In: Biotechnology ISBN 978-1-62081-071-2
Editors: L.G. Torres Bustillos and I. Garcia Peña

Chapter 15

MICROBIAL GROWTH MODELING AND ANALYSIS

Edgar Salgado-Manjarrez*, Juan S. Aranda-Barradas and Inés García-Peña
Bioprocess and Bioengineering departments, UPIBI IPN, Mexico

INTRODUCTION

Human beings have exploited the catalytic abilities of microorganisms for millennia, although it was not consciously until the discovery of microorganisms and enzymes. Since then, the number of products and services using microorganisms has multiplied geometrically. It is quite clear to us now that cells possess tremendous catalytic activities, as life can thrive in almost every corner of our planet, often forming *consortia* for degrading substances that hardly could be considered as substrates for life. These amazing capabilities make microorganisms and other cultured cells powerful catalyst that can be exploited just as they are found in nature or even improved by modifying their physiology. At the middle of the 20th century, changes in the physiology of cells were produced by mutation, a cluttered chaotic process that required lots of experimentation; latter on, by the 1970's, recombinant DNA technologies gave birth to the so called "genetic engineering", by which specific genes are introduced in a cell, modifying more specifically its physiology. However, the last 30 years have taught us that, in many cases, a "cut", "mix" and "paste" logic is not very reliable for modifying the catalytic properties of cells: there exist several control systems that often make the results unpredictable. The genomics, proteomics, metabonomics and most –omics are advancing our understanding of cells physiology at different degrees and levels. From an engineering perspective, being able to truly reengineer cell systems, predicting the results of genetic manipulations, is almost the "Holy Grail" of biotechnology.

Cell growth is an extremely complex process: thousands of reactions among thousands of substances happen simultaneously, catalyzed by thousands of enzymes, everything orchestrated by thousands of genes and other genetic control systems. Nothing yet invented by humanity reaches such complexity and yet, every year we get closer to be able to truly

* Email: esalgado@ipn.mx

reengineer cells or design novel cell systems. The complexity of the task involves the conjunction of several disciplines and the manipulation of huge amounts of data. Additionally to this biological knowledge, important concepts from reaction engineering, computer science and chemical kinetics are needed to develop a quantitative approach to cell metabolism, in order to exploit biological systems in rational biomolecule production processes. Systems Biology is considered as the part of Bioinformatics that using a systems approach, intends to unveil the interactions in biochemical systems; to some extent, it is hardly differentiated from Computational Biology, an area that uses specifically mathematical modeling and simulation to understand complex biological systems. As a part of this areas, Metabolic Engineering uses a computational, systems biology approach for manipulating biochemical and metabolic systems in a rational way through an estimation of fluxes inside the cells, accomplished by means of modeling and, more specifically, Metabolic Flux Analysis also known as Metabolic Flux Balancing. It is through this formalism that, presumably, we will in the future be able to reengineer cells or create novel cell systems.

CELL KINETICS AND DYNAMICS OF CELL CULTURES

Modeling of reacting biological systems requires a translation of basically two aspects of the system characteristics to a mathematical structure:

a) The system's changes in time considering external factors, such as homogeneity of the reaction medium, transport phenomena, pH and temperature. This behavior is depicted by the dynamical description of cell cultures. In a mathematical way, cellular culture dynamics is expressed by time-dependent differential equations that consider variations in the main features of the system (e.g. chemical species concentrations, temperature, pH) as produced by inputs, outputs, chemical reactions or any external factor to the reacting system.
b) The changes in the chemical or biochemical species that occur in the reacting system which are represented by a kinetics theory applicable to the involved chemical or biochemical reactions. In biological systems such as cell cultures aiming at producing some economically valuable metabolite, cell kinetics should comprise the suitable biological knowledge for the production process. So, kinetics of biological reactions can include only macroscopic concentrations and lumped parameters or, in an increasingly detailed depiction, data from intracellular compartments, metabolomics, reaction thermodynamics, genomics or proteomics.

The roughest form of cellular growth and production kinetics description consists in the so-called unstructured models that express simplified relations between reactive concentrations and kinetic rates. Introducing comprehensive knowledge on the main biochemical pathways allows for the analysis of metabolic fluxes that results in improved assessments of process yields. In general, the mathematical complexity of kinetic models is proportional to the description level they propose.

CELL GROWTH STOICHIOMETRY

Stoichiometry deals with the balancing of chemical reactions. Cell reactions are not different from any other reaction in this aspect, and then the methods used are equally applicable. The cell complexity arises from the number of reactions. In order to facilitate comprehension, it is both, useful and illustrating to treat the cell as the product and catalyzer of a very simple reaction, namely,

Substrates → Cell mass + Excreted metabolic products + Heat

Taking into account that C, H, O and N represent most of cell mass, cell requires that these atoms be present in the substrates. The reaction can be written considering abstract carbon and nitrogen sources and a metabolic product, as

$$\begin{aligned} &\eta_S\, C_{a_S} H_{b_S} O_{c_S} N_{d_S} + \eta_O\, O_2 + \eta_N\, C_{a_N} H_{b_N} O_{c_N} N_{d_N} \rightarrow \\ &\eta_X\, CH_{b_X} O_{c_X} N_{d_X} + \eta_P\, C_{a_P} H_{b_P} O_{c_P} N_{d_P} + \eta_C\, CO_2 + \eta_A\, H_2O \end{aligned} \tag{1}$$

where it has been assumed that growth is aerobic, hence the O_2, and CO_2 is produced, as well as H_2O as part of cell metabolism. In this equation η_j represent the stoichiometric coefficients, the number of moles of each substance.

All elements in a reaction must balance, for the carbon for instance,

$$\eta S \cdot aS + \eta N \cdot aN = \eta X \cdot 1 + \eta P \cdot aP + C \cdot 1$$

or as a summation,

$$\sum_j \left(\eta_j \cdot a_j\right) = 0 \tag{2}$$

which can be written in matrix form as:

$$\sum_j \left(\eta_j \cdot a_j\right) = \begin{bmatrix} a_S & a_N & 1 & a_P & 1 \end{bmatrix} \begin{bmatrix} -\eta_S \\ -\eta_N \\ \eta_X \\ \eta_P \\ \eta_C \end{bmatrix} = 0 \tag{3}$$

and considering every element in the reaction,

$$\begin{bmatrix} a_S & 0 & a_N & 1 & a_P & 1 & 0 \\ b_S & 0 & b_N & b_X & b_P & 0 & 2 \\ c_S & 2 & c_N & c_X & c_P & 2 & 1 \\ d_S & 0 & d_N & d_X & d_P & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -\eta_S \\ -\eta_O \\ -\eta_N \\ \eta_X \\ \eta_P \\ \eta_C \\ \eta_A \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4}$$

which can be written in compact form as

$$\mathbf{E} \cdot \boldsymbol{\eta} = \mathbf{0} \,. \tag{5}$$

where **E** is the elemental matrix with *k* rows and *n* columns, for *k* elements and *n* substances. $\boldsymbol{\eta}$ is the stoichiometric coefficients vector of size *n*. Notice that the reactants' or substrates' stoichiometric coefficients have a minus sign, since these substances are consumed in the chemical reaction.

In order to get a solution for the system of equations, such as equation (5) row number in **E** must be equal to the vector size $\boldsymbol{\eta}$. Otherwise, if there are more rows than columns and the system is underdetermined, or overdetermined, if there are more rows than columns. In terms of elements and substances, the degrees of freedom are $F = n - k$. For calculation purposes, the system can be re-written in terms of measured and calculated sub-matrices,

$$\mathbf{E} \cdot \boldsymbol{\eta} = \begin{bmatrix} \mathbf{E}^c & \mathbf{E}^m \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\eta}^c \\ \boldsymbol{\eta}^m \end{bmatrix} = \mathbf{0} \tag{6}$$

$$\mathbf{E}^c \cdot \boldsymbol{\eta}^c + \mathbf{E}^m \cdot \boldsymbol{\eta}^m = \mathbf{0} \,. \tag{7}$$

then the system can be solved as:

$$\boldsymbol{\eta}^c = -\left(\mathbf{E}^c\right)^{-1} \cdot \mathbf{E}^m \cdot \boldsymbol{\eta}^m \,. \tag{8}$$

It must be noted that $\mathbf{E}^c$ has to be invertible, so it must be square, its determinant must be different from zero and must be of full rank. If these conditions are fulfilled the system is said to be *observable*. In this way unknown coefficients in the reaction can be calculated from some known ones.

This previous stoichiometric analysis can be extended to a kinetic approach in the form:

$$\mathbf{r}^c = -\left(\mathbf{E}^c\right)^{-1} \cdot \mathbf{E}^m \cdot \mathbf{r}^m \,. \tag{9}$$

where the $\boldsymbol{\eta}$ vectors of stoichiometric coefficients are replaced by volumetric rate **r** vectors. This comprehensive approach allows for the estimation of valuable metabolites in production systems through the modeling of kinetic behavior by a stoichiometric analysis. Barrera-Martínez et al. (2011) settled a stoichiometric black box model for the assessment of bioethanol production in *Saccharomyces cerevisiae* fed-batch cultures, as illustrated in Figure 1.

By measuring the biomass accumulation, the glucose consumption and the oxygen supply volumetric rates, the estimation of produced ethanol was possible from a simplified stoichiometric description of the production system. As it can be observed, however, that the estimation error has some tendency and, in general doesn't follow quite well the experimental values. A better estimation can be obtained with a metabolic model, as explained latter.

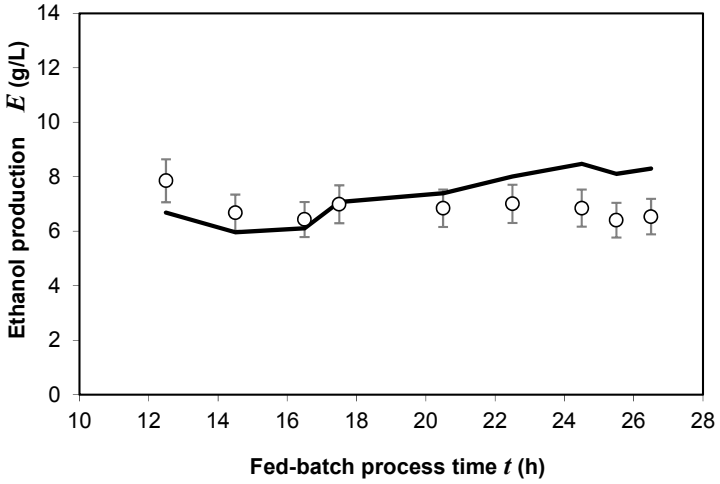

Figure 1. Ethanol production by *Saccharomyses cerevisiae* in a fed-batch system: experimental data (markers) and estimated concentrations by means of a stoichiometric black box analysis of cell growth (continuous line).

NON-STRUCTURED MODELING

Stoichiometry does not give information about the rate at which a reaction takes place. In fact, there is no way to predict this information; it has to be determined experimentally. An apparatus commonly use for both, kinetic studies and production scale is the tank reactor as depicted in Figure 2.

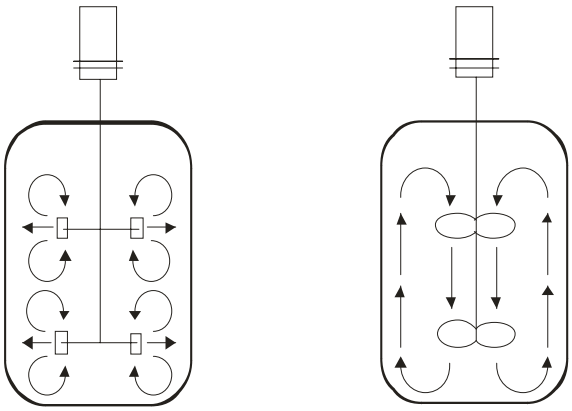

Figure 2. Classical stirred tank bioreactors.

A balance carried out over the solutes in mass or moles, **y**, in such a system reads,

$$\frac{d(V\,\mathbf{y})}{dt} = \mathbf{R} + F_0\,\mathbf{y}_0 - F\,\mathbf{y} \quad . \tag{10}$$

here **R** stands for the rate of change at which the substance is consumed or produced by the reaction. A similar balance can be written for the *mass* in the system, from which it can be concluded that as long as the density changes are negligible,

$$\frac{dV}{dt} \approx F_0 - F \tag{11}$$

as no mass is produced nor consumed in non-nuclear reactions. Combining the previous equations the final equation is as follows,

$$\frac{d\mathbf{y}}{dt} = \mathbf{r} + D\left(\mathbf{y}_0 - \mathbf{y}\right). \tag{12}$$

where **r** is the volumetric rate of change and $D = F_0/V$ is the dilution rate. This rather simple equation expresses the dynamics of solutes in the system: how the concentration of substances in a reactor changes depending on their input and output concentrations, the feed rate and the reaction rate. As such, the balance equation can be used for determining reaction rates from experimental dynamic information but it can also be used for the simulation of the behavior of the system under different conditions given some kinetic information or models.

Some simple models have been proposed that fit rather well most of the observed behaviors. For instance, the classic model of Monod,

$$\mu = \frac{r_X}{X} = \mu_{\max} \frac{S}{S + K_S}. \tag{13}$$

expresses the *specific growth rate* in function of a substrate concentration. Substrate consumption can be accounted by a model like the one of Pirt,

$$\sigma = \frac{r_S}{X} = \frac{\mu}{Y_{XS}} + m. \tag{14}$$

where σ stands for the *specific consumption rate*, Y_{XS} is the biomass yield on the substrate and m is a coefficient for accounting of all the substrate that is not used for maintenance. A similar model can also be used for the products,

$$\pi = \frac{r_P}{X} = \alpha\mu + \beta \tag{15}$$

originally proposed by Luedeking and Piret. In this equation α can be interpreted as product yield or proportionality between production and growth; and β stands for the non-growth associated production.

As these kind of models assume that biomass is a “complex” substance, they are considered non-structured. In spite of this extreme simplification, non-structured models have

been used successfully for describing a number of metabolite production systems (Noorman et al. 1993). Production of biomolecules issued from the energetic or primary metabolism, such as ethanol (Aranda and Salgado, 2002; Barrera-Martínez et al., 2011), polyols (Delia et al., 1999; Aranda et al., 2000), simple organic acids or amino acids are classically depicted through unstructured models. Processes aiming at producing secondary metabolites, such as enzymes (Martínez-Trujillo et al., 2008; Martínez-Trujillo et al., 2009) or antibiotics, can be equally described with unstructured models. Wastewater treatment processes are also successfully modeled by this unstructured approach (Rivera-Salvador et al., 2009). As a particular example, Aranda et al., (2000) used a non-structured model, based on the Monod kinetic model considering oxygen as the limiting substrate and couplet to a simple mass transfer model to demonstrate how *Candida parapsilosis* always produces xylitol, but this alcohol only accumulates under oxygen limitation. What is more, the model suggests that there exist an optimal oxygen transfer rate for xylitol production, as depicted in Figure 3.

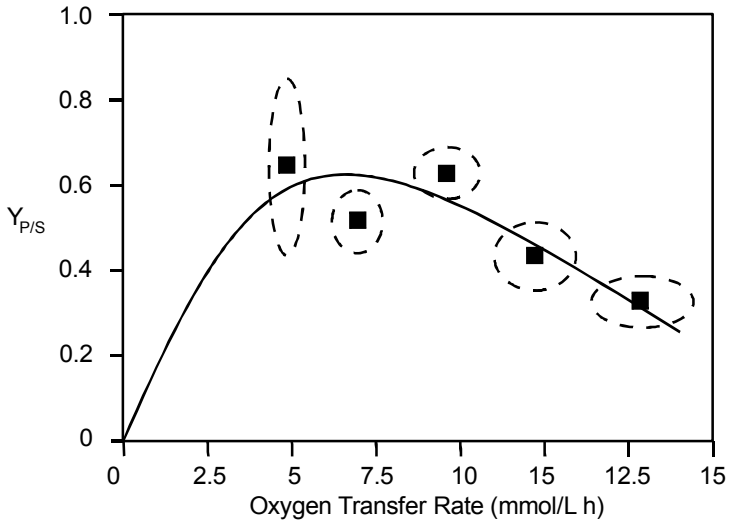

Figure 3. Xylitol production yields by *Candida parapsilosis* as a function of the oxygen transfer rate in the culture.

Indeed, non-structured models have been quite useful and its usefulness will surely continue for some years to come. However, since the 1960's a better, more realistic, approach started to develop, one that through reactions and/or substances, introduced some structure to the biomass.

METABOLIC MODELING

In bioreactors with cell cultures, catalytic reactions take place mostly inside the cells or in their surfaces. In the external medium, reactions are not usually catalyzed, what is observed in the medium is the net flow of substances from and into the cells, so, instead of considering the cells as one complex substance, it is more realistic to consider it as a *biotic* phase suspended in an *abiotic* phase or medium.

A mass balance on the abiotic phase reads,

$$\frac{d(V_A \mathbf{y}_A)}{dt} = F_{A,0}\,\mathbf{y}_{A,0} - F_A\,\mathbf{y}_A + \mathbf{R}_I \quad . \tag{16}$$

notice that concentrations are relative to the abiotic phase which volume relates to the total volume by a biotic phase volume fraction, φ, so $V_A = (1-\varphi)\cdot V$, and:

$$\frac{d\left(\left[1-\varphi\right]\mathbf{y}_A\right)}{dt} = D\left(\left[1-\varphi_0\right]\mathbf{y}_{A,0} - \left[1-\varphi\right]\mathbf{y}_A\right) + \mathbf{r}_I \quad . \tag{17}$$

as $\left[1-\varphi\right]\mathbf{y}_A = \mathbf{y}$ so the mass balance reduces to:

$$\frac{d\mathbf{y}}{dt} = D\left(\mathbf{y}_0 - \mathbf{y}\right) + \mathbf{r}_I \,. \tag{18}$$

that is identical to equation (12) except that here the kinetic term has become a transport rate, not a reaction rate. Clearly, this truly matches to reality. It should be also noted that in the previous equation volumetric exchange rates are referred to total volume, V, the sum of both volume phases.

The mass balance for the biotic phase is:

$$\frac{d(V_B \mathbf{y}_B)}{dt} = F_{B,0}\,\mathbf{y}_{B,0} - F_B\,\mathbf{y}_B + \left(\mathbf{R}_1 + \mathbf{R}_2 + \ldots + \mathbf{R}_m\right) - \mathbf{R}_I \,. \tag{19}$$

where the rate terms in parenthesis indicate the metabolic reactions that contribute to the production or consumption of a particular metabolite. Taking into account the biotic phase volume:

$$\frac{d\left(\varphi\ \mathbf{y}_B\right)}{dt} = D\left(\varphi_0\ \mathbf{y}_{B,0} - \varphi\ \mathbf{y}_B\right) + \sum_j \mathbf{r}_j - \mathbf{r}_I \quad . \tag{20}$$

as in the previous case, concentrations can be referred to the total volume,

$$\frac{d\dot{\mathbf{y}}}{dt} = D\left(\dot{\mathbf{y}}_0 - \dot{\mathbf{y}}\right) + \sum_j \mathbf{r}_j - \mathbf{r}_I \,. \tag{21}$$

the upper dot is to distinguish them from abiotic phase concentrations. Notice, however, that volumetric exchange rates and volumetric reaction rates are all referred to total volume.

Production or consumption rates of a given substance in a reaction can be expressed as the product of a stoichiometric coefficient and the reaction rate, which is usually equal to the reaction rate of a reference substance in the reaction, which has a unity stoichiometric coefficient. According to this,

$$r_i = \sum_j \alpha_{ji} r_j \tag{22}$$

so if the whole set of metabolic reactions are, in matrix form, for every substrate, metabolic product and biomass constituent,

$$\mathbf{r}_{S,i} = -\sum_j \alpha_{ji} \mathbf{r}_j = \mathbf{A}^T \cdot \mathbf{r}_M \,. \tag{23}$$

$$\mathbf{r}_{P,i} = \sum_j \beta_{ji} \mathbf{r}_j = \mathbf{B}^T \cdot \mathbf{r}_M \,. \tag{24}$$

$$\mathbf{r}_{X,i} = -\sum_j \gamma_{ji} \mathbf{r}_j = \mathbf{\Gamma}^T \cdot \mathbf{r}_M \,. \tag{25}$$

Matrices **A**, **B**, and **Γ** have *m* rows and *N*, *M* and *L* columns, respectively. Also, in these matrices the rows correspond to the reactions and the columns to the substances, this is the reason to transpose when calculating the rates for substances. The whole set of balanced metabolic reactions can be written as,

$$\mathbf{A} \cdot \mathbf{S} + \mathbf{B} \cdot \mathbf{P} + \mathbf{\Gamma} \cdot \mathbf{X} = \mathbf{0} \,. \tag{26}$$

where **S**, **P**, and **X**, stand for the substrate, metabolic products and biomass constituents (not their rates). Taking this into account,

$$\sum_{j=1}^{m} \mathbf{r}_j = \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \\ \mathbf{\Gamma}^T \end{bmatrix} \cdot \mathbf{r}_M = \mathbf{Y} \cdot \mathbf{r}_M \,. \tag{27}$$

where Y is the stoichiometric matrix, finally:

$$\frac{d\mathbf{y}}{dt} = D\left(\mathbf{y}_0 - \mathbf{y}\right) + \mathbf{Y} \cdot \mathbf{r}_M - \mathbf{r}_I \,. \tag{28}$$

the sum of all substances in this reaction is equal to the mass of the biotic phase, i.e, the biomass,

$$\frac{dX}{dt} = D\left(X_0 - X\right) + \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} - \sum_{i=1}^{n} r_I \,. \tag{29}$$

or,

$$\frac{1}{X}\frac{dX}{dt} = D\left(\frac{X_0}{X} - 1\right) + \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij} - \sum_{i=1}^{n} q_I \ . \tag{30}$$

where rates q are *specific*. By comparison of equations 29, 30 and 12, it becomes clear that,

$$\mu = \sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij} - \sum_{i=1}^{n} q_I \ . \tag{31}$$

i.e., the specific growth rate is the result of the sum of exchange rates, increasing the biomass when entering, decreasing it when exiting; plus the net conversion rates of each substance inside the cell due to every metabolic reaction.

As the reference to the total volume might be confusing, it is more convenient to rewrite this equation with concentrations and rates referred to cell mass, as follows: Biomass constituents are $\varphi \cdot \mathbf{y}_B = \varphi \cdot \rho_X \cdot \mathbf{x} = X \cdot \mathbf{x}$, where ρ_X is the biomass density and $\mathbf{x}$ are the concentrations referred to total biomass.

$$\frac{d(X\,\mathbf{x})}{dt} = D\left(X_0\,\mathbf{x}_0 - X\,\mathbf{x}\right) + \mathbf{Y}\cdot\mathbf{r}_M - \mathbf{r}_I \ . \tag{32}$$

which leads to:

$$\frac{d\mathbf{x}}{dt} = D\frac{X_0}{X}\left(\mathbf{x}_0 - \mathbf{x}\right) + \mathbf{Y}\cdot\mathbf{q}_M - \mathbf{q}_I - \mathbf{x}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij} - \sum_{i=1}^{n} q_I\right). \tag{33}$$

or,

$$\frac{d\mathbf{x}}{dt} = D\frac{X_0}{X}\left(\mathbf{x}_0 - \mathbf{x}\right) + \mathbf{Y}\cdot\mathbf{q}_M - \mathbf{q}_I - \mu\,\mathbf{x} \ . \tag{34}$$

This equation states that biotic phase composition changes in response to feed and discharges; metabolic reactions; exchange rates with the abiotic phase and concentrations are diluted proportionally to growth rate. If desired, a more succinct form of this equation can be written by grouping the metabolic reactions and exchange rates,

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_M \\ \mathbf{r}_I \end{bmatrix}. \tag{35}$$

and grouping all stoichiometric and exchange coefficients in a metabolic matrix (Noorman et al, 1993),

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix}^{-\mathbf{I}}. \tag{36}$$

so,

$$\frac{d\mathbf{x}}{dt} = D\frac{X_0}{X}\left(\mathbf{x}_0 - \mathbf{x}\right) + \mathbf{M}\cdot\mathbf{q} - \mu\cdot\mathbf{x}. \tag{37}$$

or by considering the three sets for substrates, products and biomass constituents,

$$\frac{d\mathbf{x}}{dt} = D\frac{X_0}{X}\left(\mathbf{x}_0 - \mathbf{x}\right) + \begin{bmatrix} \mathbf{A}^T & -\mathbf{I} & \mathbf{0} \\ \mathbf{B}^T & \mathbf{0} & -\mathbf{I} \\ \mathbf{\Gamma}^T & \mathbf{0} & \mathbf{0} \end{bmatrix}\cdot\mathbf{q} - \mu\cdot\mathbf{x}. \tag{38}$$

This equation sets the basis for manipulating the huge amounts of data involved in cell metabolism.

Steady State Metabolic Flux Analysis

Equations (18) and (34), state how the biomass composition changes during cell culture, the only difference between them is how the concentration is expressed: one uses total culture volume as reference, while the other uses biomass weight. Assuming steady state and sterile feed,

$$0 = -D\dot{\mathbf{y}} + \mathbf{Y}\cdot\mathbf{r}_M - \mathbf{r}_I. \tag{39}$$

$$0 = \mathbf{Y}\cdot\mathbf{q}_M - \mathbf{q}_I - \mu\mathbf{x}. \tag{40}$$

and,

$$\mathbf{Y}\cdot\mathbf{r}_M = D\dot{\mathbf{y}} + \mathbf{r}_I \tag{41}$$

$$\mathbf{Y}\cdot\mathbf{q}_M = \mathbf{q}_I + \mu\mathbf{x}. \tag{42}$$

which give a convenient way for estimating metabolic reaction rates or fluxes in function of exchange (or "conversion") rates and a culture or biomass dilution rate *if steady state is achieved.*

As not necessarily all exchange rates are measured it is more convenient to take the concise forms,

$$\mathbf{M} \cdot \mathbf{r} = D \cdot \mathbf{x} . \tag{43}$$

$$\mathbf{M} \cdot \mathbf{q} = \mu \cdot \mathbf{x} . \tag{44}$$

Mass fractions of metabolic intermediaries are usually quite low, accounting for ca. 5% of biomass weight so for these substances (for the sake of brevity, in the following only the version with specific rates is considered, but the reasoning holds for the version with volumetric rates),

$$\mathbf{M} \cdot \mathbf{q} = \mathbf{0} \tag{45}$$

Now, this equation is identical in form to equation (5), so the same reasoning applies for calculating some rates or fluxes in function of others,

$$\mathbf{M}^c \cdot \mathbf{q}^c + \mathbf{M}^m \cdot \mathbf{q}^m = \mathbf{0} \tag{46}$$

$$\mathbf{q}^c = -\left(\mathbf{M}^c\right)^{-1} \cdot \mathbf{M}^m \cdot \mathbf{q}^m . \tag{47}$$

This simple equation is the basis for steady-state metabolic flux analysis: the estimation of fluxes in a proposed biochemical network in function of others, usually external, fluxes (Vallino and Stephanopoulos, 1993).

If convenient, the metabolic matrix can be expressed as follows,

$$\begin{bmatrix} \mathbf{A}^T & -\mathbf{I} & \mathbf{0} \\ \mathbf{B}^T & \mathbf{0} & -\mathbf{I} \\ \mathbf{\Gamma}^T & \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{q}_M \\ \mathbf{q}_S \\ \mathbf{q}_P \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \\ \mathbf{\Gamma}^T \end{bmatrix} \cdot \mathbf{q}_M + \begin{bmatrix} -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{q}_S \\ \mathbf{q}_P \end{bmatrix} \tag{48}$$

or,

$$\mathbf{Y} \cdot \mathbf{q}_M = \mathbf{q}_I \tag{49}$$

$$\begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \\ \mathbf{\Gamma}^T \end{bmatrix} \cdot \mathbf{q}_M = \begin{bmatrix} \mathbf{q}_S \\ \mathbf{q}_P \\ \mathbf{0} \end{bmatrix} . \tag{50}$$

Which can be written as (Nielsen and Villadsen,1994):

$$\begin{bmatrix} \mathbf{q}_I^m \\ \mathbf{q}_I^c \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 & \mathbf{Y}_2 \\ \mathbf{Y}_3 & \mathbf{Y}_4 \\ \mathbf{Y}_5 & \mathbf{Y}_6 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{q}_M^m \\ \mathbf{q}_M^c \end{bmatrix} \tag{51}$$

where some elements of $\mathbf{q}_S$ and $\mathbf{q}_P$ are in $\mathbf{q}_I^m$ and others in $\mathbf{q}_I^c$. In $\mathbf{q}_M^m$ are found a minimum set of rates that can determine the system while $\boldsymbol{q}_M^m$ and $\mathbf{q}_M^c$ only reflects the mentioned arrangement but does not imply that any of $\mathbf{q}_M$ is actually measured. Following this rearrangement, the solution to vectors $\mathbf{q}_I^m$, $\mathbf{q}_M^c$ and $\mathbf{q}_I^c$ are:

$$\begin{aligned} \mathbf{q}_M^m &= \left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{Y}_6^{-1}\mathbf{Y}_5\right)\cdot \mathbf{q}_I^m \\ \mathbf{q}_M^c &= -\mathbf{Y}_6^{-1}\mathbf{Y}_5\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{Y}_6^{-1}\mathbf{Y}_5\right)^{-1}\cdot \mathbf{q}_I^m \\ \mathbf{q}_I^c &= \left(\mathbf{Y}_3 - \mathbf{Y}_4\mathbf{Y}_6^{-1}\mathbf{Y}_5\right)\cdot\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{Y}_6^{-1}\mathbf{Y}_5\right)^{-1}\cdot \mathbf{q}_I^m \end{aligned} \tag{52}$$

It is clear that,

$$\mathbf{q}^c = \begin{pmatrix} \mathbf{q}_M^m \\ \mathbf{q}_M^c \\ \mathbf{q}_I^c \end{pmatrix} \tag{53}$$

as in the equivalent equation (8).

Table 1. Metabolic reactions for a *Saccharomyces cerevisiae* simplified model

Flux	Condensed reaction
r_1	Glucose + $2H_3PO_4$ + 2ADP + $2NAD^+$ → 2Pyruvate + $2H_2O$ + 2ATP + 2NADH
r_2	Glucose + Pyruvate + Acetyl-CoA + NH_3 + 27.5ATP + $24.1H_2O$ → Biomass +
r_3	CoA + 27.5ADP + $27.5H_3PO_4$
r_4	Pyruvate + CoA + NAD + → Acetyl-CoA + CO_2 + NADH
r_5	Acetyl-CoA + $2H_2O$ + H3PO4 + ADP + $4NAD^+$ → $2CO_2$ + CoA + ATP +
r_6	4NADH
	Acetyl-CoA + H_3PO_4 + ADP + 2NADH → Ethanol + CoA + H_2O + ATP + $2NAD^+$
	3ADP + $3H_3PO4$ + NADH + 0.5O2 → 3ATP + NAD+ + $4H_2O$

It is important to mention that the previous analysis only provide us for a way of inferring some fluxes from others: the solution is mathematically correct, however, as everything comes from an assumed, albeit very possible, stoichiometry, the calculated fluxes can be considered very probable, but should not be assumed as certain. Any other plausible stoichiometry that makes the system observable, would give a set of fluxes that satisfies the constraints but the values can be different between two proposed stoichiometries even in some common paths. The analysis presented above was used by Barrera-Martínez et al.

(2011) for estimating the physiological state of yeast cultures with a simplified metabolic model (Table 1). They showed, for instance, that for increasing ethanol production rate, a high substrate inflow and a low oxygen one are required, whereas with a controlled glucose feed under aerobic conditions maximizes oxidative catabolism that is required for high biomass yields, as shown in Figure 4.

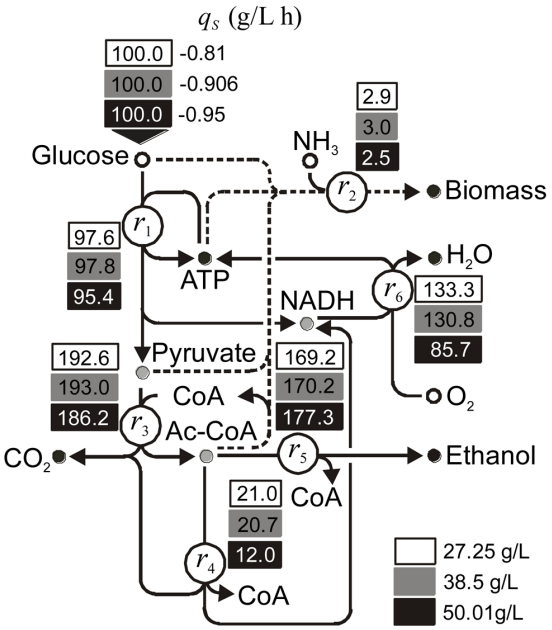

Figure 4. A minimal metabolic network for a yeast culture producing biomas and bioethanol. Metabolic fluxes for different glucose concentrations in the fed stream (27.2, 38.5 and 50.0 g/L) are referred to the glucose uptake (100%).

OVERDETERMINED SYSTEMS

When estimations for fluxes are made with the least amount of information required to specify the system, the solution is as good as the input data. Unfortunately, measurements always have experimental error so the estimated fluxes reflect the propagation of experimental error. If more rates than necessary are measured, the redundant information can be used to obtain a "best solution" under a given policy. A least squares solution, for instance, can be obtained as follows:

$$\mathbf{q}^c = -\left(\mathbf{M}^T \cdot \mathbf{M}^c\right)^{-1} \cdot \mathbf{M}^T \cdot \mathbf{M}^m \cdot \mathbf{q}^m . \tag{54}$$

Even measurements can also be corrected from the least squares solution, by introducing the latter solution in Equation (47).

$$\left(\mathbf{M}_m - \mathbf{M}_c \cdot \left(\mathbf{M}_c^T \cdot \mathbf{M}_c\right)^{-1} \cdot \mathbf{M}_c^T \cdot \mathbf{M}_m\right) \cdot \mathbf{q}_m = 0 = \mathbf{R} \cdot \mathbf{q}_m \tag{55}$$

Here **R** is the *redundancy* matrix, and usually has not full rank. When dependent rows are eliminated a reduced, full rank set is obtained,

$$\mathbf{R}_R \cdot \mathbf{q}_m = 0 \tag{56}$$

Because of experimental errors, this equality is not satisfied in practice. If the error, $\boldsymbol{\delta}$, is substracted from these measured rates with noise, $\overline{\mathbf{q}}_m$, the true values of the measured rates should be obtained,

$$\mathbf{R}_R \cdot \left(\overline{\mathbf{q}}_m - \boldsymbol{\delta}\right) = 0 \,. \tag{57}$$

from this equation the residuals, $\boldsymbol{\varepsilon}$, can be defined,

$$\mathbf{R} \cdot \overline{\mathbf{q}}_m = \mathbf{R} \cdot \boldsymbol{\delta} = \boldsymbol{\varepsilon} \neq \mathbf{0} \tag{58}$$

An estimation of the error, can be obtained as follows

$$\min_{\boldsymbol{\delta}} \Phi = \boldsymbol{\delta}^T \cdot \mathbf{V}_m^{-1} \cdot \boldsymbol{\delta} \text{ subject to } \mathbf{R} \cdot \boldsymbol{\delta} = \boldsymbol{\varepsilon} \,. \tag{59}$$

here, the measurements variance is used as weight. The solution to this problem is,

$$\hat{\boldsymbol{\delta}} = \mathbf{V}_m \cdot \mathbf{R}_R{}^T \cdot \left(\mathbf{R}_R \cdot \mathbf{V}_m \cdot \mathbf{R}_R{}^T\right)^{-1} \cdot \boldsymbol{\varepsilon} \tag{60}$$

The assessment of $\hat{\boldsymbol{\delta}}$ is useful for estimating the rate confidence intervals, as for the respiratory coefficient (RQ) in a biomass culture of *Saccharomyces cerevisiae* (Figure 5).

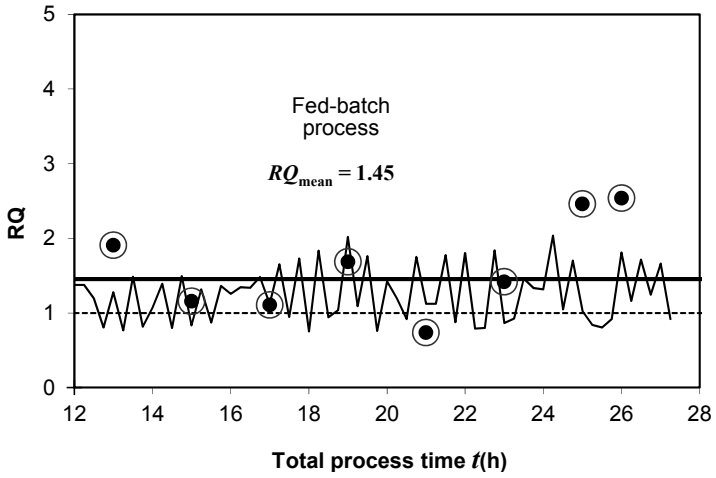

Figure 5. Estimated RQ for a yeast biomass fed-batch production process. Circles surrounding the calculated values (black markers) represent the estimated RQ confidence intervals.

UNDERDETERMINED SYSTEMS

The proposed stoichiometry is a set of constraints about what is thought the cell can do. When the measured variables are not enough to determine the system, it can yet be solved if more constraints or a suitable objective function is added. It is possible, for instance, to find the fluxes that would maximize cell or product yield.

This is carried out with linear programming methods for which it is necessary that all fluxes be positive in the equation, so the stoichiometry is rewritten splitting reversible reactions as a forward and a reverse reaction, this gives an extended system,

$$\mathbf{M}_{ex} \cdot \mathbf{q}_{ex} = \mathbf{0} \tag{61}$$

an objective function is specified and the solution found by a linear programming method, such as the simplex.

Through the use of this strategy, Aranda-Barradas et al. (2010) demonstrated that biomass yields could be estimated within an error of 12.8% of the experimental value.

Table 2. Transposed total stoichiometric matrix for the metabolic network for the xylose to xylitol bioconversion by *Candida parapsilosis*

Metabolites	Reactions													
	r_3	r_2	r_1	r_{10}	r_5	r_6	r_7	r_8	r_9	r_4	r_{11}	r_{12}	r_{13}	r_{14}
Xylose	0	−1	−1	0	0	0	0	0	0	0	0	0	0	0
O_2	0	−0.5	0	0	0	0	0	0	0	0	0	0	0	0
Xylitol	−1	0	1	0	0	0	0	0	0	0	0	0	0	0
Biomass	0	5	0	0	0	0	0	0	0	0	0	0	0	0
NH_3	0	−0.6	0	0	0	0	0	0	0	0	0	0	0	0
CO_2	0	0	0	3	0	0	0	0	0	0	0	1	0	0
H_2O	0	2	0	0	0	0	0	0	0	0	0	0	0	0
ADP	0	−3	0	0	0	0	0	1	−4	1	0	0	0	0
ATP	0	3	0	0	0	0	0	−1	4	−1	0	0	0	0
NAD^+	−1	1	0	−4	0	0	0	0	−2	0	0	0	0	0
NADH	1	−1	0	4	0	0	0	0	2	0	0	0	0	0
NADPH	0	0	−1	0	0	0	0	0	0	0	1	1	0	0
$NADP^+$	0	0	1	0	0	0	0	0	0	0	−1	−1	0	0
Erythrose-4-P	0	0	0	0	0	1	−1	0	0	0	0	0	0	0
Fructose-6-P	0	0	0	0	0	1	1	−1	0	0	−1	0	0	0
6-P-Gluconate	0	0	0	0	0	0	0	0	0	0	1	−1	0	0
Glyceraldehyde-3-P	0	0	0	0	1	−1	1	2	−1	0	0	0	0	0
Ribose-5-P	0	0	0	0	−1	0	0	0	0	0	0	0	1	0
Ribulose-5-P	0	0	0	0	0	0	0	0	0	0	0	1	−1	−1
Sedoheptulose-7-P	0	0	0	0	1	−1	0	0	0	0	0	0	0	0
Xylulose	1	0	0	0	0	0	0	0	0	−1	0	0	0	0
Xylulose-5-P	0	0	0	0	−1	0	−1	0	0	1	0	0	0	1
Pyruvate	0	0	0	−1	0	0	0	0	1	0	0	0	0	0

For instance, Aranda et al. (2010) proposed that minimizing growth would maximize xylitol production as xylose is split into these mean products. So the following objective function was proposed,

$$Min\ \mu = \sum_{i=1}^{n} \eta_i q_{M,i} - \sum_{i=1}^{n} q_I \tag{62}$$

subject to:

$$\mathbf{Y}_6 \cdot \mathbf{q}_M^c = \mathbf{0}$$
$$-q_{M,1} - q_{M,2} = q_S$$
$$-q_{M,3} - q_{M,1} = q_P$$
$$q_{M,1}, q_{M,2}, q_{M,3}, q_{M,4}, q_{M,10}, q_{M,12} > 10$$

The first constraint implies the pseudo-steady state assumption; while the 2nd and 3rd constraints are for the xylitol assimilation and production whilst the latter represents irreversibility constraints. The metabolic set is considered in Table 2, and the metabolic network is shown in Figure 6.

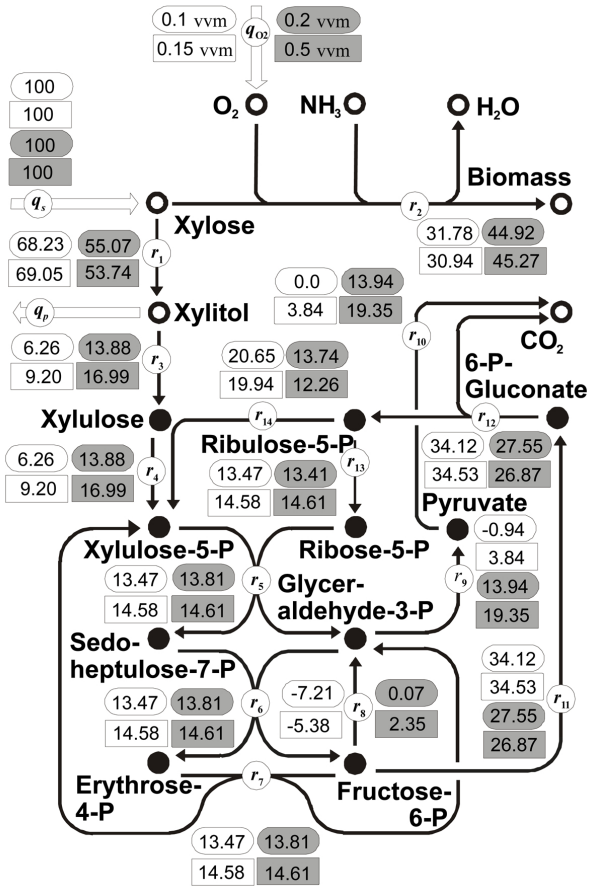

Figure 6. Metabolic fluxes r in xylose to xylitol bioconversion at four different aeration rates (○ 0.1, □ 0.15, ● 0.2, ■ 0.5 vvm) expressed as a percentage of xylose intake flow q_s. The metabolic fluxes of the reactions for the replenishing of NAD+ and NADPH are also shown. Filled circles symbolize metabolic compounds in pseudo-steady state.

^{13}C FLUX ANALYSIS

When reaction fluxes are estimated by means of measured external conversion rates, they should only be considered as a feasible solution for there is no certainty that fluxes are actually as calculated. Measuring some of the internal fluxes would confirm the calculated fluxes if coincidental, or it would increase the chances of having the actual fluxes as the information increases. Unfortunately, it is still quite difficult to measure internal fluxes. Most common methods involves the use of at least one radioactively labeled substrate at a very specific position, for instance the first carbon, and the subsequent analysis of labeled metabolites distribution once steady-state is established (Antoniewicz et al., 2007; Shastri and Morgan, 2007). As the cell metabolizes the labeled substrate, the marked atoms end at particular positions and the proportions of this distribution reveals the rates of reactions.

The isotopic labeling technique works well in heterotrophic metabolism but does not work with autotrophic organisms. In this case, as only one-carbon molecule is used as substrate, at the end all carbons in the cell are labeled so there is no way of knowing the fluxes. A transient analysis must be carried out, the mathematical complexity and the computational methods involved are much more complicated. Young et al. (2008) have proposed a simplified method that significantly reduces the computational requirements.

CONCLUDING REMARKS

This work summarizes some examples where mathematical models successfully represent different biological process. Microbial growth modelling allows us to enhance the processes and predict some changes that can drive a better performance. The development of mathematical models to describe the process and improve the understanding of the metabolism is still limited but, since this is a powerful tool for the design of optimized engineering processes, every effort is valuable.

REFERENCES

Antoniewicz MR, Kelleher JK, Stephanopoulos G. (2007) Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions, *Metabolic Engineering*, 9: 68-86.

Aranda Barradas JS, Salgado Manjarrez E (2002). Producción de biomasa de *Saccharomyces cerevisiae* y usos en el sector alimentario, *Tecnología de Alimentos*, 37: 7-15.

Aranda-Barradas J, Delia ML, Riba JP. (2000). Kinetic study and modelling of the xylitol production using *Candida parapsilosis* in oxygen-limited culture conditions, *Bioprocess Engineering*, 22: 219-225.

Aranda-Barradas JS, Garibay-Orijel C, Badillo-Corona JA, Salgado-Manjarrez E. (2010). A stoichiometric analysis of biological xylitol production, *Biochemical Engineering Journal,* 50: 1-9.

Barrera-Martínez I, González García RA, Salgado-Manjarrez E, Aranda-Barradas JS. (2011). A simple metabolic flux balance analysis of biomass and bioethanol production in *Saccharomyces cerevisiae* fed-batch cultures, *Biotechnology and Bioprocess Engineering*, 16: 13-22.

Delia ML, Aranda-Barradas J, Riba JP. (1997). Production de xylitol par *Candida parapsilosis* : Métabolisme et oxygène, *Bulletin de la Société Française de Microbiologie*, 147-156.

Martínez-Trujillo A, Aranda JS, Aguilar-Osorio G. (2008). Kinetic study on inducibility of polygalacturonases from *Aspergillus flavipes* FP-500, *Electronic Journal of Biotechnology*, 11: 1-8.

Martínez-Trujillo A, Aranda JS, Gómez-Sánchez C, Trejo-Aguilar B, Aguilar-Osorio G. (2009). Constitutive and inducible pectinolytic enzymes from *Aspergillus favipes* FP-500 and their modulation by pH and carbon source, *Brazilian Journal of Microbiology*, 40: 1-8.

Nielsen J and Villadsen J. (1994). Analysis of reaction rates, in Bioreaction Engineering Principles. Plenum Press, U.S.A., 97-162.

Noorman HJ, Heijnen JJ, Luyben, ChAM. (1993). Linear relations in microbial reaction systems: A general overview of their origin, form and use. *Biotechnology and Bioengineering*, 38: 603-618.

Rivera-Salvador V, Aranda-Barradas JS, Espinosa-Solares T, Robles-Martínez F, Toledo JU. (2009). El modelo de digestión anaeróbica IWA-ADM1: Una revisión de su evolución, *Ingeniería Agrícola y Biosistemas*, 1: 109-117.

Shastri AA and Morgan JA. (2007). A transient isotopic labeling methodology for 13C metabolic flux analysis of photoautotrophic microorganisms, *Phytochemistry,* 68: 2302–2312.

Vallino JJ and Stephanopoulos G. (1993). Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction, *Biotechnology and Bioengineering*, 41: 633-646.

Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G. (2008). An Elementary Metabolite Unit (EMU) Based Method of Isotopically Nonstationary Flux Analysis, *Biotechnology and Bioengineering*, 99:686-699.