Fed-batch optimization of recombinant α-amylase production by *Bacillus subtilis* using a modified Markov chain Monte Carlo technique

Wanwisa Skolpap*,†, Somboon Nuchprayoon**, Jeno M. Scharer***, Nurak Grisdanurak*, Peter L. Douglas***, and Murray Moo-Young***

*Department of Chemical Engineering, Thammasat University, Pathumthani, 12120 Thailand
**Department of Electrical Engineering, Chiang Mai University, Chiang Mai, 50200 Thailand
***Department of Chemical Engineering, University of Waterloo, Waterloo, N2L 3G1 Canada
(*Received 25 October 2007 • accepted 16 January 2008*)

Abstract−A modified Markov Chain Monte Carlo (MCMC) searching procedure was developed to search for an optimal set of decision variables and optimal feed rate trajectories for recombinant α-amylase expression by *Bacillus subtilis* ATCC 6051a. The bacterium also synthesizes proteases as undesirable products in fed-batch culture that need to be minimized. To maximize α-amylase productivity, a 14th-order fed-batch model was optimized by integrating Pontryagin's maximum principle with the *Luedeking-Piret* equation. The number of iterations and simulations of the proposed searching procedure were statistically examined for accuracy and acceptability of the results. It can be concluded that the proposed searching procedure increased the parameter selection opportunity near the tail ends of redefined triangular distribution. By applying a modified MCMC searching procedure with 1,500 iterations, the predicted α-amylase productivity was improved by 18% in comparison with near-optimum experimental results. This productivity was 3.5% higher than predicted by conventional MCMC optimization.

Key words: *Bacillus subtilis*, Fed-batch Culture, Nonlinear Optimization Problem, Markov Chain Monte Carlo

INTRODUCTION

Fed-batch culture offers advantages over batch processes such as minimized problems pertaining to substrate inhibition and/or catabolite repression. However, optimization of fed-batch processes is not an easy problem. The challenge of dynamic fed-batch optimization often involves the resolution of high-order, nonlinear and multimodal systems. In previous efforts, both stochastic and deterministic search techniques have been applied to maximize productivity. Evolutionary algorithms, such as genetic algorithms mimicking the principles of natural biological evolution [1-4], or ant colony algorithm mimicking food search of ants [5], have also been applied to solve for optimal feed-rate profiles. In addition to those population-based search techniques, many point-based search techniques have been also applied to determine optimal feed-rate profiles. For instance, Patkar et al. [6] successfully applied a first-order conjugate gradient algorithm for optimizing fed-batch fermentations. Cuthrell and Biegler [7] employed an orthogonal collocation-based sequential quadratic programming to optimize fed-batch culture for penicillin production (containing four state variables) that was first studied by Lim et al. [8]. Mekarapiruk and Luus [9] applied iterative dynamic programming with unspecified initial conditions to optimize feed-rate policy for producing penicillin. Their study evaluated the feed-rate profile together with the initial volume and initial substrate concentration. Three deterministic candidates given for the values of the decision variables with a fixed final time at 132 hours and equally step size were obtained from their previous work [10]. Markov Chain Monte Carlo (MCMC) procedures, the Gibbs parameter sampling and the Metropolis-Hasting algorithm, have been recently used to

estimate model parameters and decision variables in 14th-order equations for α-amylase and protease producing *Bacillus subtilis* in fed-batch culture [11] and to estimate a set of decision variables for penicillin fed-batch optimization problem by using a set of initial values given by Mekarapiruk and Luus [9,12].

In fed-batch optimization, the parameters to be optimized are embedded in a set of nonlinear differential equations. The problem observed during previous applications of the MCMC optimization technique was that parameter selection for optimal solution was often near the mean parameter value of the prior distribution. To properly select parameter values near the tail ends of the parameter distribution and also solve a set of nonlinear differential equations, excessive computing time was required. In this work, selection opportunity could be enhanced and computing time could be reduced by redefining the parameter distribution after a finite number of trials. As a result, the optimal feed-rate profile for the dual-enzyme fed-batch culture of *B. subtilis* [11] was quickly obtained and was comparable to the results obtained by using conventional searching procedure.

THE FED-BATCH MODEL

The fed-batch model represents the dynamic optimization system for α-amylase synthesis under nitrogen-limited growth of *Bacillus subtilis* ATCC 6051a [11]. Undesirable proteases were hypothesized to cause the degradation of α-amylases. The model includes mass balances of cell growth, substrate uptake, respiration, target products formation as well as byproduct (ethanol, acetate, and lactate) channeling into oxidative and overflow metabolism. Table 1 shows a list of 13 nonlinear differential equations used in this work, while the model equations are shown in Appendix A. The differential equation for the volume change as the function of the feed rate is as follows:

†To whom correspondence should be addressed.
E-mail: swanwisa@engr.tu.ac.th

Table 1. List of 13 nonlinear differential equations used in fed-batch model (see Appendix A for details)

Equation number	Description
A.1	Biomass concentration
A.2	Ammonia concentration
A.3	α-Amylase concentration
A.4	α-Amylase activity
A.5	Protease concentration
A.6	Protease activity
A.7	Ethanol concentration
A.8	Lactate concentration
A.9	Acetate concentration
A.10	Citrate concentration
A.11	Starch concentration
A.12	Isoleucine concentration
A.13	Threonine concentration

$$u = \frac{dV}{dt}. \tag{1}$$

where u is substrate feed rate and V is culture volume.

The time-dependent value of feed rate in Eq. (1) was simply derived from the proposed optimization algorithm.

It should be emphasized that α-amylase synthesis was proven to be partially growth associated and the relationship between enzyme production and growth was modeled successfully by the well-known *Luedeking-Piret* equation [13]. Due to proteolytic effect on α-amylase, the conversion from α-amylase concentration to activity was necessary for simulations. The time-averaged productivity of α-amylase was maximized by integration of Pontryagin's maximum principle and comparison with observed data, i.e., relationship between modeled product and cell growth and experimental results [11]. The set of nonlinear differential equations was solved by using the orthogonal collocation method encoded in MATLAB™ (the Math-Works, Inc., Natick, MA) by Constantinides and Mostoufi [14].

Prior to optimization the fed batch model needed to be validated by using the experimental results [11,15]. The goodness of fit was evaluated non-parametrically by employing χ^2 statistical method [16]. It signifies the quality of agreement between predicted and observed results. The χ^2 statistic is calculated as follows:

$$\chi^2 = \sum_{q}^{Q} \frac{(O_q - E_q)^2}{E_q} \tag{2}$$

where O_q is an observation in the qth row of the same column, E_q is an expectation in the qth row of the same column, and Q is a number of data points in any column.

$$df = Q - 1 \tag{3}$$

where df is a degree of freedom.

The five state variables for statistical analysis included the biomass concentration (y_1), ammonia concentration (y_2), α-amylase activity (y_4), protease activity (y_6), and starch concentration (y_{11}). The results are summarized in Table 2. According to χ^2 statistic, the observed results were comparable to the predicted results at 95% probability level. In other words, the modeled results show a good agreement with the observations.

THE OPTIMIZATION PROBLEM

The dynamic nonlinear optimization problem consists of the fed-batch model equations listed in Table 1, the objective function stated in Eq. (4), and the control variable (feed rate) given in Eq. (5).

$$\underset{u, K, t_s}{\text{Maximize}} \left(\frac{PV}{T} \right) \tag{4}$$

where P is desirable product (α-amylase) activity, V is total culture volume, T is total fermentation time period, u is feed rate of substrate, K is a specific productivity parameter, and t_s is switching time from batch to fed-batch mode.

The feed rate, u(t) is parameterized as the following:

$$u = \begin{cases} u_0 e^{K(t - t_s)} & \text{if } t > t_s \\ 0 & \text{if } t \le t_s \end{cases} \tag{5}$$

where u_0 is initial feed rate at batch to fed-batch switching time, t_s. Therefore, the decision variables are essentially initial feed rate (u_0), switching time (t_s), and the flow parameter (K).

Due to physical limitation of the pump speed used in the fed-batch experiments, the feed rate and volume are bounded as follows:

$$\begin{aligned} &0.061 \le u \le 0.15 \text{ L/h,} \\ &0.01 \le K \le 0.09, \\ &11.0 \le t_s \le 18.0 \text{ h,} \\ &4.838 \le V \le 6.038 \text{ L.} \end{aligned} \tag{6}$$

The flow parameters, the exponential constant term expressed in Eq. (5), were limited by the maximum feed rate of substrate.

PARAMETER SELECTION USING MCMC METHOD

In this work, the set of differential equations was solved simultaneously by using the orthogonal collocation method. The model parameters were estimated by using the Gibbs parameter sampling approach [17,18], and their values are summarized in Appendix B except for $Y_{X/Cit}$ because of citrate depletion prior to fed-batch fermentation.

The optimum values of the three decision variables (u_0, t_s, and K) were determined by using the Metropolis-Hastings algorithm [18]. The MCMC method is based on a number of samplings to

Table 2. χ^2 statistical test of fed-batch experiment

Parameter	Biomass conc. (g/L)	Ammonia conc. (g/L)	α-Amylase activity (U/mL)	Protease activity (U/mL)	Starch conc. (g/L)
Probability for a χ^2 statistic	0.6711	4.9247	3.8784	0.8446	0.0934
Degree of freedom	5	5	6	6	5
Critical χ^2 value at 95% probability	11.0705	11.0705	12.5916	12.5916	11.0705

find the more probable and posterior distribution of each parameter. To initiate the sampling procedure, an equilateral triangular distribution was assumed for each parameter with respect to specified bounds. Let u_0, t_s, and K be the parameters θ_1, θ_2, and θ_3, respectively. For any parameter, the mean (most likely) value (θ^{mean}) of the prior distribution was primarily estimated from its pre-specified range. The distance between the maximum (or minimum) value and the mean value of the ith parameter (δ_i) is defined as:

$$\delta_i = (\theta_i^{max} - \theta_i^{min})/2, \tag{7}$$

where θ_i^{max} and θ_i^{min} are the maximum and minimum values of parameter i, respectively.

The parameter distribution space was sampled by a Monte Carlo draw. At each iteration j, the ith parameter ($\theta_i \geq 0$), is generated by:

$$^j\theta_i = E(\theta_i) + {}^j\delta_i(r_1 + r_2 - 1), \tag{8}$$

where $E(\theta_i)$ is the expected mean value of the ith parameter determined from previous (or preliminary) estimation, r_1 and r_2 are random numbers in the range of $[-1, 1]$.

The optimum values of the three parameters are then achieved by minimizing the sum of squared differences between the objective function value obtained by substituting sampled parameters and its *benchmark* (or *baseline*) value.

To evaluate the assigned parameter distribution, the maximum theoretical specific productivity was estimated by channeling all carbon flow into biomass, α-amylase, and CO_2 production only in the metabolic flux model. This benchmark productivity, considered as the objective function value, was given as $J^{max}=9386.565$ U$_{Amy}$/h [15]. The conditional probability of θ_i can then be calculated after this estimation. By using a Bayesian approach, the conditional probability of θ_i at each trial j is as follows:

$$\Pr(\theta_i|J) \propto \left\{ \frac{(J^{max} - J|(\theta_i, t_k))^2}{s^2} \right\}^{-1}, \tag{9}$$

$$s^2 = \frac{\sum_{k=1}^{j-1}(J^{max} - J(t_k))^2}{j-1}, \tag{10}$$

where J is the objective function value, s^2 is the variance of the difference between J^{max} and J calculated from the previous iteration $(j-1)$, t_k is time when the observation k is taken.

The parameter sampling procedure was repeatedly performed for a number of times to obtain the most likely set of parameters at each trial. The probability of acceptance of a sampled value at each trial $\alpha(^j\theta_i, {}^{j-1}\theta_i)$ was estimated as follows [18]:

$$\alpha(^j\theta_i, {}^{j-1}\theta_i) = \text{Min} \left\{ 1, \frac{\pi(^{j-1}\theta_i)q(^j\theta_i|^{j-1}\theta_i)}{\pi(^j\theta_i)q(^{j-1}\theta_i|^j\theta_i)} \right\}, \tag{11}$$

where π is stationary distribution of the parameter, q is proposal distribution of the parameter, $^j\theta_i$ is the currently sampled value of parameter, and $^{j-1}\theta_i$ is the proposed value of parameter after $j-1$ iterations. The stationary distribution of the parameter, $\pi(^{j-1}\theta_i)$, was considered proportional to the inverse of the squared difference between the calculated and the benchmark productivities.

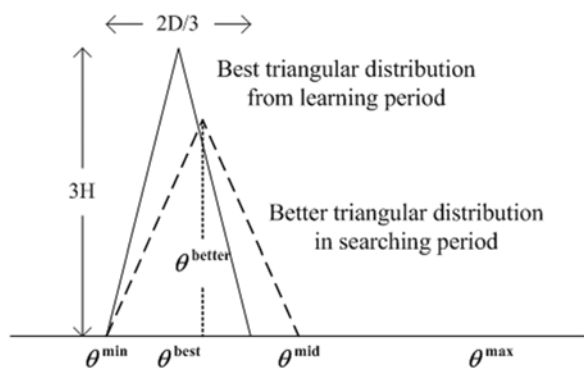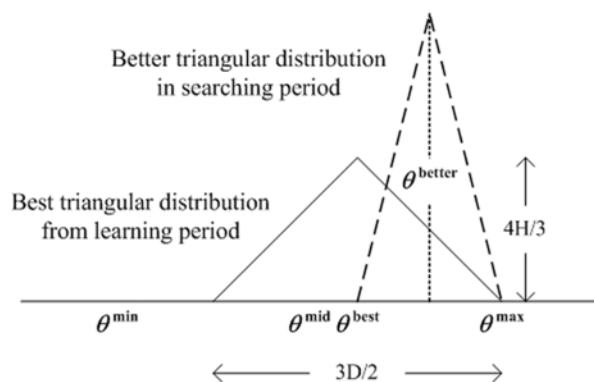**PROPOSED PARAMETER SAMPLING PROCEDURE**

Previously, the optimization algorithm using MCMC technique as sampling procedure found parameter values near the assigned mean values. The proposed MCMC sampling procedure is a modification of conventional procedures by redefining the mean of the initially assigned equilateral triangular distribution while preserving the triangular area of unity. As a result, the selection opportunity of parameters located in the newly assigned triangular distribution can be improved. In addition, an adaptive parameter sampling strategy is proposed to (1) reduce random walk behavior of the Gibbs parameter sampling procedure, (2) speed up convergence of the sampling method, and (3) ensure that optimal values are found by the sampling procedure. This is an improvement in strategy over previous procedure [15] when optimal values of parameters were found near either the upper or lower limit of the parameter sampling range. By using a single triangular distribution, it required a long time to achieve optimal parameter values.

Given a range of parameters and certain number of iterations (trials), the sampling periods may be divided into *learning* and *searching* periods. During the learning period the parameter space was sampled by using multiple triangular distributions to find the best of the proposed distributions. The triangular distribution obtained during the learning period was then sampled during the searching period to seek the optimal parameter value within the distribution space. In both periods, the parameters were sampled by a systematic approach. It is normally problem-dependent to determine the proper number of triangular distributions and the number of sampling iterations in the two periods. Generally speaking, too many triangular distributions would result in excessive costs of learning. If the number of samples during the learning period is too many or too few, it would limit the opportunity to achieve the optimal parameter value during the searching period. The procedure is illustrated in Fig. 1. The sampling periods were divided into two halves. As shown in Fig. 1(a), there is four triangular distribution in the learning period (first half). Each triangular distribution was sampled with equally probability, but the parameter value of the sampled triangular distribution was selected randomly in each trial. If, for instance, 80 trials were planned, each period would have 40 trials and each triangular distribution in the learning period would be sampled for 10 trials. It can be seen that each triangular distribution has an area of unity (DH=1) and its dimension can be derived from the maximum and minimum values of that parameter.

During the learning period multiple triangular distributions were provided for each sampled parameter. The acceptance probability of a selected parameter value was based on Eqs. (9)-(11). The objective function value corresponding to each set of sampled parameter values was then calculated and compared with the benchmark productivity. The benchmark productivity was evaluated from the maximum theoretical value of the specific productivity. The optimum set of parameter values was obtained from the parameter set that minimized the sum of squared differences between the calculated value of the objective function and the benchmark productivity. Thus, during the learning period the corresponding triangular distribution and sampled parameter value were considered as the "best location" and "best value" of that parameter, respectively. Once the learning period was completed, the best set of parameter values were identified (θ^{best}) and the corresponding triangular distributions were defined. The triangular area was again made equal to unity. Then, the

$\theta^{mid} = (\theta^{max} + \theta^{min})/2 \qquad DH = 1.0$
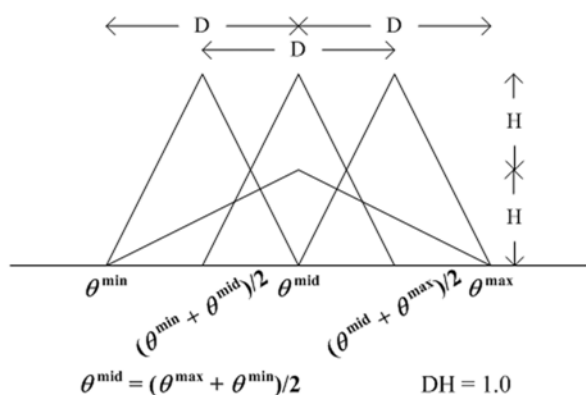
(a) Four triangular distributions to be chosen in learning period

(b) Best triangular distribution from learning period is on the left-hand side of θ^{mid}

(c) Best triangular distribution from learning period is on the right-hand side of θ^{mid}

Fig. 1. Modified MCMC parameter sampling procedure.

Table 3. Comparison of objective function values obtained from using conventional and proposed sampling procedures when switching time is at 14 hours

No. of iterations	Conventional samplings		Proposed samplings	
	Optimal value (U/h)	Standard deviation (U/h)	Optimal value (U/h)	Standard deviation (U/h)
125	6187.26	179.03	6212.19	56.13
250	6187.62	130.72	6210.45	49.26
500	6187.93	93.88	6212.60	51.49
1,000	6189.02	66.92	6213.11	52.58

Table 4. Optimization results using proposed sampling procedure when switching time is at 14 hours

Repeat	No. of iterations	Optimal value		
		u_0 (L/h)	K	J (U/h)
1	125	0.0100	0.0668	6324.6
2	125	0.0140	0.0616	6290.1
3	125	0.0140	0.0614	6290.4
4	125	0.0102	0.0642	6321.4
5	125	0.0149	0.0610	6283.1
1	250	0.0147	0.0612	6285.3
2	250	0.0110	0.0662	6321.6
3	250	0.0146	0.0613	6287.1
4	250	0.0101	0.0668	6325.7
5	250	0.0111	0.0637	6315.5
1	500	0.0102	0.0667	6326.4
2	500	0.0101	0.0643	6322.6
3	500	0.0106	0.0666	6335.7
4	500	0.0103	0.0666	6323.3
5	500	0.0100	0.0644	6323.1
1	1,000	0.0101	0.0668	6327.2
2	1,000	0.0101	0.0668	6327.2
3	1,000	0.0104	0.0666	6324.4
4	1,000	0.0101	0.0668	6327.2
5	1,000	0.0104	0.0666	6324.4

best value was taken as the median value of a new triangular distribution for the searching period. The "distance" of this new triangular distribution, previously defined in Eq. (7), was calculated as the distance between the maximum (or minimum) value and the best value. For the ith parameter (δ_i^{best}), the distance may be set as follows:

$$\delta_i^{best} = \text{Min}\{ \theta_i^{best} - \theta_i^{min}, \theta_i^{max} - \theta_i^{best}\}. \qquad (12)$$

In the searching period (second half), the parameter value was initially selected from the best location of the learning period, which becomes a proposal distribution of that trial. If the objective func-

tion value is better, a new triangular distribution is defined. Otherwise, the parameter value for the next trial is selected as the previous "best" location. For example, as shown in Fig. 1(b), suppose the best location was on the left-hand side of the median value of a parameter, a parameter value is selected (θ^{better}) and the objective function value is found to be better. Consequently, a triangular distribution is redefined. Similarly, Fig. 1(c) illustrates a new triangular distribution when the best location is on the right-hand side of the median value of a parameter.

To compare the difference between conventional and proposed sampling procedures, a simplified fed-batch problem was simulated, and the results are shown in Table 3. To avoid excessive-computational time, the set of decision variables was only u_0 and K, while t_s was set at its previous optimum of 14 hours due to its relatively high sensitivity. The number of iterations was set at 100, 250, 500, and 1,000 iterations. The conventional sampling procedure is based on

a single triangular distribution, while the proposed sampling procedure is based on multiple (four in the present case) triangular distributions. It was found that, by using a single triangular distribution, the maximum objective function value was reached only 6,185.02 U/h after running for 1,000 iterations. The range of objective function value was 661.40 U/h. When four triangular distributions were applied, the maximum objective function value was 6,212.19 U/h after running for just 100 iterations and 6,213.11 U/h after running for 1,000 iterations. The range of objective function value was 204.83 U/h after running for 1,000 iterations. These results clearly indicate that multiple triangular distributions could result in faster convergence and yield better parameter values.

Prior to solving the optimization problem, the proper number of iterations for the proposed procedure was determined by also using the simplified fed-batch problem. The simulations were carried out on an AMD Athlon 64 Processor 3000+, 1.81 GHz and 512 MB RAM using MATLAB® version 7.04 (the MathWorks, Inc., Natick, MA), and the results are in Table 4.

The certainty of simulations was investigated by varying the number of iterations and repeating simulations. As mentioned previously, the optimal values fluctuated when the number of iterations was small. With five repeat simulations, the results suggested that the number of iterations had more impact on the consistency of the results than the number of repeats.

To achieve meaningful simulated results using the modified MCMC method, the relative errors of the mean values of the objective function were calculated by using the Student t distribution. The relative errors were 2.9%, 2.1%, 1.5%, and 1.0% after 125, 250, 500, and 1,000 iterations, respectively. Thus, stationary results were obtained by performing at least 1,000 iterations per simulation run.

OPTIMIZATION RESULTS

Table 5 shows the first five best objective function values obtained from 1,000 and 1,500 iterations. The trends of the objective function values, estimated from two optimal sets of decision variables, were identical. The relative errors of the mean objective function values at 1,500 iterations were in the range of 0.36% to 0.69%. After 1,500 iterations, there was a slight improvement in the objective function values (up to 0.12%) in comparison to those after 1,000

iterations. After 1,500 iterations, the optimal decision variables yielding the optimal feed-rate profile were u_0=0.0648 L/h, K=0.0112, and t_s=14.75 h (Fig. 2). This optimal feed-rate profile is comparable to the profiles reported earlier [11]; however, the initial feed rates were slightly different. Feed rate profiles achieved by using the modified MCMC and Skolpap et al. [11] were stopped at 35 and 34 hours
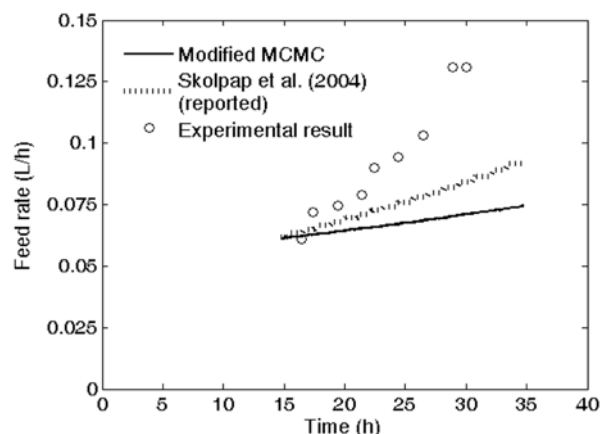

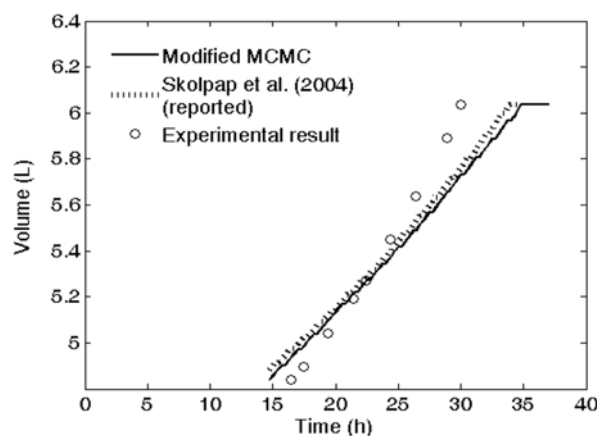Fig. 2. Comparison of optimal feed-rate profiles.
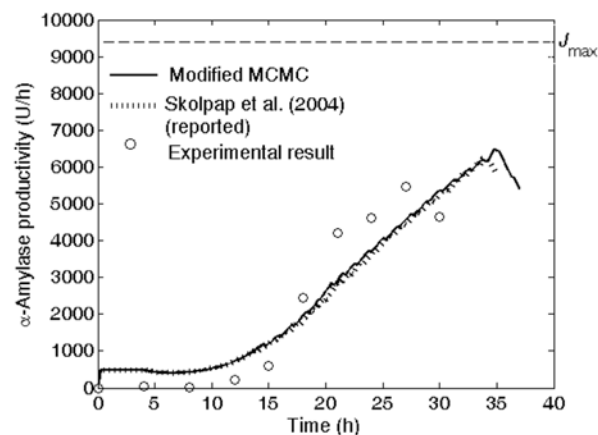
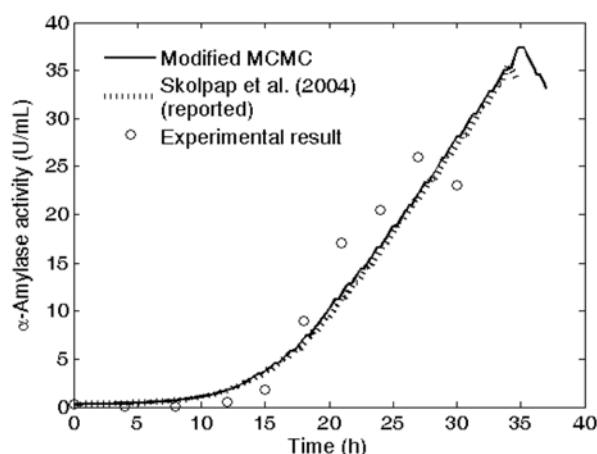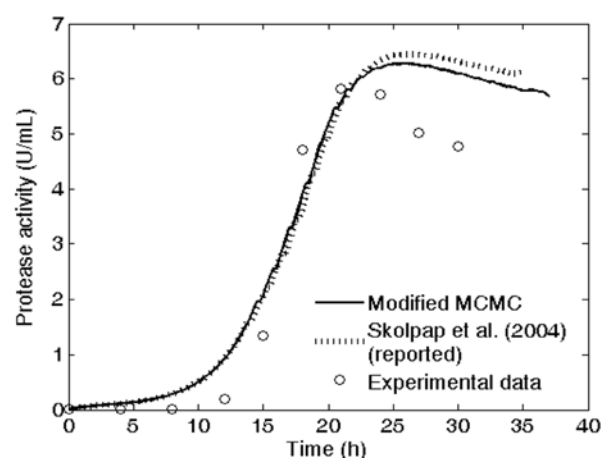

Fig. 3. Comparison of culture volume profiles.

Fig. 4. Comparison of α-amylase productivities under optimal feed-rate profiles

Table 5. Comparison of optimization results at 1,000 and 1,500 iterations

Repeat	No. of iterations	Elapsed time (sec)	u_0 (L/h)	K	t_s (h)	J (U/h)
1	1,000	7,980.41	0.0647	0.0113	14.75	6,482.2
2	1,000	7,667.74	0.0688	0.0111	15.75	6,468.7
3	1,000	8,197.38	0.0648	0.0101	14.50	6,459.1
4	1,000	8,495.84	0.0628	0.0100	13.25	6,446.9
5	1,000	7,507.69	0.0701	0.0101	16.00	6,431.9
1	1,500	11,939.20	0.0648	0.0112	14.75	6,483.8
2	1,500	11,370.55	0.0695	0.0101	15.75	6,476.6
3	1,500	12,150.91	0.0649	0.0101	14.50	6,460.7
4	1,500	12,734.53	0.0628	0.0100	13.25	6,446.9
5	1,500	11,275.45	0.0701	0.0101	16.00	6,432.3

Table 6. Comparison of maximum α-amylase productivities obtained from experiment, previous work [11], and a modified MCMC

Search technique	Optimal value				Time peaked (h)
	U_0 (L/h)	K	t_s (h)	J (U/h)	
Experimental result	0.0611	-	16.45	5474.0	27.00
Skolpap et al., 2004 (Reported)	0.0619	0.0200	14.00	6262.0	34.00
A modified MCMC	0.0648	0.0112	14.75	6483.4	34.75

Fig. 5. Comparison of α-amylase syntheses under optimal feed-rate profiles

Fig. 6. Comparison of protease syntheses under optimal feed-rate profiles

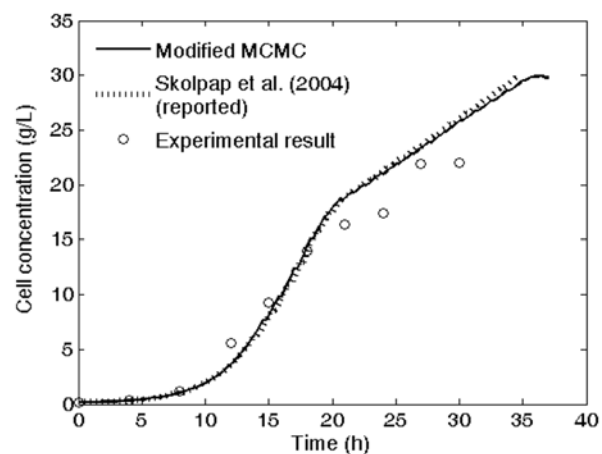Fig. 7. Comparison of biomass productions under optimal feed-rate profiles

of fermentation, respectively, since the corresponding culture volumes reached maximum values in Eq. (6) (Fig. 3).

Fig. 4 compares α-amylase productivities obtained from experimental results and simulation results in previous work [11] as well as from the proposed sampling procedure. The predicted set of decision variables and the corresponding maximum α-amylase productivities achieved from experimental results, previous work [11], and the modified MCMC are illustrated in Table 6.

The optimal feed rate trajectory obtained by using the modified MCMC achieved the maximum α-amylase productivities, the objective function, at 34.75 hour of fermentation.

As shown in Table 6 and Fig. 4, the maximum α-amylase productivity obtained experimentally, the optimal results reported in Skolpap et al. [11], achieved with a modified MCMC were 5,474 U/h at 27 h, 6,262 U/h at 34 h, and 6,483.4 U/h at 34.75 h, respectively. Apparently, the maximum α-amylase productivities attained by using the modified MCMC was improved 14.4 and 18.4% with respect to the previous work and the experimental result, respectively. The optimal feed rate profiles cannot be verified experimentally due to the limitations of pump speed adjustments. The smallest pump speed increment corresponded to a feed rate change of 0.00238 L/h, while feed rate modification predicted by the modified MCMC was 0.001125 L/h. It should be noted that the experimental observations described above were obtained as a preliminary experiment which served as a basis for fed-batch model development. The proposed modified MCMC enhanced the efficiency of seeking optimal solution over a conventional MCMC procedure. Evidently, further experimentation is needed to confirm the results.

Similarly, in Fig. 5 α-amylase activity obtained from experimental results and simulation results in the previous work [11] is compared with activity projected by the proposed sampling procedure. It was found that the proposed sampling procedure improves α-amylase activities by 35.9% and 42%, compared with those from experimental and simulation results. As shown in Fig. 6, the maximum protease activity forecast [11] obtained in this work was about 10% higher than that of experimental result. It was shown in Fig. 7 that the biomass production results along the optimal trajectories were comparable.

STATISTICAL ANALYSIS

The accuracy and acceptability of the simulated results obtained

Table 7. Statistical analysis of objective function values at 1,500 iterations (evaluated at 95% confidence interval)

t_s (h)	Mean (U/h)	Std. Dev. (U/h)	Upper Bound (U/h)	Lower Bound (U/h)	Relative error (%)
14.75	6,289.3	0.1851	6300.7	6277.8	0.48
15.75	6,283.5	0.1384	6292.1	6274.9	0.36
14.50	6,259.8	0.2054	6272.5	6247.1	0.53
13.25	6,256.0	0.2672	6272.5	6239.4	0.69
16.00	6,239.5	0.1386	6248.1	6230.9	0.36

from the proposed parameter sampling procedure and the Metropolis-Hasting algorithm were statistically examined after neglecting one-third iterations. Given 1,500 iterations, the 500 lowest objective function values were neglected and the remaining 1,000 iterations were use to estimate the relative errors of the objective function values. In Table 7, the 95% confidence interval of the objective function is shown. The calculated mean values of the objective function are all within the interval, which indicates that the optimization based on 1,500 iterations was stationary and sufficiently accurate.

CONCLUSION

The proposed modified MCMC parameter sampling procedure using multiple triangular distributions was successfully implemented in searching for a set of decision variables for optimal feed-rate profile determination of a dual-enzyme fed-batch problem. In our experience the optimal decision variables often lie in a region near the edge of the prior distributions that are rarely sampled by the Monte Carlo method. The principal advantages of the proposed method over the existing method are (a) the ability to identify proper parameter values that are located near the tail end or even outside of the parameter sampling ranges by redefining multiple parameter distributions during the sampling procedure, and (b) the ability to reduce random walk behavior by strategically dividing the sampling procedure into learning and searching periods. Given these two advantages, the optimum solution could be achieved and solution convergence could be assured. Moreover, the computational time was minimized by applying multiple distributions and relating the number of iterations with the relative error tolerance of the optimized variable.

APPENDIX A

The fed-batch model of α-amylase and protease-producing recombinant *B. subtilis* ATCC 6015a [11] is listed below. Note that the last term on the right-hand side of the differential equations is a correction for medium inflow or feed rate (u) which affects updated culture volume (V).

Biomass concentration (y_1):

$$\frac{dy_1}{dt} = (1 - e^{-k_{20}t})\left(\frac{k_1 y_2}{k_2 + y_2}\right)y_1 - k_3 y_1 y_6 - \frac{u \cdot y_1}{V} \quad \text{(A.1)}$$

where

k_1 : maximum specific growth rate [μm]
k_2 : ammonia saturation constant
k_3 : specific rate of biomass lysis

k_{20} : time lag coefficient
u : medium flow rate in fed-batch fermentations
V : bioreactor volume
y_2 : ammonium nitrogen concentration
y_6 : protease activity

Skolpap et al. [11] reported that ammonium nitrogen was a limiting substrate for the dual-enzyme system and proteolysis caused cell lysis.

Ammonium nitrogen concentration (y_2):

$$\frac{dy_2}{dt} = -\left(\frac{N_{Bio}AW_N}{MW_{Bio}}\right)\left(\frac{dy_1}{dt}\right) - \left(\frac{N_{Amy}AW_N}{MW_{Amy}}\right)\left(\frac{dy_3}{dt}\right)$$
$$- \left(\frac{N_{Pro}AW_N}{MW_{Pro}}\right)\left(\frac{dy_5}{dt}\right) - cor_X + \frac{u(y_{2,0} - y_2)}{V} \quad \text{(A.2)}$$

where

cor_X : correction term for inorganic nitrogen (ammonia) loss due to gas phase stripping
MW_{Amy} : formula weight of α-amylase
MW_{Bio} : formula weight of biomass
AW_N : atomic weight of nitrogen
MW_{Pro} : formula weight of protease
N_{Amy} : mole fraction of nitrogen in α-amylase
N_{Bio} : mole fraction of nitrogen in biomass
N_{Pro} : mole fraction of nitrogen in protease
$y_{2,0}$: ammonium nitrogen concentration in feed medium
y_3 : α-amylase concentration
y_5 : protease concentration

The well-known *Luedeking-Piret* equation [13] could be used to model the rate of α-amylase synthesis. The model comprises growth associated and non-growth associated terms as follows:

α-Amylase concentration (y_3):

$$\frac{dy_3}{dt} = k_5\frac{dy_1}{dt} + factor_{Amy}(k_6 y_1) \quad \text{(A.3)}$$

where

k_5 : growth associated α-amylase productivity constant
k_6 : non-growth associated α-amylase productivity constant

Note that $factor_{Amy} = 1 - \dfrac{Z_{Amy}}{k_{18}}$; $Z_{Amy} = \dfrac{y_3 - y_{3,0}}{y_1 - y_{1,0}}$

$factor_{Amy}$ is the ratio of non-growth associated α-amylase synthesis to maximum α-amylase yield.

where

factor$_{Amy}$: a feedback modulation term for non-growth associated α-amylase formation

k_{18} : maximum α-amylase yield

$y_{1,0}$: biomass concentration at initial time of batch period

$y_{3,0}$: α-amylase concentration at initial time of batch period

Due to the proteolytic effect only on α-amylase activity, the conversion from α-amylase concentration to activity is necessary for the simulation.

α-Amylase activity (y_4):

$$\frac{dy_4}{dt} = k_7\left(\frac{dy_3}{dt} - k_8 y_6\right) - \frac{u \cdot y_4}{V} \tag{A.4}$$

where

k_7 : conversion factor from α-amylase concentration to activity

k_8 : rate constant for α-amylase lysis by protease

The model for the biosynthesis of protease enzyme is constructed in a similar vein:

Protease concentration (y_5):

$$\frac{dy_5}{dt} = k_{17}\left(k_9 \frac{dy_1}{dt} + factor_{Pro}(k_{10} y_1)\right) \tag{A.5}$$

where

k_9 : growth associated protease productivity constant

k_{10} : non-growth associated protease productivity constant

k_{17} : rate control for protease production

Note that $factor_{Pro} = 1 - \dfrac{Z_{Pro}}{k_{19}}$; $Z_{Pro} = \dfrac{y_5 - y_{5,0}}{y_1 - y_{1,0}}$

where

factor$_{Pro}$: a feedback modulation term for non-growth associated protease formation

k_{19} : maximum protease yield

$y_{5,0}$: protease concentration at initial time of batch period

Similarly, model structure for protease activity mimics the model for α-amylase activity:

Protease activity (y_6):

$$\frac{dy_6}{dt} = k_{11}\left(\frac{dy_5}{dt} - k_{12} y_6\right) - \frac{u \cdot y_6}{V} \tag{A.6}$$

where

k_{11} : conversion of protease concentration to activity

k_{12} : non-growth associated protease productivity constant during ammonium nitrogen limitation

The formation of ethanol, lactate, and acetate strongly depended on the C : N ratio and the availability of oxygen as expressed below:

Ethanol concentration (y_7):

$$\frac{dy_7}{dt} = -k_{21} y_1 + 0.028 - \frac{u \cdot y_7}{V} \tag{A.7}$$

where

k_{21} : non-growth associated rate constant for ethanol

Lactate concentration (y_8):

$$\frac{dy_8}{dt} = k_{13} y_1 - k_{14} \frac{dy_1}{dt} - \frac{u \cdot y_8}{V} \tag{A.8}$$

where

k_{13} : non-growth associated constant for lactate

k_{14} : growth associated constant for lactate

Acetate concentration (y_9):

$$\frac{dy_9}{dt} = k_{15} y_1 - k_{16} \frac{dy_1}{dt} - \frac{u \cdot y_9}{V} \tag{A.9}$$

where

k_{15} : non-growth associated constant for acetate

k_{16} : growth associated constant for acetate

During the early stages of the fermentation, citrate served as a supplementary carbon source and was consumed as shown in the following rate expression:

Citrate concentration (y_{10}):

$$\frac{dy_{10}}{dt} = \frac{-0.5 \cdot \left(\frac{dy_1}{dt} + \left|\frac{dy_1}{dt}\right|\right)}{Y_{X/Cit}} \tag{A.10}$$

where

$Y_{X/Cit}$: a yield of biomass per gram of citrate [g biomass/g citrate]

The stoichiometric coefficients for Eq. (11) were derived by metabolic flux analysis. The carbon balance is converted to starch equivalents in the following manner:

Starch concentration (y_{11}):

$$\frac{dy_{11}}{dt} = \frac{MW_{St}}{AW_C}\Bigg\{-\frac{1}{k_4} \cdot \frac{AW_C}{MW_{St}} \cdot \frac{dy_1}{dt} - \frac{AW_C}{MW_{Amy}} \cdot \frac{dy_3}{dt} - \frac{AW_C}{MW_{Pro}} \cdot \frac{dy_5}{dt}$$
$$- \frac{AW_C}{MW_{Lac}} \cdot \frac{dy_8}{dt} - \frac{AW_C}{MW_{HAc}} \cdot \frac{dy_9}{dt} - N_{Cit}\frac{AW_C}{MW_{Cit}} \cdot Y_{Cit/St} \cdot \frac{dy_{10}}{dt}$$
$$- \frac{AW_C}{MW_{EtOH}} \cdot \frac{dy_7}{dt}\Bigg\} + \frac{u(y_{11,0} - y_{11})}{V} \tag{A.11}$$

where

k_4 : biomass yield constant ($Y_{X/S}$)

AW_C : atomic weight of carbon

MW_{Cit} : molecular weight of citrate

MW_{EtOH} : formula weight of ethanol

MW_{HAc} : formula weight of acetate

MW_{Lac} : formula weight of lactate

MW_{St} : formula weight of starch

N_{Cit} : number of carbon atoms in one mole of citrate

$y_{11,0}$: starch concentration in the fed-batch feed medium

$Y_{Cit/St}$: mass equivalent of citrate [g citrate/g starch]

An amino acid supplement of isoleucine (Ile) and threonine (Thr) is beneficial for biomass as well as α-amylase production as indicated by metabolic flux analysis [15].

Isoleucine concentration (y_{12}):

$$\frac{dy_{12}}{dt} = cor_{Ile}\left\{\left(AA_{Bio}\cdot Ile_{Bio}\cdot 0.5\cdot\left(\frac{dy_1}{dt}+\left|\frac{dy_1}{dt}\right|\right)\right)\right.$$
$$\left.+\left(Ile_{Amy}\cdot 0.5\cdot\left(\frac{dy_3}{dt}+\left|\frac{dy_3}{dt}\right|\right)\right)\right\}$$
$$+\left(Ile_{Amy}\cdot 0.5\cdot\left(\frac{dy_5}{dt}+\left|\frac{dy_5}{dt}\right|\right)\right)+\frac{u(y_{12,0}-y_{12})}{V} \quad (A.12)$$

where

AA_{Bio} : mass fraction of protein in biomass
cor_{Ile} : correction term for y_{12}
Ile_{Amy} : mass fraction of isoleucine in α-amylase
Ile_{Bio} : mass fraction of isoleucine in biomass protein
Ile_{Pro} : mass fraction of isoleucine in protease
$y_{12,0}$: isoleucine concentration in feed medium

The value of this correction factor was obtained by metabolic flux analysis. Presumably, the bacterium had no absolute external requirement for Ile and was capable to synthesize most (80%) of the Ile internally. The differential equation for Thr is:

Threonine concentration (y_{13}):

$$\frac{dy_{13}}{dt} = cor_{Thr}\left\{\left(AA_{Bio}\cdot Thr_{Bio}\cdot 0.5\cdot\left(\frac{dy_1}{dt}+\left|\frac{dy_1}{dt}\right|\right)\right)\right.$$
$$\left.+\left(Thr_{Amy}\cdot 0.5\cdot\left(\frac{dy_3}{dt}+\left|\frac{dy_3}{dt}\right|\right)\right)\right\}$$
$$+\left(Thr_{Amy}\cdot 0.5\cdot\left(\frac{dy_5}{dt}+\left|\frac{dy_5}{dt}\right|\right)\right)+\frac{u(y_{13,0}-y_{13})}{V} \quad (A.13)$$

where

cor_{Thr} : correction term for y_{13}
Thr_{Amy} : mass fraction of threonine in α-amylase
Thr_{Bio} : mass fraction of threonine in biomass protein
Thr_{Pro} : mass fraction of threonine in protease
$y_{13,0}$: threonine concentration in feed medium

As in case of Ile, the external source of Thr accounts for 15% of the total Thr requirement. Evidently, the external supply of both Ile and Thr was adequate to achieve the enhancement of α-amylase synthesis.

The differential equation for the volume change as the function of the feed rate is straightforward as expressed in Eq. (1).

APPENDIX B ESTIMATED VALUES OF MODEL PARAMETERS [3]

Model parameter	Fed-batch value	Model parameter	Fed-batch value
k_1	0.376	k_{14}	5.626×10^{-7}
k_2	0.253	k_{15}	2.481×10^{-3}
k_3	1.707×10^{-4}	k_{16}	9.970×10^{-7}
k_4	0.721	k_{17}	0.152
k_5	1.543×10^{-6}	k_{18}	0.472
k_6	1.512×10^{-2}	k_{19}	5.684×10^{-2}
k_7	10.177	k_{20}	0.399
k_8	1.208×10^{-3}	k_{21}	5.3×10^{-3}
k_9	2.689×10^{-3}	cor_{Amy}	0.880
k_{10}	0.183	cor_{Bio}	0.890
k_{11}	10.273	cor_{Pro}	0.920
k_{12}	1.068×10^{-2}	cor_X	0.000
k_{13}	1.191×10^{-4}	$Y_{X/Cit}$	NE

NE denotes an unevaluated parameter value due to citrate depletion prior to fed-batch fermentation.

NOMENCLATURE

AA_{Bio} : mass fraction of protein in biomass {0.55}
AW_C : atomic weight of carbon {12.01}
AW_N : atomic weight of nitrogen {14.01}
cor_{Amy} : correction term for ammonium nitrogen consumption of α-amylase
cor_{Bio} : correction term for ammonium nitrogen consumption by biomass formation
cor_{Ile} : correction term for y_{12} {0.2}
cor_{Pro} : correction term for nitrogen consumption of protease
cor_{Thr} : correction term for y_{13}
cor_X : correction term for inorganic nitrogen (ammonia) loss due to gas phase stripping
df : degree of freedom
E : expected parameter value
E_q : expectation in the qth row of the same column
$factor_{Amy}$: a feed back modulation term for non-growth associated α-amylase formation
$factor_{Pro}$: a feed back modulation term for non-growth associated protease formation
Ile_{Amy} : mass fraction of isoleucine in α-amylase {0.053}
Ile_{Bio} : mass fraction of isoleucine in biomass protein {0.05999}
Ile_{Pro} : mass fraction of isoleucine in protease {0.050}
J : objective function value [U/h]
J^{max} : maximum theoretical productivity value [U/h]
k_1 : maximum specific growth rate [h^{-1}]
k_2 : ammonia saturation constant [g/L]
k_3 : specific rate of biomass lysis [L/g·h]
k_4 : biomass yield constant ($Y_{X/S}$) [g cell/g starch]
k_5 : growth associated α-amylase productivity constant [g Amy/g cell]
k_6 : non-growth associated α-amylase productivity constant [h^{-1}]
k_7 : conversion factor from α-amylase concentration to activity [×10^{-3} U/g]
k_8 : rate constant for α-amylase lysis by protease [×10^3 g/U(h)]
k_9 : growth associated protease productivity constant [g Pro/g cell]
k_{10} : non-growth associated protease productivity constant [h^{-1}]
k_{11} : conversion of protease concentration to activity [×10^{-3} U/g]
k_{12} : non-growth associated protease productivity constant dur-

ing ammonium nitrogen limitation [$\times 10^3$ g/U(h)]

k_{13} : non-growth associated constant for lactate [h^{-1}]

k_{14} : growth associated constant for lactate [g Lac/g cell]

k_{15} : non-growth associated constant for acetate [h^{-1}]

k_{16} : growth associated constant for acetate [g HAC/g cell]

k_{17} : rate control constant for protease production

k_{18} : maximum α-amylase yield [g Amy/g cell]

k_{19} : maximum protease yield [g Pro/g cell]

k_{20} : time lag coefficient

k_{21} : non-growth associated rate constant for ethanol [h^{-1}]

K   : exponential constant term in feed rate equation

MW_{Amy} : formula weight of α-amylase {25.18}

MW_{Bio} : formula weight of biomass {26.08}

MW_{Cit} : molecular weight of citrate {193}

MW_{EtOH} : formula weight of ethanol {23.04}

MW_{HAc} : formula weight of acetate {30.03}

MW_{Lac} : formula weight of lactate {30.03}

MW_{Pro} : formula weight of protease {26.94}

MW_{St} : formula weight of starch {26.96}

N_{Amy} : mole fraction of nitrogen in α-amylase {0.274}

N_{Bio} : mole fraction of nitrogen in biomass {0.224}

N_{Cit} : number of carbon atoms in one mole of citrate {6}

N_{Pro} : mole fraction of nitrogen in protease {0.268}

O_q : observation in the qth row of the same column

P   : desirable product (α-amylase) concentration [g/L]

q   : proposal distribution of the parameter

Q   : a number of data points in any column

r_1, r_2 : random numbers in the range of [−1, 1]

s^2 : variance

t   : fermentation time [h]

t_k : time at observation taken [h]

t_s : switching time or starting time for feeding [h]

T   : total time period [h]

Thr_{Amy}: mass fraction of threonine in α-amylase {0.0712}

Thr_{Bio} : mass fraction of threonine in biomass protein {0.04523}

Thr_{Pro} : mass fraction of threonine in protease {0.06468}

u   : medium flow rate in fed-batch fermentations [L/h]

V   : bioreactor volume [L]

y_1 : biomass concentration [g/L]

y_2 : ammonium nitrogen concentration [g/L]

y_3 : α-amylase concentration [g/L]

y_4 : α-amylase activity [U/mL]

y_5 : protease concentration [g/L]

y_6 : protease activity [U/mL]

y_7 : ethanol concentration [g/L]

y_8 : lactate concentration [g/L]

y_9 : acetate concentration [g/L]

y_{10} : citrate concentration [g/L]

y_{11} : starch concentration [g/L]

y_{12} : isoleucine concentration [g/L]

y_{13} : threonine concentration [g/L]

$Y_{Cit/St}$: mass equivalent of citrate [g citrate/g starch] {0.48}

$Y_{X/Cit}$: yield of biomass per gram of citrate [g cell/g citrate]

$Y_{X/S}$: theoretical cell yield based on substrate consumed [g cell/

g substrate]

Greek Letters

α : the probability of acceptance of a sampled value

χ^2 : a probability of a χ^2 statistic

δ : the mean parameter

θ : the parameter value

π : stationary distribution of the parameter

Subscripts

i   : the ith parameter

q   : the qth data point of any state variable

Superscripts

better : better parameter value

best   : the best parameter value

j   : the jth iteration

min   : minimum value

max   : maximum value

REFERENCES

1. J. M. Nougués, M. D. Grau and L. Puigjaner, *Chemical Engineering and Processing*, **41**, 303 (2002).
2. L. Z. Chen and S. K. Nguang, *ISA Transactions*, **41**, 409 (2002).
3. D. Sarkar and J. M. Modak, *Chem. Eng. Science*, **58**, 2283 (2003).
4. L. Z. Chen, S. K. Nguang, X. D. Chen and X. M. Li, *Biochem. Eng. J.*, **22**, 51 (2004).
5. V. K. Jayaraman, B. D. Kulkarni, K. Gupta, J. Rajesh and H. S. Kusumaker, *Biotechnol. Prog.*, **17**, 81 (2001).
6. A. Patkar, D. H. Lee and J. H. Seo, *Korean J. Chem. Eng.*, **10**, 146 (1993).
7. J. E. Cuthrell and L. T. Biegler, *Comput. Chem. Eng.*, **13**, 49 (1989).
8. H. C. Lim, Y. J. Tayeb, J. M. Modak and P. Bonte, *Biotechnol. Bioeng.*, **28**, 1408 (1986).
9. W. Mekarapiruk and R. Luus, *Can. J. Chem. Eng.*, **79**, 777 (2001).
10. W. Mekarapiruk and R. Luus, *Ind. Eng. Chem. Res.*, **39**, 84 (2000).
11. W. Skolpap, J. M. Scharer, P. L. Douglas and M. Moo-Young, *Biotechnol. Bioeng.*, **86**, 706 (2004).
12. W. Skolpap, J. M. Scharer, P. L. Douglas and M. Moo-Young, *Songklanakarin J. Sci. Technol.*, **27**, 1057 (2005).
13. R. Luedeking and E. L. Piret, *J. Biochem. Microbiol. Technol. Eng.*, **1**, 431 (1959).
14. A. Constantinides and N. Mostoufi, *Numerical methods for chemical engineers with MATLAB applications*, Prentice-Hall Inc., New Jersey, 323-340 (1999).
15. W. Skolpap, *Bioprocess optimization using metabolic engineering approach*, Ph.D. Thesis, University of Waterloo, Waterloo, Ontario (2003).
16. D. C. Montgomery, G. C. Runger and N. F. Hubele, *Engineering statistics*, 4th Edition, John Wiley & Sons, New York (2006).
17. C.-K. Min, *Comput. Stat. Data An.*, **27**, 171 (1998).
18. W. R. Gilks, S. Richardson and D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*, Chapman & Hall, London, UK (1998).