

Short Review

Modeling cell behavior: moving beyond intuition

Mario Jolicoeur

Canada Research Chair in Applied Metabolic Engineering, Department of Chemical Engineering, École Polytechnique de Montréal, P.O. box 6079, Centre-ville Station, Montréal, Québec, H3C 3A7, Canada; Email: mario.jolicoeur@polymtl.ca; Tel: +1-514-340-4711; Fax: +1-514-340-4159.

Abstract: In the context of the launching of this new journal, we propose a forum to the community of researchers interested and involved in, or even simply questioning the why, what, how, and when of modeling cell or cell culture behavior. To start the discussion, we review some of the usual questions we are routinely asked on the pertinence of modeling cell behavior, and on who might benefit from conducting such work. To draw a global portrait, throughout this text we refer the reader to handbooks introducing the basics of modeling a biosystem, as well as to selected works that can help visualize the broad fields of applications.

Keywords: Modeling cell behavior; Mathematical model; Cell metabolism

1. Why Modeling Cell Behavior?

Each semester, I am systematically asked this question in my graduate and undergraduate courses. Since the major improvements in industrial biotech have come from cell engineering work, why should we spend time and effort developing complex mathematical models? The answer is obviously not straightforward or unique, and may differ depending on your respective research or industrial fields of interest. In fact, there are various examples to illustrate when the modeling of cell behavior is thought to provide a useful tool to address complex multifactorial problems. For instance, the Chinese Hamster Ovary (CHO) production cell platform is a very interesting case of study. This cell platform is now a key player on the industrial scene for the production of most biotherapeutics such as monoclonal antibodies (mAb) and other proteins [1,2]. Like microbial platforms, the most significant improvements in animal cell platforms' productivity are generally still obtained from

genetic engineering work. One or a set of genes that are directly involved in the biochemical reactions sequence leading to the compound of interest are selected, usually intuitively, and their expression levels are altered aiming at their overexpression or silencing [2]. However, the results are often the opposite of what was expected, and it is common to perform several steps of successive screenings for cell engineering targets as well as for parental and recombinant cell lines. In fact, intuition can be poor counsel for this type of work since the cellular metabolism is highly interlinked and regulated, making the result of a genetic or environmental intervention, from the modification of medium composition or of culture conditions, usually unpredictable. In order to be fully expressed, the improved capacity of recombinant cell lines then requires the design of specific culture protocols, implying the optimization of culture medium composition and management, with the addition of specific compounds during a culture (e.g. under fed-batch or perfusion culture modes). Moreover, successful cell engineering work, combined with the engineering of culture media and of fed-batch (or perfusion) culture strategies, has been shown to lead to increased accumulation levels in mAbs from few mg/L (~1980) to over ~10 g/L the last few decades [1,2].

Overall, the development of a highly productive bioprocess will thus rely on the cumulative success of a series of crucial steps, from the identification of a parental cell line of high productive potential and the identification of efficient target(s) of cell engineering, to the design of optimal culture media and culture management protocols enabling the maximization and maintenance of both cell productivity and viability all along a culture cascade. Furthermore, because each step has its specific set of selection criteria, the whole process is largely based on a multifactorial analysis of experimental data, which has encouraged the industry to develop high-throughput techniques to conduct massive assays (i.e., on a large scale in terms of the amount of treatments) varying experimental parameters at many levels [2]. However, although this approach makes it possible to scan an impressive set of conditions, high titer product accumulation does not guarantee the further scale-up of the process or the product quality (i.e., protein activity) [3]. Therefore, another question that arises is: “How can a complex mathematical model be of some help since high-throughput robotized screening platforms offer a massive assay capacity at a low cost per assay, allowing to screen for optimal culture media, high productive cell lines and target(s) for cell engineering works?” In this context of a multifactorial problem, attempts are now being made to develop predictive models capable of guiding interpretation of the resulting big datasets. Such models are thus expected to lead to major breakthroughs not only in the field of industrial biotechnology, but also in medical science for the study of metabolic diseases where huge multifactorial databases are available.

2. Modeling: A Matter of Objective Function

A cell evolves and interacts within an environment that stimulates or inhibits its behavior via various compounds and physico-chemical conditions (Figure 1), which all usually vary with time. In many biosystems, the use of the so-called “basic” cell culture dataset, which normally includes

biomass cell concentration or optical density, and the concentration of the limiting substrate and the product suffice to describe cell growth behavior, as well as the substrate consumption and product production rates. However, access is now available to various and complementary “omics” data enabling the characterization of almost the entire cell content, biochemical reactions stoichiometry and dynamics. The temptation can thus be great to “jump on” these big datasets, but is it always necessary? A decision map (Figure 2) can be drawn to help select a modeling approach depending on your specific objective function. The graduated levels, inspired from Bloom’s learning taxonomy, show that knowledge acquisition and mathematical development can be significant, but the model complexity you will need should always depend on your specific objective function.

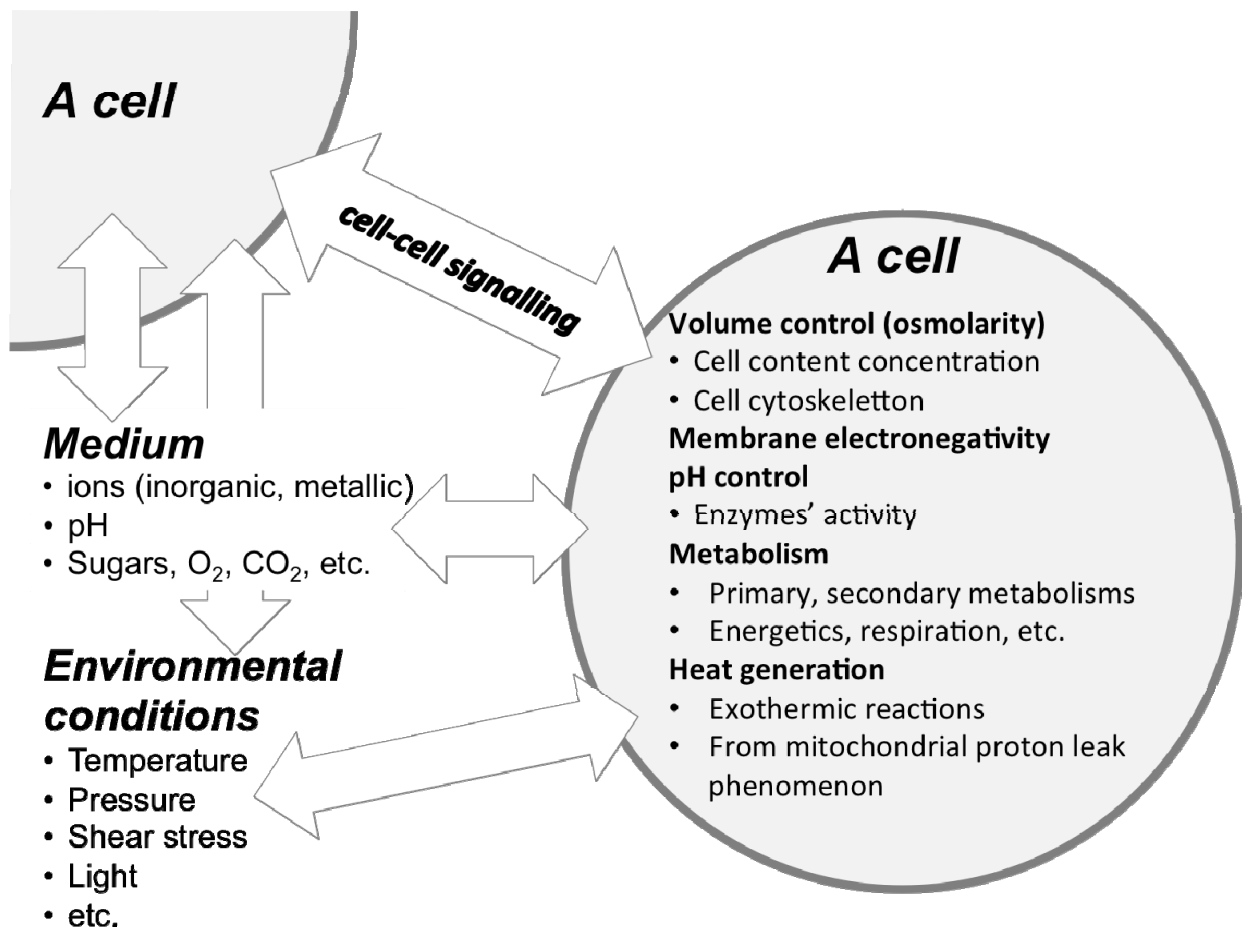


Figure 1. A cell environment.

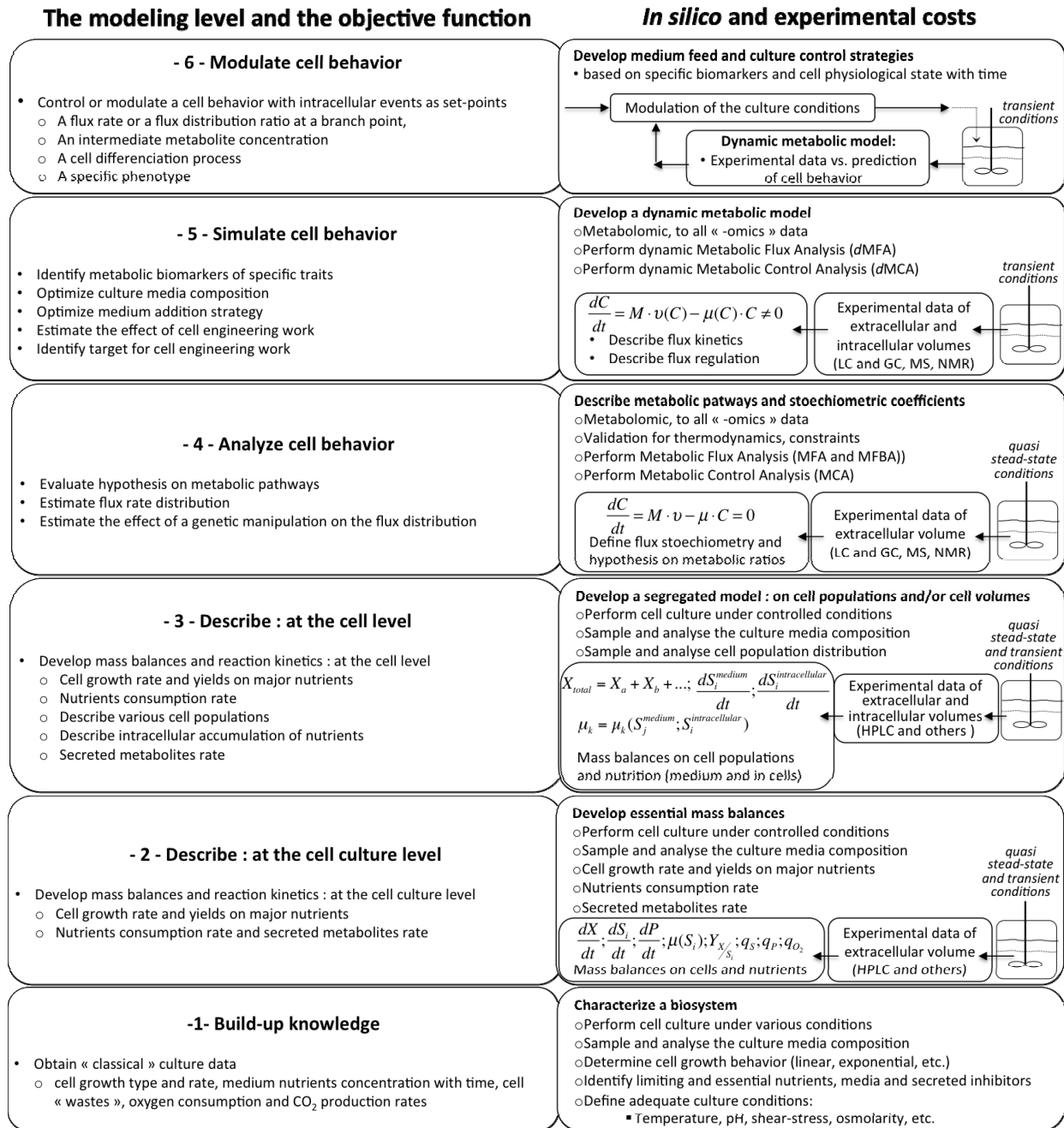


Figure 2. Modeling levels of cell behaviour.

If you are simply aiming at characterizing the effect of an environment on a cell's macroscopic behavior, you may want to start at modeling level 1, limiting your observations to traits such as cell viability, growth type and which compounds are consumed and released by the cells. Otherwise, if your goal is to characterize the interaction of cells with their environment, then you should be at level 2 and should sample and/or monitor the biosystem and acquire the experimental data necessary to compute parameters such as cell growth (or division) rate, cell yields on the major nutrients, nutrients consumption, and metabolites secretion rates (often called as wastes). There is abundant

literature available for those interested in this approach and I suggest reading the early chapters of classic handbooks such as Pirt (1968) [4] and Bailey and Ollis (1986) [5]. With these data, you can develop a mathematical system made of mass balances describing the time variations of the various culture parameters for which you have experimental data. These mass balances are essentially differential equations accounting for the consumption of nutrients, the production of the compound of interest and cell growth. Since an important issue at that level is to describe reaction kinetics, the Monod (1959) [6] equation is still pertinent and useful, along with others that have since been developed [5].

At level 2 you basically obtain a simulator of macroscopic cell behavior, which requires you to know only about initial culture conditions at time “0.” The numerical integration of the system of differential equations using commercial software will do the rest, drawing a projection in time of the state of a culture. However, although level 2 is highly sufficient for many biosystems, in the case of cells accumulating compounds (e.g. plant cells [7, 8]), or those composed of a set of sub-populations that need to be described, you may need to move on to level 3. The model is more complex at that level, but that is the cost of being able to describe complex cell populations.

If you systematically experience discrepancies between model simulations and your datasets, you may need to move further in describing a cell behavior by looking closely at the reaction level inside the cells. You are then at level 4 and focusing on cell metabolic organization. It is then crucial that the experimental design limits physiological stresses from the cell environment (Figure 1), or includes their description in the model. Knowing the stoichiometry of the enzymatic reactions you are describing, you can build up a set of mass balance equations, one per enzymatic reaction, limiting your study to steady state or quasi steady state biosystems. This approach allows for the performance of the Metabolic Flux Analysis (MFA) [9] of a biosystem, linking compounds in the cell environment to their intracellular processing through the metabolic network of enzymatic reactions. Adding adequate hypotheses or “constraints,” you can draw the space of solutions for your biosystem and thus evaluate the “potential” of the cells under study using the Metabolic Flux Balance Analysis (MFBA) [10].

When reversible biochemical reactions need to be analyzed, various labeled substrates (e.g., ^{13}C , ^{15}N , ^{31}O , etc.) can be fed to the cells and the effective flux determined [9]. If you can perform specific cell engineering work modifying an enzyme activity level, you can then perform what is called a Metabolic Control Analysis (MCA) [11], and thus evaluate the metabolic flexibility or rigidity of the cells under study. At that level, you can estimate the flux distribution within a defined metabolic network. Thus, knowing the stoichiometry of the metabolic reactions, you know where carbon flows and at what rates. This is very useful since it can tell you about potential bottlenecks limiting or diverging the carbon flow from the biosynthetic pathways leading to your product of interest or towards any of your objective functions. Combined with MCA, you can then locate potential targets of cell engineering works enabling the carbon to be redirected towards your objective function (e.g., cell growth rate, production or other). However, if you are looking for a

predictive capacity of cell behavior, MFBA appears to offer some potential [12], although you may need to move on to the next modeling level.

By predictive capacity you may understand “in time”, such as in transient processes or where perturbations occur, both cases that are normally encountered in many biosystems. You are then at level 5 where the complexity of this approach may become significant but the outcomes are worth it in many cases. The complexity of the approach involves not only mathematical work, but also, and primarily, obtaining the required experimental data. You need to quantify the cell intracellular content in metabolites and intermediate metabolites, data obtained using cell extraction protocols that must be highly efficient at quenching the cell metabolism in order to stop enzymatic reactions. The samples are then analyzed using sophisticated and expensive equipment with various detectors such as mass spectrometry [13,14] and nuclear magnetic resonance (NMR) [14,15]. With this approach, because you have experimental datasets as well as culture time, you can address the description of transient behaviors of metabolic events, although there is a high labor cost involved in describing the biochemical reaction kinetics, implying a set of kinetic parameters to be determined for each enzymatic reaction. In many cases, such parameters cannot be found in the literature or in databanks (e.g., BRENDA, *brenda-enzymes.info* [16] and references therein). And if you do find any, they may have been determined *in vitro*, meaning they may not correspond to the intracellular conditions of your biosystem.

The parameters that are not described in the literature and databanks have to be determined from estimations, while minimizing model simulation error compared to experimental data. Although the process may be quite laborious when the parameter number to be estimated is high, parameter values can still be identified following standard statistical analysis of the confidence intervals. Considering that such a model consists in a simplification of the biosystem’s metabolism, it is understood that the estimated kinetic parameter values can hardly be those of the enzymes in action; the more detailed a model described, the closer the parameter values to the real conditions.

In order to avoid or limit this problem, hybrid approaches have been proposed accounting for a limited set of transient reactions while keeping others at steady state [17,18]. However, these approaches show limited predictive capacity, requiring periodic re-scaling steps on freshly obtained experimental data. To be qualified as predictive, a model should be able to simulate cell behavior in time from defining initial conditions, including that of the cells. Such an *in silico* platform can be invaluable to perform complex tasks such as evaluating various medium compositions and medium feeding strategies. A dynamic MFC (dMFC) study can then be conducted to assess “optimal” culture strategies. For instance, the objective function can then consist in seeking to maximize carbon flow along pathways leading to a product synthesis, or for any other metabolic event. This process of simulating how each flux rate or their ratios evolve with time can also lead to the identification of specific biomarkers of a cell’s physiological state, which can then be translated into a defined combination of culture parameters that can be monitored to probe specific biosystem’s metabolic states (e.g., growth and/or production). Even from a limited set of experimental data, metabolic

changes occurring with time and culture perturbations can thus be studied [19]. Accordingly, these studies can be extremely useful for improving our understanding of the cell's intracellular dynamics against changes and perturbations in its environment [20,21].

However, the key to obtaining a predictive model may lie in the description of the various metabolic pathway regulation phenomena, which enable cell survival within a changing environment. At such a high level of modeling, dynamic MCA (dMCA) analysis can start to be performed, testing hypothesis of targets for cell engineering work. Some attempts to use such data-driven models have also been conducted to assist in defining real-time fed-batch culture strategies to control CHO [17] and plant cells' metabolic state [22,23], to study the effect of a genetic modification on potato cell behavior [24], and to study the role of energetic robustness in mice model of Parkinson's disease [25]. However, these studies were limited within known experimental spaces. As expected, the closer a model describes a cell's machinery, the better its predictive capacity, and at that complexity level the outputs (i.e., simulations) of a model should still be taken as hypothetical predictions.

If your objective function is to be able to modulate cell metabolism making “live” real-time decisions on the management of a cell environment, you then need to climb to level 6. Although at a lower level (i.e., from level 4), you can increase the model's predictive capacity by including some knowledge on proteomics, interactomics, expressomics and genomics, at level 6 this step is mandatory. It may also be required to account for physiological aspects such as those described in Figure 1. The aim here is to provide a model predictive capacity that can go beyond a known experimental space. Such a modeling level is thus expected to allow for the identification of optimal cell metabolism and cell culture management, including exploring hypotheses of therapies and biomedical approaches such as regenerative medicine.

3. Modeling Approaches vs. Biosystems

How much should modeling approaches differ among biosystems? This is an important question that arises when initiating modeling work on a cell type with which we are unfamiliar. The normal reflex is to consider that they are all specific enough to require a distinct modeling approach. Cell genetics obviously differs among species, as do the resulting metabolomics, with different intermediate metabolites and enzymes concentration and diversity, as well as the rate and regulation of metabolic fluxes. However, the cell metabolism of all cell species (to the best of our knowledge) obeys to similar kinetic laws with many pathways that have been preserved along with cell species evolution. In fact, the literature provides studies using a unique modeling approach for diverse cell platforms. However, it has been clear since the Monod model of exponential growth has been efficiently applied to every cell platform—from microbial to animal cells—that all cell types undergoing division macroscopically behave similarly, showing a direct relationship with the limiting nutrient concentration. There are of course numerous cases for which the Monod model does

not adequately apply when based on extracellular substrates. For instance, the plant cell platform, which shows a nutrient accumulation capacity, requires a model that accounts for a limiting substrate within the intracellular volume [8]. As previously mentioned, MFA, MFBA and MCA approaches, as well as their dynamic forms, have been applied to bacteria, yeasts, fungi, plant and animal cells. Therefore, after being “frame fitted” to cell species specificities, a model can be built integrating known enzyme kinetic formulations, thus accounting for the cell line specific dynamics.

4. Steady State vs. Transient: A Matter of Set-points and Regulation

As Figure 1 shows, a cell is a social entity continuously interacting with its environment to survive. This interaction, driven through the actions/reactions of the cell’s metabolic network, aims at maintaining intracellular conditions favoring the life continuum we call "viability". During its living state, a cell has thus to ensure the controlled uptake and secretion of specific compounds to enable various functions such as the regulation of intracellular osmolality and pH (which can both vary among organelles and cytosolic volumes), the uptake of precursors and co-factors to feed the pathways responsible for the biosynthesis of metabolic intermediates and energy production, and also deal with crucial signaling events. These cell "objective functions", along with others that are not mentioned here, are logically interlinked and can thus be highly challenged by the various stresses occurring along a cell culture, such as external variations of physical parameters (temperature, pH, light, osmolality, etc.), nutrients concentration or even mechanical stresses, and so on. The mechanisms of action (or reaction) a cell has to survive in a changing environment may differ among microbial, yeast, plant and mammalian cell platforms, but the literature suggests they possess a common set of tools [26].

With this global picture in mind, it becomes obvious that describing a cell population at steady state is a specific case study that provides only a partial view of a biosystem. In fact, the phenomena influencing or involved in cell behavior occur under a wide range of characteristic times (Figure 3). It is thus expected that the phenomena with a fast dynamic, which are tedious to observe because of technical limitations in sampling and monitoring, may trigger a change in the long-term cell behavior. Moreover, it is also common to perform studies under the quasi steady state hypothesis, and if the latter is mathematically verified comparing the characteristic times involved, there can be imperceptible underlying physiological changes inducing a strong bias in the conclusions that are drawn.

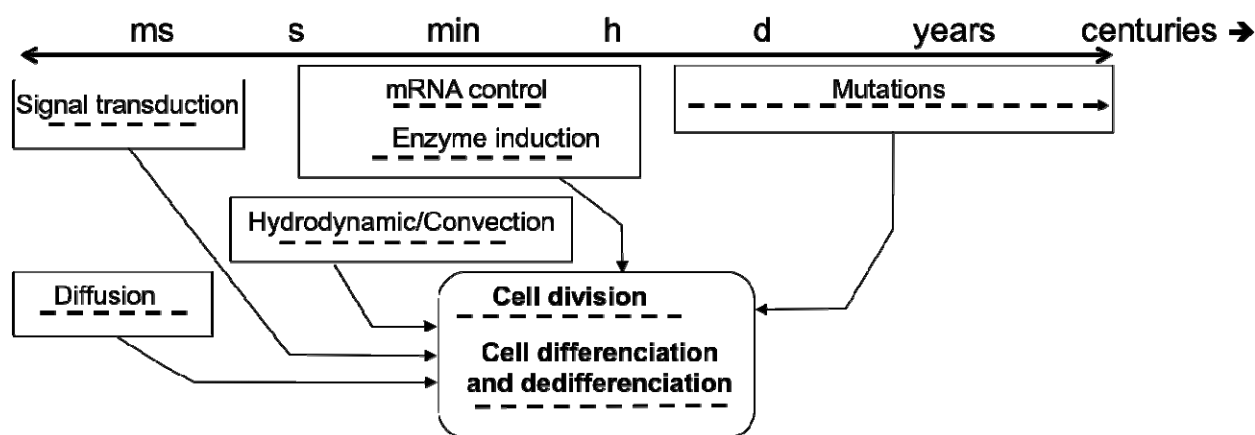


Figure 3. Time scales affecting cell behaviour.

Therefore, although it is extremely important to identify the steady states a cell is able to maintain when facing varying environmental conditions, it may be useful to characterize how a cell maintains each steady state and how the biosystem evolves within multi steady states. Moreover, it should not be forgotten that at certain steps of any bioprocess (or simply cell culture) the cells will experience transient conditions. In each bioprocess, the early stage of cell culture build-up before the production bioreactor is usually performed in a cascade of bioreactors run in batch mode, which is mostly transient. In the biomedical field, the culture of tissues or cells in regenerative medicine normally involves batch culture mode, and the cells go through highly stressful transient processes when grown onto biomaterials. In a context of producing a biotherapeutic compound, this behavior change may result in a loss in cell productivity. In a biomedical (or medical) context, it may trigger cell differentiation from a desired to a pathological phenotype. MFA and MFBA approaches have proved to be useful for estimating flux distribution in a network at steady and quasi steady state conditions, but dynamic studies (dMFA and dMFBA) are thought to allow for the evaluation of the various impacts of potential dynamic events, as well as enabling MCA studies on a biosystem, characterizing cell behavior under specific perturbations. Therefore, following a system's biology approach, basic knowledge on cell intracellular dynamics and its potential effects on cell metabolism should be integrated. This knowledge, together with the description of the mechanisms involved, will be more than helpful in the development of a descriptive modeling framework that has a predictive capacity.

5. What Should We Expect From A Mathematical Model of Cell Behavior?

The usefulness of modeling cell behavior may now be clearer, and we hope that the resulting mathematical model derived from such work will be considered as an invaluable tool to identify innovative solutions in various cases. Such *in silico* platforms can offer an innovative tool not only

for the optimization of bioprocesses, but also in medicine to simulate hypotheses on therapeutics and therapy strategies, as well as in biomedical engineering, for instance, to better understand the problem of losing a desired cell phenotype when using multipotent cells. Although these contexts differ in their research paradigm, the modeling tools developed in each case may be much more similar than might be thought.

Thus, overall, what can we expect from a metabolic model? First, it can definitely contribute to the characterization of a biosystem behavior. However, to be considered as an *in silico* platform, a model has to demonstrate a predictive capacity. As shown in Figure 2, obtaining a predictive capacity is an ultimate goal, which is expected to be costly in terms of time and effort, as well as financially, taking analytics into account. Nonetheless, such an *in silico* platform can then help save time and effort and thus accelerate a research project's success in meeting its objective function(s), such as optimizing a medium composition or fed-batch (or perfusion) culture strategy (i.e. bioreactor control) or identifying "optimal" parental cell lines or robust high-producer clones. As mentioned earlier however the use of a model to identify targets of cell engineering, to test hypotheses on the effect of a therapeutic or a therapy strategy, and to simulate cell differentiation processes may require the contribution of full "omics" technologies.

In fine, it is thought that our capacity to describe the various cell dynamics that affect cell behavior will enable a model's predictive capacity. However, the question of how important it is to remain close to a biologically relevant description of cell mechanisms while developing mathematical models still remains. The more "biological" a model, the more tedious it could be to statistically validate it with full parameter identification, but this may be the price to pay for a model that could demonstrate a predictive capacity enabling us to explore outside the comfort zone of known experimental conditions, and thus move beyond intuition.

References

1. Pavlou AK, Belsey MJ (2008) The therapeutic antibodies market to 2008. *Eur J Pharm Biopharm* 59: 389-396.
2. Wurm F (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* 22: 1393-1398.
3. Li F, Vijayasankaran N, Shen A, et al. (2010) Cell culture processes for monoclonal antibody production. *mAbs J* 2: 466-479.
4. Pirt SJ (1975) Principles of microbe and cell cultivation. Blackwell Scientific Publications.
5. Bailey JE, Ollis DF (1986) *Biochemical Engineering Fundamentals*, 2 Eds., New York: McGraw-Hill, 984.
6. Monod J (1949) The Growth of Bacterial Cultures. *Ann Rev Microbiol* 3: 371.
7. Lamboursain L, Jolicoeur M (2005) Critical influence of *Eschscholtzia californica* cells nutritional state on secondary metabolite production. *Biotechnol Bioeng* 91: 827-837.

8. Cloutier M, Bouchard-Marchand É, Perrier M, et al. (2007) A predictive nutritional model for plant cells and hairy roots. *Biotechnol Bioeng* 99: 189-200.
9. Stephanopoulos G, Aristidou, Nielsen J (1998) Metabolic Engineering. *Principles and Methodologies*, San Diego : Academic Press, 698.
10. Palsson BO (2011) Systems Biology. Simulation of dynamic network states, Cambridge UK: Cambridge University Press, 317.
11. Fell D (1997) Understanding the Control of Metabolism. In *Frontiers in metabolism*. London UK: Portland Press, 301.
12. Bordbar A, Monk JM, King ZA, et al. (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Gen* 15: 107-120.
13. Chong WPK, Thng SH, Hiu AP, et al. (2012) LC-MS-Based Metabolic Characterization of High Monoclonal Antibody-Producing Chinese Hamster Ovary Cells. *Biotechnol Bioeng* 109: 3103-3111.
14. Dunn WB, Broadhurst DI, Atherton HJ, et al. (2011) Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* 40: 387-426.
15. Ben-Tchavtchavadze M, Perrier M, Jolicoeur M (2010) A Non-Invasive Technique for the Measurement of the Energetic State of Free-Suspension Mammalian Cells. *Biotechnol Prog* 26: 532-541.
16. Scheer M, Grote A, Chang A, et al. (2011) BRENDA, the enzyme information system. *Nucleic Acids Res* 39 (Database issue <http://www.brenda-enzymes.org/>): 670-676.
17. Nolan RP, Lee K (2011) Dynamic model of CHO cell metabolism. *Metab Eng* 13: 108-124.
18. Leduc M, Tikhomiroff C, Cloutier M, et al. (2006) Development of a kinetic metabolic model: application to *Catharanthus roseus* hairy root. *Bioprocess Biosyst Eng* 28: 295-313.
19. Ghorbaniaghdam A, Henry O, Jolicoeur M (2012) A kinetic-metabolic model based on cell energetic state: study of CHO cell behavior under Na-butyrate stimulation, *Bioprocess Biosyst Eng* 36: 469-487.
20. Ghorbaniaghdam A, Chen J, Henry O, et al. (2014) Analyzing clonal variation of monoclonal antibody producing CHO cell lines using an *in silico* metabolomic platform. *PLoS One* 9: 90832.
21. Ghorbaniaghdam A, Chen J, Henry O, et al. (2014) An *in silico* study of the regulation of CHO cells glycolysis. *J Theor Biol* In press.
22. Cloutier M, Chen J, Tagte F, et al. (2009) Kinetic metabolic modelling for the control of plant cells cytoplasmic phosphate. *J Theor Biol* 259: 118-131.
23. Cloutier M, Chen J, DeDobbeleer C, et al. (2009) A systems approach to plant bioprocess optimization. *Plant Biotechnol J* 7: 939-951.

24. Valancin A, Srinivasan B, Rivoal J, et al. (2013) Analyzing the effect of decreasing cytosolic triosephosphate isomerase on *Solanum tuberosum* hairy root cells using a kinetic-metabolic model. *Biotechnol Bioeng* 110: 924-935.
25. Poliquin PO, Chen J, Cloutier M, et al. (2013) Metabolomics and in-silico analysis reveal critical energy deregulations in animal models of Parkinson's disease. *PLoS One* 8: 69146
26. Michal G, Schomburg D (2012) *Biochemical Pathways: An atlas of biochemistry and molecular biology*, 2 Eds., Wiley, 398.

© 2014, Mario Jolicoeur, licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)