A FUZZY LOGIC ALGORITHM FOR IDENTIFICATION OF THE HARVESTING THRESHOLD DURING PGA PRODUCTION BY *Bacillus megaterium*

E. R. Nucci, R. G. Silva, T. C. Gomes, R. C. Giordano and A. J. G. Cruz[*]

Departamento de Engenharia Química, Universidade Federal de São Carlos,
DEQ/UFSCar, Phone: +(55) (16) 3351-8264, Fax: +(55) (16) 3351-8266,
P.O. Box 676, 13565-905, São Carlos - SP, Brasil
E-mail: ajgcruz@power.ufscar.br

Abstract - Penicillin G acylase (PGA) is an important enzyme used as biocatalyst in the production of semisynthetic β-lactam antibiotics. Many microorganisms produce this enzyme and recombinant *Escherichia coli* has been preferred use for industrial applications. *Bacillus megaterium* is one of the microorganisms that excretes this enzyme into the medium. As a consequence, separation and purification steps are simplified. On-line measurement of enzyme activity during cultivation using in-situ sensors is a difficult task in the industrial environment due to the lack of robust and inexpensive instrumentation. This work presents the results of a fuzzy logic algorithm used to determine the moment of maximum enzyme concentration during *Bacillus megaterium* cultivations in an aerated and stirred, automated lab-scale bioreactor. The fuzzy algorithm was written in Fortran, compiled as a dynamic link library and implemented on a platform developed in MS-Visual Basic. Data were exchanged in real time between the platform and the supervisory system, which was coupled to the bioreactor. It was possible to determine the moment at which maximum enzyme activity was reached in several bioreactor assays. At this point, the end of the process was indicated to the operator. The results illustrate the importance of using reliable computational intelligence-based algorithms in biochemical reactions.
Keywords: Enzyme; State estimation; Computational intelligence; Automation; Penicillin G acylase; Bioreactor.

INTRODUCTION

Penicillin G acylase (PGA) hydrolyzes penicillin G to yield 6-aminopenicilanic acid and phenyl acetic acid. The former one is an important raw material used to produce semisynthetic β-lactam antibiotics. PGA is one of the most important immobilized enzymes in use on an industrial scale (Ospina, 1992). The microorganism most frequently reported in the literature for PGA production is recombinant *Escherichia coli*. In this case, intracellular enzyme is obtained. *Bacillus megaterium* offers the advantage of producing and secreting the enzyme, thereby simplifying the separation and purification steps.

Studies employing *Bacillus megaterium* have been reported by many researchers. Important results concerning medium optimization are found in Acevedo and Cooney (1973), Illanes et al. (1994), Gentina et al. (1997), Pinotti et al. (2000), Souza et al. (2002) and Silva et al. (*submitted to BJChE*).

Due to the great importance of this process and the increase in industrial applications of enzyme technology, studies focusing on bioprocess monitoring and control have been conducted (Chéruy, 1997). Particularly for bioreactors, there is still a lack of reliable, durable and low-cost sensors for measuring on-line key variables such as concentration of cells, substrates and products. These

---

*To whom correspondence should be addressed

variables, referred to as state variables, are essential for bioprocess monitoring, fault detection and control.

Various authors have presented results of computational algorithms that use the available on-line measurements to infer important state variables (see, for instance, Linko et al 1997 and Chéruy, 1997).

This paper deals with the design of a fuzzy logic algorithm, used to determine the time at which the maximum enzyme concentration is attained during *Bacillus megaterium* cultivations. Accurate determination of the moment to begin the harvesting operation is very important for minimizing industrial costs, reducing dead times and avoiding unnecessary degradation of the product during the final stages of a batch or fed-batch cultivation.

Variables such as molar fractions of carbon dioxide and oxygen in exhaust gases from bioreactors are directly related to the metabolic state of the process. These variables and time were used as input variables into the fuzzy algorithm.

FUZZY METHODOLOGY

Fuzzy logic was presented for the first time in 1965 at the University of California (Berkeley, USA) by Lofti A. Zadeh (von Altrock, 1995). This new approach was intended to emulate human reasoning using calculations and operations with fuzzy groups and linguistic variables (Gomide and Godwin, 1994). Fuzzy variables describe qualitative expressions such as very slow, slow, fast, very fast, and so on. These expressions correspond to the linguist values of the linguistic variable (Sousa Jr. and Almeida, 2001). In the classic group theory, a specific element is inside (1) or outside (0) the set. Zadeh proposed a different approach, defining a "degree of membership" (between 0 and 1) so that the element could assume an infinite number of values in this interval (Gomide and Godwin, 1994). The basic idea of fuzzy logic application is to combine the expertise gained through hours and hours of process observation and data analysis to obtain the response that an experienced human operator would have. The idea was to put forward a synthesis of human knowledge suitable for implementation in a computer. Thus, fuzzy sets and operations belong to the wider class of computational intelligence algorithms.

Several studies in the literature use fuzzy logic for fermentative processes. Yamada et al. (1991) used fuzzy control during the fed-batch fermentation for production of the coenzyme Q_{10}. Shiba et al. (1994) employed a fuzzy system to control the concentrations of glucose and ethanol in the medium for *Saccharomyces cerevisiae* cultivations. The objective was to improve α-amylase production in a bioreactor operated in fed-batch mode. Alfafara et al. (1993) developed a fuzzy system to control the concentration of the ethanol and identify the point of maximum production of a peptide in fed-batch yeast cultivations. Sousa Jr. and Almeida (2001) proposed a fuzzy set of reasoning rules that permitted the automated start-up of the feeding of inverted sucrose in cephalosporin C batch production.

MATERIALS AND METHODS

Production of Penicillin G Acylase

- **Microorganism**

Bacillus megaterium ATCC 14945 donated by Fundação Tropical (Campinas, SP, Brazil) was kept in cryogenic vials at –60°C using glycerol 20% v/v as cryoprotector.

- **Culture Media**

Inoculum was prepared using (in g L^{-1}) salt solution (0.2), cheese whey and a solution of a pool of 18 amino acids (10.0): alanine, aspartic acid, cystine, glycine, histidine.HCl, lysine.HCl, methionine, phenylalanine, proline, serine, threonine, tyrosine, tryptophane, valine, arginine, glutamic acid, leucine and isoleucine. The medium was produced with (g.L^{-1}) the same solution of amino acids (10.0), cheese whey (19.6) and phenyl acetic acid (2.7) (Pinotti, et al., 2000; Souza et al., 2002).

- **Enzyme Activity**

It was determined by hydrolysis of 0.2 M penicillin G at pH 8.0 (10 mM phosphate buffer) and 37 °C. The 6-APA produced was measured spectrophotometrically at 415 nm after reaction with p-dimetilaminebenzaldehyde (PDAB), in accordance with Balasingham et al. (1972). An international unit (IU) is the amount of enzyme that catalyzes the formation of 1μmol of 6-APA from penicillin G in one minute, under these conditions.

Experimental Procedure

The standard procedure for fermentation assays started with the incubation of 15 mL of cell spores in 135 mL of culture medium for 24 hours at 300 rpm and 30 °C (1000 mL shake flasks). The cells were

inoculated in the fermentation broth. The experiments were carried out in a 5-liter bioreactor (New Brunswick Scientific, Bioflo II-C). Temperature was kept at 30 °C and the aeration rate, at 2.0 L.min^{-1} (1 atm and 21.1°C). The concentration of dissolved oxygen was maintained at 20% saturation by automatic manipulation of stirrer speed (PID control). The inoculum seed was 10% of the reactor volume.

The bioreactor was coupled to a data acquisition system composed of a programmable logic controller (PLC) and a supervisory system responsible for data acquisition and process control. On-line variables were stored by the data acquisition system (time interval of 10 seconds). The following variables were monitored and obtained on-line: pH, temperature, dissolved oxygen and molar fractions of carbon dioxide in the effluent gas. The fuzzy algorithm was written in Fortran, compiled as a dynamic link library and linked to a platform implemented in MS-Visual Basic, hosted by a second computer. This computer provided data exchange in real time with the acquisition system (running in another PC computer).

FUZZY ALGORITHM

In the process of PGA production by *B. megaterium* there was a moment when the molar fraction of carbon dioxide (Y_{CO2}) reached a maximum. The behavior of this variable was compared with the evolution of enzyme activity in the medium. It could be observed that the maximum enzyme concentration was attained shortly after Y_{CO2} reached its maximum. After this point, Y_{CO2} levels began to drop, in some experiments very rapidly, and the maximum enzyme concentration was achieved

(Figure 1, a and b). The fuzzy algorithm was designed to identify this point. The first step consisted in detecting the Y_{CO2} peak. In the next step, derivates of Y_{CO2} with respect to time (dY_{CO2}/dt) were calculated and compared to a predefined value. This derivative would indicate the end of cultivation. The fuzzy controller was divided into three levels of reasoning: attention, peak identification and protection. The rules of the attention level were the following:

1) IF {early time and low Y_{CO2}} THEN {does not activate action level}.
2) IF {late time and low Y_{CO2}} THEN {does not activate action level}.
3) IF {early time and high Y_{CO2}} THEN {activates action level}.
4) IF {late time and high Y_{CO2}} THEN {activates action level}.

The action level was activated when the degree of membership exceeded a predefined value (0.8). See Figure 2a.

The rule to identify the peak of molar fraction of CO_2 was based on equation (1).

$$\Delta Y_{CO2} = Y_{CO2}(n) - Y_{CO2}(n-20) \qquad (1)$$

where: n is actual time in minutes.

Equation (1) is similar to a low-pass filter, smoothing experimental data provided in real time by the data acquisition system. The two rules employed for the identification level were the following (see Figure 2b):

5) IF {ΔY_{CO2} variation is zero or positive} THEN {does not notify Y_{CO2} peak}
6) IF {ΔY_{CO2} variation is negative} THEN {notifies Y_{CO2} peak}

(a)   (b)
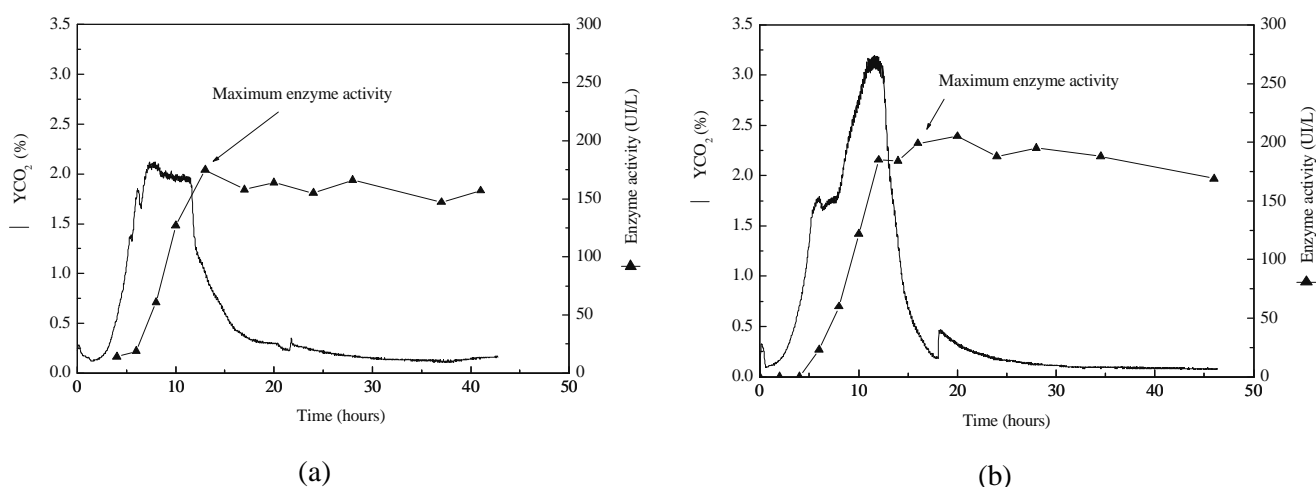
Figure 1: Molar fraction of CO_2 (Y_{CO2}) and enzyme activity during cultivations of *B. megaterium*, used to identify the fuzzy rules: (a) first experiment (b) third experiment.

(a)   (b)
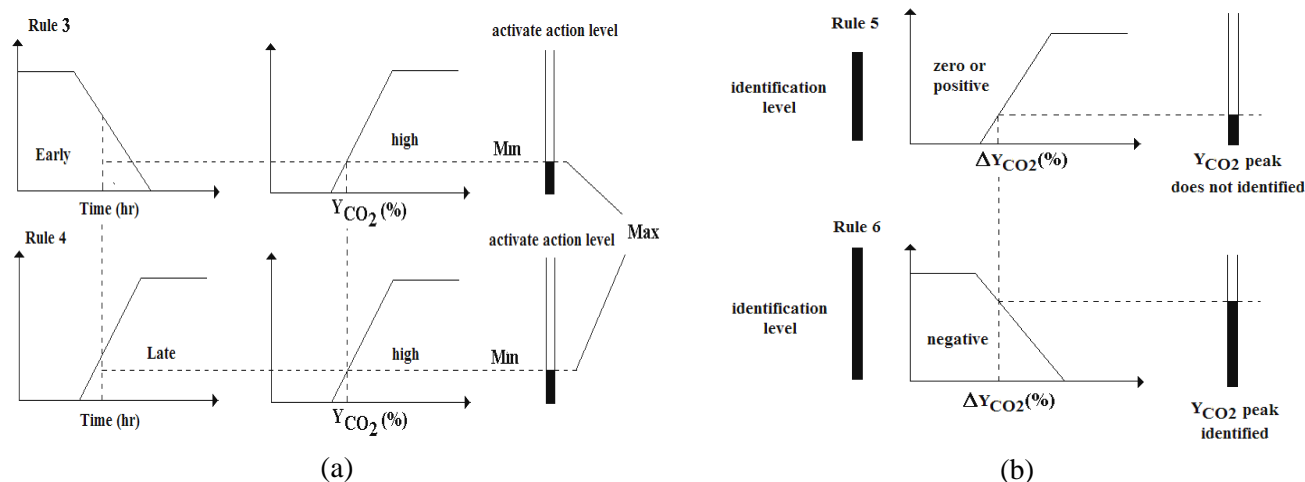
Figure 2: Schematic diagram illustrating some of the fuzzy rules implemented:
(a) rules for action level, (b) rules for identification (VC1=-0.12 and VC2=0.0).

It was also necessary to incorporate some protection rules into the fuzzy algorithm in order to avoid "false peak" identifications caused by noise-corrupted data. The following rules were included:

7) IF {time is early} AND {ΔY_{CO2} variation is negative} THEN {wait}

8) IF {time is late} AND {ΔY_{CO2} variation is negative} THEN {peak confirmed}

9) IF {time is early} AND {ΔY_{CO2} variation is positive} THEN {peak cancelled}

10) IF {time is late} AND {ΔY_{CO2} variation is positive} THEN {peak cancelled}

The threshold criterion adopted was a membership function of 0.5 or greater.

The criterion for the end of cultivation was defined based on the derivate of the molar fraction of CO_2 (dY_{CO2}/dt). When this variable reached a value of -0.20 or greater, the end of cultivation was identified. At this moment enzyme activity should be close to its maximum.

RESULTS AND DISCUSSION

The fuzzy algorithm parameters were tuned using a base set of three experiments (Runs 1-3). A set of three other experiments (Runs 4-6) was used for validation. Figure 3 shows the results of the training set. The algorithm identifies one time to finish the cultivation, which was compared with the experimental region of maximum enzyme activities.

It can be seen that the fuzzy algorithm could successfully ascertain the region of maximum enzyme concentrations.

In Run #1 the algorithm determined 12.5 hr as the end time for cultivation. Laboratory analyses showed that maximum enzyme activity (around 175 IU.L[-1]) was achieved at 13 hr.

Figure 3B shows Run #1: the end of cultivation was defined by the fuzzy algorithm as 14.2 hr, and the experimental maximum activity (207 IU.L[-1]) occurred around 16 hr. For Run #3 (Figure 3c), these times were 15.5 hr and 16 hr, respectively. Maximum activity was 200 IU.L[-1].

Three validation experiments were used to verify the robustness and efficiency of the algorithm. The algorithm was running on-line and identified the end of cultivation at the times shown in Figure 4. Nevertheless, the runs continued up to approximately 40hr to provide extra data that would confirm the adequacy of the fuzzy inference. Of course, an actual industrial reactor could have finished its run when the maximum production of enzyme was reached. The fuzzy algorithm was successful in its predictions, always providing time estimations of maximum enzyme activity within the correct experimental time span.

It should be stressed that in Run #6, the enzyme concentration at 15 hr (close to the indication of the fuzzy system) was approximately 90% of the experimental maximum. Taking into account the measurement errors and delays this is actually a very satisfactory result.

(a)

(b)

(c)
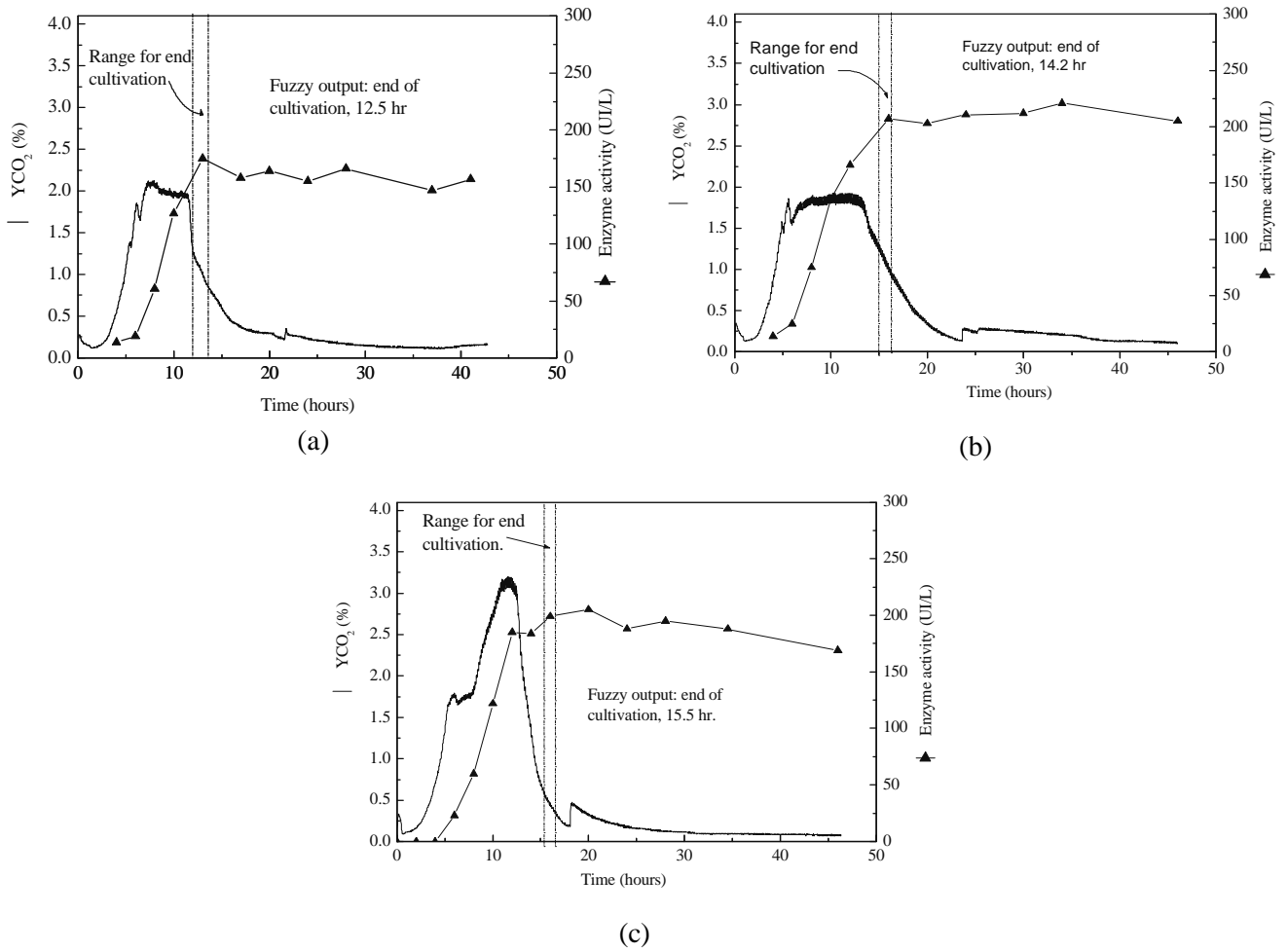
Figure 3: Training data set for tuning the fuzzy algorithm: (a) Run #1 (b) Run #2 (c) Run #3.

(a)

(b)
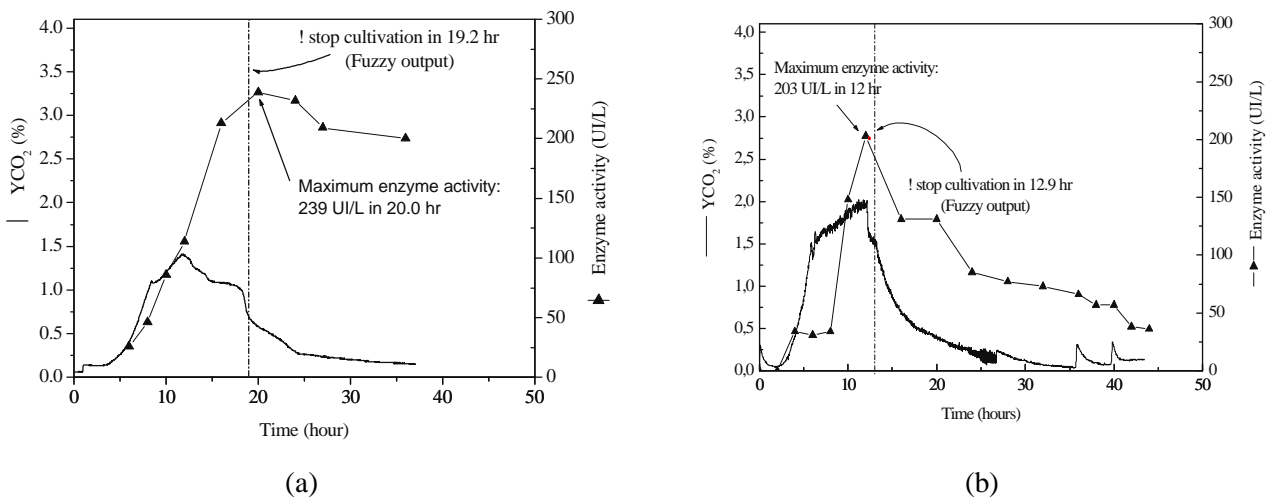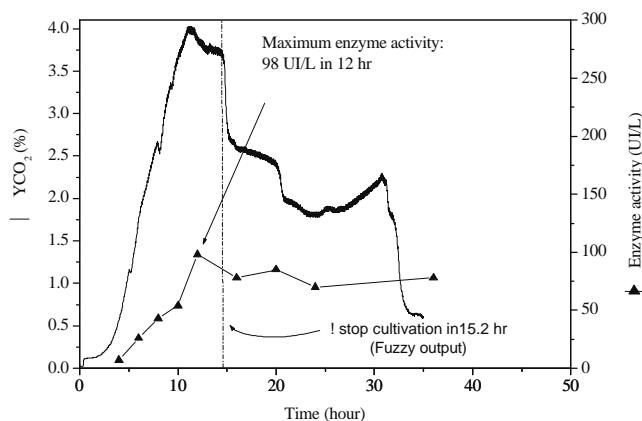
(c)

Figure 4: Validation of the fuzzy algorithm: (a) Run #4, estimated stop time 19.2 hr, maximum experimental enzyme activity 20 hr (240 IU.L^{-1}) (b) Run #5, estimated stop time 12.9 hr, maximum experimental enzyme activity 12 hr (203 IU.L^{-1}) (c) Run #6, estimated stop time 15.2 hr, maximum experimental enzyme activity 12 hr (98 IU.L^{-1}).

CONCLUSIONS

The fuzzy algorithm was validated by real experiments, accurately estimating the threshold for harvesting the bioreactor in three independent experiments which had very different behaviors. With only one measurement (molar fraction of CO_2) in the exhaust gases and the time, it was possible to identify regions of maximum enzyme activity for situations not used to tune the algorithm. Its capacity for generalization was, therefore, consistently established, providing a very positive indication of its reliability.

This approach circumvents the need for elaborate mathematical models of the complex biological system and allows the use of qualitative information as linguistic rules. The "defuzzification" of the fuzzy output may provide a degree of certainty for the decision to stop cultivation.

Certainly, phenomenological models are still important, but our conclusion in this specific case is that the computational intelligence "black box" approach provided excellent results.

ACKNOWLEDGEMENTS

REFERENCES

Acevedo, F. and Cooney, C.L., Penicillin Amidase Production by *Bacillus megaterium*. Biotechnology and Bioengineering, 15, pp. 493-503 (1973).

Alfafara, C.G., Miura, K., Shimizu, H. Shioya, S., Suga, K. and Suzuki, K., Fuzzy Control of Ethanol Concentration and Its Application to Maximum Glutathione Production in Yeast-batch Culture. Biotechnology and Bioengineering, 41, pp. 493-501 (1993).

Balasingham, K., Warburton, D. Dunnil, P. and Lilly, D., The Isolation and Kinetics of Penicillin Amidase from *Escherichia Coli*. Biochimica ET Biophysica Acta, 250-256 (1972).

Chéruy, A., Software Sensors in Bioprocess Engineering. Journal of Biotechnology, 52, pp. 193-199 (1997).

Gentina, J.C., Acevedo, F. and Villagra, M.P., Short Communication: Effect of Complex Nitrogen Sources on the Production of Penicillin Acylase by *Bacillus megaterium*. World Journal of Microbiology and Biotechnology, 13, pp. 127-128 (1997).

Gomide, F.A.C. and Godwin, R.R., Modelagem, controle, sistemas e lógica fuzzy. Controle e Automação, 4(3),pp. 97-115 (1994).

Illanes, A., Acevedo, F., Gentina, J.C., Reyes, I., Torres, R., Cartagena, O. and Ruiz, M. A., Production of Penicillin Acylase from *Bacillus megaterium* in Complex and Defined Media. Process Biochemistry, 29, pp. 263-270 (1994).

Linko, S., Luopa, J. and Zhu, Y.–H., Neural Networks as "Software Sensors" in Enzyme Production. Journal of Biotechnology, 52, pp. 257-266 (1997).

Ospina, S.S., Characterization and Use of a Penicillin Acylase Biocatalyst. Journal of Chemical Technology and Biotechnology, 53, pp.

205-214 (1992).

Pinotti, L.M., Silva, A.S., Silva, R.G and Giordano, R.L.C., Study of Different Media for Production of Penicillin G Acylase from *Bacillus megaterium* ATCC 14945. Applied Biochemistry and Biotechnology, 84-86, pp. 655-663 (2000).

Shiba, S., Nishida, Y., Parque, Y.S., Iijima, S. and Kobayashi, T., Improvement of Cloned α-Amilase Gene-expression in Fed-batch Culture of Recombinant *Saccharomyces cerevisiae* by Regulation Both Glucose and Ethanol Concentrations Using a Fuzzy Controller. Biotechnology and Bioengineering, 44, pp. 1055–1063 (1994).

Silva, R.G., Souza, V.R., Nucci, E.R., Pinotti, L.M., Cruz, A.J.G., Giordano, R.C., Giordano, R.L.C., Using a Medium of Free Amino Acids to Produce Penicillin G Acylase in Fed-Batch Cultivations of *Bacillus megaterium* ATCC 14495. Submitted to Brazilian Journal of Chemical Engineering. (August, 2004).

Sousa Jr, R. and Almeida, P.I.F., Design of Fuzzy System for the Control of a Biochemical Reactor in Fed-batch Culture. Process Biochemistry, 37, pp. 461-469 (2001).

Souza, V.R., Pinotti, L.M., Inocêncio, G.E.M. and Giordano, R.L.C., Influence of Amino Acids Composition in the Production of Penicillin G Acylase by *Bacillus megaterium* ATCC 14945. 7th SHEB (Seminar on Enzymatic Hydrolysis of Biomass), Maringá, Brazil. (2002).

Yamada, Y., Haneda, K., Murayama, S. and Shiomi, S., Application of Fuzzy System to Coenzyme Q10 Fermentation. Journal of Chemical Engineering Japan, 24, pp. 94-99 (1991).

von Altrock, C., Fuzzy Logic and Neuro Fuzzy Applications Explained. Prentice Hall Inc., Upper Saddle River, New Jersey, USA (1995).