0273-2289/95/5152—0449$06.60

Optimization for a Recombinant *E. coli* Fed-Batch Fermentation

Q. CHEN,[1,3] W. E. BENTLEY,[1,2] AND W. A. WEIGAND*,[1]

[1]*Department of Chemical Engineering; and* [2]*Center for Agricultural Biotechnology, University of Maryland, College Park, MD 20742;* [3]*Current address: Cray Research, Eagan, MN 55121-1500*

ABSTRACT

The operating strategy that produces the maximum foreign protein expression for a fed-batch process is desired. This is achieved by using a feasible quadratic programming (FSQP) algorithm with a structured model that describes cell growth and product formation for recombinant *E. coli*. Optimization calculations for a fed-batch culture have not been performed with a model of this complexity up to this point. A constraint on the maximum cell concentration was included. For a fixed value of batch time, the results show that the optimal time profile of feed flow rate can increase the yield of foreign protein by 12–29% over a constant feed rate policy. Also, it was found that the computation time for the FSQP algorithm can be reduced significantly by considering suboptimal profiles of the feed rate, with a minor effect on calculated protein yield.

Index Entries: Optimization; recombinant *E. coli*; fed-batch reactor; FSQP.

INTRODUCTION

Protein production using recombinant *E. coli* may be enhanced through appropriate cultivation strategies and genetic manipulations. Development of such strategies requires careful consideration of the behavior of the recombinant system, which includes the protein synthesis activity of the host cells in response to environmental and cellular changes, the inhibitory effect of substrate level on cell growth, product accumulation,

*Author to whom all correspondence and reprint requests should be addressed.

host–plasmid interactions, and so forth. The objective of this work is to formulate and optimize a fed-batch fermentation process based on a structured model that describes protein expression by a recombinant strain of *E. coli*.

To design and operate a fed-batch fermenter, which commonly yields the highest product concentration as compared with batch or continuous reactors, it is important to determine the optimal feed rate policy to produce the maximum yield. Numerous articles have dealt with the determination of the optimal feed rate profiles for fed-batch fermentations, e.g., Weigand et al. (*1*) and Modak and Lim (*2*). For more complex fermentations, a numerical computation method is required to optimize the system. There are many numerical methods available to solve nonlinear optimal control problems. With Pontryagin's maximum principle, the optimal control problems are transformed into two boundary value problems, which can then be solved iteratively by control vector iteration (CVI), quasilinearization, or boundary condition iteration (BCI). These methods, however, are not well suited to problems with final state inequality constraints or for singular optimal problems. Optimal control problems can also be transformed into finite dimensional optimization problems by using parameterization schemes. Such schemes have been widely recognized for some time and allow the solution of the optimal control problem by the standard techniques of mathematical programming. There has been recent progress in solving the general optimal control problems, where any of the functions in the performance index, the system equations, or the final state constraints can be nonlinear. Cuthrell and Biegler (*3*) proposed a nonlinear programming (NLP) method for a fed-batch penicillin fermentation process by discretizing both the state and control variables.

If a structured model is used for fed-batch fermentation optimization, the determination of the optimal control policy can be quite difficult if methods based on Pontryagin's maximum principle are used. In this article, a nonlinear programming approach is used for optimization. In particular, a Feasible Sequential Quadratic Programming (FSQP) technique (*4*) is used for an optimization study of protein production. The computation time for this problem is very high. This is caused by the complexity of the structured model. Consequently, a simplified suboptimal profile is proposed to reduce the computation time for optimization.

DESCRIPTION OF THE PROCESS

A model describing the growth of recombinant strains of *E. coli* able to produce foreign protein has been developed by Bentley and Kompala (*5*). This structured model is based on a kinetic approach to describe the fermentation reactions, and it calculates the instantaneous growth rate. The equations describing the fed-batch fermentation are given by the following mass balances:

Biomass: $$(dX / dt) = \mu X - (F / V) X \quad (1)$$

where X is bulk biomass concentration (g dry wt/L), F is feed flow rate (L/h), V is fermentor volume (L), and μ is instantaneous specific growth rate (h^{-1}).

Substrate: $$(dS / dt) = -(\mu X / Y_{x/s}) + (F / V)(S_f - S) \quad (2)$$

where S is bulk substrate concentration (g/L), S_f is substrate concentration in the feed (g/L), and $Y_{x/s}$ is substrate yield based on glucose (g dry wt/g).

Foreign protein: $$(dP_f / dt) = \mu_4 [A / (K_{PfA} + A)] RG_f - K_{TP}P_f - \mu P_f) \quad (3)$$

where P_f is foreign protein mass fraction (g/g dry wt), A is amino acid mass fraction (g/g dry wt), G_f is DNA mass fraction (g/g dry wt), R is RNA mass fraction (g/g dry wt), μ_4 is kinetic constant for foreign protein synthesis (g dry wt/g·h), K_{PfA} is the saturation constant for amino acid (g/g dry wt), and K_{TP} is the kinetic constant for foreign protein turnover (h^{-1}).

Volume: $$(dV / dt) = F \quad (4)$$

The numerical values of the model parameters are listed in Table 2 in Appendix 2. The structured model is depicted in Appendices 1 and 2, and described in detail in the original reference (*5*). This metabolic model lumps the intracellular constituents of *E. coli* into eight different pools: protein, P, foreign protein, P_f, ribosome, R, chromosomal DNA, G, plasmid DNA, G_f, lipids, L, nucleotides, N, and amino acids, A. The relationship between these different pools and the method to obtain specific growth are described in detail by Bentley and Kompala (*5*).

OPTIMAL CONTROL PROBLEM OF FED-BATCH PROCESSES

The use of piecewise constant control segments to discretize the control policy is employed in this optimization problem. Piecewise constant control is simple to implement and has been widely used to solve nonlinear optimal control problems.

Let us consider the system described by the vector differential equation

$$(dx / dt) = f(x, u, p, t) \quad (5)$$

where the state vector is $x(t)$, the control vector is $u(t)$, and the parameter vector is p. The initial conditions are given by:

$$x(0) = x_0 \quad (6)$$

The optimal control problem is to find the control policy $u(t)$ and the parameter vector p that minimize the general scalar performance index.

$$Min\, J(p, x_f, t_f) \quad (7)$$

while taking the systems from a given initial state to the target region. Also, the control policy $u(t)$ and the parameter vector p are subject to constraints, which are expressed as simple upper and lower bounds:

$$u_j^L \leq u_j(t) \leq u_j^U \qquad j = 1, 2, \ldots, r \tag{8}$$

$$p_k^L \leq p_k \leq p_k^U, \qquad k = 1, 2, \ldots, m \tag{9}$$

MATHEMATICAL PROGRAMMING FORMULATION

To transform the continuous optimal control problem into a finite dimensional optimization problem, we shall use a piecewise constant control policy as was done by Tremblaya and Luus (*6*). First we divide the time interval into N sample intervals, and

$$\sum_{i=1}^{N} t_i = t_f \tag{10}$$

We define the input policy during these intervals as:

$$u(t) = c_i, \qquad t_{i-1} \leq t < t_i \tag{11}$$

It is clear that if N is chosen sufficiently large, the piecewise constant optimal control policy will be sufficiently close to the continuous optimal control policy.

The vector of unknown variables to be determined by using NLP is:

$$y^T = (c_1^T, c_2^T, \ldots c_p^T, p^T) \tag{12}$$

The resulting NLP problem for this optimization is to determine the y^T, which maximizes the total foreign protein in the reactor at the end of the fed-batch culture. In terms of the system variables, the performance index to be maximized is given by:

$$J = X(t_f)\, P_f(t_f)\, V(t_f) \tag{13}$$

where t_f is fixed. The system is controlled by manipulating the substrate feed flow rate, which is bounded by:

$$0 \leq u(t) \leq u_{max} \tag{14}$$

and the system is subjected to constraints on the volume:

$$V \leq V_m \tag{15}$$

and maximum cell mass concentration:

$$X \leq X_m \tag{16}$$

The FSQP algorithm (*4*) is used as a nonlinear programming technique for optimization of protein production. The FSQP algorithm is based on a software package, Sequential Quadratiac Programming (SQP), developed

at the University of Bayreuth, Germany, to generate a feasible path during the optimization search (*7*). FSQP is a set of FORTRAN subroutines for the minimization of the maximum of a set of smooth objective functions subject to nonlinear equality and inequality constraints, linear equality and inequality constraints, and simple bounds on the variables. If the initial guess for the vector of unknown variables provided by the user is not feasible for nonlinear inequality constraints and linear constraints, FSQP first generates a point satisfying all of these constraints by iterating on the problem of minimizing the maximum of the constraints. Then nonlinear equality constraints are turned into inequality constraints (*4*) and the original objective function is replaced by a modified objective function. Therefore, the resulting optimization problem only involves linear constraints and nonlinear inequality constraints. The user must provide subroutines that define the objective functions and constraint functions. For this transformed problem, FSQP implements algorithms that are described and analyzed by Panier and Tits (*8*), Bonnans et al. (*9*), and Zhou and Tits (*10*). These algorithms are based on an SQP iteration modified to generate feasible iterates. Details on FSQP can be found in a user's guide written by Zhou and Tits (*4*). More information can be obtained from Andre L. Tits of the Electrical Engineering Department at the University of Maryland at College Park.

OPTIMIZATION STUDIES

The main part of the computation time for a dynamic optimization problem is the time needed to solve the differential equations in order to evaluate the objective function and constraints. The differential equations were integrated using the 5/6 formula pair routine IVPAK, as implemented in the IMSL routine, with a local tolerance error of 10^{-6}. All the computations were carried out on a Sun workstation.

Because the structured model of this process is very stiff, this optimization problem is challenging and difficult. The average computation time for solving the differential equations for this model is about 25 min. Also, the computation time spent evaluating the gradient is a substantial part of the total computation time. If the number of sampling intervals is high enough to approximate closely the continuous control policy, the computation time becomes very large. Therefore, as a suboptimal way to solve this optimization problem, reducing the computation time by using smaller values of N is proposed.

As a first example, the total fed-batch time, t_f, is chosen as 50 h, the inlet substrate concentration, S_f, is 160 g/L, and the initial and final reactor volumes are V_0 = 1 L and V_{max} = 5 L, respectively. As a base case, we show the profiles produced by a constant feed rate of 0.08 L/h. Figure 1 shows the resulting constant feed rate profiles of substrate concentration

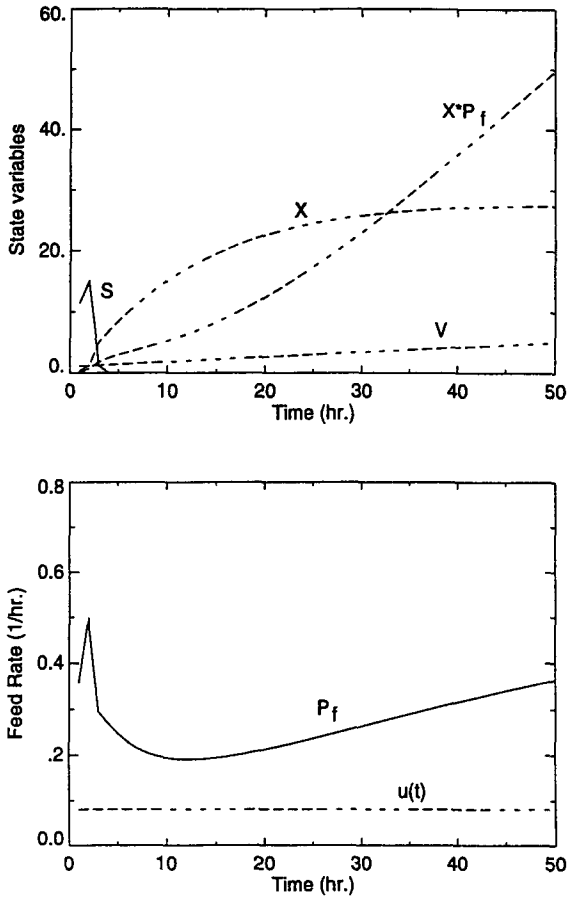

Fig. 1. The constant feeding substrate control profiles for a fed-batch recombinant fermentation.

(*S*), cell mass concentration (*X*), fermentor volume (*V*), and intracellular protein ($X * P_f * V$). The constant feed rate policy corresponds to $N = 1$. When the nonlinear programming approach to optimization is used with three piecewise segments ($N = 3$), we obtain the profiles shown in Fig. 2. The constant feed rate profiles in Fig. 1 are used as the initial values for the nonlinear programming optimization calculation. The results for $N = 3$ show that the final intracellular mass of protein reaches 55.7 g, an increase of 11.5% over the constant feed rate value of 49.2 g. In Fig. 3, we show the profiles corresponding to $N = 2$. In this case, we have a more suboptimal solution, with final protein of 54.9 g. However, the computation time is significantly lower. The effect of the sampling interval, *N*, on the optimization results and computation time are shown in Table 1. As expected, an increased number of sampling intervals produces a higher value of protein yield, but at the cost of a significant increase in computation time. In terms of bioreactor performance, the optimal solution derived at by this technique balances the requirement for high cell mass produc-

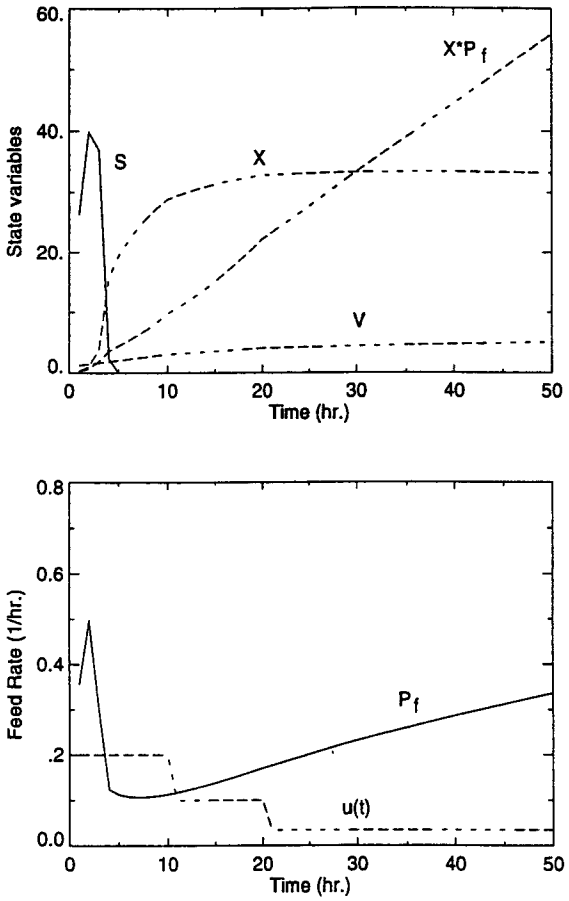

Fig. 2. The optimal piecewise constant control profiles for a fed-batch recombinant fermentation with $N = 3$ and $t_f = 50$ h.

tion with high specific protein yield. High cell production is achieved by maintaining high cell growth along with a high level of total substrate consumption, both of which are favored by a high volumetric flow rate of substrate. High specific protein yield, however, is obtained at low growth rates. Previously, Bentley and coworkers demonstrated that for constitutive promoter expression systems, the higher specific yields of foreign protein were obtained at lower cell growth rate (*11*). Among several explanations that were presented, perhaps the most important is simply the gene dosage effect: higher plasmid copy number produces higher protein yield on a per-cell basis. The plasmid replication mechanism is essentially growth-rate-independent in the range of this work (*12*), so that at low growth rates (or long doubling times), there is more time for plasmid replication and, hence, higher copy number growing per cell. Thus, in cases where the plant is constrained by reactor time (i.e., multiple products run in few reactors), the optimization scheme demonstrated here is quite important.

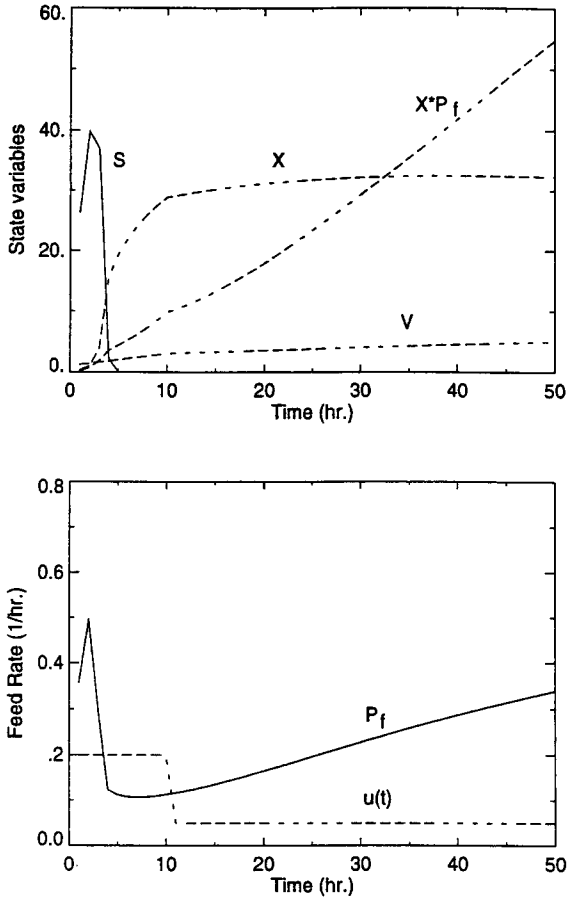

Fig. 3. The optimal piecewise constant control profiles for a fed-batch recombinant fermentation with $N = 2$ and $t_f = 50$ h.

Table 1
The Optimization Results with the Effect of the Sample Interval

Sampling interval, N	Optimization performance	Number of function evaluations	Average computation time, h
1	49.21	1	0.41
2	54.92	72	30.00
3	55.70	125	52.08
8	55.92	1290	527.5

When comparing the constant feed rate case (Fig. 1) with the suboptimal feeding profiles (Figs. 2 and 3), we see that the major effect of the optimal profiles is to increase the cell mass formed. That is, the fraction of foreign protein in the cell is actually decreased in Figs. 2 and 3 (≈ 0.33) when compared to Fig. 1 (≈ 0.37), whereas the increase in cell mass in Fig. 2 (≈ 33 g/L) and Fig. 3 (≈ 32 g/L), when compared to Fig. 1 (≈ 27.5),

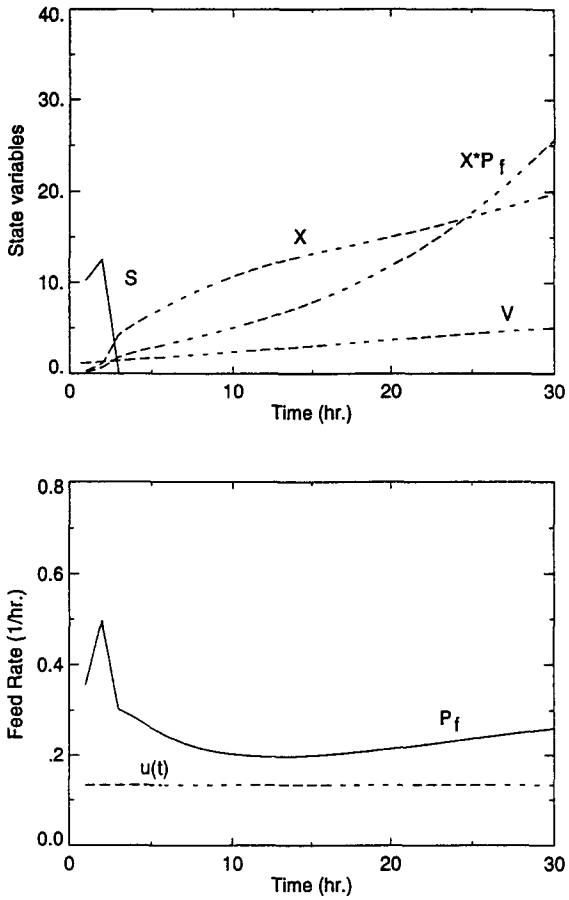

Fig. 4. The constant feeding substrate control profiles for a fed-batch recombinant fermentation with t_f = 30 h.

causes the total protein to increase, that is 55.7 g in Fig. 2 and 54.9 g in Fig. 3 vs 49.9 g in Fig. 1.

Next, the effect of changing the fermentation time from 50 to 30 h is examined. The constant feed rate profiles are shown in Fig. 4, whereas the profiles for two piecewise segments are shown in Fig. 5. When the total fermentation is decreased to 30 h, the effect of even the simplest optimal profile (N = 2) is to increase the protein yield by 28.8%, that is, from 26.2 to 33.5 g. The results indicate the need to consider the benefit of high specific productivity at the onset of the fed-batch part of the fermentation that is obtained by a drop in the feed rate.

DISCUSSION

A structured kinetic model has been used to simulate the fed-batch fermentation system for recombinant protein synthesis by *E. coli*. This

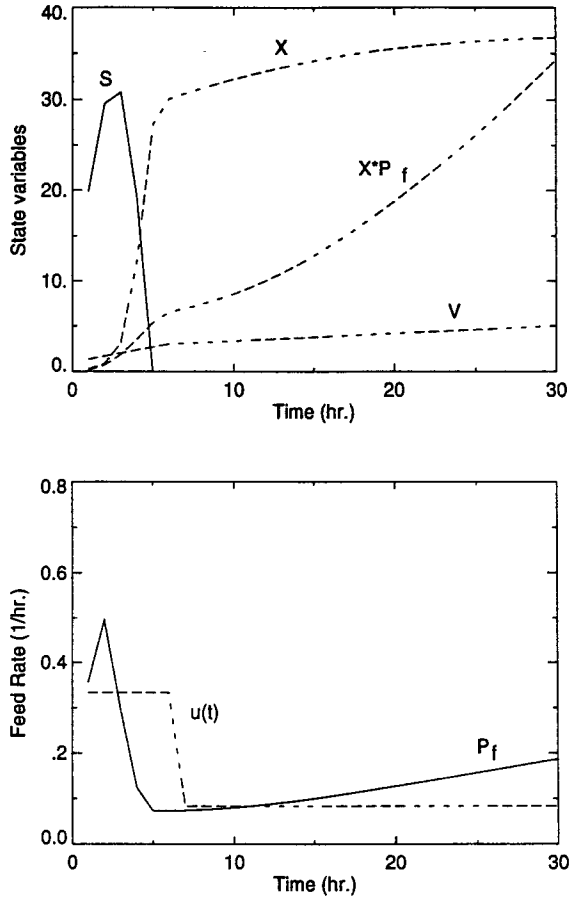

Fig. 5. The optimal piecewise constant control profiles for a fed-batch recombinant fermentation with $N = 2$ and $t_f = 30$ h.

structured model is used to calculate the specific growth rate and the fraction of protein in the cell, and it contains eight state variables. When this model is combined with the cell mass and volume balances for the fed-batch reactor, we have a system with 11 state variables. In addition, fed-batch fermentations are subject to final state inequality constraints, and singular optimal inputs when the optimal feed rate profile is derived. This type of problem is difficult to solve when using numerical methods based on Portryagin's maximum principle. As an alternative, this optimal control problem was transformed into a finite element optimization problem that can be solved by nonlinear programming. In particular, we used the FSQP method. Nonlinear programming methods have been used before (3), but not for a system with a complex structured model or with this high number of state variables. However, because of the complicated kinetic model used, the computation time required with feasible sequential quadratic programming was very high. A major part of the computation time is used for function evaluations for the gradient calculation.

Therefore, a suboptimal approach that decreases the number of sampling intervals was used.

The results (Table 1) indicate that the suboptimal protein yield is only slightly decreased from optimization with higher N, but the computation time requirement was significantly decreased. These optimized profiles show protein yield increased by 11.5 and 28.8% for fed-batch culture times of 50 and 30 h, respectively. A more detailed examination of this process will require the inhibitory effect of acetate level on cell growth and protein synthesis to be considered. This, in turn, will require a more complex model and an expanded number of state variables to describe the acetate synthesis. The results shown here indicate that optimization with structured models that more closely describe the fermentation system is possible, if a finite element optimization method is employed. Since structured models better describe the fermentation system, when compared to the unstructured models used in prior optimization studies, the methods presented here could lead to the application of mathematical culture optimization methods to actual fermentation processes.

REFERENCES

1. Weigand, W. A., Lim, H. C., Creagan, C., and Mohler, R. (1979), *Biotechol. Bioeng. Symp. Ser.* **9,** 335–348.
2. Modak, J. M.. and Lim, H. C. (1984), *Biotechnol. Bioeng.* **30,** 528–540.
3. Cuthrell, J. E. and Biegler, L. T. (1987), *AICHE J.* **33,** 1257.
4. Zhou, T. J. and Tits, A. L. (1993), *User's Guide for FSQP Version 3.3,* University of Maryland.
5. Bentley, W. E. and Kompala, D. S. (1989), *Biotechnol. Bioeng.* **33,** 49–61.
6. Tremblaya, M. D. and Luus, R. (1989), *Can. J. Chem. Eng.* **67,** 494–502.
7. Schittkowski, K. (1986), *QLD: A FORTRAN Code for Quadratic Programming, User's Guide,* Mathematisches Institut, Universitat Bayreuth, Germany.
8. Panier, E. R. and Tits, A. L. (1993), *Math. Programming* **59,** 261–276.
9. Bonnans, J. F., Panier, E. R., Tits, A. L., and Zhou, J. L. (1992), *SIAM J. Numer. Anal.* **29,** 1187–1202.
10. Zhou, J. L. and Tits, A. L. (1993), *J. Optim. Theory Appl.* **76,** 455–476.
11. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., and Kompala, D. S. (1990), *Biotechnol. Bioeng.* **35,** 668–681.
12. Bentley, W. E. and Quiroga, O. E. (1993), *Biotechnol. Bioeng.* **42,** 222–234.

APPENDIX 1
Dynamic Equations for Constituent Pools

$$(dA / dt) = [(dA / dt)]_s - \epsilon_1 [(dN / dt)]_s - \epsilon_2 [(dL / dt)]_s - \gamma_1 [dP / dt)]_s - \gamma_1 [(dP_f / dt)]_s - \mu A \quad (17)$$

$$(dN / dt) = [(dN / dt)]_s - \gamma_2 [(dG / dt)]_s - \gamma_2 [(dG_f / dt)_s - \gamma_2 [(dR / dt)]_s - \mu N \quad (18)$$

$$(dP / dt) = [(dP / dt)]_s - \mu P \quad (19)$$

$$(dP_f / dt) = [(dP_f / dt)]_s - \mu P_f \quad (20)$$

$$(dG / dt) = [(dG / dt)]_s - \mu G \quad (21)$$

$$(dG_f / dt) = [(dG_f / dt)]_s - \mu G_f \quad (22)$$

$$(dL / dt) = [(dL / dt)]_s - \mu L \quad (23)$$

$$(dR / dt) = [(dR / dt)]_s - \mu R \quad (24)$$

CALCULATION OF GROWTH RATE

$$\mu = (dA / dt)_s + (1 - \epsilon_1)(dN / dt)_s + (1 - \epsilon_2)(dL / dt)_s + (1 - \gamma_1) [(dP / dt)_s + (dP_f / dt)_s] + (1 - \gamma_2)(dG / dt)_s + (dG_f / dt)_s + (dR / dt)_s] \quad (25)$$

APPENDIX 2
Synthesis Rate Expressions for Constituent Pools

$$(dA / dt)_s = k_1 [K_A / (K_A + A)] [S / (K_{AS} + S)] [A / (K_{2A} + A)] \quad (26)$$

$$(dN / dt)_s = k_2 [K_N / (K_N + N)] [A / (K_{NA} + A)] [S / (K_{NS} + S)] \quad (27)$$

$$(dP / dt)_s = \mu_1 [A / (K_{PA} + A)] RG - K_{TP} P \quad (28)$$

$$(dP_f / dt)_s = \mu_4 [A / (K_{P_f A} + A)] RG_f - K_{TP} P_f \quad (29)$$

$$(dG / dt)_s = \mu_2 [N / (K_{GN} + N)] \quad (30)$$

$$(dG_f / dt)_s = \mu_5 [N / (K_{G_f N} + N)] \quad (31)$$

$$(dL / dt)_s = \mu_3 [S / (K_{LS} + S)] [A / (K_{LA} + A)] \quad (32)$$

$$(dR / dt)_s = \mu_6 [N / (K_{RN} + N)] [A / (K_{RA} + A)] G - K_{TR} R - K'_{TR} R [K_{TR_S} / (K_{TR_S} + S)] \quad (33)$$

Table 2
Stoichiometric, Maximum Rate, and Saturation Constants for Amino Acids, Nucleotides, Protein, DNA, Fatty Acids and Lipids, Ribosomal RNA, Foreign Protein, and DNA

Amino acids			
k_1	1.69 h^{-1}	K_{AS}	0.01 g/L
ϵ_1	0.5485 (*g A / g N*)	K_A	0.125
ϵ_2	0.1418 (*g A / g F*)	K_{2A}	0.001
γ_1	1.167 (*g A / g P*)		
Nucleotides			
k_2	1.19 h^{-1}	K_{NS}	0.01 g/L
γ_2	1.056 (*g N / g G*)	K_{NA}	0.026
K_N	0.125		
Protein			
μ_1	154.4 (*gm / ggRgGhr*)	K_{TP}	0.08 h^{-1}
K_{PA}	0.002		
DNA			
μ_2	0.078 h^{-1}	K_{GN}	0.01
Fatty acids and lipids			
μ_3	0.52 h^{-1}	K_{LA}	0.026
K_{LS}	0.01 g/L		
rRNA			
μ_6	19.64 (*g / ggGhr*)	K_{TP_S}	0.01 g/L
K_{RN}	0.026	K_{TR}	0.179 h^{-1}
K_{TR}	0.147 h^{-1}	K_{RA}	0.001
Foreign protein			
μ_4	800 h^{-1}	$K_{P_{fA}}$	0.002
Foreign DNA, plasmids			
μ_5	0.0005 h^{-1}	$K_{G_{fN}}$	1×10^{-9}