Practical exercises CompSysBio course: modelling integrated networks of metabolism and gene expression

Hidde de Jong

September 27, 2017

1 Carbon catabolite repression in bacteria

All free-living bacteria have to adapt to a changing environment. Specific regulatory systems respond to particular stresses, but the most common decision bacteria have to make is the choice between alternative carbon sources, each sustaining a specific, maximal growth rate. Many bacteria have evolved a strategy that consists in utilizing carbon sources sequentially, in general favouring carbon sources that sustain a higher growth rate. As long as a preferred carbon source is present in sufficient amounts, the synthesis of enzymes necessary for the uptake and metabolism of less favourable carbon sources is repressed. This phenomenon is called *Carbon Catabolite Repression (CCR)* and the most salient manifestation of this regulatory choice is *diauxic growth* (Figure 1) [4, 6, 11, 14, 16, 23].

CCR, occupying such a central position in the regulation of bacterial metabolism, has been intensely studied for more than 50 years. The underlying regulatory system involves a complex interplay between metabolism, signaling by metabolites and proteins, and the regulation of gene expression, in the context of global constraints on cell physiology. In order to explain how the observed behavior of a bacterial cell emerges from networks of biochemical reactions and regulatory interactions, and predict the response of this system to specific experimental perturbations, mathematical models have been found useful in systems biology [1, 10].

A variety of models has been proposed for CCR, focusing on different aspects of the phenomenon. Here, we review these different modeling approaches and illustrate their capacity to predict the hallmark feature of CCR, diauxic growth. Following [11], we propose a highly simplified representation of diauxic growth, in order to explain and compare the salient features of the models that have been proposed in the literature. We will see that to some extent, the overall logic of diauxic growth can be captured by all modeling approaches.

2 Model definition

Bacterial metabolism is conventionally viewed as a system of biochemical reactions, converting external substrates into biomass and by-products. This system can be modelled by coupled ordinary differential equations (ODEs) describing how the reactions, occurring at a specific rate v_j, change the metabolite concentrations x_i over time. x and v represent the vectors of metabolite

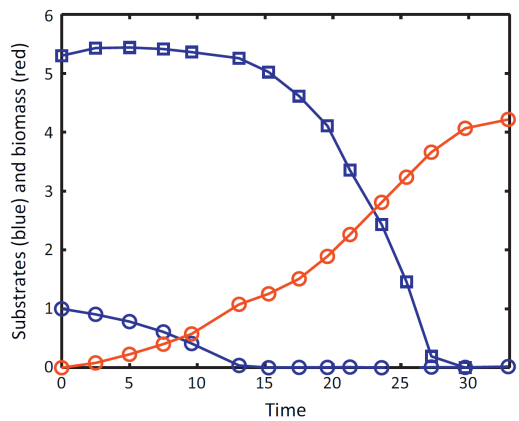

Figure 1: Carbon catabolite repression and diauxic growth [11]. When in the presence of two different growth substrates, the bacterium first metabolizes the substrate sustaining the highest growth rate. After exhaustion of the preferred substrate, the enzymes necessary for the utilization of the second substrate are synthesized, leading to a temporary growth lag, after which slower growth resumes on the second substrate. The experimental data for glucose (blue circles), lactose (blue squares) and biomass (red circles) are taken from [3]. Carbon catabolite repression refers to the different mechanisms that bring about the above-mentioned changes in enzyme and metabolite levels and metabolic fluxes.

concentrations and reaction rates, respectively. The *stoichiometry matrix* N couples the intracellular metabolites to the reactions, by indicating which metabolites are produced and consumed in the reaction and at which relative ratios:

$$\dot{x} = N \cdot v, \qquad x(0) = x_0. \tag{1}$$

A simple metabolic network fueling growth from two different substrates, shown in Figure 2, can be written in the above form by defining $x = [X_1, X_2, M]'$ [mmol gDW^{-1}], $v = [v_1, v_2, v_3, v_4, v_5]'$ [mmol gDW^{-1} h^{-1}], and

$$N = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 4 & 0 & 1 & -10 \end{bmatrix}. \tag{2}$$

Notice that at this level of description the dependency of the reaction rates on metabolite and enzyme concentrations is not explicitly taken into account. By convention, the concentrations are expressed in units mmol gDW^{-1}, bearing in mind that the volume of a growing cell population is usually assumed to be proportional to the quantity of biomass, expressed in units gram dry weight (gDW).

Exercise 1 *Write down the reactions of the system corresponding to N. Why is the stoichiometry coefficient in the biomass reaction consuming M much higher than the stoichiometry coefficients in the uptake and metabolic reactions converting S_1 and S_2 to M?*

The model for internal cellular processes is coupled to differential equations describing substrate uptake and biomass growth over time:

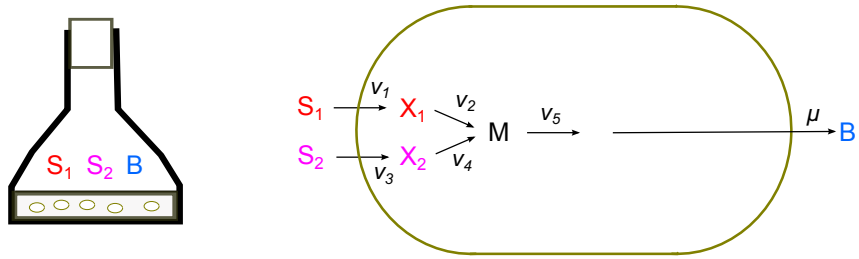

Figure 2: Simple metabolic network for growth on two different substrates. The concentrations of the substrates in the growth medium are denoted by S_1 and S_2 [mmol L^{-1}], wherease the concentrations of the metabolites in the cell population are denoted X_1, X_2, and M [mmol gDW^{-1}]. The uptake reactions occur at rates v_1 and v_3, the internal reactions at rates v_2 and v_4, and the conversion of intermediary metabolite M into biomass at concentration B [gDW L^{-1}] at a rate v_5 [mmol gDW^{-1} h^{-1}]. The latter reaction gives rise to a growth rate μ [h^{-1}].

$$\dot{S}_1 = -v_1 \cdot B, \qquad S_1(0) = S_{1,0}, \tag{3}$$
$$\dot{S}_2 = -v_3 \cdot B, \qquad S_2(0) = S_{2,0}, \tag{4}$$
$$\dot{B} = \beta \cdot v_5 \cdot B, \qquad B(0) = B_0, \tag{5}$$

where S_1 and S_2 [mmol L^{-1}] are the substrate concentrations and B [gDW L^{-1}] is the biomass concentration. Notice that the extracellular concentrations of substrates and biomass have a different unit than the intracellular concentrations appearing in Eq. 1, since they have a different reference volume (the volume of the growth medium [L] rather than the volume or biomass of the growing cell population [gDW]). Because of this difference in units, the rate of incorporation of M into the biomass occurs needs to be multiplied with the conversion constant β [gDW $mmol^{-1}$].

Let V_{medium} [L] and $V_{population}$ [L] denote the volume of the medium and the cell population growing in the medium, respectively. The relations between the units can then be expressed as follows:

$$\alpha \cdot V_{population} = B \cdot V_{medium}, \tag{6}$$

where α is the biomass density in the growing cell population.

Exercise 2 *What is the unit of α? Explain the meaning of Eq. 6 in words.*

The growth rate of the cell population is defined as

$$\mu = \frac{\dot{V}_{population}}{V_{population}}. \tag{7}$$

Exercise 3 *Show with Eqs 3-7 that $\mu = \beta\, v_5$, and therefore that Eq. 5 can be rewritten as:*

$$\dot{B} = \mu \cdot B, \qquad B(0) = B_0, \tag{8}$$

3 Dynamic flux balance analysis

At steady state, Eq. 1 becomes

$$0 = N \cdot v. \tag{9}$$

This *flux balance equation* is underdetermined, as there are generally more reactions than metabolites. For example, the stoichiometry matrix of Eq. 2, has three rows (metabolites) and five column (reactions). Additional constraints on the fluxes can be defined, based on measurements of uptake or secretion fluxes, limits on enzyme capacity, or thermodynamic constraints. *Flux balance analysis* aims at selecting solutions from the flux cone of the equation that optimize a certain criterion, such as biomass production or ATP production [18, 19].

While classical flux balance analysis considers the network at one specific (quasi-)steady state, *dynamic flux balance analysis* allows the (quasi-)steady state to vary over time as a function of changing substrate concentrations and other growth conditions. At each time-point, the metabolic fluxes are defined as the solution(s) of a flux balance optimization problem and the concentrations of external substrates, products, and biomass evolve in accordance with the optimized exchange fluxes [13].

In other words, when describing the dynamics of the system by Eqs 3-5, at each time instant t the following optimization problem is solved:

$$\text{Find } v_{opt}(t) = \arg\max_{w(t)} \ \beta \cdot w_5(t), \tag{10}$$

such that $w(t)$ satisfies $0 = N \cdot w(t)$ and the following inequality constraints:

$$-\infty < w_1(t) \leq k_1 \cdot \frac{S_1(t)}{K_1 + S_1(t)}, \tag{11}$$

$$-\infty < w_2(t) < \infty, \tag{12}$$

$$-\infty < w_3(t) \leq k_3 \cdot \frac{S_2(t)}{K_2 + S_2(t)}, \tag{13}$$

$$-\infty < w_4(t) < \infty, \tag{14}$$

$$0 \leq w_5(t) < \infty, \tag{15}$$

where k_1, k_2 [mmol gDW^{-1} h^{-1}] are kinetic rate constants and K_1, K_2 [mmol gDW^{-1}] half-saturation constants. The solution of this problem is then used as an external input in the system of Eqs 3-5 by setting at each time instant $v(t) = v_{opt}(t)$.

Exercise 4 *Run the dynamic flux balance model using the file* `dynamicFBASimpleModel2.m`*. The file initializes the COBRA toolbox, declares the model defined by the equations in the previous section and some simulation parameters, and launches the function* `dynamicFBA` *of the COBRA toolbox. Does dynamic FBA predict diauxic growth? How do you explain this result?*

Exercise 5 *Which regulatory constraint could be added to the model to allow dynamic FBA to predict diauxic growth?*

4 Kinetic modeling

Flux balance analysis allows predictions of the network dynamics to be made from very little information, aided by the assumption that some objective function, for example growth rate, is optimized. The approach has several limitations though. First, it may not be clear what is the most appropriate choice for an objective function [17, 20, 21, 22]. Second, the fluxes are chosen as the free variables, but this does not make it possible to explicitly model regulatory interactions and predict metabolite concentrations.

An alternative approach is the use of *kinetic models* [7]. Kinetic models take into account kinetic expressions for the reaction rates v_j as a function of the intra- and extracellular concentrations of metabolites, enzymes, and cofactors, thus providing a full description of the networks dynamics. In order to transform the model of the simple diauxic growth network of Fig. 2 into a kinetic model, we need to define the reactions rates v as a function of the metabolite concentrations x, *i.e.*, $v \equiv v(x)$. Here, we will assume Michaelis-Menten kinetics for the uptake rates and simple first-order mass-action kinetics for the intracellular metabolic reactions. This gives rise to the following equations:

$$v_1(S_1) = k_1 \cdot \frac{S_1}{K_1 + S_1}, \tag{16}$$

$$v_2(X_1) = k_2 \cdot X_1, \tag{17}$$

$$v_3(S_2) = k_3 \cdot \frac{S_2}{K_2 + S_2}, \tag{18}$$

$$v_4(X_2) = k_4 \cdot X_2, \tag{19}$$

$$v_5(M) = k_5 \cdot M, \tag{20}$$

where $k_1, \ldots, k_5$ [mmol gDW^{-1} h^{-1}] are kinetic rate constants and K_1, K_2 [mmol gDW^{-1}] half-saturation constants.

Exercise 6 *Run the kinetic model stored in* `metabolicModel.m` *by means of the file* `simulateMetabolicSystem.m`*. Compare the results with those obtained using dynamic FBA.*

A possible regulatory interaction favoring diauxic growth is the repression of the uptake of secondary carbon sources when the preferred substrate is available. In bacteria such regulatory interactions have been identified and are known as *inducer exclusion* [4, 6, 11]. For instance, in *E. coli* inducer exclusion involves the phosphotransferase system (PTS), responsible for the uptake of glucose and other carbon sources. In the presence of glucose, the preferred carbon source, the glucose-specific component of the PTS inhibits the activity of several transporters and enzymes, thus preventing the uptake and metabolism of alternative carbon sources. Fig. 3*A* is a schematic illustration of the effect of this regulatory interaction.

Exercise 7 *Adapt the kinetic model of the previous exercise so as to integrate the regulatory interaction in Fig. 3A. Call the resulting files* `simulateMetabolicSystemRegulation.m` *and* `metabolicModelRegulation.m`*. Does the model predict diauxic growth? Test the sensitivity of the model predictions to the values of the parameters characterizing the uptake inhibition interaction.*

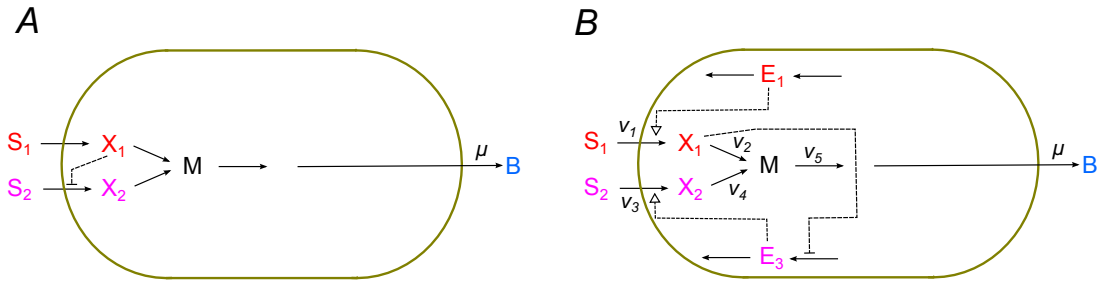

Figure 3: Regulatory interactions responsible for diauxic growth. *A* Regulation on the metabolic level. When the preferred substrate (X_1) is present, the uptake of the secondary substrate (S_2) from the growth medium is inhibited. *B* Regulation on the gene expression level. When the preferred substrate (X_1) is present, the expression of the gene encoding the protein E_3 involved in the uptake of the secondary substrate S_2 is inhibited. E_1 is required for the uptake of X_1, but we assume that it is constitutively expressed here. For simplicity, we ignore the enzymes catalyzing the other reactions (v_2, v_4, v_5).

Another level of regulation of diauxic growth involves the enzymes catalyzing metabolic reactions and the proteins making up substrate transport systems. In many bacteria, the expression of genes encoding enzymes and transporters necessary for the assimilation of secondary carbon source is repressed when the preferred carbon source is available [4, 6, 11]. In *E. coli*, this again involves the glucose-specific component of the PTS, called EIIA^{Glc}. When glucose is available, EIIA^{Glc} inhibits the enzyme producing the signalling molecule cAMP, required for the expression of genes involved in the uptake and metabolism of alternative carbon sources, such as lactose or arabinose. Fig. 3*B* shows the corresponding regulatory interaction in the diauxic growth network. For simplicity, we only introduce genes encoding the transporters, called E_1 and E_3 in the figure, and ignore the genes that code for the enzymes associated with the other metabolic reactions.

In order to account for gene regulation in diauxic growth, we adapt the equations defining the reaction rates v_1 and v_3 given above:

$$v_1(S_1, E_1) = k_1 \cdot E_1 \cdot \frac{S_1}{K_1 + S_1}, \tag{21}$$

$$v_3(S_2, E_3) = k_3 \cdot E_3 \cdot \frac{S_2}{K_2 + S_2}, \tag{22}$$

where E_1 and E_3 [mmol gDW^{-1}] are the enzyme concentrations. We use the same parameters as in Eqs 16 and 18, but notice that the units, and therefore the meaning, are different (h^{-1} instead of mmol gDW^{-1} h^{-1}).

The full kinetic model also needs equations to describe the dynamics of the enzyme concentrations E_1 and E_3:

$$\dot{E}_1 = c_1 - g_1 \cdot E_1, \qquad E_1(0) = E_{1,0}, \tag{23}$$

$$\dot{E}_3 = c_3 \cdot \frac{L_2^2}{L_2^2 + X_1^2} - g_3 \cdot E_3, \qquad E_3(0) = E_{3,0}, \tag{24}$$

where c_1 and c_3 [mmol gDW^{-1} h^{-1}] are protein synthesis constants, g_1 and g_3 [h^{-1}] protein degradation constants, and L_2 [mmol gDW^{-1}] a regulation constant. The first term in the right-hand side of Eq. 24 accounts for the regulation of the expression of the gene encoding E_3, or more precisely the synthesis of the protein E_3.

Exercise 8 *Run the kinetic model stored in* `metabolicModelGeneRegulation.m` *by means of the file* `simulateMetabolicSystemGeneRegulation.m`*. Compare the results of regulation on the metabolic and gene expression level. What is the effect of increasing or decreasing the exponent 2 (the cooperativity constant) in the expression of the regulation of E_3 synthesis by X_1 (Eq. 24)? What happens when the substrate is absent from the growth medium?*

5 Modeling integrated cellular networks: metabolism, gene expression, and growth

The previous section showed that in order to obtain diauxic growth for the simple network of Fig. 2, it is necessary to introduce regulatory interactions. We considered both regulation on the metabolic and genetic level, reminiscent of actual regulatory interactions that have been identified in bacteria. In all of the above models, the growth rate is taken proportional to the rate of the biomass reaction consuming precursor metabolites M. However, this approach does not relate the growth rate to the molecular contents of the cell making up the biomass (enzymes, transporters, metabolites, ...). Moreover, it does not take into account that the enzymes and transporters necessary for the assimilation of substrates for growth are themselves produced from precursor metabolites, thus ignoring an important feedback loop in the system. Recently, there has been a regained interest in this global control of cellular behavior [2, 5, 9, 8].

Fig. 4 is an extension of the simple diauxic growth network of Fig. 3*B*, addressing some of the issues outlined above. In particular, it shows that the proteins E_1 and E_3 are synthesized from the precursor metabolite M, through reactions with rates r_1 and r_3, respectively, and that the proteins are diluted by growth. We assume that the synthesis of E_1 and E_2 costs 5 molecules of M per enzyme.

The above considerations result in the following general form of the kinetic model

$$\dot{x} = N \cdot v(x) - \mu\, x, \qquad x(0) = x_0, \tag{25}$$

where the additional term in the right-hand side represents growth dilution of the intracellular metabolites and proteins.

Exercise 9 *Define $X = x/V_{population}$ and $V = v/V_{population}$. Bearing in mind that $\dot{X} = N \cdot V$, derive Eq. 25.*

For the example of Fig. 4, we have $x = [X_1, X_2, M, E_1, E_3]'$, $v = [v_1(S_1, E_1), v_2(X_1), v_3(S_2, E_3), v_4(X_2), r_1(M)$ and

$$N = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 4 & 0 & 1 & -5 & -5 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{26}$$

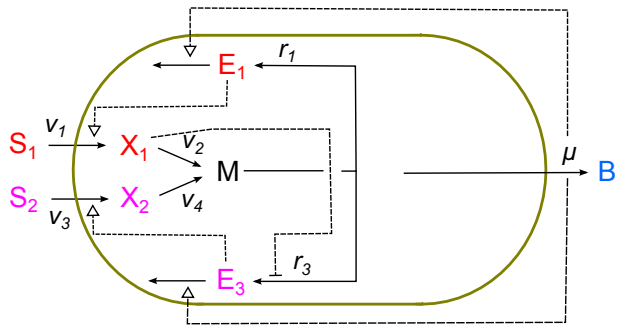

Figure 4: Global control of network responsible for diauxic growth. The simple network of Fig. 3*B* is extended with additional reactions describing how precursor metabolites M are used for the synthesis of proteins, in particular the transport proteins E_1 and E_3. The scheme also indicates that the growth rate influences the concentration of E_1 and E_3 through growth dilution. For simplicity, the interactions have been omitted for the enzymes catalyzing the other reactions (v_2, v_4, v_5).

In this model we have ignored the degradation of the enzymes, which is often negligible in comparison with growth dilution due to the high stability of most proteins in bacterial cells [12, 15]. Moreover, the biomass reaction of the previous sections is no longer necessary here, since we explicitly model the demand for precursors through their incorporation in proteins.

Assuming further that protein synthesis from M is a first-order process, we have

$$r_1(M) = c_1 \cdot M \tag{27}$$

$$r_3(M) = c_3 \cdot \frac{L_2^2}{L_2^2 + X_1^2} \cdot M. \tag{28}$$

Notice that the constants c_1 and c_3 [h^{-1}] do not have exactly the same meaning and units as in Eqs 23 and 24, but for notational efficiency we keep the same symbols. The rate equations for $v_1(\cdot), \ldots, v_5(\cdot)$ are the same as used in previous sections.

The explicit modeling of the incorporation of metabolites in proteins also allows for a more principled way of defining the growth rate, by setting the biomass equal to the total mass of molecules relative to the mass of M:

$$B = (4X_1 + X_2 + M + 5E_1 + 5E_3) \cdot \alpha \cdot V_{population} \cdot \beta / V_{medium}. \tag{29}$$

This yields the following expression for the growth rate:

$$\mu = \frac{\dot{B}}{B} = \frac{4v_1 + v_3}{4X_1 + X_2 + M + 5E_1 + 5E_3}. \tag{30}$$

Exercise 10 *How can you explain Eq. 29? Derive Eq. 30 from Eq. 29 and the extended model of Eqs 25 and 26.*

Exercise 11 *Extend the model so as to explicitly describe the dependency of the reaction rates v_2 and v_4 on the concentrations of enzymes E_2 and E_4. Add differential equations for the dynamics of these enzyme concentrations.*

The extended model can be used to simulate diauxic growth, including the synthesis of enzymes from precursors.

Exercise 12 *Run the kinetic model stored in* `metabolicModelGeneRegulationGrowthControl-NewBiomass.m` *by means of the file* `simulateMetabolicSystemGeneRegulationGrowthControl-NewBiomass.m`*. Compare the results with those obtained using the models in Section 4. In particular, which differences do you observe for enzyme concentrations that are initially 0? Can you explain the differences?*

References

[1] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* Chapman & Hall/CRC, Boca Raton, FL, 2007.

[2] S. Berthoumieux, H. de Jong, G. Baptist, C. Pinel, C. Ranquet, D. Ropers, and J. Geiselmann. Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.*, 9:634, 2013.

[3] K. Bettenbrock, S. Fischer, A. Kremling, K. Jahreis, T. Sauter, and E. D. Gilles. A quantitative approach to catabolite repression in *Escherichia coli. J. Biol.Chem.*, 281:2578–2584, 2006.

[4] J. Deutscher, C. Francke, and P. W. Postma. How Phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.*, 70(4):939–1031, 2006.

[5] L. Gerosa, K. Kochanowski, M. Heinemann, and U. Sauer. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.*, 9:658, 2013.

[6] B. Görke and J. Stülke. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.*, 6(8):613–624, 2008.

[7] R. Heinrich and S. Schuster. *The Regulation of Cellular Systems.* Chapman & Hall, New York, 1996.

[8] L. Keren, O. Zackay, M. Lotan-Pompan, U. Barenholz, E. Dekel, V. Sasson, G. Aidelberg, A. Bren, D. Zeevi, A. Weinberger, U. Alon, R. Milo, and E. Segal. Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.*, 9:701, 2013.

[9] S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375, 2009.

[10] A. Kremling. *Systems Biology: Mathematical Modeling and Model Analysis.* CRC Press, Boca Raton, FL, 2014.

[11] A. Kremling, J. Geiselmann, D. Ropers, and H. de Jong. Understanding carbon catabolite repression in *Escherichia coli* using quantitative models. *Trends Microbiol.*, 23(2):99–109, 2015.

[12] K.L. Larrabee, J.O. Phillips, G.J. Williams, and A.R. Larrabee. The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*. *J. Biol. Chem.*, 255(9):4125–30, 1980.

[13] R. Mahadevan, J. S. Edwards, and F. J. Doyle. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.*, 83(3):1331–1340, 2002.

[14] J. Monod. *Recherches sur la Croissance des Cultures Bactériennes*. Hermann et Cie, Paris, 1942.

[15] R.D. Mosteller, R.V. Goldstein, and K.R. Nishimoto. Metabolism of individual proteins in exponentially growing *Escherichia coli*. *J. Biol. Chem.*, 255(6):2524–32, 1980.

[16] A. Narang. Quantitative effect and regulatory function of cyclic adenosine 5'-phosphate in *Escherichia coli*. *J. Biosci.*, 34(3):445–463, 2009.

[17] E.J. O'Brien, J.A. Lerman, RL. Chang, D.R Hyduke, and B.O. Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.*, 9:693, 2013.

[18] J.D. Orth, I. Thiele, and B.O. Palsson. What is flux balance analysis? *Nat. Biotechnol.*, 28(3):245–8, 2010.

[19] B.O. Palsson. *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press, Cambridge, 2nd edition, 2015.

[20] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.*, 3:119, 2007.

[21] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, 2012.

[22] S. Schuster, T. Pfeiffer, and D. A. Fell. Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.*, 252(3):497–504, 2008.

[23] A. Ullmann. Catabolite repression: a story without end. *Res. Microbiol.*, 147(6-7):455–458, 1996.