MODEL-BASED PREDICTION OF PRODUCT FORMATION FOR PENICILLIN FERMENTATION

J.Q. Yuan[a], M.J.C. Lauwerijssen[b], L. Sang[c], Y.Q. Shen[d] and P.A. Vanrolleghem[e]

[a] *Department of Automatic Control, Shanghai Jiaotong University, 1954 Huashan Road, 200030 Shanghai, PR China. Email: jqyuan@mail.sjtu.edu.cn*
[b] *Food and Bioprocess Engineering Group, Wageningen University, P.O. box 8129, 6700 EV Wageningen, The Netherlands. Email:maarten.lauwerijssen@95.student.wau.nl*
[c] *Department of Mathematics, Dalian University, 116622 Dalian, PR China*
[d] *Harbin Pharmaceutical Company, 150086 Harbin, PR China*
[e] *BIOMATH Department, Ghent University, Coupure Links 653, B-9000 Gent, Belgium E-mail: Peter.Vanrolleghem@rug.ac.be*

Abstract: The combined cell age and kinetic model for penicillin fermentation, which was proposed earlier, was tested with 20 industrial scale charges of fed-batch cultivations. After lumped identification of the parameters, model simulations fitted the measurements well even for the widely fluctuated charges. Corresponding to the high fitting accuracy, it was found that only two parameters out of ca. twenty are charge-dependent. This encourages the further study on the model-based prediction of the most important state variable - product concentration. A rolling identification-prediction technique was used for this purpose. Pseudo on-line prediction with a predictive horizon up to 40h-ahead was carried out and the average prediction accuracy was found as high as 2-4%. *Copyright 2001 IFAC*

Keywords: Penicillin fermentation; parameter identification; prediction

1. INTRODUCTION

Penicillin fermentation belongs to one of the most widely investigated bioprocesses during last fifty years. Although the biggest improvements in titre increase from 2 u ml^{-1} of Fleming's original isolate to 70000 u ml^{-1} of modern strains owed mainly to strain mutation and selection techniques (Penalva, *et al*., 1998), process control is never less important for optimal operation in a commercial plant (Calam, 1987). In this context, process modeling has been regarded as a powerful tool. Bajpai and Reuss (1980, 1981) proposed an unstructured model for penicillin fermentation, in which biomass was considered as a lumped state variable. Further studies focused on the differentiation of biomass into metabolically distinguishing compartments, resulting more segregated models. For example, Nestaas and Wang (1983) distinguished three cell types: growing tips, non-growing but producing, and non-producing degenerated cells. Paul and Thomas (1996) divided the filamentous hyphae into four distinct regions, i.e., active growing, nongrowing or penicillin producing, vacuoles and degenerated or metabolically inactive regions. This model was well tested in prediction of the penicillin G production during fed-batch fermentation of an industrial *P. chrysogenum* strain under different glucose feeding regimes. The morphologically structured model proposed by Nielsen (1993) described penicillin fermentation in a simple way that the filamentous hyphae were divided into three morphological forms: apical, subapical and hyphal compartments. With only 14 parameters, the model simulations were well in agreement with the

fed-batch fermentations by using a high yielding strain *P. chrysogenum* (Zangirolami *et al*., 1997). Similarly, Tiller *et al*. (1994) proposed a segregated kinetic model for penicillin fermentation in which biomass was subdivided into growing, nongrowing and lysed parts. Both growing and nongrowing parts were assumed to be capable of producing penicillin. The model of Tiller *et al*. distinguished itself from others by incorporation of the mean cultivation age, which resulted in an age-dependent transition from one morphological part to another. Also, this model was well tested with the data of penicillin fed-batch fermentations. The model of Tiller *et al*. formed basis of the study shown in this article. In order to describe commercial penicillin fermentations with a titre exceeding 50000 u ml^{-1} under industrial medium compositions, the original model of Tiller *et al*. has been modified in such a way that substrate was subdivided into direct usable, hydrolysis-requiring and unusable sugars. Besides, a cell cycling model was established which delivered the average age of different cell populations. Furthermore, the penicillin formation rate was assumed to be dependent on the average age of producing cell population. This resulted in a combined age and segregated kinetic model (Yuan *et al*., 1999). Since the aeration and agitation assured the dissolved oxygen above 30% of the saturation value, the gas-liquid mass transfer was not taken into account. The combined model system contained 18 parameters. This paper dealt with the industrial application oriented aspects of it. Firstly, real industrial data were used to test the model so as to confirm that the model is structurally correct, i.e., well fitting of the model simulations to the data of large amount of historical charges both in tendency and in accuracy (in the latter case, charge-wise lumped identification of model parameters is implied). Secondly, the predictive ability - the most important property of a successful model - was tested. An accurate 40h-ahead prediction of penicillin concentration was achieved by using rolling identification - prediction (RIP) method. Detailed description of the on-line applicable RIP as well as the testing results were given.

2. VERIFICATION OF THE COMBINED AGE AND SEGREGATED KINETIC MODEL

Twenty fed-batch fermentations were carried out with the same commercial strain of *P. chrysogenum* in a bioreactor which has a volume of over 50,000 l. The inoculation ratio was ca. 10% for all cultivations. The feeding during fed-batch operation included carbon source, phenyl acetic acid and ammonium sulfate solutions. Foam was controlled by adding vegetable oil. As intermediate harvesting, the spent medium was periodically withdrawn after 60 h. No significant autolysis of biomass was observed during the scheduled cultivation period. Samples were usually taken every 4 hours for assays of penicillin concentration and every 24 hours for assays of residual sugar concentration. The dry biomass concentration was not measured. Fig. 1 shows the time course of the penicillin concentration of these industrial charges. Where, the fluctuation of product concentration at the end of fermentation was as high as 17%. For model testing, only the original recorded measurements were used throughout the text.

Off-line model testing revealed that, although there are nearly twenty model parameters in the combine age and segregated kinetic model for *P. chrysogenum* (Yuan *et al.*, 1999), the number of strongly charge-dependent parameters was only two, i.e. the maximal specific growth rate on sugar μ_{Smax} and the maximum specific penicillin formation rate ν_{max}. Table 1 lists the identification results of the relatively constant model parameters for all twenty charges. The Simplex method proposed by Nelder and Mead (1965) was then used for the lumped identification of the two charge-dependent parameters. The goal of identification was to find the best parameter combination, for which the sum of square errors between model simulated and measured penicillin and sugar concentrations, respectively, is minimized. Table 2 shows the identified charge-dependent parameters.

To quantify the model residuals, a relative fitting error for penicillin concentration at i^{th} sampling point, $e_{fit}(i)$, is calculated by Eq. (1), where, P_{ms} and P_m represent the model simulated and the measured penicillin concentration, respectively. Subsequently, a quality-index for the model validation is defined as the average of relative fitting errors $\bar{e}_{fit}$, see Eq. (2), where q is the number of measurement points.

$$e_{fit}(i) = \frac{P_{ms}(i) - P_m(i)}{P_m(i)} \tag{1}$$

$$\bar{e}_{fit} = \sqrt{\frac{\sum_{i=1}^{q} e_{fit}(i)^2}{q}} \tag{2}$$

Furthermore, the industrial-scale penicillin fermentation was divided into two phases: (1) the earlier phase of cultivation, for which T_f < 96 h and (2) the later phase of cultivation, for which $T_f \geq 96$ h. The average of relative fitting errors corresponding to these phases are differentiated as $\bar{e}_{fit,1}$ and $\bar{e}_{fit,2}$, respectively. Table 2 also shows $\bar{e}_{fit}$, $\bar{e}_{fit,1}$ and $\bar{e}_{fit,2}$ as calculated after parameter identification. For the twenty testing charges, it can be seen that the mean of $\bar{e}_{fit}$ is 7.5 %, to which the earlier phase of cultivation contributes the main part: $\bar{e}_{fit,1}$ amounts to 12.1 %. This can be mainly caused by the fluctuations of inocula both quantitatively and qualitatively as well as constant changes in the medium ingredients (Calam, 1987). However, the later phase of fermentation is described very well by the model with a mean $\bar{e}_{fit,2}$ of only 2.0 %.

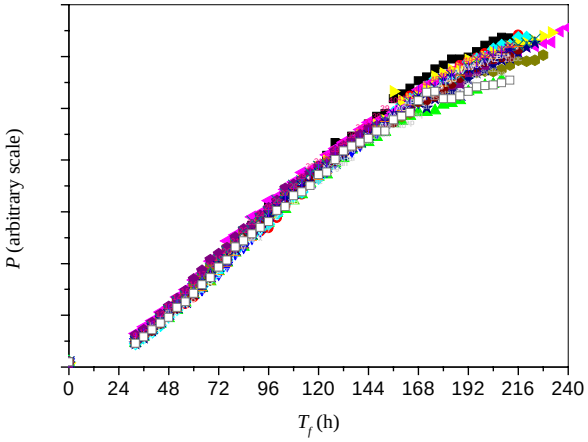

Fig. 1. Time course of penicillin concentration of twenty randomly selected industrial charges.

Table 1 Relatively constant model parameters valid for charge 1-20. Meanings of the parameters see Yuan *et al*. (1999)

Parameter	Unit	Value
f_{12}	h^{-2}	0.000085
f_{vac}	h^{-2}	0.000075
K_A	h^{-2}	0.00008
K_{PM}	g l^{-1}	2.0
K_S	g l^{-1}	0.4
K_{Ss}	g l^{-1}	0.2
K_h	h^{-2}	0.0015
m	h^{-1}	0.01
R_S	-	0.05
R_U	-	0.01
r_{sgmax}	h-1	0.0032
Y_{PS}	g g^{-1}	1.2
Y_{XS}	g g^{-1}	0.61
Y_{XPM}	g g^{-1}	0.60
μ_P	h^{-1}	0.035
μ_{PMmax}	h^{-1}	0.05

Table 2. μ_{Smax} and ν_{max} obtained by lumped Identification with the Simplex method and the resulting fitting errors for the penicillin concentration.

No	μ_{Smax} (h^{-1})	ν_{max} (h^{-1})	$\bar{e}_{fit}$ (%)	$\bar{e}_{fit,1}$ (%)	$\bar{e}_{fit,2}$ (%)
1	0.061	0.0075	9.0	14.2	1.4
2	0.069	0.0065	10.1	16.6	2.0
3	0.061	0.0074	8.1	12.3	3.6
4	0.060	0.0073	10.7	16.6	3.7
5	0.069	0.0067	5.8	8.9	1.9
6	0.061	0.0079	6.3	10.2	1.2
7	0.060	0.0079	10.2	17.0	2.3
8	0.061	0.0076	8.8	14.2	1.5
9	0.061	0.0076	11.0	18.7	1.8
10	0.061	0.0078	7.6	12.4	1.5
11	0.069	0.0064	4.7	7.3	1.9
12	0.065	0.0064	4.6	6.9	2.0
13	0.068	0.0069	8.6	14.0	2.0
14	0.068	0.0069	7.3	12.0	1.5
15	0.068	0.0065	2.8	3.8	1.9
16	0.069	0.0068	9.5	15.2	3.2
17	0.075	0.0062	7.2	12.0	1.5
18	0.069	0.0066	7.0	11.0	2.1
19	0.069	0.0066	5.1	8.6	1.0
20	0.069	0.0066	6.7	10.5	2.1
Mean	0.066	0.0070	7.5	12.1	2.0

Figs. 2 and 3 show the comparison of the model simulations of the state variables *S* and *P* and their measurements for two selected charges: one with a high fitting error (charge 4) and one with an average fitting error (charge 5). No measurements for biomass concentration *X* were plotted since they are not available. For reasons of confidentiality, real ordinate scales have been removed. These figures show that the combined model system fits the penicillin concentration and residual sugar concentration very well, which indicates the off-line follow-up ability of the model to the changing cultivation conditions. The validity of the model structure is therefore confirmed.

It is interesting to evaluate the modeled penicillin concentration plotted in Figs. 2 and 3 more elaborately. For charge 4 (see Fig. 2) the modeled concentration is too high in comparison to the measurements until 144 h and afterwards becomes too low. For charge 5 (see Fig. 3) the fitting results are fine until 168 h. Subsequently, the fitting becomes too high. Such kind of discrepancy is referred as systematic error of the model. To more or less extent, similar discrepancy is observed for most other charges (data not shown). Such a systematic error may be reduced if some parameters are adopted as time-variant and correspondingly on-line identified. This, as demonstrated in the next section, may result in an adaptive model which improves not only the

average fitting accuracy, but also the average prediction accuracy. The latter is especially emphasized for on-line applications of the model.

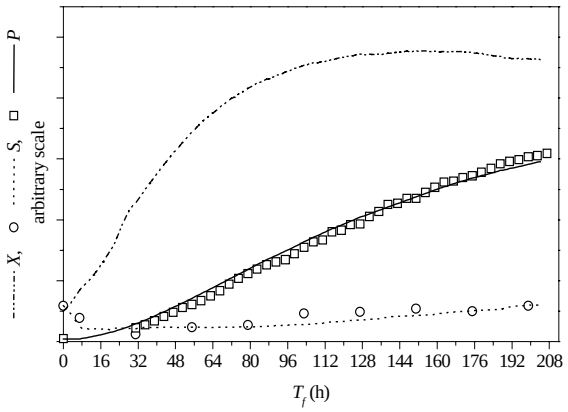

Fig.2. Simulated state variables (lines) and their measurements (symbols) for charge 4. Parameters see Table 1 and 2.

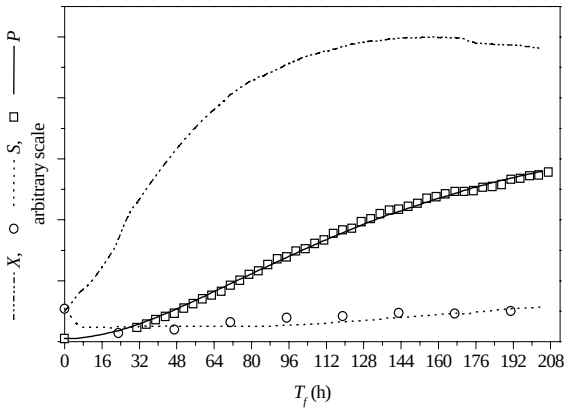

Fig.3. Simulated state variables (lines) and their measurements (symbols) for charge 5. Parameters see Table 1 and 2 .

3. MODEL-BASED PREDICTION OF PENICILLIN CONCENTRATION

3.1 Rolling Identification-Prediction

Penicillin fermentation shows different stages. Initially, penicillin formation is limited, whereas biomass concentration increases rapidly. Subsequently, there is a period in which penicillin concentration increases almost linearly. Afterwards, a decrease in product formation can be observed. During the last two stages, biomass concentration is normally rather stationary (see simulations of X in Fig.2 and 3, which has also been validated by off-line assays). The length and the transient trajectories of these stages differ from charge to charge. In the present model, the time-variant kinetics of the process was described by the time-variant and charge-dependent model-parameters μ_{Smax} and ν_{max}. To on-line identify these parameters, a variation of the rolling identification-optimization technique proposed by Guo *et al.* (1995) was introduced, the so-called RIP - rolling identification-prediction. This method implies the use of the moving data windows' technique. The available measurements of penicillin and residual sugar concentration in the data window with width T_D were used for the parameter re-identification. Determination of T_D is a compromise between the ability to follow process dynamics of the model and the reliability of the re-identified parameters; shorter T_D results in more quickly responding parameters, but also higher influence of measurement noise and vice versa. Each data window was accompanied by a prediction window with a width T_P, which had a maximum horizon of 40 h. This maximum prediction horizon was subdivided into 5 steps with a step length of 8 h, i.e. prediction of 8 h, 16 h, 24 h, 32 h and 40 h ahead corresponding to 1st through 5th step, respectively. For industrial applications, the wider prediction horizon of $T_P > 24$ h is particularly interesting for optimal scheduling purposes. Since the actual penicillin concentration was sampled every 4 h in industrial fermentation, moving the data window with accompanying prediction window forward with a shift-time T_M of 4 h assures that the data window contains the latest available information for parameter re-identification. For a historical charge with a fermentation time of T_f, the maximal number of RIP-cycles, N, can be calculated by using Eq. (3).

$$N = \frac{T_f - T_D - T_P}{T_M} + 1 \qquad (3)$$

In addition to the two parameters already mentioned, the initial penicillin concentration of the k^{th} data window $P|_{T_{D(k,0)}}$ was also included in re-identification instead of taking the measurement directly, so as to avoid a systematic error caused by possibly great measurement noise. Besides the penicillin concentration, each data window needs an initial state that describes all other state variables in the beginning. The initial state for the k^{th} data window consists of an array of the state variables, such as segregated biomass concentrations and their average age, substrate concentration and liquid volume in the fermenter. For the first data window, these variables took the measurements or pre-set values. Subsequently, the values for the k^{th} data window were obtained by taking the values of model simulation with the re-identified parameters in the $(k-1)^{th}$ data window at the relevant point of time. The input-variables needed for model calculation were feeding rate of substrates and the withdrawal rate. The values for these variables can be directly read from the industrial feeding and withdrawal files.

For the k^{th} prediction window, the initial state takes the model-simulated values at the right border of the k^{th} data window. The input-variables (feeding rate of substrates and the withdrawal rate) should be provided in advance for model simulation. However, for a charge in progress, the actual values of these variables are not yet available. In this case, they are estimated by taking the average values of the feeding rates during the latest 8 hours. This is basically in accordance with the practical situation since the feeding rates were quite constant after 60 h of cultivation. Withdrawal is usually carried out periodically, so on the basis of the last point of time of the recorded withdrawal, withdrawal during the prediction period is scheduled. The whole procedure described above is schematically depicted in Fig.4. All the other parameters in the kinetic model except for μ_{Smax} and ν_{max} took the values as listed in Table 1. Again, the Simplex method was used for the rolling identification of the two parameters as well as the initial penicillin concentration for each data window. The goal of identification was to minimize the sum of square errors between model simulated and measured penicillin and sugar concentration, respectively, in the related data window.

For quantitative evaluation of the prediction, an average relative prediction error is introduced. The relative prediction error of penicillin concentration that is predicted m-step ahead in the k^{th} RIP-cycle is specified as $e_{pred}(k,m)$, see Eq. (4). The average of relative prediction errors corresponding to the predictions m-step ahead in q RIP-cycles is defined as $\bar{e}_{pred}(m)$, see Eq. (5).

$$e_{pred}(k,m) = \frac{P_{pred}(k,m) - P_m(k,m)}{P_m(k,m)} \qquad (4)$$

$$\bar{e}_{pred}(m) = \sqrt{\frac{\sum_{k=1}^{q} e_{pred}(k,m)^2}{q}} \qquad (5)$$

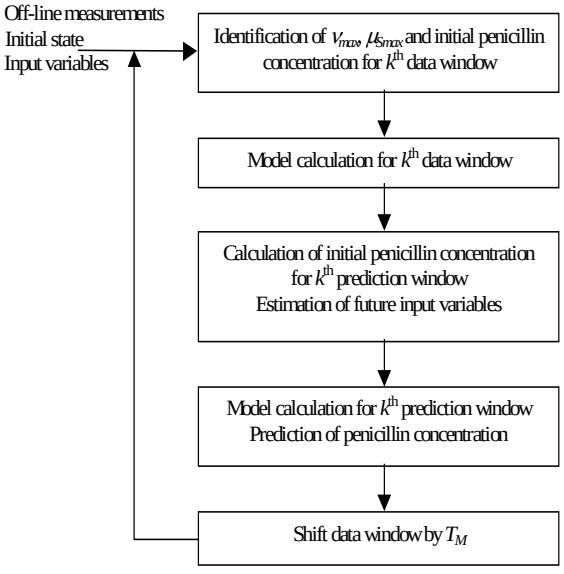

Fig.4. Rolling identification-prediction (RIP) outline for penicillin fed-batch fermentation.

3.2 Prediction Results

For the industrial fermentation data available, the optimal width of the data window T_D was found to be 56h. Table 3 shows the average prediction errors for the first ten charges by using the RIP-method as represented in Fig. 4. It is clear from the table that the average prediction error is very low even at the widest prediction horizon of 40 h (in that case, the mean of $\bar{e}_{pred}(5)$ is 3.2 %). In Fig. 5, 32h-ahead prediction of penicillin concentration is compared to the measurements for three charges: charge 1 with a high prediction error of 3.2 %, charge 9 with an average prediction error of 2.5 % and charge 10 with a low prediction error of 1.4 %. The predicted lines start from 88 h (T_D + T_P = 56 h + 32 h). The prediction error of the rest ten charges was found to be comparable with that of the first ten charges (data not shown).

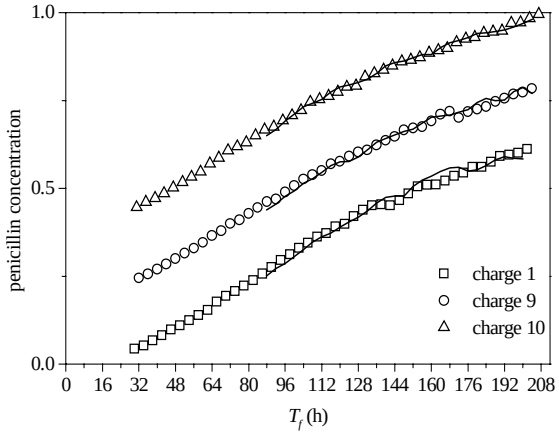

Fig.5. Comparison between rolling identification-prediction (32 h ahead) and measurements of penicillin concentration for charges 1, 9 and 10. Symbols are measured data, lines are predictions. Charges 9 and 10 are plotted respectively 0.2 and 0.4 arbitrary units higher for clarity.

Table 3 Average of relative prediction errors with RIP-method (%)

Charge No	Prediction horizon (h) +8	+16	+24	+32	+40
1	2.0	2.3	2.5	3.2	4.1
2	1.3	1.7	2.3	2.7	2.9
3	2.6	2.6	2.6	3.2	3.6
4	1.9	1.9	2.0	2.7	3.6
5	1.4	1.7	2.1	2.6	3.0
6	1.5	1.6	1.9	2.1	3.2
7	1.6	1.8	2.1	2.3	3.2
8	1.5	1.4	1.8	2.0	2.4
9	1.9	1.9	2.2	2.5	3.5
10	1.3	1.5	1.1	1.4	2.4
Mean	1.7	1.8	2.0	2.5	3.2

4. DISCUSSION AND CONCLUSION

The combined age and segregated kinetic model proposed earlier has been confirmed to be able to describe industrial-scale penicillin fermentation well. This was proven by low average fitting error of 2.0 % during the second half of fermentation (see $\bar{e}_{fit,2}$ in Table 2). Then, for the purpose of on-line applications, model-based prediction of the penicillin concentration was investigated. Rolling identification-prediction (RIP) method has been shown to be an effective way for accurate and reliable prediction. Using RIP, the charge-dependent and time-variant parameters μ_{Smax} and ν_{max} were on-line identified with the data during the latest 56h. Then, based on the re-identified model parameters, prediction of penicillin concentration up to 40h-ahead was generated. The average prediction accuracy was typically 2.0-4.0%.

Implementation of the RIP-prediction software requires reasonable computation time. On a Pentium 233-computer with 32 MB RAM, each RIP-cycle took about 30 seconds. The prediction software is also convenient in use. For each charge, three files should be supplied: an initial state file, a feeding and withdrawal file and a file with the sugar and penicillin measurements. The initial state file contains information about state variables such as sugar and penicillin concentration, starting volume *etc*. at the beginning of fermentation. The feeding and withdrawal file contains the most up-to-date information about the feeding rates and concentrations and the withdrawal in the latest data window. The file containing the off-line sampling assays is needed for parameter identification and must be updated each time new measurements become available. Therefore, the latter two files should be updated on-line. The output files of the software contain information about the predicted penicillin concentration and the profit function. The profit function is an economic performance index which is closely related to the product concentration. Detailed description of the profit function will be found elsewhere (Yuan *et al*., 1997). The predicted profit function forms the basis for optimal scheduling.

5. ACKNOWLEDGEMENTS

This work is in the frame China-Flanders Scientific Corporation Program. JQ Yuan is grateful to Natural Science Foundation of China (No. 68774021) and the Open Project Program of the State Key Laboratory of Bioreactor Engineering/China for the financial support.

6. REFERENCES

Bajpai, R.K. and M. Reuss (1980). A mechanistic model for penicillin production. *J. Chem. Tech. Biotechnol.*, **30**, 332-224.

Bajpai, R.K. and M. Reuss (1981). Evaluation of feeding strategies in carbon-regulated secondary metabolite production through mathematical modeling. *Biotechnol. Bioeng*., **23**, 717-738.

Calam, C. T. (1987). *Process development in antibiotic fermentations*, pp. 149-210. Cambridge Unversity Press.

Guo S.R., J.Q. Yuan and K.-H. Bellgardt (1995). Simulation of profit optimization for baker's yeast continuous production. In: *Computer Applications in Biotechnology, IFAC Proc. Series* (A Munack and K. Schuegerl. (Eds)), pp. 255-258. Pergamon Press.

Nelder, R., J.A. Mead (1965). A simplex method for function minimization. *Comput. J.*, **7**, 308-313.

Nestaas, E. And D.I.C. Wang (1983). Computer control of the penicillin fermentation using the filtration probe in conjunction with a structured process model. *Biotechnol. Bioeng*., **25**, 781-796.

Paul, G.C. and C.R. Thomas (1996). A structured model for hyphal differentiation and penicillin production using *Penicillium chrysogenum*. *Biotechnol. Bioeng.*, **51**, 558-572.

Penalva, M. A., R. T. Rowlands and G. Turner (1998). The optimization of penicillin biosythesis in fungi. *Trends in Biotechnology*, **16**, 483-489.

Nielsen, J. (1993). A simple morphologicaly structured model describing the growth of filamentous microorganisms. *Biotechnol. Bioeng.*, **41**, 715-727.

Tiller, V., J. Meyerhoff, D. Sziele, K. Schuegerl, K.-H. Bellgardt (1994). Segregated mathematical model for the fed-batch cultivation of a high-producing strain of Penicillin chrysogenum. *J. Biotechnol.*, **34**, 119-131.

Yuan J.Q., S.R. Guo, K.-H. Bellgardt and K. Schuegerl (1997). Profit optimization for mycelia fed-batch cultivation, *J. Biotechnol.*, **54**, 175-193.

Yuan J.Q., P.A. Vanrolleghem, Y.Q. Shen and K.-H. Bellgardt (1999). Combined age and segregated kinetic model for penicillin fed-batch cultivation. In: *Proceedings of the 14th World Congress of IFAC*, Vol. O, pp 415-420.

Zangirolami, T. C., C. L. Johanson, J. Nielsen, S. B. Jorgensen (1997). Simulation of penicillin production in fed-batch cultivations using a morphologically structured model. *Biotechnol. Bioeng.*, **56**, 593-604.