COMBINED METABOLIC AND CELL POPULATION MODELING FOR YEAST BIOREACTOR CONTROL

Michael A. Henson * **Dirk Müller** **
Matthias Reuss **

* *Department of Chemical Engineering, University of Massachusetts, Amherst, MA 01003, USA*
** *Institute of Biochemical Engineering, University of Stuttgart, 70569 Stuttgart, Germany*

Abstract: Several investigators recently have explored the use of cell population balance equation (PBE) models for the design of biochemical reactor control strategies. A major disadvantage of the PBE modeling approach is that the incorporation of intracellular reactions needed to accurately describe cellular processes leads to substantial computational difficulties. We investigate an alternative modeling technique in which the cell population is constructed from an ensemble of individual cell models. The average value or the number distribution of any intracellular property captured by the cell model can be computed from ensemble simulation data. To illustrate the basic procedure, a single cell model of yeast glycolytic oscillations is used to construct large cell ensembles for the investigation of cell population synchronization. The potential use of cell ensemble models for bioreactor controller design are discussed.

Keywords: cell metabolism, population balance modeling, bioreactor control.

1. INTRODUCTION

The cell population balance equation (PBE) has been developed to describe heterogeneities in large cell populations (Eakman *et al.*, 1966). Most PBE models are based on a single internal state such as cell age (Hjortso and Nielsen, 1995) or cell mass (Zhu *et al.*, 2000). Cell PBE models with a vector of internal states corresponding to intracellular concentrations also can be constructed. The incorporation of intracellular reactions within the PBE framework is facilitated by utilizing a distribution function that represents the mass fraction of cells with a particular internal state (Nielsen and Villadsen, 1994). In addition to difficulties associated with modeling cell cycle events, an inherent limitation of the mass fraction PBE formulation is that a detailed intracellular description leads to a computationally intractable model due to the high dimension of the internal cell state.

Shuler and co-workers (Ataai and Shuler, 1985; Kim and Shuler, 1990; Schuler and Domach, 1983) have developed an alternative modeling approach for heterogeneous cell populations. Rather than formulate the governing PBE, the cell population is described by an ensemble of single cell models which differ according to key properties such as the division size. The number distribution function with respect to any property captured by the single cell model can be calculated from ensemble simulation data. Ensembles with approximately 250 individual cells have been used to predict steady-state and transient size distributions for aerobic (Schuler and Domach, 1983) and anaerobic (Ataai and Shuler, 1985) continuous

cultures of *E. coli* as well as plasmid instability in a genetically modified *E. coli* strain (Kim and Shuler, 1990). We have not found any recent developments or applications of this promising modeling approach. In this paper, we outline the construction of a cell ensemble model for predicting population synchronization associated with yeast glycolytic oscillations (Henson *et al.*, 2002). The results are used to access the utility of cell ensemble models for bioreactor controller design.

2. YEAST GLYCOLYTIC OSCILLATIONS

Glycolysis is the cellular process by which glucose is metabolized to generate stored energy in the form of ATP. Under certain laboratory conditions oscillations have been observed in glycolytic intermediates and extracellular species. Experimental studies (Ghosh and Chance, 1964) suggest that an autocatalytic reaction in the glycolytic chain is responsible for single cell oscillations. Additional experimental work has focused on characterizing the cellular mechanisms which cause synchronization of individual cells such that they oscillate in phase, thereby producing oscillations at the cell population level. These studies suggest that excreted acetaldehyde is the extracellular species which mediates synchronization (Richard *et al.*, 1996).

2.1 Single Cell Model

A single cell model derived from the glycolytic reaction pathway shown in Figure 1 is used for the computational studies presented in this paper. The following model equations (Wolf and Heinrich, 2000) are obtained for an arbitrary cell i:

$$\frac{dS_{1,i}}{dt} = J_0 - v_{1,i} \tag{1}$$

$$\frac{dS_{2,i}}{dt} = 2v_{1,i} - v_{2,i} - v_{6,i} \tag{2}$$

$$\frac{dS_{3,i}}{dt} = v_{2,i} - v_{3,i} \tag{3}$$

$$\frac{dS_{4,i}}{dt} = v_{3,i} - v_{4,i} - J_i \tag{4}$$

$$\frac{dN_{2,i}}{dt} = v_{2,i} - v_{4,i} - v_{6,i} \tag{5}$$

$$\frac{dA_{3,i}}{dt} = -2v_{1,i} + 2v_{3,i} - v_{5,i} \tag{6}$$

where: S_1, S_2, S_3, S_4, N_2 and A_3 denote the intracellular concentrations of the species shown in Figure 1; J_0 is the flux of glucose into the cell; and J_i is the net flux of acetaldehyde/pyruvate out of the i-th cell. The intracellular reaction rates v_2–v_6 depend linearly on the species involved in each reaction (Wolf and Heinrich, 2000). The reaction rate v_1 includes an additional nonlinear factor that accounts for autocatalytic behavior:

$$v_1 = k_1 S_{1,i} A_{3,i} \left[1 + \left(\frac{A_{3,i}}{K_I}\right)^q\right]^{-1} \tag{7}$$

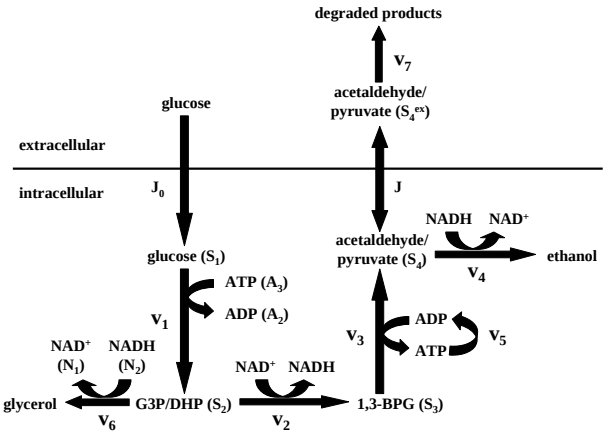

Fig. 1. Yeast glycolysis reaction pathway.

where: k_1, K_I and q are kinetic parameters. The flux of acetaldehyde/pyruvate from the i-th cell into the extracellular environment is modeled as:

$$J_i = \kappa(S_{4,i} - S_{4,ex}) \tag{8}$$

where: $S_{4,ex}$ is the extracellular acetaldehyde/pyruvate concentration; and κ is a coupling parameter related to the cell permeability.

2.2 Cell Population Balance Equation Model

A population balance equation (PBE) model based on the glycolytic reaction network in Figure 1 is formulated to demonstrate the associated computational difficulties. The PBE model is derived using a distribution function that represents the mass fraction rather than the number fraction of cells with a particular internal state because this formalism allows intracellular reactions to be incorporated in a straightforward manner (Nielsen and Villadsen, 1994). The PBE is written as:

$$\frac{\partial \Psi(x,t)}{\partial t} + \sum_{j=1}^{J} \frac{\partial}{\partial x_j} \left[R_j(x,t)\Psi(x,t)\right] = \left[\mu(x,t) - \mu(t)\right] \Psi(x,t) \tag{9}$$

where: x is the internal state vector; $\Psi(x,t)dx$ represents the mass fraction of cells with internal state in the range $[x, x+dx]$ at time t; J is the number of intracellular species; x_j is the intracellular concentration of species j; and $R_j(x,t)$ is the net rate of formation of species j. The function $\mu(x,t)$ represents the specific growth rate of cells with internal state x, while $\mu(t)$ is the average specific growth rate. Both quantities can be calculated directly from the intracellular reaction rates given the associated stoichiometry (Nielsen and Villadsen, 1994). For the yeast glycolytic pathway

depicted in Figure 1, the number of intracellular species $J = 6$ where $x^T = [S_1\ S_2\ S_3\ S_4\ N_2\ A_3]$.

From a computational perspective, the key point is that the internal state x which characterizes the intracellular concentrations of each cell is of dimension six. Assume the model is to be solved numerically by discretization in each of the six internal coordinates (Zhu *et al.*, 2000). This procedure will yield a set of nonlinear ordinary differential equations (ODEs) with time as the only independent variable. If the same number of discretization points m is used for each coordinate, then the total number of ODEs is $n = m^6$. Even with a coarse discretization where $m = 10$, this procedure results in one million ODEs. The dimension of the internal state vector clearly places severe limitations on the complexity of the intracellular reaction network that can be utilized. Furthermore, the mass fraction PBE formalism is not useful for modeling cell cycle events such as budding and mitosis (Nielsen and Villadsen, 1994).

3. CELL ENSEMBLE MODEL

The ensemble modeling technique allows single cell behavior to be described with an appropriate level of detail and circumvents the computational problems inherent in the PBE modeling approach. Furthermore, the number distribution with respect to any property captured by the single cell model can be calculated from ensemble simulation data. Below we construct a cell ensemble model to investigate the synchronization phenomenon associated with yeast glycolytic oscillations.

3.1 Model Formulation and Solution

The dynamics of the i-th cell in the population are represented by (1)–(8). A mass balance on extracellular acetaldehyde/pyruvate is derived under the assumption that the cell volume density (φ) remains constant as the number of cells M is varied (Wolf and Heinrich, 2000):

$$\frac{dS_{4,ex}}{dt} = \frac{\varphi}{M}\sum_{1}^{M} J_i - v_7 = \frac{\varphi}{M}\sum_{1}^{M} J_i - kS_{4,ex} \quad (10)$$

where k is the kinetic constant of the acetaldehyde/pyruvate degradation reaction. The total number of ODEs (n) in the cell ensemble model increases linearly with the number of intracellular species (6) and the number of individual cells (M): $n = 6M+1$. This is to be contrasted with the PBE model (9) where the number of ODEs obtained from discretization increases as the power of the number of intracellular species.

The model parameter values used in the subsequent simulations are identical to those listed in our original paper (Henson *et al.*, 2002). For these values, the cell ensemble model possesses a single stable periodic solution in which all the cells oscillate in phase and with the same amplitude regardless of the cell number. Substantially more complex oscillatory solutions are obtained when each cell is subject to random perturbations in the intracellular kinetic parameters. The dynamic simulation code was developed in FORTRAN using the variable step ODE solver `DVODE` (Brown *et al.*, 1989). Efficient solution of large cell ensembles was achieved by approximating the full Jacobian matrix with a banded Jacobian matrix. The actual Jacobian matrix is not banded due to the presence of the acetaldehyde/pyruvate flux J_i in (4) and (10). When these flux terms are neglected in the Jacobian calculation, the problem becomes highly banded. We found that this simplification reduced computation time by at least an order of magnitude. A typical one hour dynamic simulation with 1000 cells required less than 10 minutes of CPU time on a Pentium III 700 MHz processor.

3.2 Calculation of Distribution Properties

Numerical integration of the cell ensemble model produces a data matrix which contains the intracellular concentrations of each cell and the extracellular acetaldehyde/pyruvate concentration at each sampling point in time. This problem of computing cell size distributions from ensemble data was investigated previously for *E. coli* (Schuler and Domach, 1983). Below we present a simple algorithm for computing the cell number distribution with respect to any intracellular variable.

Let $z(t)$ represent the intracellular variable for which the cell number distribution function $N(z,t)$ is to be calculated. Consider discretization of the internal coordinate z into L intervals of width $\Delta z_l = z_l - z_{l-1}$ where $z_0 = z_{min}$ and $z_L = z_{max}$. By definition of the distribution function:

$$\int_0^\infty N(z,t)dz \cong \sum_{l=1}^{L} N_l(t)\Delta z_l = 1 \quad (11)$$

where $N_l(t)$ represents the average value of $N(z,t)$ over the interval Δz_l. Denote $\tilde{z}_i(t_k)$ as the value of the intracellular variable z produced by the i-th cell at the discrete time t_k. For an ensemble consisting of M individual cells, the mean value of the intracellular variable z at time t_k is:

$$\bar{z}(t_k) = \frac{1}{M}\sum_{i=1}^{M} \tilde{z}_i(t_k) \quad (12)$$

The distribution function is computed by partitioning the ensemble into the discrete intervals:

$$\tilde{n}_l(t_k) = \sum_{i=1}^{M} \{S[\tilde{z}_i(t_k) - z_{l-1}] - S[\tilde{z}_i(t_k) - z_l]\} \quad (13)$$

where: $\tilde{n}_l(t_k)$ represents the number of cells with intracellular state z in the range $[z_{l-1}, z_l)$; $l \in [1, L]$; and $S(x)$ is the unit step function. The discretized approximate number distribution function is calculated as:

$$\tilde{N}_l(t_k) = \frac{\tilde{n}_l(t_k)}{M\Delta z_l}, \qquad l \in [1, L] \tag{14}$$

If the discretization interval is sufficiently small, then a smooth continuous number distribution $\tilde{N}(z, t_k)$ can be computed from the discrete distribution values $\tilde{N}_l(t_k)$ by polynomial interpolation.

Resolution of the population behavior is determined primarily by the number of cells M in the ensemble. As M is increased, the number of intervals L also can be increased such that each interval is populated with a sufficient number of cells to produce a smooth distribution function. If L is chosen too small relative to M, resolution is unnecessarily lost. Conversely, the distribution function will be noticeably non-smooth if L is chosen too large relative to M.

4. SIMULATION STUDY

The cell ensemble modeling approach is applied to the problem of yeast glycolytic oscillations. We focus on NADH concentration dynamics to allow comparisons to experimental data where the average NADH concentration was continuously measured by fluorometry (Ghosh and Chance, 1964; Richard *et al.*, 1996). The ensemble model produces non-trivial number distributions (*i.e.*, cells with different NADH concentrations) only if there is some source of randomness in the individual cell models. Two possible sources of randomness are investigated in the following simulations.

The first test involves an ensemble of 1000 cells in which the initial conditions of each individual cell are perturbed according to a Gaussian distribution with zero mean and a variance of 2.25. Figure 2 shows the NADH concentration evolution of each cell. Due to the large variance used, initially the cell population is disorganized and exhibits no temporal structure indicative of a synchronized culture. A highly synchronized population in which the cells oscillate in phase and with a period of approximately one minute is observed after 60 minutes. Figure 3 shows the ensemble average NADH concentration dynamics and the NADH number distributions computed with $z_0 =$ 0 mM, $z_L = 0.3$ mM and $\Delta z = 0.005$ mM at three times during the simulation. The computed distributions show the presence of two distinct cell subpopulations which eventually become synchronized and converge into a single population that produces fully developed oscillations. Slower synchronization is observed when less cell models are included in the ensemble (Henson *et al.*, 2002).

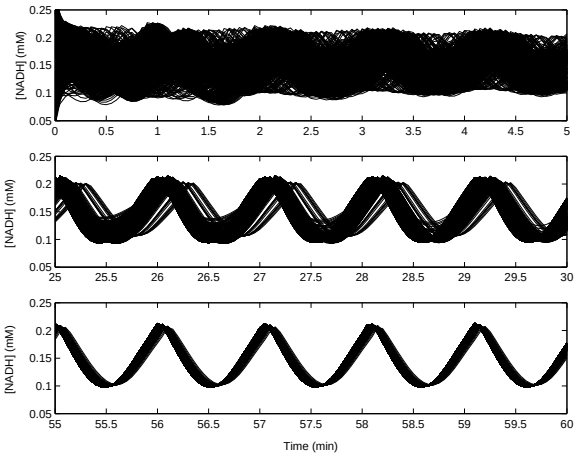

Fig. 2. Cell population synchronization for randomized initial cell state.

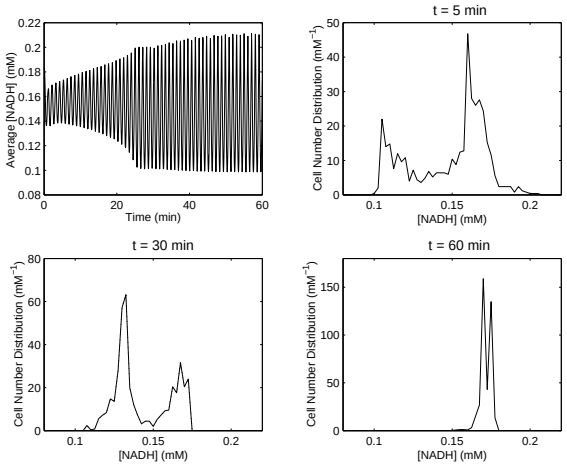

Fig. 3. Ensemble average NADH concentration dynamics and NADH number distributions corresponding to Figure 2.

The next test involves a 1000 cell ensemble in which the intracellular reaction rate parameters of each cell are randomly perturbed with zero mean and variance of 6.25×10^{-4}. Figure 4 shows the evolution of the NADH concentration of each individual cell at the end of a 150 minute dynamic simulation. While the dynamic behavior is not easily characterized, the random variations appear to produce three distinct subpopulations. This behavior is more clearly evident in Figure 5 where the dynamics of the average NADH concentration and the NADH concentration of three individual cells near the end of the 150 minute simulation are shown. The individual cells have been chosen to show the oscillatory dynamics of representative cells from the three subpopulations observed in Figure 4. Each subpopulation must contain a sufficient number of cells to yield accurate predictions of average and distribution properties. This is possible only if the cell ensemble is sufficiently large. The top plot in Figure 6 shows the complete evolution of the average NADH concentration. The bottom plot shows the NADH number distributions computed at at 0 (—), 10 ($\cdots$) and 150 ($- - -$)

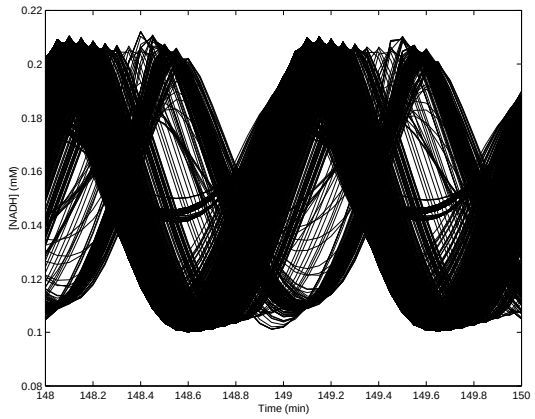

Fig. 4. Cell population dynamics for randomized intracellular kinetic parameters.

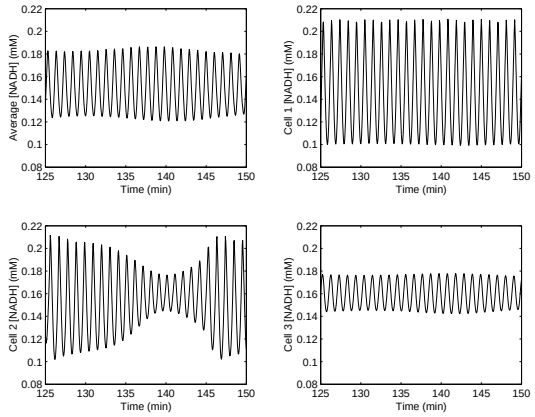

Fig. 5. NADH concentration dynamics for the total ensemble and three representative cells corresponding to Figure 4.

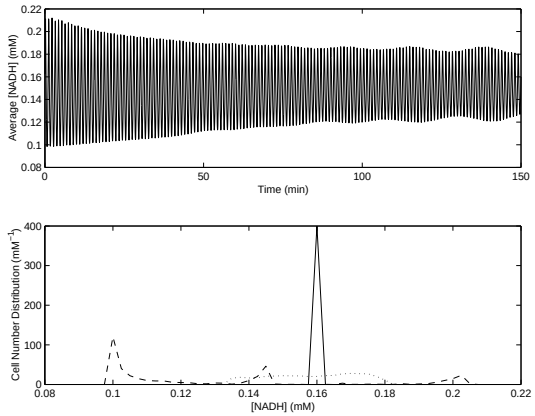

Fig. 6. Ensemble average NADH concentration dynamics and NADH number distributions corresponding to Figure 4.

minutes. The sharp initial distribution is indicative of a highly synchronized cell population. As random cell variations lead to desynchronization, the NADH distribution becomes increasingly dispersed. The final distribution clearly shows the existence of the three distinct cell subpopulations.

5. BIOREACTOR CONTROLLER DESIGN

The cell ensemble modeling approach allows a detailed single cell model to be incorporated within a cell population description. As a result, cell ensemble models are well suited for predicting complex population dynamics observed in biochemical reactors. The development of cell population control strategies based on such detailed models offers the potential to enhance bioreactor productivity, especially with respect to extracellular metabolites that are excreted during a specific phase of the cell cycle as a result of complex interactions between cellular metabolism and cell cycle events (Alberghina *et al.*, 1991).

As compared to simpler modeling techniques commonly used for bioreactor simulation and control (Daoutidis and Henson, 2002), an obvious shortcoming of the cell ensemble approach is model complexity. We have shown that the total number of model equations increases linearly both with the number of equations in the single cell model and the number of cells included the ensemble. However, the simple cell model used in this paper yields an ensemble model of 6000 differential equations. More complex cell models (Domach and Shuler, 1984; Ataai and Shuler, 1985) and/or larger ensembles will yield population models that are difficult to solve efficiently. This raises serious questions about the possible utility of these models for bioreactor control.

Clearly the development of control strategies based on high dimensional cell ensemble models will be facilitated by continuing improvements in computing technology. Parallel advances in single cell modeling and cell ensemble solution algorithms also will be required. For example, judicious simplification of intracellular reaction networks can reduce the computational burden associated with the individual cell model. Only the reaction pathways most relevant to the cellular behavior being studied need be incorporated. Furthermore, linear reaction pathways often can be lumped into a single reaction without loss of model fidelity to reduce the number of dependent variables (Nielsen and Villadsen, 1994). The cell model studied in this paper only contains lumped reactions in the upper part of the glycolytic pathway because this level of detail is sufficient to describe the oscillatory dynamics.

Model complexity also may be reduced by using a relatively small number of single cells to construct the ensemble model. In this paper, we used large ensembles of 1000 cells to achieve fine resolution of the computed cell number distributions. Previous work on *E. coli* suggests that smaller ensembles comprised of a few hundred cell models can be sufficient to resolve the population behavior (Domach and Shuler, 1984; Ataai

and Shuler, 1985). Furthermore, the ensemble size required to achieve satisfactory prediction of cell population dynamics within a feedback control strategy may be considerably less than that needed to generate high fidelity simulation results.

The development of control strategies based on cell ensemble models will require the formulation and solution of large-scale state estimation problems. The first step in this direction is analysis of cell ensemble model observability for given sets of intracellular and extracellular measurements. Extensive distribution measurements of intracellular concentrations that are required to achieve observability typically will not be available. In this case, estimation of unobservable state variables via an open-loop observer should be possible. Regardless of the problem formulation, the nonlinear state estimator will require the development of customized numerical solution techniques. The computation time required to solve the cell ensemble model studied in this paper was reduced by an order of magnitude by exploiting the approximately banded structure of the model equations. The same simplification should be applicable to other ensemble models in which the individual cells interact via a limited number of species in the extracellular environment. For example, cell cycle dependent oscillations observed in continuous yeast bioreactors are believed to be mediated by ethanol excreted into the extracellular environment (Nielsen and Villadsen, 1994).

Despite these possible improvements, cell ensemble models will remain complex and difficult to utilize for model-based control. The incorporation of such models within nonlinear optimization-based control strategies appears to infeasible. We intend to pursue linear model predictive control and simple nonlinear control techniques such as feedback linearization. More specifically, population control strategies will be developed for continuous yeast bioreactors to determine the productivity improvements which result from using an ensemble model based on a detailed single cell model instead of population balance equation model in which the intracellular state is characterized only by cell mass (Zhu *et al.*, 2000).

6. ACKNOWLEDGEMENTS

The first author would like to acknowledge the Alexander von Humboldt Foundation (Germany) for partial financial support of this research.

REFERENCES

Alberghina, L., B. M. Ranzi, D. Porro and E. Martegani (1991). Flow cytometry and cell cycle kinetics in continuous and fed-batch fermentations of budding yeast. *Biotech. Prog.* **7**, 299–304.

Ataai, M. M. and M. L. Shuler (1985). Simulation of CFSTR through development of a mathematical model for anaerobic growth of *Escherichia coli* cell population. *Biotech. Bioeng.* **27**, 1051–1055.

Brown, P. N., G. D. Byrne and A. C. Hindmarsh (1989). VODE: A variable coefficient ODE solver. *SIAM J. Sci. Stat. Comput.* **10**, 1038–1051.

Daoutidis, P. and M. A. Henson (2002). Dynamics and control of cell populations in continuous bioreactors. *AIChE Symposium Series* **326**, 274–289.

Domach, M. M. and M. L. Shuler (1984). A finite representation model for an asynchronous culture of *E. coli*. *Biotech. Bioeng.* **26**, 877–884.

Eakman, J. M., A. G. Fredrickson and H. H. Tsuchiya (1966). Statistics and dynamics of microbial cell populations. *Chem. Eng. Prog. Symp. Ser.* **62**, 37–49.

Ghosh, A. and B. Chance (1964). Oscillations for glycolytic intermediates in yeast cells. *Biochemical and Biophysical Research Communications* **16**, 174–181.

Henson, M. A., D. Muller and M. Reuss (2002). Cell population modeling of yeast glycolytic oscillations. *Biochemical J.* **368**, 433–446.

Hjortso, M. A. and J. Nielsen (1995). Population balance models of autonomous microbial oscillations. *J. Biotech.* **42**, 255–269.

Kim, B. G. and M. L. Shuler (1990). A structured, segregated model for genetically modified *Escherichia coli* cells and its used for prediction of plasmid stability. *Biotech. Bioeng.* **36**, 581–592.

Nielsen, J. and J. Villadsen (1994). *Bioreaction Engineering Principles*. Plenum Press. New York, NY.

Richard, P., B. M. Bakker, B. Teusink, K. van Dam and H. V. Westerhoff (1996). Acetaldehyde mediates the synchronization of sustained glycolytic oscillations in populations of yeast cells. *Eur. J. Biochem.* **235**, 238–241.

Schuler, M. L. and M. M. Domach (1983). Mathematical models of the growth of individual cells. In: *Foundations of Biochemical Engineering* (H. W. Blanch, E. T. Papoutsakis and G. Stephanopolous, Eds.). Chap. 5, pp. 93–133. ACS. Washington D.C.

Wolf, J. and R. Heinrich (2000). Effect of cellular interaction on glycolytic oscillations in yeast: A theoretical investigation. *Biochem. J.* **345**, 321–334.

Zhu, G.-Y., A. M. Zamamiri, M. A. Henson and M. A. Hjortso (2000). Model predictive control of continuous yeast bioreactors using cell population models. *Chem. Eng. Sci.* **55**, 6155–6167.