*BIO*AUTOMATION, 2009, **13** (4), 197-210

Mathematical Modelling of the Process for Microbial Production of Branched Chained Amino Acids

Georgiev T.[2], Todorov K.[1], Ratkov A.[1*]

[1] *Institute of Microbiology, Bulgarian Academy of Science*
26 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
E-mail: ratkoval@microbio.bas.bg
[2] *Technical University of Sofia, Faculty of Automatics*
Automation of Continuous Processes
8 Kliment Ohridski St., 1000 Sofia, Bulgaria

Summary: This article deals with modelling of branched chained amino acids production. One of important branched chained amino acid is L-valine. The aim of the article is synthesis of dynamic unstructured model of fed-batch fermentation process with intensive droppings for *L-valine* production. The presented approach of the investigation includes the following main procedures: description of the process by generalized stoichiometric equations; preliminary data processing and calculation of specific rates for main kinetic variables; identification of the specific rates takes into account the dissolved oxygen tension; establishment and optimisation of dynamic model of the process; simulation researches. MATLAB is used as a research environment.

Keywords: Modelling, Optimisation, Branched chained amino acids, Fed-batch process with droppings, L-Valine

1. INTRODUCTION

In terms of market volume, development over the last 20 years has been tremendously bullish in the so-called feed amino acids *L-lysine*, *L-valine*, *DL-methionine*, *L-threonine*, and *L-tryptophan*. These four amino acids constitute the largest share of total amino acid market, estimated in 2005 at approximately US $ 4.5 billion (Fig. 1) [4]. Recently, much attention has been paid to development of technologies for production of other amino acids including *L-valine*. As expected, the bioprocess technologies for their production were significantly improved by application of different cultivation methods with more sophisticated feeding rates and profiles and new automation strategies [3].

[*] *Corresponding author*

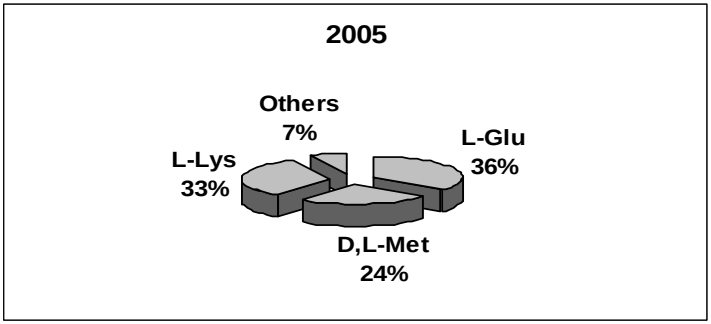

Fig. 1 Amino acids world supply

The synthesis of mathematical models for biotechnological processes in principal is known to be the major task for the application of modern control science for their optimization. The models normally involve two kinds of parameters: the yield coefficients, which rely on the structure of the generalised stoichiometric reactions, and the kinetic rates, which rely on the specific metabolism pathways [1].

Some approaches for dynamic modelling of *L-valine* fed-batch fermentation process have been described in our previous articles [2, 10] including modelling of fed-batch fermentation process with droppings. This article aims to present an approach for development of dynamic unstructured model for *L-valine* fed-batch fermentation process with intensive droppings of the culture broth as well as investigation of the specificity of the process and its reflection on the obtained mathematical model. This approach presents extended variant of generalized stoichiometric equations and takes into account the dropping conditions by sequential reactions. The specific rates: growth rate (μ), substrate utilization rate (ν), production rate (ρ) and dissolved oxygen utilization rate (γ) are estimated by optimisation procedure taking into account the real life experimental data.

Identification procedure, applied to estimation of the model structure and coefficients, taking into consideration the specificity concerning dropping procedure. The important stage of this procedure is the parametric optimisation of the model. The procedures for identification, optimisation and simulation researches are realized by MATLAB and STAGRAPHICS packages [5-9].

Main approaches and steps, used for development of mathematical models, are described in details in our previous articles [2, 10].

2. EXPERIMENTAL RESULTS

Experimental procedures and analysis

The variable volume fed-batch fermentation process is carried out at laboratory scale fermentor with 7 litres of total volume. *Corynebacterium glutamicum B-023* is used as a producer. Analytical methods used for the characterisation of the process are as follows: biomass is measured as dry cell mass, $g \cdot l^{-1}$; sugar concentration – as reducible compounds, $g \cdot l^{-1}$; *L-valine* – by chromatographic method, $g \cdot l^{-1}$. During the process on-line measurements of differed physical-chemical variables are done by proper sensors.

The experimental data are shown in Fig. 2. Dissolved oxygen concentration $mMol \cdot l^{-1}$ is denoted as DO in Fig. 2.

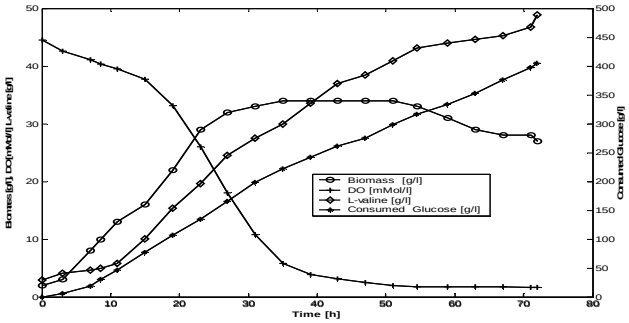

Fig. 2 Time course of the experimental data

Primary data processing

The primary data processing of the experimental data includes the following stages: transformation the different measurable units to unit $g \cdot l^{-1}$ taking into account changes (as strict balance) of the working volume during the process; calculation of the specific rates: growth rate (μ) h^{-1}, substrate utilization rate (ν) h^{-1}, production rate (ρ) h^{-1}, dissolved oxygen utilization rate (γ) h^{-1}. To reach the final aim of primary data processing - calculation of the specific rates –

the appropriate optimal approximation of the experimental data is done by spline functions. The derivatives of variables describing the process kinetics are also calculated. The specific growth rate, the rate of substrate utilisation, and the specific production rates are calculated by the equations:

$$\hat{\mu} = \frac{\dot{X}_t}{X_t}, \hat{v} = \frac{\dot{S}_c}{X_t}, \hat{\rho} = \frac{\dot{L}_{vt}}{X_t}, \hat{\gamma} = \frac{\dot{C}}{X_t} \tag{1}$$

where: X_t – total biomass produced at proper time, expressed as g·l^{-1}; S_c – total sugar consumed at proper time expressed as g·l^{-1}; L_{vt} – total *L-valine* produced at proper time expressed as g·l^{-1}; C – dissolved oxygen concentration expressed as mMol·l^{-1}, and derivatives of these variables denotes as follows $\dot{X}_t, \dot{S}_c, \dot{L}_t, \dot{C}$. The total biomass produced (X_t) in calculation of the specific dissolved oxygen utilization rate (γ) is expressed as mMol·l^{-1}.

Formulas (1) present an initial estimation of the values and shapes of specific rates. The specific rates must be calculated subject to initial and boundary conditions. These conditions are reflected to the inner behaviour of the kinetics variables.

3. BASIC RESULTS

The identification procedure includes the following main stages:

- determination of the set of generalized stoichiometric equations;
- identification of the specific rates;
- description of the dynamic model subject to dropping conditions;
- parametric optimisation of the obtained model, and
- simulation.

Generalized stoichiometric equations

The fermentation processes could be described by the following scheme of generalized stoichiometric equations:

$$
\begin{aligned}
&S_C \xrightarrow{\varphi_X} X \\
&C + S_C \xrightarrow{\varphi_S} X \xrightarrow{\varphi_{OUT}} X \\
&S_R \xrightarrow{\varphi_S} S_C \xrightarrow{\varphi_{OUT}} S_C \\
&C + S_C + X \xrightarrow{\varphi_{LV}} L_V \xrightarrow{\varphi_{OUT}} L \\
&V_0 \xrightarrow{\varphi_F} V_f \xrightarrow{\varphi_{OUT}} V
\end{aligned}
\tag{2}
$$

where: φ_X, φ_G, φ_S, φ_L, φ_F, φ_{OUT} are rates of the generalized stoichiometric reactions, $g \cdot l^{-1} \cdot h^{-1}$; V_0 – initial volume of the culture broth, V_f – final volume of the culture broth, l; X – biomass concentration, $g \cdot l^{-1}$; S – substrate concentration as a sugar remain concentration – *Sr* or sugar consumed concentration – *Sc*, $g \cdot l^{-1}$; L_V – *L-valine* concentration, $g \cdot l^{-1}$; C – dissolved oxygen concentration (DO), $mMol \cdot l^{-1}$.

The hypotheses concerning the specific rates of the amino acids biosynthesis are utilised as follows:

$$\hat{\mu} = \hat{\mu}(S_C, C), \hat{v} = \hat{v}(\hat{\mu}, X), \hat{\rho} = \hat{\rho}(\hat{\mu}, X), \hat{\gamma} = \hat{\gamma}(\hat{\mu}, X, C) \tag{3}$$

The proposed hypothesises are revealed below.

Dynamic model of the fed-batch process with dropping

The dynamic model subject to material balance of the process could be describe as follows

$$\frac{d\xi}{dt} = \mathbf{K}\varphi(\xi) - D_L^{IN}\xi + F - D_L^{OUT}\xi \tag{4}$$

$$y = \mathbf{P}\xi$$

where $\boldsymbol{\xi}^T = [X, S_R, Sc, L_V, C, V]^T$ – state space vector; $\boldsymbol{\varphi}^T = [\varphi_X, \varphi_G, \varphi_S, \varphi_{Lv}, \varphi_F, \varphi_{OUT}]^T$ – vector of reaction's rates; $\mathbf{F}^T = [0, 0, f_1, 0, f_2, 0]^T$ – vector of input streams with elements $f_1 = D_L^{IN} S_{IN}$ and $f_2 = Q(t)(C^* - C) + (D_L^{IN} + D_L^{OUT})C$; $\mathbf{P}$ – matrix; y – measurable output from sensors; S_{IN} – concentration of the feeding solution, $g \cdot l^{-1}$, $\mathbf{K}$ – matrix of yield coefficients as follows:

$$\mathbf{K}=\begin{bmatrix} k_{11} & k_{12} & 0 & -k_{14} & 0 & \pm k_{16} \\ 0 & 0 & -k_{23} & 0 & 0 & 0 \\ -k_{31} & -k_{32} & k_{33} & -k_{34} & 0 & \pm k_{36} \\ 0 & 0 & 0 & k_{44} & 0 & \pm k_{46} \\ 0 & -k_{52} & 0 & -k_{54} & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{65} & \pm k_{66} \end{bmatrix}$$

$$\mathbf{P}=\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Identification procedure includes the following stages:

- Model structure identification.
- Estimation of the coefficients, which are expressed as yield and reaction rate coefficients.

The model structure identification is obtained by reduction of the state space vector by one variable $\boldsymbol{S_R}$ and the following substitutions:

$$\begin{aligned} & k_{11}\varphi_X + k_{12}\varphi_G - k_{14}\varphi_{L_V} = K_1\mu(S,C)X \;;\; k_{23}=0 \\ & -k_{31}\varphi_X - k_{32}\varphi_G + k_{33}\varphi_S - k_{34}\varphi_{L_V} = K_2\nu(\mu,X)X \\ & k_{44}\varphi_{L_V} = K_3\rho(\mu,X)X \\ & -k_{52}\varphi_G - k_{54}\varphi_{L_V} = -K_4\gamma(\mu,X,C)X \\ & k_{65}\varphi_F = 2D_L^{IN}V \end{aligned} \tag{5}$$

The rates of the generalized stoichiometric reactions are reviled by the specific rates as it is shown by the equalities (5). The models of the specific rates are discussed below.

Dropping conditions

It is assumed that at discrete time moments of the dropping, the derivatives of the kinetic variables are equal to zero. The dropping conditions subject to model (4) – (5) are satisfied as follows:

- Dropping conditions for growth

$$F_{OUT} = K_1 \mu V(t_k) - F_{IN} \tag{6}$$

- Dropping conditions for L-lysine production

$$F_{OUT} = K_3 \rho \frac{X(t_k)}{L(t_k)} V(t_k) - F_{IN} \tag{7}$$

- Dropping conditions for substrate utilization

$$F_{OUT} = K_2 v \frac{X(t_k)}{S(t_k)} V(t_k) + F_{IN} \left(\frac{S_{IN} - S_C(t_k)}{S_C(t_k)} \right) \tag{8}$$

The comparison of the expressions (7) - (8) obtains the equality

$$K_3 \rho \frac{X(t_k)}{L(t_k)} = K_2 v \left(\frac{X(t_k)}{S_C(t_k)} \right) + \frac{F_{IN}}{V} \left(\frac{S_{IN}}{S_C(t_k)} \right) \tag{9}$$

It could be emphasized that these conditions are satisfied at the discrete time moments.

Models of the specific rates

The identification procedure of the specific rates includes the following steps:

- Application of linear regression for selection of suitable set of predictors and initial estimates of the parameters.
- Non-linear regression application for final parameter estimation.
- Simulation in MATLAB environmental.

The results of the identification procedure are derived by **STATGRAPHICS** facilities as follows.

Table 1. Models of the specific rates

Model	R^2	
$\hat{\mu} = EXP(a_0 + a_1 S_C + a_2 S_C^2 + a_3 C + + a_4 C^2 + a_5 C^3 + a_6 (S_c C^2) + + a_7 (S_c^2 C) + a_8 (S_c C^3) + a_9 (S_c^3 C))$	R^2 = 0.740	(10)
$\hat{v} = EXP(b_0 + b_1 \mu + b_2 \mu^2 + b_3 \mu^3 + + b_4 X + b_5 X^2 + b_6 X^3 + b_7 X^4$	R^2 = 0.800	(11)
$\hat{\rho} = EXP(c_0 + c_1 \mu + c_2 \mu^2 + c_3 \mu^3 + + c_4 \mu^4 + c_5 X + c_6 X^2 + c_7 X^3 + + c_8 X^4 + c_9 (\mu X^2) + c_{10} (\mu^2 X))$	R^2 = 0.995	(12)
$\hat{\gamma} = \exp(f_0 + f_1 \mu + f_2 \mu^2 + f_3 \mu^3 + + f_4 X + f_5 X^2 + f_6 X^3 + f_7 C + + f_8 C^2 + + f_9 C^3)$	R^2 = 0.991	(13)

The acceptation of the specific rate models is based on the value of the determination coefficient ($\mathbf{R^2}$) obtained by non-linear regression. Investigation of the residuals shows that the majority of them are normally distributed but the serial correlation could be seen.

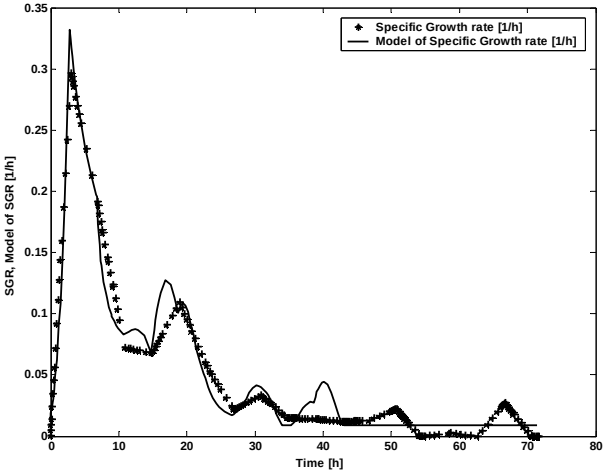

Fig. 3 Time course of the specific growth rate

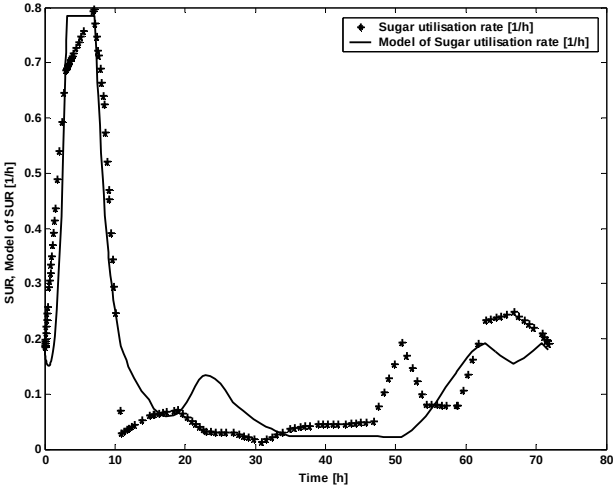

Fig. 4 Time course of the specific substrate utilisation rate

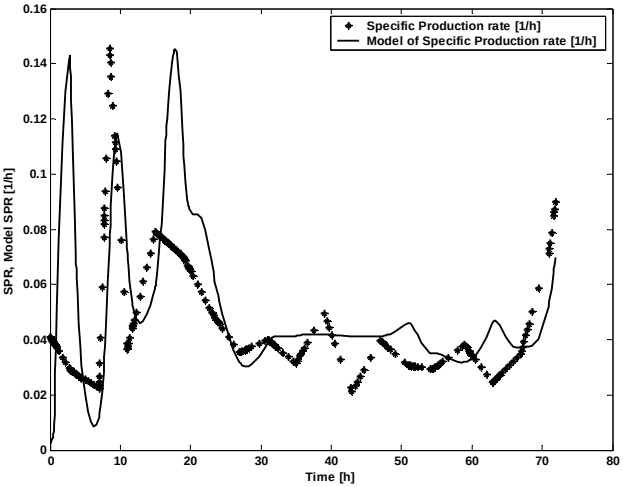

Fig. 5 Time course of the specific production rate

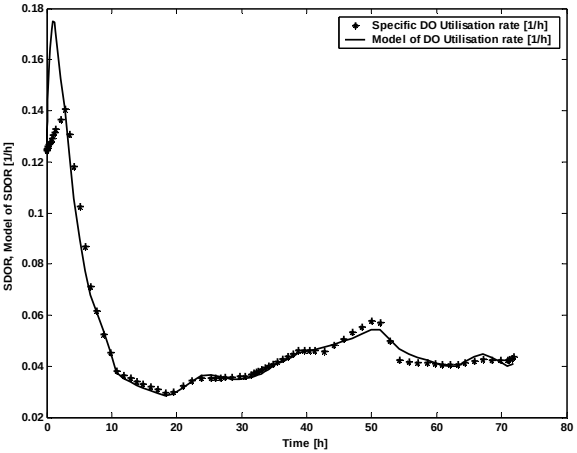

Fig. 6 Time course of the specific DO utilisation rate

Optimization and simulation

The second stage of this procedure is connected with the parametric optimization of the model. The Levenberg – Marquardt and trust-region algorithms with least squares objective function are used for optimization. The results are shown in Table 2.

Table 2. Estimated parameters

K_1 = 1.0999987654321	K_3 = 0.876309112534133
a_0 = -945.63131569056	c_0 = -0.157374725070751
a_1 = 6.89345283456696	c_1 = -139.959162072509
a_2 = -0.0126724764083549	c_2 = 3438.7037377048
a_3 = 2035.6832209993	c_3 = -16288.1457419416
a_4 = -1094.4874073787	c_4 = 23118.302418634
a_5 = -9.75247928506224	c_5 = -2.00499400225319
a_6 = -0.382479938032688	c_6 = 0.230698115470715
a_7 = 0.0153188509816506	c_7 = -0.00878436378243785
a_8 = 2.89744029830562	c_8 = $10.7703913641139 \times 10^{-5}$
a_9 = $-1.90056261571754\ 10^{-5}$	c_9 = 0.0943875758578959
K_2 = 0.55	c_{10} = -55.3316752540464
b_0 = -4.0056815196058	K_4 = 3.7963175784579
b_1 = -9.36254877866809	f_0 = 2.94735646406437
b_2 = 60.8258709464606	f_1 = 5.60837948712245

b_3 =	-77.8933745418553	f_2 =	-44.8301170576686
b_4 =	1.5038299287602	f_3 =	76.1872394355565
b_5 =	-0.176369912283121	f_4 =	-0.0676848581735653
b_6 =	0.00733311171420185	f_5 =	$9.680361319374\ 10^{-5}$
b_7 =	-0.000107756907072379	f_6 =	$5.98817266274896\ 10^{-7}$
		f_7 =	-0.136981833222773
		f_8 =	0.00955220039012647
		f_9 =	-0.000194754080760628

During the parametric optimization the experimentally established optimal modes of the feeding rate and oxygen saturation are applied. The concentration S_{IN} is also obtained as a function by an optimization based on the experimental data.

The results are shown below.

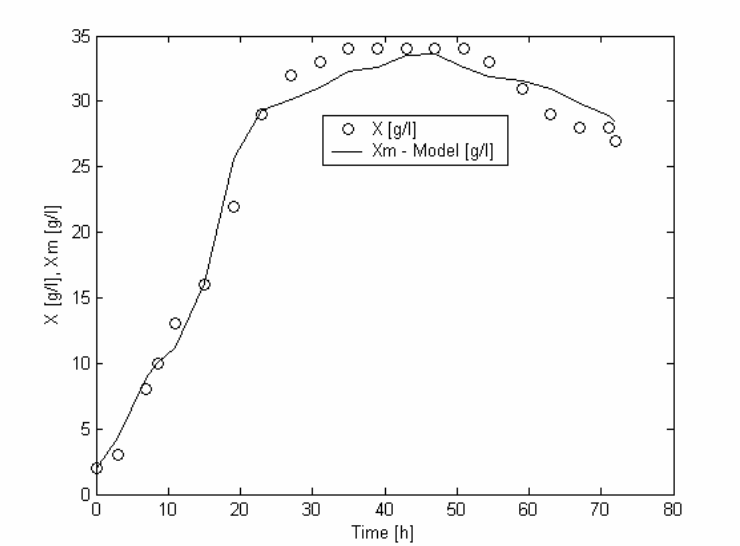

Fig. 7 Time course of the experimental biomass concentration (*X*) and model (*Xm*)

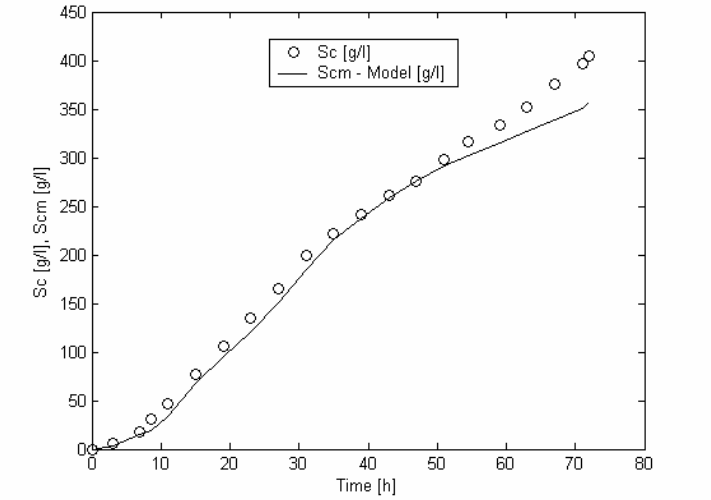

Fig. 8 Time course of the sugar consumed concentration (*Sc*) and model (*Scm*)

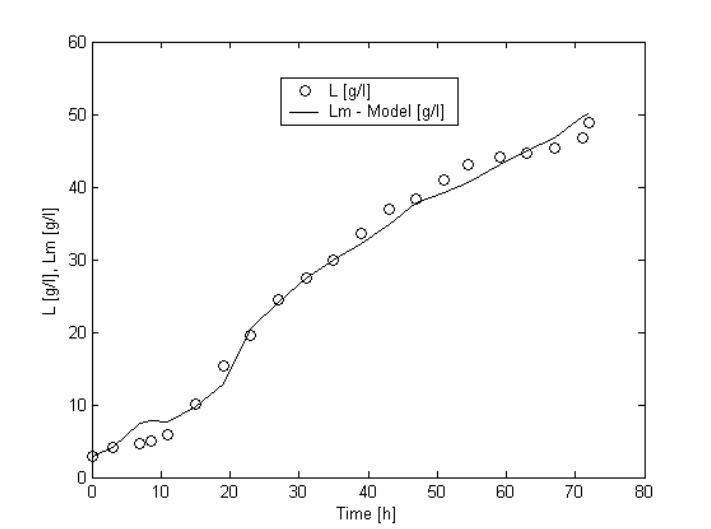

Fig. 9 Time course of the L-valine concentration (*L*) and model (*Lm*)

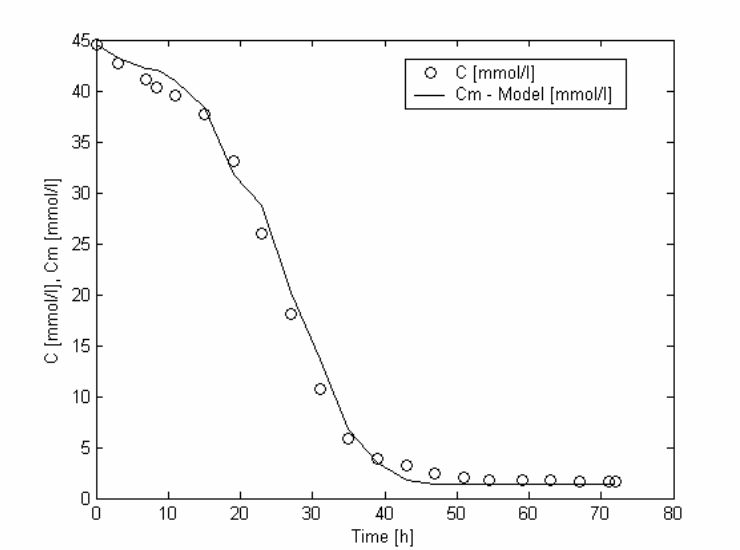

Fig. 10 Time course of the dissolved oxygen concentration (*C*) and model (*Cm*)

4. CONCLUSIONS

The following considerations and conclusions can be derived from the investigation, connected with the development of a mathematical model of the discussed process.

1. The models of the specific rates are obtained by linear and non-linear regression. These procedures permit investigation of the model structure and estimation of the initial parameters. Additional simulation shows the properties of the derived models (see Fig. 3 – 6)
2. The final stage of the investigation is connected with the parametric optimisation of the model through the non-linear optimisation procedure under the confidence intervals of the parameters using Optimisation Toolbox. The Levenberg-Maquardt and trust-region algorithms with least squares objective function are used for optimisation.
3. The mathematical model describes the trend of the investigated experimental data and specific rates of the main kinetic variables (Fig. 7 – 10).

REFERENCES

1. Bastin G., L. Chen, V. Chotteau, Can We Identify Biotechnological Processes?, *Proceedings of IFAC, Modelling and Control of Biotechnological Processes*, Colorado, USA; 83–88.
2. Georgiev Tz., Al. Ratkov, St. Tzonkov (1997), Mathematical modelling of fed-batch fermentation processes for amino acid production, *Mathematics and Computers in Simulation*, 44: 171–285.
3. Herman T., Industrial production of amino acids by coryneform bacteria, *Journal of Biotechnology*; 104: 155–172.
4. Leuchtenberger W., K. Huthmacher, K. Drauz, Biotechnological production of amino acids and derivatives: current status and prospects. *Appl. Microbiol. Biotechnology*; 69: 1–8.
5. MathWorks Inc., MATLAB User's Guide, 2007.
6. MathWorks Inc., System Identification Toolbox, User's Guide, 2007
7. MathWorks Inc. SIMULINK, Using SIMULINK, 2007.
8. MathWorks Inc., Optimisation Toolbox, User's Guide, 2007.
9. STATGRAPHICS, Version 5.1 Plus For Windows, User Manual, Magnugistics Inc. USA.
10. Todorov K., Tz. Georgiev, V. Ivanova, J. Kristeva, I. Dimov, Al. Ratkov, Time Series Analysis of Fed-batch Fermentation Process for L-valine Production, *Bioautomation*, Vol.4., April, ISSN 1312-451X pp. 24–29.