# SUPPLEMENTARY INFORMATION

## Biosynthesis flux model

Nicolás A. Vaccari,[†] Dahlin Zevallos-Aliaga,[†] Tom Peeters,[‡] and Daniel G. Guerra[*,†]

†*Unidad de Moléculas Individuales, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia, Lima, Perú*

‡*Open BioLab Brussels, Erasmushogeschool Brussel, Brussels, Belgium*

E-mail: daniel.guerra@upch.pe

Phone: +51 319-0000 ext.233217

# SUPPLEMENTARY INFORMATION: MODEL BUILDING

## Initial definitions

### Proteome fractions

A particular set of proteins $i$, will be defined here as a proteome's fraction, $\varphi_i$, by dividing the total mass of said group of proteins, $m_i$, over the mass of the complete proteome, $m$:

$$\varphi_i(t) = \frac{m_i(t)}{m(t)} \tag{1}$$

Where:

$$m(t) = \sum_i m_i(t) \tag{2}$$

To study how $\varphi_i$ evolves in time, we need to define some extra terms. The first one is the anabolic flux which is the velocity at which the cell produces proteins:

$$\frac{dm(t)}{dt} = \rho(t) \cdot m(t) \tag{3}$$

We define this as the product of the total anabolic rate ($\rho$), and the mass of the total proteome ($m$). Also, as mentioned in the main text, the total anabolic flux, can be divided into subfluxes that produce the different protein groups:

$$\frac{dm_i(t)}{dt} = f_i(t) \cdot \rho(t) \cdot m(t) \tag{4}$$

Where $f_i$ is the expression fraction, which represents the fraction of the total anabolic flux dedicated to produce a set $i$ of proteins. From equation (2), (3) and (4) we can deduce that:

$$\rho(t) = \sum_i f_i(t) \cdot \rho(t) \tag{5}$$

Then, we can define each corresponding production rate for each of the protein sets as follows:

$$\rho_i(t) = f_i(t) \cdot \rho(t) \tag{6}$$

## Biosynthesis allocation theorem

To study proteome fractions and their change over time, we start by unfolding the derivative of $\varphi_i$ as the derivation of a division:

$$\frac{d\varphi_i(t)}{dt} = \frac{d\left(\frac{m_i(t)}{m(t)}\right)}{dt} \tag{7}$$

Expanding and resolving:

$$= \frac{m(t) \cdot \frac{dm_i(t)}{dt} - m_i(t) \cdot \frac{dm(t)}{dt}}{m^2(t)}$$

$$= \frac{m(t) \cdot f_i(t) \cdot \rho(t) \cdot m(t) - m_i(t) \cdot \rho(t) \cdot m(t)}{m^2(t)}$$

$$= \rho(t) \cdot \left( f_i(t) - \frac{m_i(t)}{m(t)} \right)$$

$$\frac{d\varphi_i(t)}{dt} = \rho(t) \cdot [f_i(t) - \varphi_i(t)] \tag{8}$$

The evolution of $\varphi_i$ depends on the anabolic rate $\rho$ and the difference between the $f_i$ and $\varphi_i$. This expression is very similar in form to that of potential-induced currents, where $\frac{1}{\rho}$ can be interpreted as the resistance, and the difference between $f_i$ and $\varphi_i$ can be understood as a sort of phenotypic potential differential, where $f_i$ represents the proteome fraction value that the cell intends, and $\varphi_i$ represents the value it actually has. Overall, our analysis provides a framework for an easy understanding of the dynamics of proteome fractions.

Because measuring the anabolic rate can be challenging, we will look for a particular, more useful form of the theorem to express it in function of the growth rate. When the ratio protein mass-to-biomass is constant the following identity is true:

$$\frac{dm(t)}{dt} \propto \frac{db(t)}{dt} \tag{9}$$

Where $b$ is biomass and follows the accumulation equation:

$$\frac{db(t)}{dt} = \mu(t) \cdot b(t) \tag{10}$$

Where $\mu$ is growth rate, a classical parameter easily measured as the change of optical

density in time. Using equations 3, 9 and 10 we obtain that as long as the protein to biomass ratio remains constant:

$$\mu(t) = \rho(t) \tag{11}$$

From experimental measures it is known that this condition is met during the exponential phase of bacterial growth[1]. Rearranging the equation (8), for this particular case we obtain:

$$\frac{d\varphi_i(t)}{dt} = \mu(t) \cdot [f_i(t) - \varphi_i(t)] \tag{12}$$

# Regulation function and expression fraction

## Definitions

In general, it is possible to represent the rate of expression of a set $i$ of proteins as:

$$\widehat{\rho_i}(t) = r_i(t) \cdot \rho(t) \tag{13}$$

Where $r_i$ can represent a repression term, induction term or a constitutive term. We will refer to $r_i$ as the regulation function because it modifies the behavior of the protein production flux for specific a set of proteins.

Here, we will provide functions to represent three different modes of regulations.

The first case considers the production of a set of proteins $i$ that is repressed by a known concentration of a transcription factor $TF$, with a repression constant $k_{TF}$. We can model this regulated production flux as follows:

$$\widehat{\rho_1}(t) = \theta_1 \cdot \frac{k_{TF}}{TF(t) + k_{TF}} \cdot \rho \tag{14}$$

The second case represents the expression as induced by a known concentration of an inducer $I$, with an induction constant $k_I$:

$$\widehat{\rho_2}(t) = \theta_2 \cdot \frac{I(t)}{I(t) + k_I} \cdot \rho(t) \tag{15}$$

In both equations, $\theta$ is a coefficient that represents the maximum activation capacity of the regulatory circuit that governs the set of proteins. In other words, it indicates what fraction of the total protein production is directed towards this set of proteins when the system is perfectly derepressed ($TF = 0$) or perfectly induced ($I \to \infty$).
Finally, a third case is a constitutive mode of expression, where there is no repressor nor inducer modifying the production flux. This mode of regulation is represented only by the maximum activation capacity coefficient $\theta$:

$$\widehat{\rho_3}(t) = \theta_3 \cdot \rho(t) \tag{16}$$

In all cases, the regulation function, $r_i$, is an abstraction that only represents the behaviour of the overall regulatory mechanism of a set of proteins. This regulatory function can be very complex or relatively simple as the three cases defined above. Compiling various proteins under a single set and generalizing a unified $r_i$ for them may necessitate the inclusion of several environmental factors and intracellular transduction pathways.

By definition, the sum of all the fluxes of all proteome fractions must be equal to the total anabolic flux. This means that the sum of all regulation functions can be normalized:

$$\frac{r_1(t) \cdot \rho(t) + r_2(t) \cdot \rho(t) + r_3(t) \cdot \rho(t) + \ldots + r_n(t) \cdot \rho(t)}{[r_1(t) + r_2(t) + r_3(t) + \cdots + r_n(t)]} = \rho(t) \tag{17}$$

From this expression, we deduce that the production rate of a proteome fraction, $\varphi_i$, is:

$$\rho_i(t) = \frac{r_i(t)}{\sum_j r_j(t)} \cdot \rho(t) \tag{18}$$

Recurring to equation (4) and using this last equation, we finally define the expression fraction $f_i$, which as it was said before represents the fraction of the total anabolic flux dedicated to produce a set $i$ of proteins, as:

$$f_i(t) = \frac{r_i(t)}{\sum_j r_j(t)} \tag{19}$$

## Regulation functions for $\varphi_Q$, $\varphi_R$ and $\varphi_C$.

Having defined the regulation function and the expression fraction; we can assign a particular regulation mechanism to each proteome fraction to build a model of the cell. This model will be able to simulate the expression of each proteome fraction and the cell's contents in time. To emphasize this suitability for simulation, we will denote with (t) all the variables that will be calculated.

In previous models, it has been assumed the presence of a set of structural proteins that doesn't change through time[2,3]. Here, we will define this group of proteins as a baseline proteome fraction, denoted as $\varphi_Q$, that follows a regulation function equal to a constant, $Q$ (a constitutive mode). In consequence, under stable conditions, $\varphi_Q$ will remain constant:

$$r_Q(t) = Q \tag{20}$$

To model the regulation of the set of proteins that function as nutrient fixators, $\varphi_C$, we take inspiration from the work done in Towbin et al.[3], where it was proposed that the production this proteome fraction is downregulated by the accumulation of anabolic precursors $x$:

$$r_C(t) = C \cdot \frac{k_c}{x(t) + k_c} \tag{21}$$

This regulatory mode makes $x$ an inhibitor for the expression of $\varphi_C$. Here, $C$ is the maximum activation capacity of the nutrient fixator, and $k_C$ is the repression constant for this fraction expression by the pool of anabolic precursors, $x$.

Previously, it has been proposed that the pool of metabolite $x$ regulates the expression of $\varphi_R$ and $\varphi_C$ through the ppGpp and CRP-cAMP systems, respectively[4]. To represent these opposite regulations, we approximate that the sum of both fractions will equal to a constant, at least during exponential growth. This means that taking the derivatives of this sum with respect to time, we obtain:

$$\frac{d\varphi_R(t)}{dt} + \frac{d\varphi_C(t)}{dt} = 0 \tag{22}$$

What this equation means is that the sum of the regulation functions for $\varphi_R$ and $\varphi_C$ must be constant. Using this piece of information and the definition for proteome fraction differential in equation (8), and the equation (21) we the sum of $r_R$ and $r_C$ should look something like:

$$r_R(t) + r_C(t) = C + \varepsilon \tag{23}$$

Where $\varepsilon$ represents part of $\varphi_R$ regulation that is not influenced by $\varphi_C$ regulation. However, for our purposes we will assume this term to be zero. Taking this into consideration, if we expand $r_C$, using equation (23) and then isolating $r_R$, we obtain:

$$r_R(t) = C \cdot \frac{x(t)}{x(t) + k_R} \tag{24}$$

Also, to reduce the number of constants we use the fact that the sum of $C$ and $Q$ equals 1. We then replace in $r_C$ and $r_R$:

$$r_C(t) = (1 - Q\ ) \cdot \frac{k_C}{x(t) + k_C} \tag{25}$$

$$r_R(t) = (1 - Q\ ) \cdot \frac{x(t)}{x(t) + k_R} \tag{26}$$

Then, the allocation theorem equations for all proteome fractions can be deduced.

For $\varphi_C$ evolution we have:

$$\frac{d\varphi_C(t)}{dt} = \mu(t) \cdot \left( (1 - Q\ ) \cdot \frac{k_C}{x(t) + k_C} - \varphi_C(t) \right) \tag{27}$$

For $\varphi_R$ evolution we have:

$$\frac{d\varphi_R(t)}{dt} = \mu(t) \cdot \left( (1 - Q\ ) \cdot \frac{x(t)}{x(t) + k_R} - \varphi_R(t) \right) \tag{28}$$

And for $\varphi_Q$ we have:

$$\frac{d\varphi_Q}{dt} = \mu(t) \cdot \left( Q - \varphi_Q(t) \right) \tag{29}$$

## Introducing a heterologous protein

The T7 expression system is a convenient regulatory circuit to study heterologous genetic circuits. It provides a well-known regulation mechanism for the protein being expressed under its control. Moreover, it is a system that includes its own RNA polymerase, which is not directly influenced by any metabolite or additional transcription factor. Finally, when activated by IPTG, its regulatory function can be considered constant because IPTG is a non-metabolizable ligand whose intracellular concentration, once it reaches equilibrium, remains stable:

$$r_H(t) = H \cdot \frac{I}{I + k_i} \tag{30}$$

Here $I$ is the concentration of an inducer, $H$ is the maximum expression activation number for the heterologous circuit, and $k_i$ is its induction constant. Following the theorem expressed in Equation (8), we obtain the change in time of the proteome fraction represented by the heterologous protein during the exponential phase:

$$\frac{d\varphi_H(t)}{dt} = \mu(t) \cdot \left( \frac{H \cdot \frac{I}{I+k_i}}{1 + H \cdot \frac{I}{I+k_i}} - \varphi_H(t) \right) \tag{31}$$

Future models could include a function for the internalization of a non-constant inducer $I(t)$, or its production and consumption inside the cell.

# Kinetic modelling

In this chapter, we include the units corresponding to each one of the equations and definitions to facilitate their dimensional analysis. International units are used for mass (g), and volume (L). Because the rates are linked to growth rate, time is presented in hours (h).

## Catabolic activity

Previously we have defined a proteome fraction as the fraction of the proteome that represent a group of protein that share a similar genetic regulation and a common function or activity. In this section we will discuss the latter. To describe the activity of the sets we have defined, we will borrow a nutrients uptake function described by Towbin *et al.* and modify it to be dependent on the total mass of the nutrient-fixation proteins $m_C$. Additionally, we will add an extra term for the equation to take the shape of an enzyme's rate equation with allosteric regulation. Then, we obtain:

$$-V \cdot \frac{dS(t)}{dt} = \beta \cdot m_C(t) \cdot \frac{S(t)}{S(t) + k_s} \cdot \frac{k_1}{x(t) + k_1} \cdot N(t) \qquad \left( g \cdot h^{-1} \right) \tag{32}$$

Where $\beta$ is the average turnover number for the uptake and fixation of nutrients, $N$ is the cell count, $V$ is the culture volume, $S$ is the concentration of extracellular nutrients; $k_s$ is the constant related to the affinity of the nutrient fixators for their substrate, the extracellular nutrients; $k_1$ is the inhibition constant of $x$, acting as an acompetitive inhibitor of the nutrient fixator.

Additionally, similarly to the classical equations described by Monod[5], the total substrate for a single cell can be described as:

$$-V \cdot \frac{dS(t)}{dt} = Y(t) \cdot m(t) \cdot N(t) \qquad \left(g \cdot h^{-1}\right) \tag{33}$$

$$Y(t) \cdot m(t) = \beta \cdot m_C(t) \cdot \frac{S(t)}{S(t) + k_s} \cdot \frac{k_1}{x(t) + k_1} \qquad \left(g \cdot h^{-1}\right) \tag{34}$$

Isolating $Y$, we obtain that the substrate consumption rate is given by:

$$Y(t) = \beta \cdot \varphi_C(t) \cdot \frac{S(t)}{S(t) + k_s} \cdot \frac{k_1}{x(t) + k_1} \qquad \left(h^{-1}\right) \tag{35}$$

## Anabolic activity

The anabolic fraction corresponds to the set of proteins involved in the consumption of anabolic precursor $x$ for the synthesis of new protein mass. To model the protein production flux in a cell, we adapt the carbon uptake equation proposed by Towbin et al.[3] and relate it to the total ribosomal protein mass $m_R$ as an ordinary enzymatic rate equation. As a simplification, we will consider no recycling activity.

$$\frac{dm(t)}{dt} = \gamma \cdot m_R(t) \cdot \frac{x(t)}{x(t) + k_2} \qquad \left(g \cdot h^{-1}\right) \tag{36}$$

Here, $\gamma$ is the average turnover number for ribosomal activity, and $k_2$ is the affinity constant of the biomass producers for these anabolic precursors. All the protein production will be assumed to correspond to the building of a new cell. Similar to the substrate

consumption case, we used Monod's equation for biomass accumulation to model the protein production flux as detailed in equation (3), using this with our last result we get:

$$\rho(t) \cdot m(t) = \gamma \cdot m_R(t) \cdot \frac{x(t)}{x(t) + k_2} \qquad \left( g \cdot h^{-1} \right) \tag{37}$$

Isolating $\rho$, we get that:

$$\rho(t) = \gamma \cdot \varphi_R(t) \cdot \frac{x(t)}{x(t) + k_2} \qquad \left( h^{-1} \right) \tag{38}$$

## Metabolite pool

In our low-dimensional model, there is only one cellular metabolite. The anabolic precursor $x$ regulates the entire system, and it is also the building block for biomass production. $x$ is produced from the substrate uptake flux $\frac{dS(t)}{dt}$ and it is depleted by the biomass production flux:

$$v \cdot \frac{dx(t)}{dt} = \xi \cdot \frac{dS(t)}{dt} - \frac{1}{\zeta} \cdot \frac{dm(t)}{dt} \qquad \left( g \cdot h^{-1} \right) \tag{39}$$

In this equation, $\xi$ is the cell efficiency constant for converting extracellular nutrients/substrate into metabolite pool, and $\zeta$ is the efficiency constant for producing biomass from it. Rearranging:

$$\frac{dx(t)}{dt} = \left[ \xi \cdot Y(t) - \frac{1}{\zeta} \cdot \mu(t) \right] \cdot \frac{m(t)}{v(t)} \qquad \left( g \cdot l^{-1} \cdot h^{-1} \right) \tag{40}$$

Because we assume that the mass density within the cell, hence the protein density as well, is close to constant[6] we can simplify the expression by solving for the product of the cell density and the efficiency expressions:

$$\frac{dx(t)}{dt} = \xi' \cdot Y(t) - \frac{1}{\zeta'} \cdot \mu(t) \qquad \left( g \cdot l^{-1} h^{-1} \right) \tag{41}$$

# Model constants

## Substrate uptake

We looked for conditions where the Equation (35) could be equated to the uptake of extracellular substrates quantified in experimental reports. For example, in the chemostat studies by Carlson et al.[7]. We used these data combined with the definitions and values reported by Scott et al.[2], the RNA-protein ratio of the ribosome, and the nutritional capacity of a medium. Also, due to out work tryes to describe the contents of a single cell, we used the information from the NEB web page on protein data[1] for the dry weight of a single cell, $2.8 \cdot 10^{-13} g$, and the cell's volume ( $1.55 \cdot 10^{-15}$ $l$). Using this data and some further try-and-error tweaking, we obtained an estimation of the upper limit of uptake activity, $\beta$:

$$\beta = 2791.2 \ g \cdot l^{-1} \cdot h^{-1}$$

The value of $k_s$ was set to the arbitrary value of 1 mM to represent the cell's affinity to nutrients in general (we took glucose as the reference substate).

$$k_s = 1 \ g \cdot l^{-1}$$

## Biomass production

We used data from steady-state growth values to solve the growth rate (using equation (38)). For a strain growing at 30°C, the $\gamma \cdot \frac{x}{x+k_2}$ term is approximately 2.6721 $h^{-1}$ [2]. In this case, we will assume that $x$ value is near $k_2$ so that $\varphi_R$ is half-saturated of precursors:

$$\frac{x}{x+k_2} \approx 0.5$$

[1] *https://international.neb.com/tools-and-resources/usage-guidelines/protein-data*

However, when simulations were ran the the outcome was not a realistic one, so we also tweaked this value. The value that finally worked as intended wqas 3 times greater than the previous one:

$$\gamma = 17.1\ h^{-1}$$

Similarly, as we said on both the uptake and growth rate equations, $x$ value is close to $k_1$ and $k_2$, thus making these two constants also similar between them. This idea is also supported by a function fitting that Towbin et al.[3] applied to their data. For this "Enzyme", we assume a generic $K_m$ of 0.1 g $\cdot l^{-1}$. Considering that the molecular weight of an average substrate/metabolite is 100 $g \cdot mol^{-1}$:

$$k_C = k_R = k_1 = k_2 = 0.1\ g \cdot l^{-1}$$

$$\varphi_Q = 0.45$$

For the intracellular carbon variation equation (equation (41)), we also need the constants $\zeta'$ and $\xi'$, which represent, but is not exactly equal, the amount of $x$ produced per $S$ and the ammount of $x$ required to produce one gram of protein respectively. Then $\zeta'$ and $\xi'$ are:

$$\zeta' = \frac{1}{269.6}$$

$$\xi' = 0.5$$

For the External substrate depletion equation, we need the result of dividing the average cell protein mass $m$, arround 1.55 $10^{-15}$ g , and the culture volume $V$, which is 200 $\mu l$ for our microcultures:

$$\frac{m}{V} = \frac{1.55 \cdot 10^{-15}\ g}{2 \cdot 10^{-4}\ l} = 0.775 \cdot 10^{-11}$$

### Induction of heterologous protein production

Finally, in the regulation of the $\varphi_H$, we set $H$, $ki$ and $n$ to the experimental values obtained from the hill fitting:

$$H = 0.2$$

$$k_i = 0.02\ \mu M$$

$$n = 4$$

## References


(1) Dennis, P. P.; Bremer, H. Macromolecular composition during steady-state growth of Escherichia coli B-r. *Journal of Bacteriology* **1974**, *119*, 270–281.

(2) Scott, M.; Gunderson, C. W.; Mateescu, E. M.; Zhang, Z.; Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science (New York, N.Y.)* **2010**, *330*, 1099–1102.

(3) Towbin, B. D.; Korem, Y.; Bren, A.; Doron, S.; Sorek, R.; Alon, U. Optimality and sub-optimality in a bacterial growth law. *Nature Communications* **2017**, *8*, 14123.

(4) Erickson, D. W.; Schink, S. J.; Patsalo, V.; Williamson, J. R.; Gerland, U.; Hwa, T. A global resource allocation strategy governs growth transition kinetics of Escherichia coli. *Nature* **2017**, *551*, 119–123, Number: 7678.

(5) Monod, J. *Recherches sur la croissance des cultures bactériennes*; Actualités scientifiques et industrielles; Hermann & cie: Paris, 1942; OCLC: 6126763.

(6) Borkowski, O.; Ceroni, F.; Stan, G.-B.; Ellis, T. Overloaded and stressed: whole-cell

considerations for bacterial synthetic biology. *Current Opinion in Microbiology* **2016**, *33*, 123–130.

(7) Carlson, R.; Srienc, F. Fundamental Escherichia coli biochemical pathways for biomass and energy production: creation of overall flux states. *Biotechnology and Bioengineering* **2004**, *86*, 149–162.