Biotechnology Letters **23:** 911–916, 2001.

An artificial intelligence tool for bioprocess monitoring: application to continuous production of gluconic acid by immobilized *Aspergillus niger*

N.V. Sankpal, J.J.S. Cheema, S.S. Tambe & B.D. Kulkarni*
Chemical Engineering Division, National Chemical Laboratory, Pune 411008, India
**Author for correspondence (Fax: +91-20-5893041; E-mail: bdk@ems.ncl.res.in)*

Received after revisions 2 April 2001; Accepted 3 April 2001

Key words: artificial intelligence, *Aspergillus niger*, continuous fermentation, gluconic acid, immobilization, monitoring, symbolic regression

Abstract

Experimental data on continuous fermentation of sucrose and glucose solution at low pH to gluconic acid by *Asprgillus niger* immobilized on cellulose fabric show complex dynamic behaviour including a decline in yield. The data have been analyzed using an artificial intelligence based symbolic regression technique to provide a mathematical model for predicting values of conversion 5, 10 and 15 h ahead values of conversion. These predictions can be used during continuous operations to monitor the bioprocess and adjust the residence time of fermentation to get complete and more efficient conversion of sucrose or glucose to gluconic acid.

Introduction

Commercial production of a wide range of metabolites use fungal mycelia cultured as free mycelia, pellets/flocks or as artificially bound/entrapped cells (Romero-Gómez *et al.* 2000). During fermentation, the type of morphology strongly influences metabolite production and its effects remain unexplored and relatively unclear (Braun & Vecht-Lifshitz 1991). In a fermenter, mycelia are subjected to external and internal mass- and heat-transfer stresses (Karel & Robertson 1989), and mechanical agitation can help to remove such limitations. The mechanical forces, induced by turbulent flow (by aeration or agitation) may however lead to: (i) pellet erosion owing to loss of the pellet surface (Van Sujidam & Metz 1981, Cui *et al.* 1997), (ii) pellet breakage (Bhavaraju & Blanch 1976, Metz & Kossen 1977), (iii) cell lysis/disintegration (Wittler *et al.* 1986), and (iv) cell breakage or surface erosion (Taguchi *et al.* 1968). These forces produce a spontaneous or gradual loss of conversion efficiency due to loss in cellular activities (May *et al.* 1929, Sakurai *et al.* 1989, Corrado *et al.* 1997). Immobilization of biocatalysts, use of modified reactor configuration etc. are now routinely practiced to alleviate such difficulties (Birnbaum *et al.* 1983, Roehr *et al.* 1996, Sankpal & Sahasrabudhe 2001). Notwithstanding these improvements, the problem still continues and the concomitant loss of enzyme activity or conversion efficiency continues to pose practical difficulties.

In this work, we have used an Artificial Intelligence (AI) based tool performing symbolic regression (hereafter referred to as 'AI Based Symbolic Regression', AIBSR) (Goldberg 1989, Holland 1992, Tambe *et al.* 1996, Kulkarni *et al.* 1999) – to monitor a bioprocess. The AIBSR technique allows us to obtain automatically a mathematical expression which accurately fits a given set of experimental data. For the sake of illustration, experimental data on the conversion efficiency of sucrose and glucose to gluconic acid by *Aspergillus niger* is used as a time series to develop the functional form of an expression for predicting, one step ahead (5 h, 10 h, 15 h) scenario. The predictions of the model are validated against the experimental data and afford an opportunity to take remedial measures when loss in conversion efficiency is predicted. The strategy is adopted for real experiments on fermentation to gluconic acid and results presented.

Materials and methods

Organism and growth

Aspergillus niger NCIM 545 (National Collection of Industrial Microorganisms NCL, Pune) was grown on potato/dextrose/agar slopes (PDA). Experiments were carried out in a fermenter set-up consisting of an enclosed chamber (70 × 35 × 30 cm) equipped with a thermohygrometer, humidifier, temperature controller and an inlet and outlet for air. A woven cellulosic fabric support, commonly known as 'Turkish Toweling', was used to immobilize *A. niger*. A piece of this fabric (68 × 8 × 0.5 cm) comprising 100 ml void volume (interstitial space/volume between microfibrils) was utilized. Fermentation medium was dripped continuously on the support by a peristaltic pump at a desired flow rate. Immobilization of *A. niger* mycelia by adsorption technique, its cellular adaptation and behaviour at low pH, complete sporulation inhibition, effects of various medium compositions, support configuration, and analysis of reactants and products have been described in more detail in our earlier work (Sankpal *et al.* 1999, 2000).

Increasing the interfacial area of mycelia on porous support will increase the mass transfer and O_2 diffusion during fermentation of glucose to produce free gluconic acid of high productivity. Sucrose, an alternative cheap substrate, also produces gluconic acid and fructose as cost-effective products. During fermentation, a pH gradient exists on the fabric support. For the feed pH of 6, the exit pH of 1.9 to 2.3 was monitored during a typical run. The pH profile on the support indicates incremental formation of gluconic acid in successive zones of the fabric support. The average specific rate of gluconic acid formation is computed to be 1.3 and 2.4 g gluconic acid g^{-1} mycelia.h using sucrose and glucose, respectively.

Loss in the conversion efficiency during continuous fermentation

During continuous bioconversion of glucose and sucrose to gluconic acid, the specific rate of gluconic acid formation dropped as fermentation progressed. A typical conversion profile for glucose and sucrose is shown in Figure 1. The drop in conversion values (Figures 1a and 1c) is an indication of loss of productivity and overall bioconversion efficiency of fermentation. Based on the estimate of local reaction rate, substrate concentration etc. one can conjecture an increase in residence time (RT) to compensate for the loss in local conversion efficiency (Figures 1b and 1d). Figure 2 shows the conversion – time plot for a typical run where the RT was adjusted such that the conversion efficiency remains above 90%. The RT values indicated in the Figure 2 are at the instant of time where marked drop in the efficiency is observed. Thus, the restoration of higher conversion was achieved by progressively raising the residence time (values increasing from 7.8 to 10.4 h). This avoids the presence of unreacted glucose or sucrose in mother liquor as well as the load on the purification step. Online determination of local reaction rates, substrate concentration etc. involves cumbersome procedures and may not be always feasible. We have, therefore, used the experimental runs in Figures 1b and 1d to develop a mathematical expression predicting conversion as a function of time. Towards this end, we use the AIBSR technique that is described below in details. Once a reasonable expression predicting the conversion vs. running time is obtained, the same can be used as a process-monitoring tool obviating the need for too-frequent costly experimental measurements.

Artificial intelligence based symbolic regression

The methodology and applications of the AIBSR technique for time series analysis has been discussed in details (Yadavalli *et al.* 1999, Sankpal 2000) and has been modified for use here. The stepwise AIBSR implementation is presented below in brief for reasons of being self-contained.

Stepwise AIBSR implementation

For a given time series $\{x_1, x_2, \ldots, x_t, \ldots, x_T\}$ (training set) of length T, the AIBSR searches for the fittest expression in the form of a one-step-ahead predictor defined as:

$$x_{t+1} = f\{x_t, x_{t-1}, \ldots, x_{t-\alpha}\}; \quad \alpha \le t \le (T-1); \quad 0 \le \alpha \le L, \tag{1}$$

where t denotes discrete time; α refers to the number of lags; L represents the maximum permissible lags, and f exemplifies the function relating the current (x_t) and the past (lagged) values $\{x_{t-1}, \ldots, x_{t-\alpha}\}$, with the one-step-ahead value (x_{t+1}) in the time series. In the following, the AIBSR procedure for obtaining the best fitting form of function f is elaborated.

Step 1 (Initialization): randomly generate an initial population of candidate strings (size = N_{init}) using the postfix strategy (Kulkarni *et al.* 1999, Yadavalli *et al.*

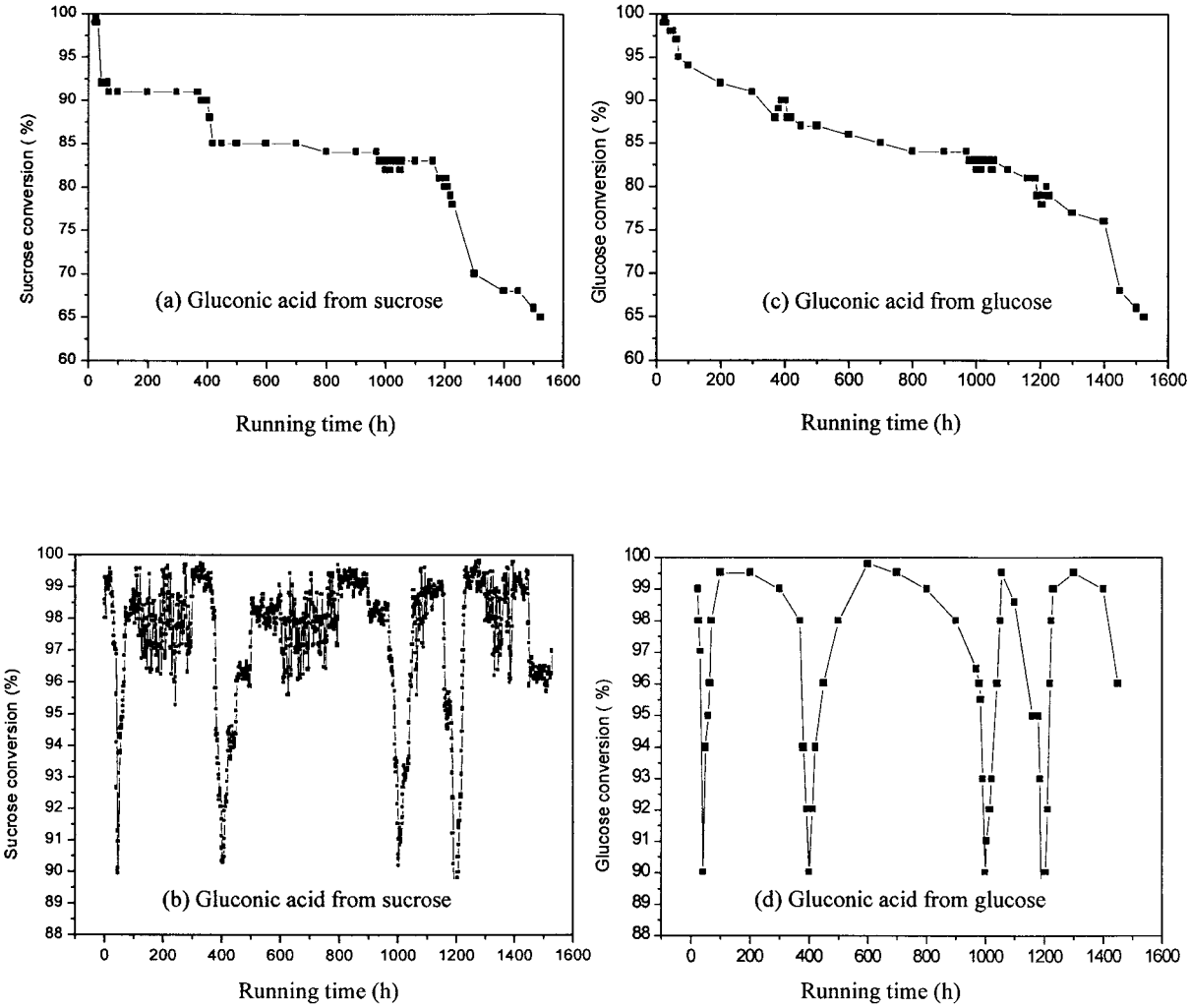

Fig. 1. Experimental conversion (100% = 100 g l^{-1}) profiles during continuous production of gluconic acid by immobilized *A. niger* at residence time of 7.8 h: (a) for sucrose, (c) for glucose. Residence time was adjusted to achieve bio-conversion (b) for sucrose, and (d) for glucose respectively. Experiments were carried out under steady conditions of temperature and humidity, aeration.

1999). Each string, of length l_{str}, is partitioned into a number of blocks comprising a sequence of randomly selected symbols representing operands and operators. Operands include constants of an expression and variables representing the current and the lagged values whereas operators indicate addition, subtraction, multiplication and division.

Step 2 (Fitness evaluation and population ranking): the fitness of a string is a measure of its data fitting ability and further survival. It is expressed using a sum-squared error (SSE) dependent fitness function:

$$R_l^2 = 1 - \frac{\Delta_l^2}{\sum_{t=\alpha+1}^{T} (x_t - \bar{x})}; \; l = 1, 2, \ldots, N_{\text{pop}}, \quad (2)$$

where R_l^2 denotes the fitness value of lth string; $\bar{x}$ represents the mean of all x_t's in the time series and N_{pop} denotes the number of strings. The Δ_l^2 representing SSE is estimated according to

$$\Delta_l^2 = \sum_{t=\alpha+1}^{T} (x_{t+1}^l - x_{t+1})^2. \quad (3)$$

After evaluating fitness of all the strings in the present population, they are ranked in the decreasing order of their fitness values.

Table 1. AIBSR parameter values and results of statistical analysis.

N_{init}	l_{str}	N_{gen}	α	p_{mut}	$[Z_1, Z_2]$	R^2	CC
50	35	35	2	0.02	[−10, 10]	0.9896 (trn)	0.9976 (trn)
						0.9905 (tst)	0.9984 (tst)

R^2: fitness score; CC: correlation coefficient; trn: training set; tst: test set; N_{init}: initial number of strings; l_{str}: length of each string; N_{gen}: number of generations; α: number of lags; p_{mut}: probability of mutation; $[Z_1, Z_2]$: range of constants.

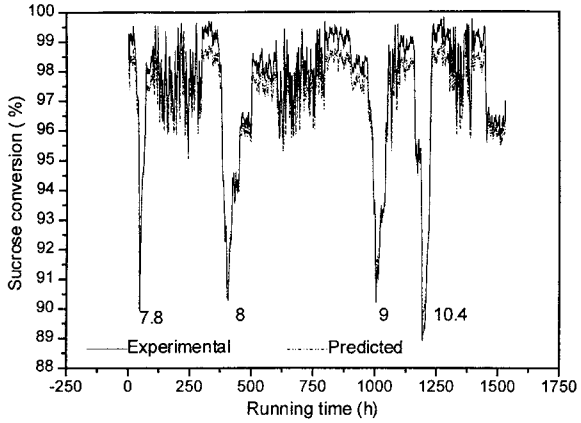

Fig. 2. Experimental and one-step-ahead (10 h) predicted values for continuous bio-conversion of sucrose (100% = 100 g l^{-1}), Relative humidity: 90%, residence time: 7.8 to 10.4 h, biomass on support 1.12 g.

Table 2. Fitness score (R^2) and correlation coefficients (CC) for multi-step ahead predictor.

Prediction step (h)	R^2	CC
5	0.9789 (trn)	0.9897 (trn)
	0.9793 (tst)	0.9905 (tst)
10	0.9654 (trn)	0.9821 (trn)
	0.9672 (tst)	0.9838 (tst)
15	0.9532 (trn)	0.9765 (trn)
	0.9571 (tst)	0.9772 (tst)

trn: training set; tst: test set.

Step 3 (Parent selection): for producing the next generation of strings, the present population is mated for which a mating pool of fitter strings is formed. In this study, we are employing 'elitist selection' scheme, where fitness-driven string selection procedure makes it less likely that the low fitness strings will be mated. Here, a certain number $(N_{\text{init}}/2) + 1$, of top ranking strings are selected as a mating pool following which parent-pairs such as [Strings 1 and 2], [Strings 2 and 3]...[Strings $(N_{\text{ini}}/2)$ and $(N_{\text{ini}}/2 + 1)$] are formed. The top ranking strings are more likely to contain the blocks that would finally yield us the best fitting expression.

Step 4 (Reproduction): the strings in the mating pool are copied to the next generation and remaining ones are discarded. This operation allows the fitter strings in the present population to be part of next generation.

Step 5 (Crossover): each parent-pair from the mating pool forms two offspring; a block from each parent string is randomly selected, and exchanged mutually. The resulting N_{ini} number of new strings together with $(N_{\text{ini}}/2 + 1)$ cloned parents strings form a new generation whose size (N_{pop}), equal to $(1.5 \times N_{\text{ini}}) + 1$, is maintained through the subsequent generations. New strings are created by mutual exchange of randomly selected string blocks and represent a global search as new regions of the solution space are explored.

Step 6 (Mutation): a fraction of string blocks is mutated randomly, with a tiny probability of mutation (P_{mut}) (range 0.001–0.05). Mutation involves replacing a randomly chosen string element by another of same type i.e., an operator (operand) replaces operator (operand).

The various operations of fitness evaluation to mutation (steps 2 to 6) are iteratively performed until AIBSR evolves over a pre-specified number of generations (N_{gen}) and where the fitness of the best string no longer increases. As creation of initial population, crossover and mutation are stochastic procedures, the final solution depends upon the sequence of random numbers generated during the search. In view of this, it becomes necessary to repeat the AIBSR procedure many times by employing different random number generator seeds for arriving at an overall optimal solution. Once a highly fit string is secured, its simplification may become necessary to obtain the shortest but mathematically equivalent expression.

Results and discussion

As described in general methodology, several equations simulating the conversion of sucrose vs. time dynamics were obtained. We report here the best fitting expression as follows:

$$x_{t+1} = (x_t + 0.714x_t) - (1.16112(x_t + 10.7121)) - (3.04(x_{t-1} - x_t)(x_t + 0.3224)). \quad (4)$$

High fitness score (R^2) and correlation coefficient (CC) magnitudes (≈ 1.0) were obtained using this expression for training and test sets (Table 1). The empirical equation searched by the AIBSR has thus captured the inherent dynamic relationship between the current and one-step-ahead values. The efficacy of the proposed Equation (4) was also tested for multistep ahead prediction in recurrent mode. This was done by substituting the x_t by actual experimental data points, after every 5, 10 and 15 h depending upon the prediction step size. The R^2 and CC of the multi-step ahead prediction (5, 10 and 15 h) were in the acceptable limit of accuracy (Table 2). However, for the sake of clarity the 10-h ahead prediction curve is shown in Figure 2.

Equation (4) can be used on-line during fermentation as a bioprocess monitor. It provides the scenario 5, 10 and 15 h ahead giving sufficient time to take remedial measures such as adjustment of RT etc to ensure complete conversion of substrate and bioconversion efficiency. This is especially important since on-line measurement of concentrations of many substances is either not possible or quite often very cumbersome. We have successfully used this strategy to run the fermenter for over a period of 2 months, setting the cut-off efficiency at 90%. The time at which the fermenter would reach this level was predicted on-line and adjustment of residence times was carried out ahead of time to avoid fall in conversion efficiency below this level. A significant improvement in throughput processing was thus possible.

Acknowledgement

The authors acknowledge financial support provided by Council of Scientific and Industrial Research (CSIR) for carrying out these investigations.

References

Bhavaraju SM, Blanch HW (1976) A model for pellet breakup in fungal fermentation. *J. Ferment. Technol.* **54**: 466–468.

Birnbaum S, Larsson P, Mosbach K (1983) Immobilized cells. In: Scouten WH, ed. *Solid Phase Biochemistry: Analytical and Synthetic Aspects.* New York: John Wiley and Sons, pp. 679–762.

Braun S, Vecht-Lifshitz SE (1991) Mycelia morphology and metabolite production. *Trends Biotechnol.* **9**: 63–68.

Corrado S, Mario Z, Attilio C, Marco D (1997) Continuous citric acid fermentation in a supported biomass reactor. *Chem. Eng. Technol.* **20**: 348–353.

Cui YQ, Van der Lans RG JM, Luyben KC AM (1997) Effect of agitation intensities on fungal morphology of submerged fermentation. *Biotechnol. Bioeng.* **55**: 715–726.

Goldberg DE (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley.

Holland JH (1992) *Adaptation in Natural and Artificial Systems*. 2nd edn. Ann Arbor: University of Michigan Press.

Karel SF, Robertson CR (1989) Autoradiographic determination of mass transfer limitations in immobilized cell reactors. *Biotechnol. Bioeng.* **34**: 320–336.

Koza J (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* Cambridge, MA: The MIT Press.

Kulkarni BD, Tambe SS, Dahule RK, Yadavalli VK (1999) Consider genetic programming for process identification. *Hydrocarb. Process.* **78**: 89–97.

May OE, Herric AJ, Moyer AJ, Hellbach R (1929) Semi-plant scale production of gluconic acid by mold fermentation. *Ind. Eng. Chem.* **21**: 1198–1203.

Metz B, Kossen NFW (1977) The growth of molds in the form of pellets: literature review. *Biotechnol. Bioeng.* **19**: 781–799.

Roehr M, Kubicek CP, Kominek J (1996) Gluconic acid. In: Rehm HJ, Reed G, eds. *Biotechnology*, Vol. 6, 2nd edn. USA: VCH, pp. 348–362.

Romero-Gómez SJ, Augur C, Viniegra-González G (2000) Invertase production by *Aspergillus niger* submerged and solid-state fermentation. *Biotechnol. Lett.* **22**: 1255–1258.

Sakurai H, Lee HW, Sato S, Mukataka S, Takahashi J (1989) Gluconic acid production at high concentrations by *Aspergillus niger* immobilized on a nonwoven fabric. *J. Ferment. Bioeng.* **67**: 404–408.

Sankpal NV (2000) *Surface and Submerged Fermentation by Filamentous Fungi: Applications to Production of Gluconic Acid, Citric Acid and Dextran Gum*. Ph.D. Thesis, Chapter 4. Pune, India, University of Pune, pp. 99–117.

Sankpal NV, Sahasrabudhe NA (2001) Production of organic acids and metabolites on fungi and applications in food industry. In: Khachatourians GG, Arora DK, eds. *Applied Mycology and Biotechnology,* Vol. 1. Amsterdam: Elsevier Science, pp. 436–448.

Sankpal, NV, Joshi AP, Kulkarni BD (2000) Nitrogen dependent regulation of gluconic acid and/or citric acid production by *Aspergillus niger*. *J. Microbiol. Biotechnol.* **10**: 15–55.

Sankpal NV, Joshi AP, Sutar II, Kulkarni BD (1999) Continuous production of gluconic acid by *Aspergillus niger* immobilized on a cellulosic support: study of low pH fermentative behavior of *A. niger*. *Process. Biochem.* **35**: 317–325.

Taguchi H, Yoshida T, Tomita Y, Teramoto S (1968) The effects of agitation on disruption of the mycelial pellets in stirred fermenters. *J. Ferment. Technol.* **46**: 814–822.

Tambe SS, Kulkarni BD, Deshpande PB (1996) *Elements of Artificial Neural Networks with Selected Applications in Chemical Engineering and Biological Sciences.* Louisville, KY: Simulation and Advanced Controls, Inc.

Van Suijdam JC, Metz B (1981) Fungal pellet breakup as a function of shear in fermenter. *J. Ferment. Technol.* **59**: 329–333.

Wittler R, Baumgartl H, Lubbers DW, Schugerl K (1986) Investigations of oxygen transfer into *Penicillium chrysogenum* pellets by microprobe measurements. *Biotechnol. Bioeng.* **28**: 1024–1036.

Yadavalli VK, Dahule RK, Tambe SS, Kulkarni BD (1999) Obtaining functional form for chaotic time series evolution using genetic algorithm. *Chaos* **9**: 789–794.