Bioprocess and Biosystems Engineering 24 (2001) 83–91
DOI 10.1007/s004490100229

Generic model control of induced protein expression in high cell density cultivation of *Escherichia coli* using on-line GFP-fusion monitoring

M.P. DeLisa, H.J. Chae, W.A. Weigand, J.J. Valdes, G. Rao, W.E. Bentley

Abstract A model-based control algorithm (generic model control) is presented for fed-batch cultivation of recombinant *Escherichia coli* producing either transcriptional or translational fusion products. With the recent development of translational and operon fusions using green fluorescent protein (GFP) [Albano et al. (1998) Biotechnol Prog 14:351–354] along with an on-line GFP sensor [Randers-Eichhorn et al. (1997) Biotechnol Bioeng 55:921–926], real-time measurements of foreign protein level are now possible. A mathematical model is presented that is both accurate and simple so as to ensure that all state variables remain observable during cultivation. A balance between model accuracy and mathematical tractability was obtained to facilitate the formulation of the control algorithm. Generic model control (GMC) is a process model-based control algorithm incorporating a process model directly within the control structure. GMC was desirable since linearization of the process model was not necessary and robust performance could be obtained despite process disturbances or plant/model mismatch. Furthermore, a time-delay compensator was built into the control law to account for the observed 90-min lag associated with GFP fluorescence. The feasibility of the GMC algorithm was demonstrated by simulations.

Keywords Model-based control, Generic model control, High cell density fermentation, Green fluorescent protein, *Escherichia coli*, On-line sensor

List of symbols

d	process disturbance(s)
E'	corrected error signal
F_I	feed rate of inducer (l h^{-1})
$F_{I_{max}}$	maximum allowable inducer feed rate (l h^{-1})
F_S	feed rate of substrate (glucose; l h^{-1})
i	index for controller constants K_1 and K_2
I	inducer concentration (g l^{-1})
I_F	inducer feed concentration (g l^{-1})
I_{max}	maximum allowable inducer concentration (g l^{-1})
k	index for discrete control intervals
k_1	induced biosynthesis rate constant
k_2	biodegradation rate constant (h^{-1})
K_1	proportional controller constant
K_2	integral controller constant
K_{I_0}	induction constant in protein synthesis rate (g l^{-1})
K_{I_1}	induction constant in protein synthesis rate (g $^{-1}$)
K_{IC}	inducer saturation constant (arabinose system; g l^{-1})
K_S	substrate saturation constant (g l^{-1})
K_{SI}	substrate inhibition constant (g l^{-1})
P_f	foreign protein (CAT) concentration (g l^{-1})
P_f^*	optimal foreign protein (CAT) trajectory (g l^{-1})
q_I	specific inducer consumption rate (*ara* system; h^{-1})
$q_{I_{max}}$	maximum specific inducer consumption rate (*ara* system; h^{-1})
S	substrate concentration (g l^{-1})
S_F	substrate feed concentration (g l^{-1})
t_{GMC}	initiation of GMC controller (h)
t_f	final time of fermentation (h)
u	process input(s)
V	culture volume (l)
V_{max}	maximum allowable culture volume (l)
x	process state variable(s)
X	biomass concentration (g l^{-1})

Received: 12 January 2000 / Accepted: 10 December 2000
Published online: 19 July 2001

M.P. DeLisa, H.J. Chae, W.E. Bentley (✉)
Center for Agricultural Biotechnology,
University of Maryland Biotechnology Institute,
College Park, MD 20742, USA
E-mail: bentley@eng.umd.edu
Tel.: +1-301-4054321
Fax: +1-301-3159075

W.A. Weigand
Department of Chemical Engineering, University of Maryland,
College Park, MD 20742, USA

J.J. Valdes
Edgewood Research, Development and Engineering Center,
Aberdeen Proving Ground, MD 21010, USA

G. Rao
Medical Biotechnology Center, University of Maryland
Biotechnology Institute, College Park, MD 20742, USA

M.P. DeLisa, H.J. Chae, W.E. Bentley
Department of Chemical Engineering, University of Maryland,
College Park, MD 20742, USA

G. Rao
Department of Chemical and Biochemical Engineering,
University of Maryland, Baltimore County, MD 21228, USA

Partial financial support was provided by the National Science Foundation ECSEL Coalition (to the University of Maryland, College Park) and the United States Army Engineering, Research and Development Center, Edgewood, MD (Grant no. DAAM01-96-0037; to William E. Bentley).

y	process output variable(s)
y^*	desired process output
y_m	modeled process output
$Y_{X/S}$	biomass yield coefficient (g g^{-1})
α	protein attenuation factor
μ	specific growth rate (h^{-1})
μ_{max}	maximum specific growth rate (h^{-1})
π	specific foreign protein production rate (h^{-1})
θ	time delay (h)
τ	speed of response parameter
ξ	shape of response parameter

1 Introduction

Maximization of active recombinant product is a central concern facing researchers of microbial fermentation. Since most recombinant proteins in *Escherichia coli* are synthesized intracellularly, the overall productivity of a recombinant *E. coli* fermentation is a function of both the cell density and the specific yield. One way in which bioprocess researchers have improved protein productivity is by optimizing cell mass concentration during cultivation. Simple pre-determined substrate feeding profiles [1, 2, 3] as well as more sophisticated substrate feeding optimization algorithms based on Pontryagin's maximum principle [4, 5, 6, 7], evolutionary programming [8] and neural network modeling [9] have been implemented to achieve high cell densities. However, simply maximizing cell concentration does not always guarantee high productivity. Indeed, increased cell concentration can also be accompanied by a decline in recombinant protein productivity [10, 11]. Additionally, using off-line pre-determined profiles can lead to the deterioration of process performance in the presence of modeling errors or unmodeled disturbances. These are mitigated, however, by employing on-line parameter and state variable estimation [12, 13].

The use of on-line, automated control algorithms has been instrumental in improving foreign protein production. On-line sensors for glucose, acetate, and dissolved oxygen tension have led to the development of control strategies that adjust the substrate feed rate based on feedback from these measurements [11, 14, 15]. The drawback of these approaches is that protein synthesis is controlled in an indirect manner, typically via the maximization of cell mass. Efforts to control productivity directly, such as optimization of inducer feed rate using structured models [16] and on-line optimal control of foreign protein production [17], have been limited due to the lack of an on-line sensor for protein assay. Recent advances utilizing green fluorescent protein (GFP) fusion constructs coupled with on-line GFP monitoring should help overcome these limitations.

Since its discovery by Shimomura [18], GFP has had a remarkable impact in the area of bioprocessing due to some of its unique characteristics. Unlike other fluorescent tags, GFP does not require cofactors to fluoresce or fixation techniques that can be deleterious to cells [19]. Among its many recent applications, GFP was shown as a quantitative marker of foreign protein in low [20, 21] and high cell density recombinant *E. coli* fermentations [22]. When coupled with a low-cost optical probe, in-vivo on-line monitoring during cultivation was achieved [22, 23].

With the recent development of on-line GFP monitoring, control algorithms that respond directly to the level of product can now be implemented. Previously, an accurate and tractable mathematical model was proposed and validated for recombinant protein expression in high cell density cultivation of *E. coli* [13]. In the present paper, a nonlinear model-based control algorithm is demonstrated for a recombinant product expressed simultaneously with a GFP marker in a nonlinear fed-batch production process. This construction allows for the implementation of a model-based control algorithm that is product specific, such that the control actions are made based on product concentrations and not on external state parameters such as dissolved oxygen tension, glucose, or acetate concentration. In the process industries, model-based controllers such as dynamic matrix control (DMC) [24], model algorithmic control (MAC) [25], and internal model control (IMC) [26] have been used to control a melange of applications in order to meet production demands. These controllers typically rely on linear approximations of the process, whereas the vast majority of biochemical processes behave in a nonlinear fashion, and whose behavior can change rapidly and severely over a period of time. Generic model control (GMC) [27, 28] was utilized here because the nonlinear dynamic structure of the model could be directly embedded in the control law. Consequently, a linearized approximation of the model was not necessary, allowing for stable operation around a transient trajectory during fed-batch operation.

An advantage of GMC was that it was a single-step control law where solution of the last control step was a good approximation to the solution of the current control step. Additionally, the controller was tuned using straightforward controller parameters. These features allowed for robust control performance, even in the presence of unmodeled disturbances or plant/model mismatch. Development of a model-based controller that can be used to control a recombinant *E. coli* fermentation is shown here. Specifically, this paper first presents the underlying structure of GMC and then demonstrates the feasibility of this control structure via simulation. This is one of the first critical steps in the development of product-based bioreactor control.

2 Generic model control

A model-based control algorithm was explored here through computer simulation prior to implementation with GFP probe (currently under investigation). Specifically, GMC [27] was chosen as it allows incorporation of nonlinear, multivariable process models directly in the control algorithm. A generalized schematic of the GMC control scheme is depicted in Fig. 1a. The process was described by a dynamic model comprised of a set of ordinary differential equations:

$$\mathbf{x} = f(\mathbf{y}, \mathbf{u}, \mathbf{d}, t) \tag{1}$$

$$\mathbf{y} = g(\mathbf{x}) \tag{2}$$

where **x** is the state vector of dimension m, **u** is a vector of process inputs of dimension n, **d** is a vector of process disturbances of dimension l, **y** is the output vector of dimension n, and t is time. The functions f and g are nonlinear and it was assumed that all states and outputs were measurable. From (1) and (2):

$$\mathbf{y} = \frac{\partial g}{\partial \mathbf{x}^T} f(\mathbf{x},\mathbf{u},\mathbf{d},t) \tag{3}$$

Noting that the desired trajectory (setpoint) of the outputs is **y***, it was desired that the rate of change of **y***, (**y**)*, be proportional to the deviation from the setpoint while eliminating offset. The closed-loop trajectory was therefore expressed as follows:

$$(\mathbf{y})^* = \mathbf{K}_1(\mathbf{y}^* - \mathbf{y}) + \mathbf{K}_2 \int (\mathbf{y}^* - \mathbf{y})\mathrm{d}t \tag{4}$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are diagonal $n \times n$ matrices and were determined using the following expressions:

$$\mathbf{K}_{1(i,i)} = \frac{2\xi_i}{\tau_i}, \quad \mathbf{K}_{2(i,i)} = \frac{1}{\tau_i^2} \tag{5}$$

where ξ_i and τ_i determine the shape and speed, respectively, of the desired closed-loop trajectory. For a step change in the setpoint, the closed-loop trajectory of Eq. (4) yields a pseudo-second-order response [29]. A more detailed account of system responses can be found elsewhere [30]. Finally, to ensure that the rate of change of the process outputs tracks the desired reference system, Eq. (3) was set equal to Eq. (4), yielding:

$$\frac{\partial g}{\partial \mathbf{x}^T} f(\mathbf{y},\mathbf{u},\mathbf{d},t) = \mathbf{K}_1(\mathbf{y}^* - \mathbf{y}) + \mathbf{K}_2 \int (\mathbf{y}^* - \mathbf{y})\mathrm{d}t \tag{6}$$

In practice, the exact process is rarely known and an approximate model is used to determine the control law:

$$\frac{\partial \hat{g}}{\partial \hat{f}\,\hat{\mathbf{x}}^T} \hat{f}(\mathbf{y},\mathbf{u},\mathbf{d},t) = \mathbf{K}_1(\mathbf{y}^* - \mathbf{y}) + \mathbf{K}_2 \int (\mathbf{y}^* - \mathbf{y})\mathrm{d}t \tag{7}$$

where $\hat{g}$ and $\hat{f}$ are approximations to the true model. Solution of Eq. (7) for the manipulated inputs, **u**, results in an algebraic control law to be solved at every sample time. Any inaccuracies introduced by this approximation will result in $y(t)$ not tracking $y^*(t)$ but the integral term in the control algorithm compensates for this.

Following the format for GMC, a SISO (single input–single output) nonlinear control law was developed by equating Eqs. (A5) and (7) such that the input/output model (Eq. A5) exactly matched the reference model (Eq. 7). Algebraic solution of the resulting equation for the manipulated process input, $u(t)$, lead to the final form of the GMC control law. Initially, F_I was selected as the manipulated variable that would be used to control the level of recombinant protein expression. However, the relatively low inducer feed rate values, which arose from the use of highly concentrated inducer feed solutions, led to sluggish control performance when F_I was implemented as the manipulated variable (not shown). It was noted, however, that the expression $[I/(K_I+I)]$ explicitly accounted for the connection between the inducer, I, and the cloned gene concentration, P_f. Therefore, the control input, $u(t)$, was defined as:

$$u(t) = \left[\frac{K_{I_0} + I}{K_{I_1} + I}\right] \tag{8}$$

and was subject to the following constraint:

$$u(t) \geq 0 \tag{9}$$

In addition, constraints on state variables were imposed during all simulations as follows:

$$0 \leq V \leq V_{\max} \tag{10}$$

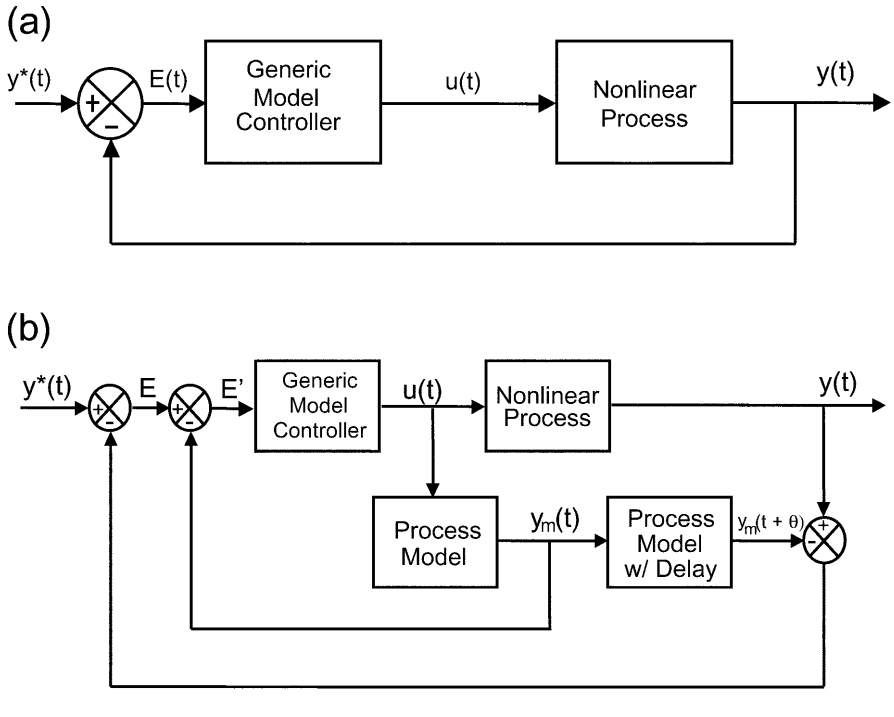

Fig. 1. **a** Generic model control (GMC) block diagram, and **b** block diagram of GMC with incorporation of time-delay compensation strategy

$$0 \leq F_I \leq F_{I\max} \tag{11}$$

$$0 \leq I \leq I_{\max} \tag{12}$$

At every control interval, the manipulated variable, $u(t)_k$, was used to calculate the corresponding value of inducer concentration, I_k, for that interval. Finally, solution of Eq. (A3) for the inducer feed rate, F_{I_k}, determined the control action to be implemented on the process. In Eq. (A3), the term dI/dt was needed and was subsequently estimated using a discrete approximation as follows:

$$\frac{\mathrm{d}I}{\mathrm{d}t} \cong \frac{\Delta I}{\Delta t} = \frac{[I_k - I_{k-1}]}{\Delta t} \tag{13}$$

where the subscript k denotes the current control interval, while k–1 refers to the value from the previous interval, and Δt is the integration interval.

3 Time-delay compensation

Experimentally, GFP fusion constructs allowed for on-line tracking of a model protein (i.e., CAT) during cultivation of *E. coli*. However, previous reports have indicated that GFP fluorescence intensity lags cloned gene expression by approximately 90 min due to chromophore cyclization [21]. Shifting fluorescence data by a constant 90 min was necessary to track foreign protein levels. Accordingly, levels of the foreign protein (chloramphenicol acetyltransferase) were determined directly from the following relationship [22]:

$$P_f\left(g\ l^{-1}\right)|_{t-1.5} = 0.134 \times FI(V)|_t - 0.012 \tag{14}$$

where FI is the fluorescence intensity measured in volts by the on-line optical sensor. Importantly, Eq. (14) provided values for the cloned gene level, utilizing real-time fluorescence measurements, that were delayed by 90 min. This time delay was viewed as a threat to control of the *E. coli* bioprocess since it is a significant fraction of the total processing time and is large compared to the doubling time of the bacteria.

In linearized process control methods, a common technique employed to compensate for time delay is the Smith predictor [31, 32]. In our system, an internal loop was integrated into the GMC control block diagram (Fig. 1b) that calculated an error signal, E', which would occur if no time delay were present. In this scenario, the control output, $u(t)$, is sent to the actual process and to the process model. Upon receiving the control action, the model predicted the effect of this action on the process, calculating an undelayed output value, $y_{\mathrm{m}}(t)$. The controller compared this value to the desired or optimal trajectory value at the current time step and calculated the necessary control action to be implemented at the next time step. Consequently, the error signal was the difference between the desired output value (y^*) and the undelayed output as predicted by the model (y_{m}). Additionally, the model also calculated the process output delayed by θ, $y_{\mathrm{m}}(t+\theta)$, and compared this value to the actual process output, $y(t)$, to provide a correction for modeling errors and disturbances. Finally, the corrected error was written:

$$E'(t) = y^*(t) - y_m(t) - [y(t) - y_m(t+\theta)] \tag{15}$$

which simplifies to the following if the process model is perfect and there are no significant disturbances:

$$E' = y^*(t) - y_m(t) \tag{16}$$

where $y^*(t)$ is the optimal trajectory, $y_{\mathrm{m}}(t)$ is the undelayed model output, $y(t)$ is the actual process output, and $y_{\mathrm{m}}(t+\theta)$ is the model output delayed by θ.

4 Results and discussion

Simulations were performed implementing GMC as the control law used to determine the feed profile of inducer necessary to maintain the foreign protein level (P_{f}) at or near a pre-determined foreign protein trajectory (P_{f}^*). All simulations were performed using Advanced Continuous Simulation Language (ACSL; MGA Software, Concord, MA) and utilized model parameters listed in Table 1 and 2. During each iteration, the measured value of P_{f} was compared to a desired value, P_{f}^*, and a control action, $u(t)$, was calculated. At each time interval, the calculated control input, $u(t)$, was used to determine the value of inducer concentration, I, required to keep the foreign protein concentration, P_{f}, on the desired profile, P_{f}^*. From this value of I, the feed rate of inducer was calculated by solving Eq. (3) for the variable, F_I. Therefore, the final control action resulted in a change of the inducer feed rate

Table 1. Model parameters and other simulation conditions. All data obtained from experiment (DeLisa et al. [22], Chae et al. [13])

Parameter	Strain		Variable	Strain	
	JM105 [pBAD-GFP::CAT]	JM105 [pTrcHis-GFPuv/CAT]		JM105 [pBAD-GFP::CAT]	JM105 [pTrcHis-GFPuv/CAT]
$\mu \mathrm{m}_{\max}$ (h^{-1})	0.55	0.36	S_F (g l^{-1})	400	400
K_S (g l^{-1})	0.23	0.13	I_F (g l^{-1})	50	50
K_{SI} (g l^{-1})	103.8	99.0	$X(0)$ (g l^{-1})	0.025	0.025
$Y_{X/S}$ (g g^{-1})	0.52	0.45	$S(0)$ (g l^{-1})	20	20
α	0.15	0.15	$P(0)$ (g l^{-1})	0	0
$q_{I_{\max}}$ (h^{-1})	0.005	0	$V(0)$ (l)	2	2
K_{IC} (g l^{-1})	0.015	N/A[a]	$V_{\max}$ (l)	3	3
t_{GMC} (h)	20	20	$F_{I_{\max}}$(g l^{-1})	5.0	0.6
t_{f} (h)	30	30	$I_{\max}$ (g l^{-1})	5.0	0.6

[a]N/A: not applicable

to the fermentor. In practice, this control action would be implemented on the process directly. However, in this report we have evaluated GMC in this nonlinear time-delayed system via simulation in order to gauge the general feasibility of the control algorithm.

For all simulations, an optimal profile for P_f^* was determined via solution of an open-loop optimization using an SQP-based algorithm [13] which, in turn, was similar to profiles obtained experimentally. Additionally, an exponential feed rate of glucose, F_S, was pre-determined according to Paalme et al. [3]. This feeding policy was identical to the pre-determined profile used in both modeled simulations and experiments, and ensured attainment of high cell densities. Growth-related parameters and cultivation conditions for both arabinose- (JM105 [pBAD-GFP::CAT]) and IPTG-inducible (JM105 [pTH-GFPuv/CAT]) systems are displayed in Table 1. Since arabinose is metabolized by *E. coli*, increases and decreases in inducer concentration are possible within the fermentor enabling much tighter process control, compared to an IPTG-based system. Correspondingly, simulations utilizing IPTG were less satisfactory and hence excluded here. Lastly, product-related and GMC parameters are shown in Table 2.

4.1
Tuning of generic model controller

A method for tuning GMC controllers based on selection of a desired profile for the controlled variable, $y(t)$, is outlined in Lee and Sullivan [27]. Briefly, the shape of this profile is characterized by the two parameters, ξ and τ. A figure of GMC reference trajectories is included in Lee and Sullivan, which presents relative control performances for different combinations of ξ and τ. Following selection of ξ and τ, calculation of K_1 and K_2 is straightforward from Eq. (5). Ideally, control constants should be tuned for minimal offset between P_f^* and P_f. However, previous reports have indicated that rapid induction can be deleterious and might not be modeled accurately using a simple model [33]. Thus, control constants were selected that avoided severe overshoot in P_f while still suppressing offset. In tuning the GMC controller, since overshoot was undesirable, ξ was set to 4.0. From the Lee and Sullivan tuning chart, with ξ=4.0, the speed of response was chosen as τ=60 min.

4.2
Performance of GMC controller

For this system, it was known that the GFP fluorescence signal lagged the expression of cloned genes by approximately 1.5 h (θ=1.5 h) and was modeled using a first order transfer function with a time constant of 1.5. The effect of this time delay was evidenced when control was performed without the assistance of time-delay compensation and the aforementioned values were used for ξ and τ, respectively. The closed-loop response for this scenario yielded an initial lag of the optimal trajectory followed by a severe overshoot and ended with slight oscillations (data not shown). This is because in time-delay systems, the effect of the control action will not be immediately felt by the process thereby adding a phase lag to the feedback loop [32]. Consequently, the gain must be decreased to maintain stable operation resulting in sluggish control compared to that obtained when there is no time delay. The sluggish response behavior will persist unless steps to circumvent the time delay are incorporated into the control loop.

A representative simulation using the aforementioned control law for a high cell density cultivation of *E. coli* JM105 [pBAD-GFP::CAT] is depicted in Fig. 2 (simulated time course of cell mass, glucose concentration and glucose feed) and Fig. 3. For GMC control results (Figs. 3a, b, and c), it was first assumed that plant/model mismatch and unmeasured disturbances were negligible. In addition, no time-delay compensation was implemented but the control parameters were retuned to be less aggressive so as to minimize overshoot. Figure 3a, b, and c demonstrate the tracking of the optimal profile, P_f^*, by the actual foreign protein expressed, P_f, which was made possible by manipulating the inducer feed rate, F_I, as determined by the GMC controller. In order to simulate the actual industrial setting, the controller was initiated at approximately 20 h. This enabled attainment of sufficient cell mass prior to induction and supplemental glucose feed.

The effect of this time delay on control is evident in Figs. 3a and b, which illustrate that initially there was sluggish control of the foreign protein concentration between 20 and 22.5 h. In Fig. 3a, the closed-loop response improved slightly by hour 25, but significant offset between the optimal trajectory and the simulated expression level were observed over the final 5 h. Importantly, no overshoot of the target trajectory was observed. In Fig. 3b, retuning the constants for slightly more aggressive control (implemented as increases in K_1 and K_2) led to a decrease in the offset observed over the final 5 h, but the simulated protein product level still deviated from the optimal tra-

Table 2. Estimated foreign production-related model parameters and GMC constants

Model parameter	Value	Control parameter	Value
k_1	8.48±0.48	ξ	4.0
k_2	0.60±0.11	τ	1.0
K_{I_0}	0.65±0.05	K_1	8.0
K_{I_1}	2508.1±8.0	K_2	1.0

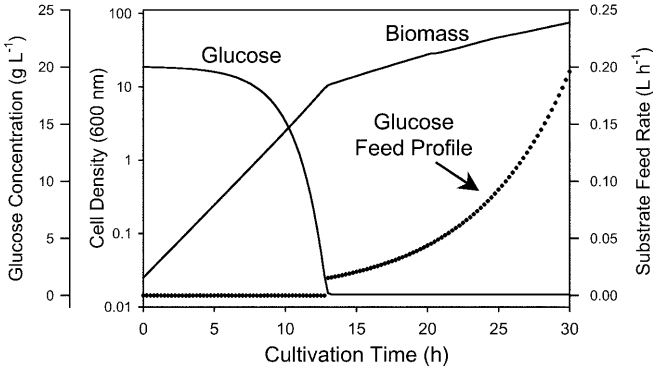

Fig. 2. A time profile of simulated cell density, glucose concentration, and glucose feed rate for *E. coli* JM105 [pBAD-GFP::CAT] fermentation

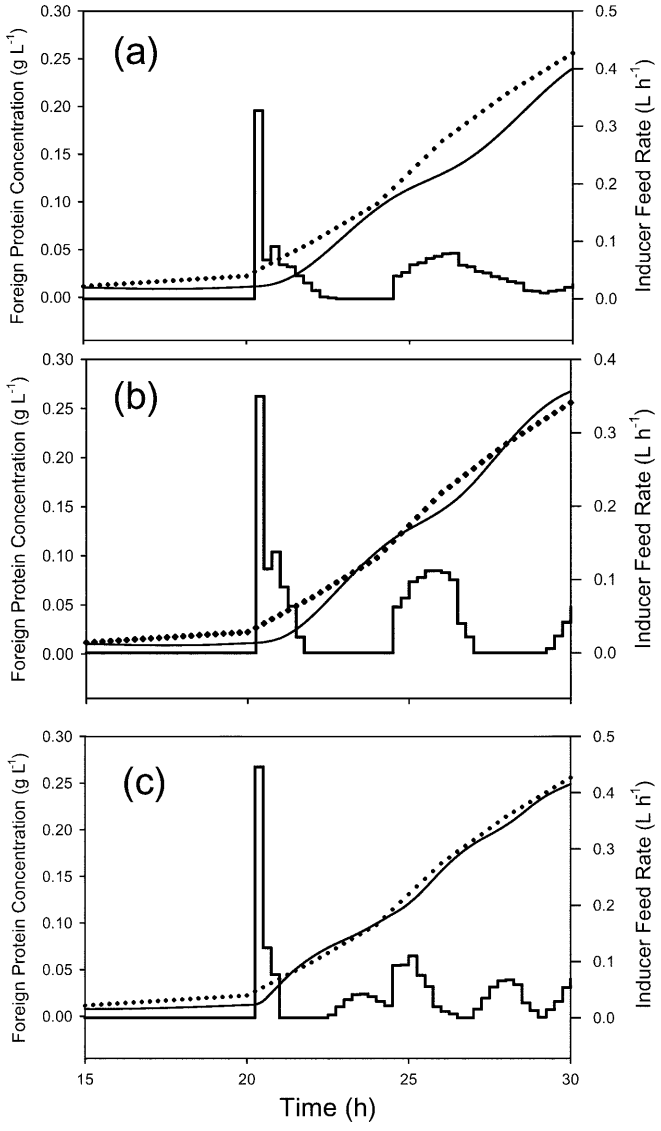

Fig. 3. **a** Optimal trajectory (●), foreign protein concentration (–), and inducer feed rate (–) obtained during on-line GMC simulation (ξ=4.0, τ=2.0) with no time-delay compensation; **b** optimal trajectory, foreign protein concentration and inducer feed rate for GMC simulation (ξ=4.0, τ=1.5) with no time-delay compensation; and **c** optimal trajectory, foreign protein concentration, and inducer feed rate obtained during on-line GMC simulation (ξ=4.0, τ=1.0) with time-delay compensation strategy. Plant/model mismatch and unmodeled disturbances assumed negligible. Controller initiated at 20 h

jectory over the same period. Notably, over the final 5 h of the simulation, an oscillatory response was observed culminating with overshoot of the optimal trajectory. In Fig. 3a, there was significant offset but no overshoot whereas in Fig. 3b the offset has been reduced at the expense of some overshoot, illustrating a trade-off in control performance.

Subsequently, simulations demonstrating GMC using time-delay compensation to account for the lag in GFP fluorescence were performed and the results are depicted in Fig. 3c. Shown are the foreign protein expression level, the optimal trajectory of foreign protein, and the inducer feed rate obtained with the controller constants tuned as mentioned previously. Again, the model was assumed devoid of error and there were no unmodeled disturbances, hence GMC performance was quite good. Additionally, larger and therefore more aggressive K_1 and K_2 values were used while still maintaining robust control. The ability to use higher gains without jeopardizing stability, even when a time delay exists, provides for much more robust control especially when modeling error and disturbances are encountered.

4.3 Robustness evaluation

In practice, the exact process model is rarely known and approximations of the actual process are used to describe the system as accurately as possible. Unfortunately, this approach invariably leads to some amount of plant/model mismatch, as the true process is always more complex than the model (structural mismatch) and the true process parameter values can never be precisely determined (parameter mismatch) [29]. Provided the modeling errors are within approximately ±30% of the actual values, the time-delay compensator scheme has demonstrated improved results over process control with no time-delay compensation [34]. As a full robustness analysis of the GMC formulation is beyond the scope of this paper, the robustness properties of the controller were investigated via simulation. Changes in process conditions (Table 3) from their nominal values were made and the ability of the controller to handle these discrepancies was evaluated [35].

To illustrate the controller's ability to handle modeling error, a ±30% discrepancy between the value of k_1 in the model and that observed experimentally was implemented. This was achieved by randomly generating a ±30% error in the value of k_1 at each control interval. Since k_1 was determined to be the most sensitive of the seven production-related parameters [13], this error posed the greatest threat to robust closed-loop control. As shown by Fig. 4, a random 30% error in k_1 did not significantly affect the tracking of the optimal trajectory. Additionally, even with notable modeling error, the closed-loop response obtained with time-delay compensation was an improvement over the control observed with no modeling error and without any time-delay compensation. As a further comparison, the same random error in k_1 was introduced to a simulation using the GMC law with no time-delay compensation. As expected, when modeling error was present, the closed-loop response was extremely poor as the cloned gene level was considerably offset from the optimal trajectory (data not shown).

Table 3. Robustness evaluation; *RE*: random error

Model parameter	Case 1	Case 2	Case 3
k_1	±30% RE	±20% RE	
k_2		±10% RE	
K_{I_0}		±10% RE	
K_{I_1}		±10% RE	
S_F			800 g l^{-1}
I_F			100 g l^{-1}
$X(0)$			0.035 g l^{-1}
$S(0)$			30 g $^{-1}$
$V(0)$			1.0 l

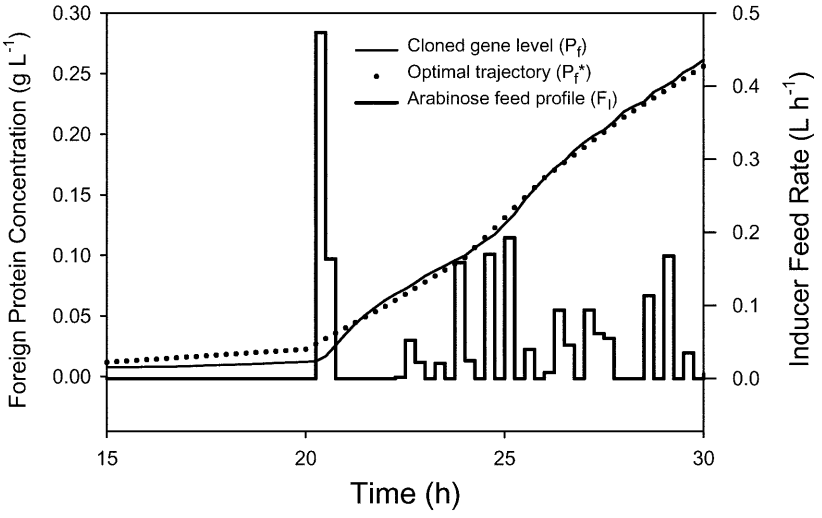

Fig. 4. Optimal trajectory, foreign protein concentration, and inducer feed rate obtained during on-line GMC simulation (ξ=4.0, τ=1.0) with time-delay compensation strategy. Plant/model mismatch implemented as a ±30% random error in the biosynthesis rate constant, k_1. Controller initiated at 20 h

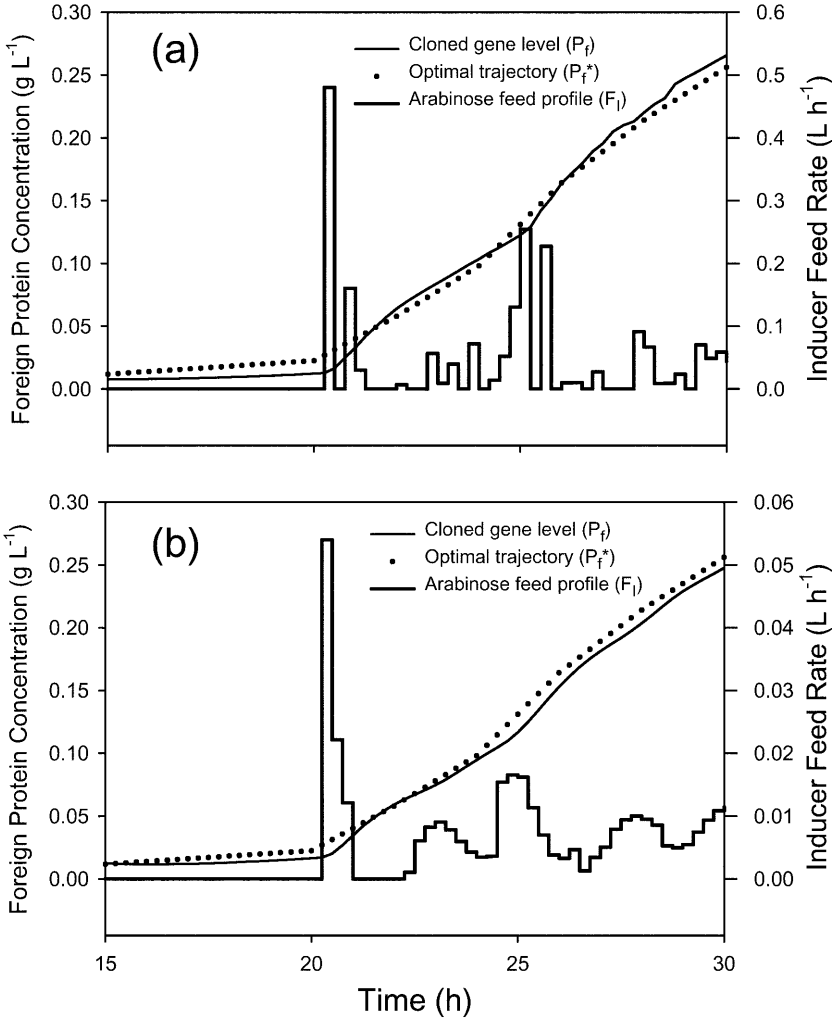

Fig. 5. **a** Optimal trajectory, foreign protein concentration, and inducer feed rate for on-line GMC simulation with time-delay compensation and random error in production-related parameters k_1, k_2, K_{I_0} and K_{I_1}; **b** optimal trajectory, foreign protein concentration and inducer feed rate for on-line GMC simulation with time-delay compensation and changes in five simulation conditions. Controller initiated at 20 h

A more realistic scenario was explored by extending the error beyond the product synthesis rate (k_1) and introducing a random 10% error in the production-related parameters k_2, K_{I_0} and K_{I_1} in addition to a 20% error in k_1. The tracking of the optimal trajectory by the actual foreign protein level as well as the value of the manipulated variable, F_I, for this case is depicted in Fig. 5a. Using controller constants tuned as described previously, the GMC performance was quite good prior to approximately 27 h, however, a slight overshoot of the desired setpoint was observed beyond 27 h. Overall, the GMC controller did an adequate job tracking the setpoint despite the extensive error in all four production-related parameters.

A final test of the controller's robustness was performed by changing five simulation conditions from their nominal

values as originally given in Table 1. The initial values for X, S, and V were changed to 0.035 g l^{-1}, 30 g l^{-1}, and 1.0 l, respectively, while the feed concentrations of glucose and inducer were increased to 800 and 100 g l^{-1}, respectively. Again, controller constants were tuned for nominal operation and additional modeling error and disturbances were assumed negligible. GMC simulation results, incorporating the changes in simulation conditions, are depicted in Fig. 5b. GMC controller performance was again quite good even with significant changes in initial conditions, which can often be detrimental when using simple process models.

5
Conclusions

Demonstrated here is a model-based control algorithm capable of determining inducer feed profiles that allow foreign protein concentration to accurately track an optimal foreign protein trajectory. Generic model control was chosen as the control algorithm as it imbedded the process model within the control law, enabling the nonlinear dynamic structure of the model to be incorporated. Importantly, GMC performed well despite process disturbances and plant/model mismatch (modeling error). With the recent availability of a low-cost optical sensor for detection of recombinant GFP-fusion products [22, 23], the need for on-line feedback controllers in bioprocessing has arisen. Additionally, the development of accurate and tractable mathematical models coupled with improved computational processors has opened the door for rapid, on-line calculations and simulations. Previously, Chae et al. [13] proposed a simple, unstructured model for cloned gene expression in recombinant cultures of *E. coli*. Incorporation of this model within the GMC architecture has led to the development of a powerful tool for controlling foreign protein production based directly on foreign protein levels and not on other indirect process variables (i.e., dissolved oxygen tension, biomass or glucose concentration). Further, coupling a time-delay compensation scheme to the GMC algorithm was demonstrated to sustain robust closed-loop control despite the 90-min time lag associated with GFP fluorescence. Lastly, the proportional and integral reference trajectory was capable of controlling protein expression in the face of significant plant/model mismatch (±30% modeling error in parameter k_1) and changes in five simulation conditions. Overall, the feasibility of on-line control using GMC and GFP-fusion monitoring was demonstrated by simulation and the simulation results presented here have provided the groundwork for implementing bioprocesses with a generic scheme for monitoring and controlling levels of a recombinant protein.

Appendix A

A1
Mathematical model

In order to successfully implement GMC, a simple but descriptive model was necessary. As the model was recently shown in the literature [13], it is reproduced here in the Appendix. Moreover, the model was constructed so that most of the state variables were readily available via existing on-line techniques. The model was defined as follows:

$$\frac{dX}{dt} = \mu X - \frac{X}{V}(F_S + F_I) \tag{A1}$$

$$\frac{dS}{dt} = \frac{F_S}{V}(S_F - S) - \frac{\mu X}{Y_{X/S}} - \frac{S}{V}(F_I) \tag{A2}$$

$$\frac{dI}{dt} = \frac{F_I}{V}(I_F - I) - \frac{I}{V}(F_S) - q_I X \tag{A3}$$

$$\frac{dV}{dt} = F_S + F_I \tag{A4}$$

$$\frac{dP_f}{dt} = \pi X - k_2 P_f - \frac{P_f}{V}(F_S + F_I) \tag{A5}$$

where X, S, I, V, and P_f are the concentrations (g l^{-1}) of biomass, substrate, glucose, inducer, and foreign product, respectively. The terms F_S and F_I are the feed rates (g h^{-1}) of glucose and inducer, while S_F and I_F are the concentrations (g l^{-1}) of glucose and inducer in the feed streams. Lastly, $Y_{X/S}$ is the growth yield coefficient (g cell mass/g substrate consumed), k_2 is the product degradation constant (h^{-1}), and π and q_I are the specific foreign production and inducer consumption rates (h^{-1}), respectively.

Foreign production models that consider induction effects have been documented [36, 37] and have successfully predicted the shock and recovery dynamics of IPTG-induction on cell growth. The following expression describing specific foreign protein expression in *E. coli*, adopted from Bentley et al. [36] and Lee and Ramirez [37], was investigated:

$$\pi = k_1 \mu \left[\frac{K_{I_0} + I}{K_{I_1} + I}\right] \tag{A6}$$

where K_{I_0} and K_{I_1} are the inducer saturation constants (g l^{-1}) and k_1 is the induced foreign protein biosynthesis rate.

The inducible systems derived from the *lac* promoter were regulated by addition of isopropyl-β-D-thiogalactopyranoside (IPTG), a gratuitous, non-metabolized inducer. Alternatively, the inducible system derived from the *ara* promoter was controlled by introduction of arabinose, which was readily metabolized by the host cell. Both systems were utilized experimentally, although the *ara* promoter system was hypothesized to be more appropriate for process control as the inducer concentration could be both raised (input) and decreased (consumption) without significant biomass dilution. For the *ara* promoter system, an inducer consumption rate was included that accounted for the loss of arabinose due to host cell metabolism and was based on the *E. coli* arabinose uptake mechanism [38]:

$$q_I = q_{I_{max}} \left[\frac{I}{K_{IC} + I}\right] \tag{A7}$$

where $q_{I_{max}}$ is the maximum specific consumption rate (h^{-1}) and K_{IC} is the arabinose saturation constant (g l^{-1}).

During the batch mode of operation, the specific growth rate could be modeled according to the classical

Monod equation with substrate inhibition [39]. Unfortunately, this expression was inadequate when cloned genes were expressed. The metabolic burden placed on the cell during recombinant protein production [40, 41, 42] is often manifested as a reduction in growth rate [10, 36]. Therefore, in order to model the mitigating effect that foreign protein expression has on the specific growth rate, the Monod expression was modified as follows:

$$\mu = \left(\frac{\mu_{\max} S}{K_S + S + \frac{S^2}{K_{SI}}} \right) exp(-\alpha P_f) \qquad \text{(A8)}$$

where $\mu_{\max}$ is the maximum specific growth rate (h^{-1}), K_S is the substrate saturation constant (g l^{-1}), K_{SI} is the substrate inhibition constant (g l^{-1}) and α is the protein attenuation factor (dimensionless).

References

1. Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer W-D (1995) Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. J Biotechnol 39:59–65
2. Hellmuth K, Korz DJ, Sanders EA, Deckwer W-D (1994) Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high cell density cultivation of *Escherichia coli* TG1. J Biotechnol 32:289–298
3. Paalme T, Tiisma K, Kahru A, Vanatalu K, Vilu R (1990) Glucose-limited fed-batch cultivation of *Escherichia coli* with computer-controlled fixed growth rate. Biotechnol Bioeng 35:312–319
4. Weigand WA (1981) Maximum cell productivity by repeated fed-batch culture: constant yield case. Biotechnol Bioeng 23:249–266
5. Lee J, Ramirez WF (1994) Optimal control of induced foreign protein production by recombinant bacteria in fed-batch reactors. AIChE J 40:899–907
6. Modak JM, Lim HC, Tayeb YK (1986) General characteristics of optimal feed rate profiles for various fed batch fermentation processes. Biotechnol Bioeng 28:1396–1407
7. Lim HC, Tayeb YK, Modak JM, Bonte P (1986) Computational algorithms for optimal feed rates for a class of fed-batch fermentation: numerical results for penicillin and cell mass production. Biotechnol Bioeng 28:1408–1420
8. Fogel DJ (1995) Evolutionary computation: toward a new philosophy of machine intelligence. IEEE Press, New York
9. Chaudhuri B, Modak JM (1998) Optimization of fed-batch bioreactor using neural network model. Bioprocess Eng 19:71–79
10. Zabriskie DW, Arcuri EJ (1986) Factors influencing productivity of fermentations employing recombinant micororganisms. Enzyme Microb Technol 8:705–717
11. Shimizu N, Fukuzono S, Fujimori K, Nishimura N, Odawara Y (1988) Fed-batch cultures of recombinant *Escherichia coli* with inhibitory substance monitoring. J Ferment Technol 66:187–191
12. Gee DA, Ramirez WF (1996) On-line state estimation and parameter identification for batch fermentation. Biotechnol Prog 12:132–140
13. Chae HJ, DeLisa MP, Cha HJ, Weigand WA, Rao G, Bentley WE (2000) A framework for on-line optimization of recombinant protein expression in high cell density *Escherichia coli* cultures using GFP-fusion monitoring. Biotechnol Bioeng 29(3):275–285
14. Turner C, Gregory ME, Thornhill NF (1994) Closed-loop control of fed-batch cultures of recombinant *Escherichi coli* using on-line HPLC. Biotechnol Bioeng 44:819–829
15. Konstantinov K, Kishimoto M, Seki T, Yoshida T (1990) A balanced DO-stat and its application to the control of acetic acid and excretion by recombinant *Escherichia coli*. Biotechnol Bioeng 36:750–758
16. Chen Q, Bentley WE, Weigand WA (1995) Optimization for a recombinant *E. coli* fed-batch fermentation. Appl Biochem Biotechnol 51/52:449–461
17. Lee J, Ramirez WF (1996) On-line optimal control of induced foreign protein production by recombinant bacteria in fed-batch reactors. Chem Eng Sci 51:521–534
18. Shimomura O, Johnson FH, Saiga Y (1962) Extraction, purification and properties of Aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol 59:223–239
19. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. Science 263:802–805
20. Albano CR, Randers-Eichhorn L, Bentley WE, Rao G (1998) Green fluorescent protein as a real time quantitative reporter of heterologous protein productoin. Biotechnol Prog 14:351–354
21. Albano CR, Randers-Eichhorn L, Chang Q, Bentley WE, Rao G (1996) Quantitative measurement of green fluorescent protein expression. Biotechnol Tech 10:953–958
22. DeLisa MP, Li J, Rao G, Weigand WA, Bentley WE (1999) Monitoring fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. Biotechnol Bioeng 65:54–64
23. Randers-Eichhorn L, Albano C R, Sipior J, Bentley WE, Rao G (1997) On-line green fluorescent protein sensor with LED excitation. Biotechnol Bioeng 55:921–926
24. Cutler CR (1983) Dynamic matrix control: an optimal multivariable algorithm with constraints. PhD Thesis, University of Houston, Texas
25. Richalet J, Rault A, Testud JL, Papon J (1979) Model predictive heuristic control: applications to industrial processes. Automatica 14:413
26. Garcia CE, Morari M (1982) Internal model control: 1. A unifying review and some new results. Ind Eng Chem Proc Res Dev 21:308
27. Lee PL, Sullivan GR (1998) Generic model control (GMC). Comput Chem Eng 12:573–580
28. Lee PL (1993) Nonlinear process control: applications of generic model control. Springer, Berlin Heidelberg New York
29. Signal PD, Lee PL (1992) Generic model adaptive-control. Chem Eng Commun 115:35–52
30. Bartusiak RD, Georgakis C, Reilly MJ (1989) Nonlinear feedforward/feedback control structures designed by reference system synthesis. Chem Eng Sci 44:1837–1851
31. Smith OJM (1959) A controller to overcome deadtime. ISA J 6:28
32. Ogunnaike BA, Ray WH (1994) Process dynamics, modeling, and control. Oxford University Press, New York
33. Ramirez D, Bentley WE (1999) Characterization of stress and protein turnover from protein overexpression in fed-batch cultures. J Biotechnol 71:39–58
34. Seborg DE, Edgar TF, Mellichamp DA (1989) Process Dynamics and control. Wiley, New York
35. Cott BJ, Macchietto S (1989) Temperature control of exothermic batch reactors using generic model control. Ind Eng Chem Res 28:1177–1184
36. Bentley WE, Davis RH, Kompala DS (1991) Dynamics of CAT expression in *E. coli*. Biotechnol Bioeng 38:749–760
37. Lee J, Ramirez WF (1992) Mathematical modeling of induced foreign protein production by recombinant bacteria. Biotechnol Bioeng 39:635–646
38. Lin ECC (1996) Dissimilatory pathways for sugars, polyols and carboxylates. In: Neidhardt FC (ed) *Escherichia coli* and Salmonella, 2nd edn, vol 1. ASM Press, Washington, DC, pp 307–342
39. Andrews JF (1968) A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. Biotechnol Bioeng 10:707–723
40. Bentley WE, Mirjalii N, Anderson DC, Davis R, Kompala DS (1990) Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol Bioeng 35:668–681
41. Glick BR (1995) Metabolic load and heterologous gene expression. Biotechnol Adv 13:247–261
42. Ko YF, Bentley WE, Weigand WA (1995) The effects of cellular energetics on foreign protein production. Appl Biochem Biotechnol 50:145–159