Population Balance Model for Cellular Processes in Biological Systems: Biochemical and Biomedical Applications

Charles David Immanuel
Dept. of Chemical Engineering and Chemical Technology, Imperial College, London SW7 2AZ, UK
c.immanuel@imperial.ac.uk; Phone: +44 (0)20 7594 5594; Fax: +44 (0)20 7594 6606

Abstract

In this paper, the generic problem of the development of a population balance model for those biological systems involving cellular processes is discussed. Such models are of interest for bioreactor optimisation and control, as well as for different biomedical problems including the study of cancer. The generic model and its features are presented, serving partly to consolidate the studies that have been reported in the literature already. The major challenges and issues with this problem are highlighted. One of the challenges is with respect to the solution technique for these complex models, which issue is addressed here. The algorithm proposed for solving these complex models is based on a hierarchical two-tier solution strategy that has been proposed previously (Immanuel and Doyle, III, Chem. Eng. Sci., 2003). The two-tier strategy involves the calculation of the rates of sub-processes that contribute to the population dynamics in the first tier, with the population itself being updated in the second tier. This in turn enables performing a major portion of the calculations off-line (once at the start of the simulation), thereby achieving a major reduction in the computational load. The algorithm is implemented on a six-dimensional population balance bioreactor model.

Introduction

The interest in the development of detailed mathematical models for biological processes is undergoing a tremendous boost in recent years among the Chemical Engineers. This is because of the various purposes that a mathematical model of biological processes can serve [16]. The primary purpose is the growing interest in a biochemical route to the manufacture of commodity chemicals, pharmaceuticals, food products, and agricultural products. Apart from the bioreactors, the models will be very versatile tools for the analysis of biomedical problems and in therapeutical studies. Very detailed studies on the various individual aspects of biological systems is carried out by the biologists, biochemists, and the researchers in the medical fields. These individual aspects need to be consolidated to form a comprehensive picture, to study the effect of these individual mechanisms on the overall process, and hence to bring these individual studies to further fruition. A comprehensive model will provide a framework to do this, and will in most cases be the only tool available to explain the various intricate experimental observations. The mathematical model can aid in the study of the underlying processes themselves, by suggesting appropriate and effective experimentation. The net result is a symbiotic improvement in the process understanding and model development. The resultant model will be in a form suitable for the different end applications, be it to perform design, optimisation and control of bioreactors or for therapeutical studies, as the case may be.

The focus of this study is on biological systems that involve cellular processes. The most common and simple methodology for the mathematical modelling of such systems is based on the so-called unstructured models, which do not distinguish the individual cells in the aqueous medium. However, the complexity of the bioprocesses warrents the introduction of structure into the model, to obtain a more rigorous representation of the processes. *A more rigorous modelling approach will account for the heterogenous nature of the reaction medium, the non-uniformity of the cell populations with respect to their mass and internal metabolites, the cell cycles (the processes of growth, division, death, and in certain cases even recombination), and the interaction of the cell cycle with the environment. This can be naturally done employing the population balance framework.*

Although the advantages of pursuing a population balance approach for bioreactor modelling has been recognised since the 1960's by Fredrickson and Tsuchiya [5], the actual application of these models in reactor optimisation and control has not been pursued until much recently [17, 10]. This is partly attributed to a lack

of sufficient depth of understanding of the underlying mechanisms. The complexity of the metabolic pathway, and hence of the mathematical model, has also been a major deterent in this problem. Although these deterents have been sufficiently overcome in recent years, the control studies reported are based on total-mass-structured cell population balance models, with a single internal coordinate. However, the complexity of the process, with the large number of metabolites and reactions involved, warrants and even necessitates the introduction of further structure into the models. The cell cycle phases, and the growth and division rates, are dependent not only on the total mass of the cells, but more so on the mass of the different metabolites within the cells. Thus, there is a need to develop higher dimensional population balance models, by accounting for the distribution of different properties (and not just the total cell mass) among the cell populations.

On a different level of time scales and to a certain extent the model complexity, the modelling of cell cycles assumes importance for characterising the physiological conditions and understanding several medical problems. For instance, the cells that constitute the various organs in the human body are involved in the processes of growth, division and death, and their interaction with the surroundings and the other cells. Alterations to the cell cycles, due to perturbations either internally within the cells or externally in the surrounding, lead to cancerous growth of the cells and to tumours. Even though modelling studies on cancer have appeared in the literature (such as the Marchuk's Model for general immune response), these studies do not account for the cell cycle phases. Instead, they mainly account for the interaction between normal cells, cancer cells and the antigens, resulting in relatively simple (three state) models (similar in structure to the first models on HIV [12]). Since the major disease factor lies in the cell cycle phase, the best model for cancer has to account for the cell cycles. Thus, again, a population balance model becomes the most appropriate modelling tool for cancer. In a similar vein, *population balance models of microbial populations also assume importance for the study of diseases caused by bacterial populations*, which co-exist and interact with each other as well as with bacteriophages. In this case again, unlike in the case of the relatively short-spanned bioreactors, the interactions between mutation and recombination (crossover) will assume a critical factor.

Measurement of the cell populations and their distribution will be a critical factor in model development, to ensure that the right level of detail is present in the model to provide it with the predictive capabilities. *Flow cytometry is a boon for the development of cell population balance models.* Although it stems from the age-old concept of gram-staining of the cells, it has been considerably developed over the years to be amenable to current requirements [15]. Flow cytometers can also be adapted for on-line measurement of cell populations, with sample times of 18 minutes being reported [1]. These developments are very positive aspects for model development. The instrument can be used to monitor the distribution of select individual metabilites, or to monitor the total mass distribution of the population.

In this paper is presented a general, yet realistic formulation of the population balance-based models for the different biological problems highlighted above. Further, the major challenges in implementing this approach are identified. *Preliminary results that overcome some of these challenges are presented, particularly with respect to the numerical solution of bioreactor models.*

Opportunities

Cell cycle and regulation

As indicated in the introduction, a realistic model of the cellular biological processes has to account for the cell cycles in the heterogeneous medium. The cells in the population undergo a cycle of changes through their life time. The individual cells synthesise metabolites including proteins, RNA and DNA, and thereby grow in size. Upon the duplication of the genetic carriers – the DNAs, the cells undergo cell division into two usually identical daughter cells, contributing new and younger members to the population. The cell cycle is divided into two regimes – the interphase and the mitosis phase (M-phase). The interphase is that regime of the cell cycle that characterises the synthesis of metabolites, cell growth and DNA replication. The M-phase characterises the cell division. In the case of the Eukaryotes, the interphase is further divided into the G1 phase, the S phase, and the G2 phase. The G1 phase accounts for the synthesis of the metabolites required for DNA replication, the S phase accounts for DNA synthesis, the G2 phase accounts for the gap between DNA synthesis and cell division that serves to ensure proper replication of the DNA. There could also be a G0 phase, which represents an additional gap during which the cells are at rest. The cell division (the M phase) is itself divided into several phases (Prophase, Metaphase, Anaphase and Telophase) that represent the processes of separation of the DNA, cytokinesis and formation of the cell membrane dividing the daughter cells. The cells usually cease to divide (senescent state) after a certain number of generations of reproduction. Cell senescence usually sets in after 60-80 generations. The cells also

undergo apoptosis and necrosis (cell death). The duration of the cell cycle depends upon the organism, and in the human body, the cell cycle duration also depends upon the organs that the cells constitute.

The G1 and S phases are characterised by the complex metabolic reaction scheme that occurs within the cells. The amount of information available on the various pathways (glycolysis pathway, TCA cycle pathway *etc.*) far exceeds the amount that can be modelled reasonably with the available computational resources. Thus, in developing models for bioreactors and for biomedical applications, the metabolic pathways that are the most relevant will need to be adopted. Techniques such as the cybernetic modelling concepts [8] can be exploited to account for both the simplification that is introduced into the metabolic pathways, and more importantly to account for the selectivity and regulations that are inherent within the cells.

The passage from one phase to the other is dictated by several regulatory mechanisms (check points), some of which are triggered by external impulses. The most important regulation is the RAS signalling, which regulates growth and proliferation as well as survival of the cell populations. The RAS signalling operates through several pathways such as the RAF pathway and the Mitogen-Activated Protein Kinase (MAPK) pathway, which determine the transition of the cells into the synthesis phase of the cell cycle [3]. The protein p53, a type of cyclin activated kinase inhibitor, is crucial to maintain normal cell cycles by supressing cell proliferation and regulating cell growth. Mutants of p53 and the perturbations that appear in these regulatory mechanisms are the major reasons behind the occurence of cancer and malignancy, as has been borne out by several studies [11, 13]. In addition, the cell senescence is contingent on telomere shortening (telomeres being precursors to replicated DNAs), which is perturbed by the expression of the enzyme telomerase that leads to 'immortality' of cells (return from the senescent state back to a proliferative state) and hence to cancer [11]. Similarly, cell apoptosis is regulated by the transforming growth factor TGF-β [14]. Population balance provides a natural framework to model these various mechanisms.

Model formulation

The general population balance model as applicable to biological systems is given in Equation (1). In this equation, $F^n(\mathbf{x},\mathbf{t})$ accounts for the population density of the cells at generation n (after n divisions), $\mathbf{x}$ accounts for the internal coordinates (the concentration of the metabolites within the cells), and N is the number of internal coordinates considered in the problem. $\frac{dx_i}{dt}$ accounts for the rate of change of the metabolites due to kinetic pathways within the cells. $\Re_{div}^{n-1}(\mathbf{x},\mathbf{t})$ accounts for the formation of cells in generation n by cell division in the previous generation $n-1$. Likewise, $\Re_{div}^{n}(\mathbf{x},\mathbf{t})$ accounts for the depletion of cells in the current generation n due to cell division. $\Re_{div}^{n}(\mathbf{x},\mathbf{t}) = \mathbf{0}$ for n beyond the threshold generation for cell senescence to occur. In bioreactor models, cell senescence is usually not of significance as the residence time of the cells is much shorter. But this assumes importance for cancer modelling. $\Re_{death}(\mathbf{x},\mathbf{t})$ accounts for cell death and $\Re_{cell-cell}(\mathbf{x},\mathbf{t})$ can be employed to account for cell-cell interactions such as the recombination aspects in bacterial populations.

$$\frac{\partial}{\partial t}F^n(\mathbf{x},\mathbf{t}) + \mathbf{\Sigma_{i=1}^{N}}\frac{\partial}{\partial \mathbf{x_i}}(\mathbf{F^n}(\mathbf{x},\mathbf{t})\frac{\mathbf{dx_i}}{\mathbf{dt}}) = \mathbf{\Re_{div}^{n-1}}(\mathbf{x},\mathbf{t}) - \Re_{div}^{n}(\mathbf{x},\mathbf{t}) - \mathbf{\Re_{death}}(\mathbf{x},\mathbf{t}) + \mathbf{\Re_{cell-cell}}(\mathbf{x},\mathbf{t}) \quad (1)$$

Employing the assumption of a deterministic process, the various phases are mutually exclusive in terms of the concentration of the cell metabolites, as clearly explained in a recent article [4]. Based on this argument, one need not introduce further distinction into the population to distinguish among the various phases of the cells. However, one might need different growth kernels ($\frac{dx_i}{dt}$) for the different cell cycle phases. In the above cited paper, in the final phase (M-phase), the age of the cells in this phase is also taken into account in order to account for the variable time delay associated with the processes of cytokinesis and the development of the cell partition before the cell division. In the model presented in equation (1), this distinction is not made to keep the model simple, and to maintain the right level of detail.

On the other hand, one also needs to model the effect of the environment on the cell cycle phases, such as through regulatory mechanisms. As indicated previously, the modelling of the regulation will be critical mainly for medical problems. In modelling the regulatory processes, one might adopt the models for these mechanisms that have been published in the literature. For example, the MAPK pathway has been recently modelled [2]. While a strategy similar to the one proposed in that study can be employed in our population balance formulation, a strategy based on the extension of the cybernetic approach will also be a very viable alternative (both in terms of simplifying the model and in terms of avoiding difficulties in representing the regulatory aspects of the pathway).

Modelling for Bioreactor control

Due to the relatively small life time of the cells within the reactors, cell senescence aspects will not be important. Thus, the model presented above can be simplified

substantially by not distinguish between the different generations of the cells. Also, the cell-cell interaction effects can be ignored. In effect, the PBE reduces to the form shown in Equation (2).

$$\frac{\partial}{\partial t}F(\mathbf{x},\mathbf{t}) + \mathbf{\Sigma_{i=1}^{N}}\frac{\partial}{\partial \mathbf{x_i}}(\mathbf{F}(\mathbf{x},\mathbf{t})\frac{\mathbf{dx_i}}{\mathbf{dt}}) = \mathbf{\Re_{div}}(\mathbf{x},\mathbf{t}) - \Re'_{div}(\mathbf{x},\mathbf{t}) - \mathbf{\Re_{death}}(\mathbf{x},\mathbf{t}) \quad (2)$$

The cell division terms are modelled as in Equations (3) and (4). The extracellular metabolites are modelled as in Equation (5), accounting for the diffusion out of the cells and any deactivation in the aqueous medium.

$$\Re'_{div}(\mathbf{x},\mathbf{t}) = \mathbf{\Gamma}(\mathbf{x})\mathbf{F}(\mathbf{x},\mathbf{t}) \quad (3)$$

$$\Re_{div}(\mathbf{x},\mathbf{t}) = \int_{\mathbf{x_1}}^{\mathbf{x_{1,max}}}\int_{\mathbf{x_2}}^{\mathbf{x_{2,max}}}\int_{\mathbf{x_3}}^{\mathbf{x_{3,max}}}\int_{\mathbf{x_4}}^{\mathbf{x_{4,max}}}\int_{\mathbf{x_5}}^{\mathbf{x_{5,max}}}\int_{x_6}^{x_{6,max}}\Gamma(x)F(x,t)dx_1dx_2dx_3dx_4dx_5dx_6 \quad (4)$$

$$\frac{dx_{ex}}{dt} = \int_{x_1}\int_{x_2}\int_{x_3}\int_{x_4}\int_{x_5}\int_{x_6}\kappa(x-x_{ex})F(x,t)dx_1dx_2dx_3dx_4dx_5dx_6 - k_w x_{ex} \quad (5)$$

Modelling for Cancer Studies

A more detailed population balance model becomes important for this application, in view of the larger 'residence time' of the process. In particular, the model will need a detailed account of the regulatory mechanisms (check points). One may need to account for a larger number of internal metabolites in this case, and moreover, the generation number of the cell populations will need to be monitored.

The model developed will aid in the study of the sensitivity of the different regulatory mechanisms, and hence to identify the most sensitive mechanism that one should target for treatment. For example, the effect that the RAS signalling and the other individual pathway have on the overall process of tumour growth can be analysed using the mathematical model (even if only on a qualitative level). These sensitivity informations can then be used to direct further experimentation and research along the most promising path for therapeutic studies. For therapeutic studies, the antibody dynamics and its interaction with both the normal cells and the cancerous cells will need to be incorporated into the model.

Challenges

The challenges in the development of population balance models for biological systems can be categorised into three areas. The first category pertains to the complexity of the process, and the challenges in identifying an appropriate level of detail for the various applications at hand and a suitable mathematical formulation for each mechanism. The knowledge base with regard to biological processes has reached such a stage that one has to be eclectic in choosing the aspects and details that one needs to incorporate into the model. This issue was illustrated briefly above in selecting the models for bioreactors. In this case, there is much less need to incorporate the regulatory mechanisms. Also, the portion of the intracellular metabolic pathway that is relevant to the process can be focussed upon. On the other hand, for biomedical applications, one needs to have a more comprehensive metabolic pathway, and also the regulatory aspects assume great importance. Extensions of the cybernetic approach can be employed at different stages of the model.

The second category pertains to the identification of the appropriate kernels for the various portions of the model. While the development of the growth kernels might be an easy task, the same cannot be said about the division kernel. Also, the partition functions for non-symmetric divisions might be another challenge to obtain. However, the ability to measure and monitor individual portions of the process through flow cytometry provides great promise for the identification of the appropriate kernels.

The third category pertains to the solution of the models. The need is a generic yet realistic and efficient numerical solution technique to solve these complex models. Such a technique will segregate the effect of solution capabilities on the level of details in the models, and enable attaining the full potential in the model that reflects current process knowledge.

Representative model simulation

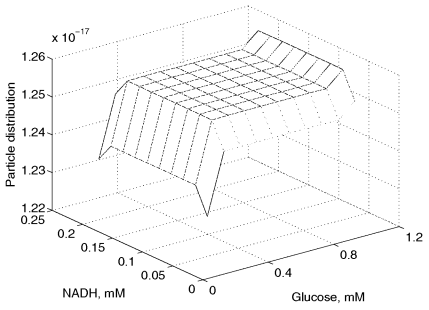
In this paper, a representative model for a bioreactor is presented. The yeast system described in Henson *et al.* [6] is chosen as the model system of study. A six-dimensional population balance model is developed, as recommended there. The six internal coordinates chosen are the intra-cellular concentrations of glucose; glyceraldehyde 6-phosphate; 1,3 biphospho glycerate; pyruvate; NADH; and ATP. The intra-cellular metabolic pathways and the associated reaction kinetics are modelled employing a variation of the Monod kinetics [6]. These kinetics will form the basis for modelling the cell growth phenomenon in Equations (1) and (2). The nutrient glucose is a continuous input to the cells, while pyruvate is a metabolite produced by the cells that is excreted to the aqueous medium (with mass transfer coefficient κ).

Applying traditional solution techniques such as the discretisation methods or the methods of weighted resid-

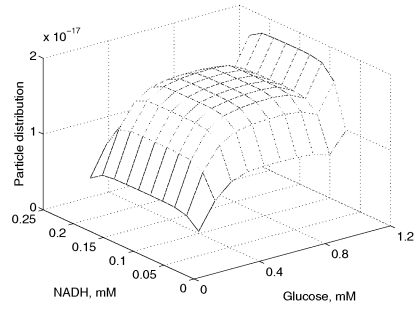
uals results in a large system of equations, even with a modest choice of discretisation parameters. This aspect, combined with the stiffness of the model equations, leads to very large computational requirements. The largest system for which a method of weighted residuals has been applied is a 3-dimensional population balance model [9], with much simplified kinetics. For a six-dimensional system, Henson *et al.* [6] report a constant-N Monte Carlo-like solution method. This particular method is based on the superposition of multiple single cell models rather than on the solution of the population balance equation itself. Although this is a viable starting point, these techniques do not allow a rigorous characterisation of the population dynamics. *In the present study, a realistic, feasible and efficient solution technique is employed for the population balance models.* The hierarchical two-tier technique that was developed for single-dimensional models [7] is extended to this 6-D case. In this technique, the population balance equation is explicitly formulated in terms of the individual phenomena of the nucleation, growth, division and death. As indicated earlier, the continuous growth (or shrinkage) process is modelled using partial differentials, while the discrete processes that include the aggregation (coagulation) and breakage processes are modelled in terms of complex integrals. The evaluation of these integrals require very high computational loads by themselves. In the hierarchical two-tier solution technique, the individual rates of the nucleation, growth and death phenomena are computed in the first tier of the algorithm, and the population is updated in the second tier. To reduce the high computational demand of modelling the aggregation/breakage integrals, an off-line solution strategy is employed. In this strategy, the complex quadratures are solved semi-analytically in terms of the (unknown) population densities *a priori*, thereby substantially reducing the on-line computational load.

For bioreactor problems, cell aggregation is not important. However, one must account for cell division. The derivation of the *off-line semi-analytical solutions for the integral terms representing the cell division phenomenon* is relatively easier than the corresponding aggregation integral. Cell division is assumed to be connected with the concentration of glucose within the cell, with the cells dividing upon reaching a threshold glucose concentration. Also, the division is assumed to be binary, with a uniform split of each of the metabolites.

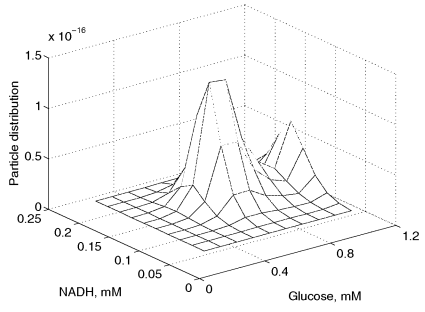
The algorithm is based on a finite-volume discretisation of the distribution domains. The particle population is assumed to be uniform within each of the finite volumes. The growth phenomenon is modelled based



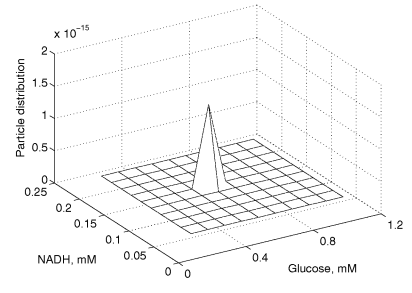
(a) t=6 × 10^{-3} s



(b) t=0.6 s



(c) t=6 s



(d) t=1800 s

Figure 1: Simulated evolution of the cell populations, depicted with respect to two of the internal co-ordinates. The internal coordinates are the intra-cellular concentrations of glucose and nicotinamide adenide dinucleotide hydrogen (NADH). As the time progresses, a synchronisation of the population is observed.

on the metabolic reaction rates at the six-dimensional boundaries of the finite volumes. As explained earlier, the 6-dimensional integrals that characterise cell division are reduced to simpler multiplication and addition of terms, which while being accurate as analytical solutions, substantially reduces the on-line computational load. Space constraints prevent the inclusion of the detailed algorithm and the model equations here, but will be elaborated elsewhere.

Since a rigorous population balance formulation is adopted here, the modelling of the interaction of the cells with the aqueous medium can be done in a straightforward manner, without having to resort to unnecessary assumptions and simplifications that would be necessary in the statistical methods. In the later stages of the project, the *few* macroscopic parameters in the model will be determined by fitting the simulation results to experimental data from flow cytometry. Preliminary simulation results are presented in Figures 1. This indicates the synchronisation of the cell populations, which effect has been observed in other studies [6]. Parametric sensitivity studies also indicate an ability to capture complex dynamics such as the oscillatory and limit cycle behaviour that have been observed experimentally.

Conclusion

A general and realistic framework was presented for the modelling of cellular processes in biological systems, employing the population balance ideas. The model will have ramifications for biochemical studies (design, optimisation and control of bioreactors) as well as for biomedical studies (in particular, to understand the occurence of cancer and to identify the most sensitive regimes of the cellular processes to target for therapeutic investigation). The major challenges that confront model development were highlighted. One of these is the challenge in the solution of these complex models in an efficient manner. It is imperative to have an efficient solution technique, so as to remove the computational burden from being a determining factor for the model features and details. A hierarchical solution strategy is presented, which enables performing a major portion of the computations *a priori*. Despite being based on a rigorous six-dimensional population balance formulation, the technique gives solution times of about an hour for a 1 hour run. This will be adequate for design and optimisation of bioreactors, and also for analysis of medical problems. For on-line control purposes, one can resort to simpler, one-dimensional population balance models or the faster cell ensemble solution technique [6]

References

[1] N. Abu-absi, A. Zamamiri, J. Kacmar, S. J. Balogh, and F. Srienc. Automated Flow Cytometry for Acquisition of Time-Dependent Population Data. *Cytometry Part A*, 51A:87–96, 2003.

[2] A. R. Asthagiri and D. A. Lauffenburger. A Computational Study of Feedback Effects on Signal Dynamics in a Mitogen-Activated Protein Kinase (MAPK) Pathway Model. *Biotechnol. Prog.*, 17:227–239, 2001.

[3] J. Downward. Targetting RAS Signalling Pathways in Cancer Therapy. *Nature Reviews: Cancer*, 3:11–22, 2003.

[4] A. G. Fredrickson. Population Balance Equations for Cell and Microbial Cultures Revisited. *AIChE J.*, 49:1050–1059, 2003.

[5] A. G. Fredrickson and H. M. Tsuchiya. Continuous Propagation of Microorganisms. *AIChE J.*, 9(4):459–468, 1963.

[6] M. A. Henson, D. Muller, and M. Reuss. Cell population modelling of yeast glycolytic oscillations. *Biochem J.*, 368:433–446, 2002.

[7] C. D. Immanuel and F. J. Doyle III. Computationally-Efficient Solution of Population Balance Models Incorporating Nucleation, Growth and Coagulation. *Chem. Eng. Sci.*, 58(16):3681–3698, 2003.

[8] D. S. Kompala, D. Ramkrishna, and G. T. Tsao. Cybernetic Modeling of Microbial Growth on Multiple Substrates. *Biotech. Bioeng.*, 26:1272–1281, 1984.

[9] N. V. Mantzaris, P. Daoutidis, and F. Srienc. Numerical solution of multi-variable cell population balance models: III. Finite elements methods. *Comput. Chem. Eng.*, 25:1463–1481, 2001.

[10] N. V. Mantzaris, F. Srienc, and P. Daoutidis. Nonlinear productivity control using a multi-staged cell population balance model. *Chem. Eng. Sci.*, 57:1–14, 2002.

[11] N. F. Mathon and A. C. Lloyd. Cell Senescence and Cancer. *Nature Reviews: Cancer*, 3:203–213, 2001.

[12] J. E. Mittler, B. Sulzer, A. U. Neumann, and A. S. Perelson. Influence of Delayed Viral Production on Viral Dynamics in HIV-1 Infected Patients. *Math. Biosci.*, 152:143–163, 1998.

[13] D. Ruggero and P. P Pandolfi. Does the Ribosome Translate Cancer? *Nature Reviews: Cancer*, 3:179–192, 2003.

[14] P. M. Siegel and J. Massague. Cytostatic and Apoptotic Actions of TGF-beta in Homeostasis and Cancer. *Nature Reviews: Cancer*, 3:807–820, 2003.

[15] F. Srienc. Cytometric data as the basis for rigorous models of cell population dynamics. *J. Biotechnol.*, 71:233–238, 1999.

[16] G. Stephanopoulos. Invited comment: Chemical and Biological Engineering. *Chem. Eng. Sci.*, 58:4931–4933, 2003.

[17] G. Zhu, A. Zamamiri, M. A. Henson, and M. A. Hjortso. Model predictive control of continuous yeast bioreactor using cell population balance models. *Chem. Eng. Sci.*, 55:6155–6167, 2000.