ease of comparison, the abbreviations used here to express the formula for total amino acid count are the same as used in the (1): P_M stands for protein/mass expressed in units of $10^{17}aa/OD_{460}$, $M_C(\mu g)$ stands for μg of cell dry weight per 10^9 cells, and M_C stands for OD_{460} units per 10^9 cells. We compute $N_{AA/CDW}$ in the following way:

$$N_{AA/CDW} = \frac{P_M}{\frac{M_C(\mu g)}{M_C}} \qquad [\frac{\#AA}{\mu gCDW}] \tag{4}$$

We finally obtain a value for total concentration of amino acids:

$$C_{AA/CDW} = 4.98 \qquad [\frac{mmol.AA}{gCDW}] \tag{5}$$

for the growth rate of $\mu = 1[h^{-1}]$.

We compute the ribosome concentration by assuming that the cell has as many ribosomes as needed to translate the flux of total protein at steady state:

$$R = \frac{\mu C_{AA/CDW}}{k_T \times p_{R_a}} \tag{6}$$

where k_T is the ribosome efficiency in $[\frac{AA}{h}]$ and p_{R_a} is the percentage of active ribosomes expressed as $e^{\mu\tau}$, τ being the maturation time of ribosomes, set to be $5min$ (7). The ribosome concentration thus computed is:

$$R = 7.24 \times 10^{-5} \qquad [\frac{mmol}{gCDW}] \tag{7}$$

Now we can compute the cytosolic density for $\mu = 1[h^{-1}]$ from equation 1 to be:

$$D_{cyt} = 4.89 \qquad [\frac{mmol.AA}{gCDW}] \tag{8}$$

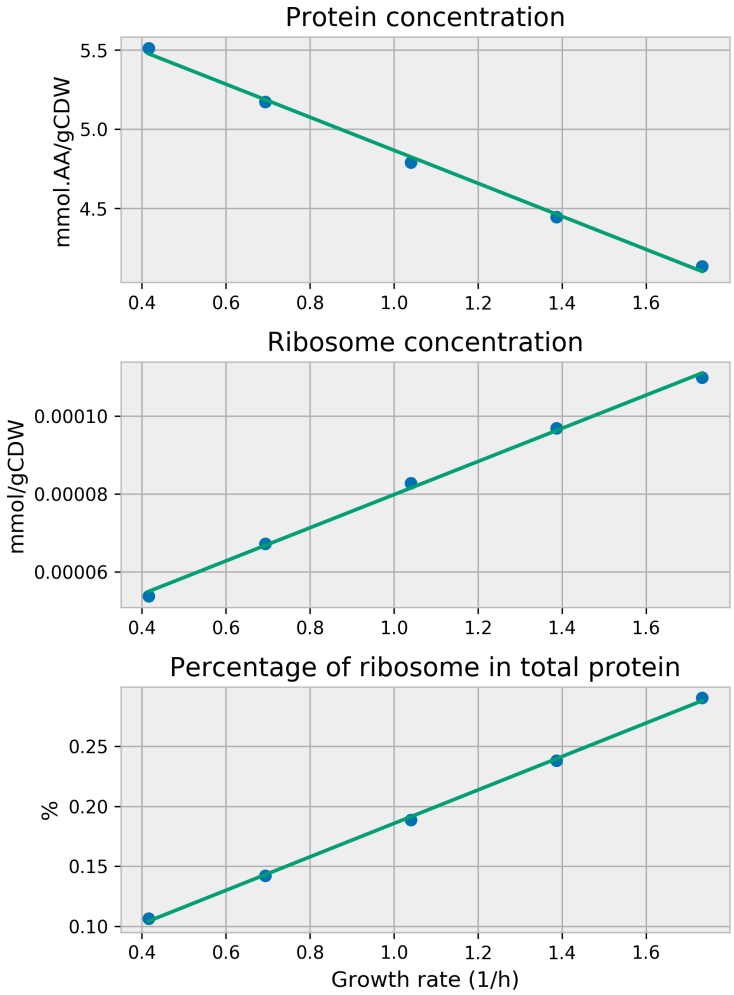

Figure 1: Computed values for total amino acid concentration, ribosome concentration and percentage of ribosome in total protein.

2.2 Total amino acid concentration

We compute total cytosolic concentration as described in the supplementary Text E1 in (4), by solving the following system of equations:

$$\mu C_{AA/CDW} = k_T e^{\mu\tau} R \tag{9}$$

$$p_{cyt} C_{AA/CDW} + d_{r/AA} R = D_{cyt} \tag{10}$$

for growth rates available in (1). We used the computed values to estimate the linear dependency of amino acid concentration to the growth rate:

$$C_{AA/CDW} = -1.04\mu + 5.91 \tag{11}$$

The results of the estimation can be seen on figure 1.

3 Machinery efficiencies

The efficiencies of process machineries were estimated using proteomics datasets from (12) for growth on 12 different carbon sources and 1 supplemented with 20 amino acids. We basically estimate the total amino acid concentration flux that needs to be processed by a particular machinery, and divide it by the abundance of the machinery, obtaining process machinery efficiency in $[\frac{\#AA}{h}]$.

3.1 Folding

We consider two major chaperoning systems (GroEL/S and DnaJK) in exponential growing cells. We describe them as one cellular process having single machinery composed of the two chaperones in their right stoichiometries that needs to fold, as a rough measurement, 10% (2) of all protein. We estimate the total concentration of amino acids that needs to be folded per unit time using the estimation obtained in the section 2.

$$\nu_{P_{fold}} = \mu \times 0.1 \times P_{tot}(\mu) = 0.31 \qquad [\frac{mmol.AA}{gCDW}] \tag{12}$$

The total number of amino acids in the chaperone complex consisting of all subunits in their correct stoichiometries (tig, dnaJ, dnaK, groL, groS, grpE) is $N_{AA/ch} = 10829$. Number of measured amino acids of the same complex is $N_{AA/ch/mes} = 3.8 \times 10^7$. Efficiency of the chaperone complex becomes:

$$k_{CH} = \frac{\nu_{P_{fold}}}{\frac{N_{AA/ch/mes}}{N_{AA/ch}}} \tag{13}$$

The folding efficiency as a linear function of the growth rate is:

$$k_{CH}(\mu) = 7.2\mu + 1.59 \qquad [s^{-1}] \tag{14}$$

and can be seen on Figure 2a.

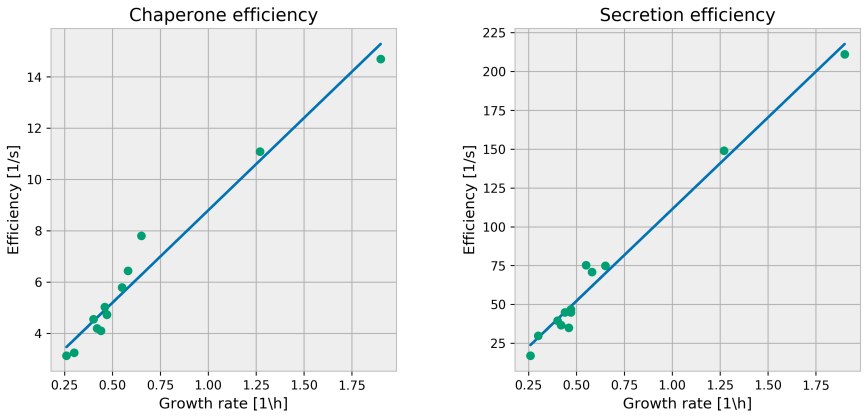

Figure 2: Efficiency of cellular process machineries (folding and secretion) estimated from proteomics data

3.2 Secretion

We model the general secretory pathway (sec) of *Escherichia coli*, since most non-cytosolic proteins are translocated to their compartments via this pathway (11). Thus our concetration of amino acids to be secreted per unit time will be:

$$\nu_{P_{sec}} = \mu \times (1 - p_{cyt}(\mu)) \times P_{tot}(\mu) \tag{15}$$

The rest of the procedure is the same as in the case of folding, and the final linear relation between the growth rate and secretion efficiency is

$$k_{SEC}(\mu) = 118.23\mu - 6.94 \qquad [s^{-1}] \tag{16}$$

and can be seen on Figure 2b.

4 Default values

Default value chosen for enzymatic efficiency is 12.5 s^{-1}, and was obtained as a best fit for predicted growth rates to growth rates determined experimentally for cells grown in batch cultures for 12 different media (12).

On Figure 3, we show the differences in growth rate prediction as a consequence of change in the default enzyme efficiency value.

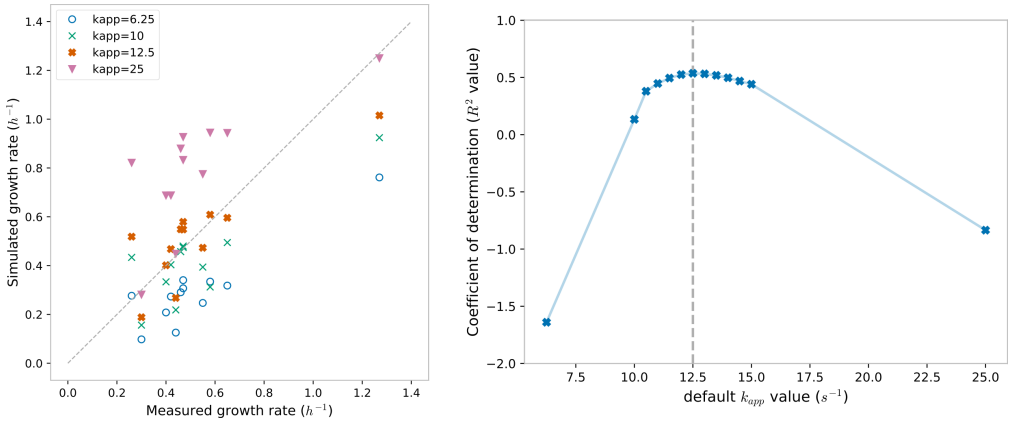

Figure 3: Left: Predictions of growth rate for 4 different values for default enzyme efficiency. Right: Change in goodness of fit as a consequence of change of default enzyme efficiency value.

References

[1] Bremer, H. and Dennis, P. P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1).

[2] Deuerling, E., Patzelt, H., Vorderwülbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., Mogk, A., Schulze-Specking, A., Langen, H., and Bukau, B. (2003). Trigger factor and dnak possess overlapping substrate pools and binding specificities. *Molecular microbiology*, 47(5):1317–1328.

[3] Goelzer, A., Fromion, V., and Scorletti, G. (2011). Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):1210–1218.

[4] Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., Mariadassou, M., Aymerich, S., Hecker, M., Noirot, P., et al. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic engineering*, 32:232–243.

[5] Kubitschek, H., Baldwin, W., and Graetzer, R. (1983). Buoyant density constancy during the cell cycle of escherichia coli. *Journal of bacteriology*, 155(3):1027–1032.

[6] Kubitschek, H. E., Baldwin, W. W., Schroeter, S. J., and Graetzer, R. (1984). Independence of buoyant cell density and growth rate in escherichia coli. *Journal of bacteriology*, 158(1):296–299.

[7] Lindahl, L. (1975). Intermediates and time kinetics of the in vivo assembly of escherichia coli ribosomes. *Journal of molecular biology*, 92(1):15–37.

[8] Marr, A. G. (1991). Growth rate of escherichia coli. *Microbiological reviews*, 55(2):316–333.

[9] Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990). *Physiology of the bacterial cell: a molecular approach*, volume 20. Sinauer Sunderland.

[10] Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., and Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of escherichia coli metabolism—2011. *Molecular systems biology*, 7(1):535.

[11] Pugsley, A. P. (1993). The complete general secretory pathway in gram-negative bacteria. *Microbiological reviews*, 57(1):50–108.

[12] Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent escherichia coli proteome. *Nature biotechnology*, 34(1):104–110.