IMPLEMENTATION OF FUNCTIONAL STATE APPROACH FOR MODELLING OF ESCHERICHIA COLI FED-BATCH CULTIVATION

O. Roeva1, T. Pencheva1, Y. Georgieva2, B. Hitzmann3, S. Tzonkov1
Centre of Biomedical Engineering "Prof. Ivan Daskalov", Bulgarian Academy of Sciences, Sofia, Bulgaria1
Technical College – Yambol, Yambol, Bulgaria2
Universität Hannover, Institut für Technische Chemie, Hannover, Germany3

ABSTRACT
This paper presents the implementation of functional state approach to modelling of Escherichia coli fed-batch cultivation. Due to the complex metabolic pathways of microorganisms, the accurate modelling of bioprocesses is rather difficult. The functional state approach of a process is an alternative concept which helps in modelling and control of complex processes. The approach main idea is developing of models based on multiple submodels for each functional states (operating regime). In each functional state the process is described by a conventional type of model, called the local model, which is valid in this state. For parameter identification of the model the genetic algorithms are used. Genetic algorithms are directed random search techniques, applying the mechanics of natural selection and natural genetics, which can find the global optimal solution in complex multidimensional search spaces. Based on the available experimental data and simulations of E. coli fed-batch cultivation it is shown how this process can be divided into functional states and how the model parameters can be obtained on the basis of genetic algorithms. By simulation and comparison between the results and experimental data, can be seen how the concept of functional state approach works and how effective is the proposed identification scheme.

Introduction
Model formulation for a bioprocess is traditionally performed under conditions of a well-defined medium with single-substrate limitations, conditions that do not apply to most industrial fermentations which typically include a complex medium. In many cases, conventional numeric models, which describe the overall process behaviour, i.e. are globally valid, cannot be used in on-line monitoring and control, either because they do not describe the process well enough or contain too many poorly known parameters. Fermentation processes are characterized by a complicated structure of organization and independent characteristics, which determines their nonlinearity and non-stationary. Simple unstructured models, which account for key process variables (cell density, product and substrate concentrations), do not reflect metabolic changes and are unsuitable for many tasks. Model predictions can be improved by using of structured models, but the models incorporating too many equations and unknown parameters provide a qualitative, rather than quantitative description of the process. The structured model of a bioprocess is normally so complicated that it is difficult to be used for industrial scale production. Therefore, alternative modelling methods for monitoring and control are
needed.

The functional state approach to a process is a concept, which helps in modelling and control of complex processes such as bioprocesses. In [6] is presented a number of approaches to modelling and control problems arising from the work with systems of ever-increasing complexity and associated nonlinearity. The authors describe an approach which embraces a wide range of methods for developing complex models and controllers based on multiple submodels.

Zhang [9] introduces the functional state concept to describe and analyze the current biological state of bioprocesses, and applies the approach in expert system-based fault diagnosis and in control of bioprocesses. The main idea is to use a two-level hierarchy where at the first level the process is divided into macrostates, called functional states, according to behavioural equivalence. In each functional state the process is described by a conventional type of model, called a local model, which is valid in this functional state only. In each functional state, certain metabolic pathways are active enough to dominate the overall behaviour of the process. The biological behaviour is quite similar in each functional state. At the second hierarchical level some numeric detection algorithms and/or rules based on expert knowledge can be used for the recognition of the functional states and state transitions. In many batch-type processes, the functional states would naturally be identified with the different phases of the process. In a fed-batch or continuous process, the situation is more complex, but some functional states can be recognised and some functional state model can be used.

This paper illustrates the concept of functional state approach to modelling of cultivation of E. coli. The process is divided into several operating regimes. The process dynamics in each functional state is described by a simple local model. In principle, the structure of local models in different regimes can be different. Here the basic forms of the local models are the same, and only the parameters of the models have different numerical values. A set of local models together with functional states ‘dynamics’ can be used to describe, monitor and control the overall fed-batch cultivation of E. coli.

The estimation of the models' parameters is made using of Matlab Genetic Algorithms Toolbox procedures. The Multi-population genetic algorithms for parameter estimation is used. Today the most common direct methods used for global optimization are evolutionary algorithms such as genetic algorithms. Genetic algorithms (GAs) are directed random search techniques, based on the mechanics of natural selection and natural genetics, which can find the global optimal solution in complex multidimensional search spaces. These algorithms proved to be very suitable for the optimization of highly nonlinear problems with many variables. Recently, genetic algorithms have been used extensively in solving many optimization-searching problems [5, 7]. Compared with conventional optimization methods, GAs do not assume that the search space is differentiable or continuous. Also GAs do not require linearity in the parameters which is needed in iterative searching optimization techniques. Therefore, genetic algorithms are more workable in use for parameter identification of fermentation models.

Materials and Methods

Cultivation of recombinant micro-organisms e.g. Escherichia coli, in many cases is the only economical way to produce pharmaceutical biochemicals such as interleukins, insulin, interferons, enzymes and growth factors. To maximize the volumetric productivities of bacterial cultures it is important to grow E. coli to high cell concentration [3, 4, 8]. The use of fed-batch cultivation in the fermentation
industry takes advantage of the fact that residual substrate concentration may be maintained at a very low level in such a system. One of the most frequent used substrate for the cultivation of microorganisms is glucose. A low residual level of substrate may be advantageous in:

- Removing repressive effects of rapidly utilized carbon sources and maintaining conditions in the culture within the aeration capacity of the fermenter;
- Avoiding the toxic effects of a medium component.

The culture used in this work is *Escherichia coli* MC4110. The cultivation of *E. coli* is performed in a 2 l bioreactor (Bioengineering, Switzerland), using a mineral medium [1], in Institut für Technische Chemie, Universität Hannover. Before inoculation a glucose concentration of 2.5 g/l is established in the medium. Glucose in feeding solution is 100 g/l. Initial liquid volume is 1350 ml, pH is controlled at 6.8 and temperature is kept constant at 35 °C. The aeration rate is kept at 275 l/h air, stirrer speed at start 900 rpm, after 11 h the stirrer speed is increased in steps of 100 rpm and at end is 1500 rpm. Oxygen is controlled around 35%.

**Off-line analysis.** For off-line glucose measurements as well as biomass and acetate concentration determination samples of about 10 ml are taken roughly every hour. Off-line measurements are performed by using the Yellow Springs Analyzer (Yellow Springs Instruments, USA).

**On-line analysis.** For on-line glucose determination a flow injection analysis (FIA) system has been employed using two pumps (ACCU FM40, SciLog, USA) for a continuous sample and carrier flow rate. To reduce the measurement noise the continuous-discrete extended Kalman filter is used [1].

**Process modelling using the functional state approach**

Any model will have a limited range of operating conditions in which it is sufficiently accurate or performance sufficiently well in order to serve its purpose. This range may be restricted by several factors, such as modelling assumptions, stability properties, or experimental conditions. A model that is useful in a region less than the full range of operating conditions is called local model, as opposed to a global model which is useful over the full range of operating conditions [9].

The basis of this approach is a decomposition of the system's full range of operation into a number of possibly overlapping operating regimes (functional states). In each functional state, a simple local model is applied. These local models are then combined in some way to yield a global model. Hence, model development within this approach typically consists of the following tasks:

- Decompose the system's full range of operation into functional states.
- Select simple local model structures within each functional state. These structures will often be determined by the relevant system knowledge that is available under different operating conditions, as well as the intended purpose of the model.
- The local model structures are usually parameterized by certain variables that must be determined.

The following assumptions are made in developing the local models of the fed-batch cultivation of *E. coli*:

- The bioreactor is completely mixed.
- The main products in a cultivation of *E. coli* are biomass, water, carbon dioxide and, under some conditions, acetate.
- The substrate glucose mainly is consumed oxidatively and its consumption can be described by Monod kinetics.
- The acetate production rate is assumed to be directly proportional to the formation of biomass.
- Variation in the growth rate, acetate production and substrate consumption do not significantly change the elemental com-
position of biomass, thus balanced growth conditions are only assumed.

- Parameters except for the substrate and product concentrations, e.g. pH and temperature, are controlled to certain acceptable constant values during the process.

The rates of cell growth, sugar consumption, acetate production and dissolved oxygen concentration in a cultivation of *E. coli* are commonly described for all operating regimes according to the mass balance as follows:

\[
\frac{dX}{dt} = \mu X - \frac{F}{V} X
\]

\[
\frac{dS}{dt} = -q_s X + \frac{F}{V} (S_{in} - S)
\]

\[
\frac{dA}{dt} = q_A X - \frac{F}{V} A
\]

\[
\frac{dO_2}{dt} = -q_0 X + k_L a (O_{2}^* - O_2)
\]

\[
\frac{dV}{dt} = F
\]

\[
\mu = \mu_{max} \frac{S}{k_s + S}, \quad q_s = \frac{1}{Y_{S/X}} \mu, \quad q_A = \frac{1}{Y_{A/X}} \mu,
\]

\[
q_0 = \frac{1}{Y_{O_2/X}} \mu,
\]

where: \(X, S, S_{in}, A, O_2\) are the concentrations of, respectively, biomass, substrate (glucose), influent glucose, acetate and dissolved oxygen, \([\text{g/l}]\); \(F\) is the influent flow rate, \([\text{l/h}]\); \(V\) is the bioreactor volume, \([\text{l}]\); \(k_L a\) is the volumetric oxygen transfer coefficient, \([\text{h}^{-1}]\); \(Y_{S/X}, Y_{A/X}, Y_{O2/X}\) are yield coefficients, \([\text{gg}^{-1}]\), and \(\mu, q_s, q_A, q_0\) are the specific rates of, respectively, growth, substrate utilization, product (acetate) formation and oxygen consumption, \([\text{h}^{-1}]\).

Mentioned above specific rates vary in connection with the functional states.

The whole *E. coli* growth process can be divided into at least four functional states.

In each state, the *E. coli* metabolism is dominated by certain metabolic pathways. The process is defined to be in the first functional state when the glucose concentration is above the critical level (\(S_{crit}\)) and there is sufficient dissolved oxygen. The process enters in the second state when the glucose concentration decreases to be equal to or below the critical level and there is sufficient dissolved oxygen in the broth. The process is defined to be in the third functional state when no acetate is available, the glucose concentration is no more than critical level and the dissolved oxygen is above its critical level (\(O_{2crit}\)). In the fourth state acetate is available but no glucose is in the broth, and the dissolved oxygen concentration is above the critical level.

The rules for recognition of the functional states are summarized in Table 1.

For cultivation of *E. coli* three functional states - I, II and IV are recognized. The parameter functions of the local models in the three states are listed in Table 2. The structures of the models are identical, only the parameters values are different, depending on the state.

### Estimation of the models parameters

For estimation of the model parameters offline measurements of biomass, substrate (glucose), acetate and dissolved oxygen are used. The parameter functions of the local models are shown in Table 2.

Under Matlab 5.3 environment a Simulink model of the *E. coli* fed-batch

<table>
<thead>
<tr>
<th>State</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>(S &gt; S_{crit}) and (O_2 &gt; O_{2crit});</td>
</tr>
<tr>
<td>II</td>
<td>(S \leq S_{crit}) and (O_2 \geq O_{2crit}) and (A &gt; 0);</td>
</tr>
<tr>
<td>III</td>
<td>(S \leq S_{crit}) and (O_2 \geq O_{2crit}) and (A = 0);</td>
</tr>
<tr>
<td>IV</td>
<td>(S \leq S_{crit}) and (O_2 &lt; O_{2crit}) and (A &gt; 0);</td>
</tr>
</tbody>
</table>

### TABLE 1

Rules for recognition of the functional states
TABLE 2

Parameter functions of the local models in *E. coli* cultivation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Functional state I</th>
<th>Functional state II</th>
<th>Functional state IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\mu_{I, \text{max}}^I \frac{S}{k_I S^I + S}$</td>
<td>$\mu_{II, \text{max}}^II \frac{S}{k_I S^II + S}$</td>
<td>$\mu_{IV, \text{max}}^IV \frac{S}{k_I S^IV + S}$</td>
</tr>
<tr>
<td>$q_S$</td>
<td>$\frac{1}{Y_{S/X}} \mu_{I, \text{max}}^I \frac{S}{k_I S^I + S}$</td>
<td>$\frac{1}{Y_{S/X}} \mu_{II, \text{max}}^II \frac{S}{k_I S^II + S}$</td>
<td>$\frac{1}{Y_{S/X}} \mu_{IV, \text{max}}^IV \frac{S}{k_I S^IV + S}$</td>
</tr>
<tr>
<td>$q_L$</td>
<td>$\frac{1}{Y_{A/X}} \mu_{I, \text{max}}^I \frac{S}{k_I S^I + S}$</td>
<td>$\frac{1}{Y_{A/X}} \mu_{II, \text{max}}^II \frac{S}{k_I S^II + S}$</td>
<td>$\frac{1}{Y_{A/X}} \mu_{IV, \text{max}}^IV \frac{S}{k_I S^IV + S}$</td>
</tr>
<tr>
<td>$q_{A2}$</td>
<td>$\frac{1}{Y_{O_{2}/X}} \mu_{I, \text{max}}^I \frac{S}{k_I S^I + S}$</td>
<td>$\frac{1}{Y_{O_{2}/X}} \mu_{II, \text{max}}^II \frac{S}{k_I S^II + S}$</td>
<td>$\frac{1}{Y_{O_{2}/X}} \mu_{IV, \text{max}}^IV \frac{S}{k_I S^IV + S}$</td>
</tr>
<tr>
<td>$k_L a$</td>
<td>$k_L^I a$</td>
<td>$k_L^II a$</td>
<td>$k_L^IV a$</td>
</tr>
</tbody>
</table>

fermentation has been developed. The *Simulink* model describes the differential equations (1)-(4), parameters and initial values. In order to identify the parameters in the *Simulink* model, a script containing the necessary instructions for the Genetic Algorithms Toolbox has been developed.

To implement the genetic algorithms, the models’ parameters have to be parameterised in terms of chromosomes. Each chromosome corresponds to one different objective function value. The objective function is used to provide a measure of how individuals have performed in the problem domain. In the case of minimization problem, the fitted individuals will have the lowest numerical value of the associated objective function. This raw measure of fitness is only used as an intermediate stage in determining the relative performance of individuals in genetic algorithms. The selection algorithm chooses individuals for reproduction on the basis of their relative fitness. The genetic algorithms are terminated when some criteria are satisfied. In the literature the most often used optimization criterion is defined as a modelling error, i.e. the mean square deviation between the model output and the corresponding data obtained during the fermentation. The optimization criterion is presented as follows:

$$ J = a_1 \sum (X_d - X_m)^2 + a_2 \sum (S_d - S_m)^2 + $$

$$ a_3 \sum (A_d - A_m)^2 + a_4 \sum (O_{2d} - O_{2m})^2 \rightarrow \min (6) $$

where $X_d$, $S_d$, $A_d$ and $O_{2d}$ are the vectors of experimental data and $X_m$, $S_m$, $A_m$ and $O_{2m}$ are the vectors of simulated (model) data. Weight coefficient $a_i$ vary in connection with the functional states in the following borders:

- $a_1 = 0.5 \div 1$;
- $a_2 = 1 \div 2$;
- $a_3 = 5 \div 10$;
- $a_4 = 0.01 \div 0.05$.

Initial tests with genetic algorithms are done with a large number of individuals and generations, exactly 1500 individuals and 250 generations. The initial values of other parameters and functions are taken from [2].

Several optimizations have been made when the initial parameters related to the genetic algorithms have been changed. After different runs some of parameters have been changed:

- GGAP - Generation gap (how many new individuals are created in every generation) - 0.97;
- PRECI - Precision of binary representation - 20;
- SEL_F - Selection function - Roulette.
Table 3

Estimated values of the local model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>State I Value</th>
<th>Parameter</th>
<th>State II Value</th>
<th>Parameter</th>
<th>State IV Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{I_{max}}^I$</td>
<td>0.57 h⁻¹</td>
<td>$\mu_{I_{max}}^II$</td>
<td>0.59 h⁻¹</td>
<td>$\mu_{I_{max}}^IV$</td>
<td>0.47 h⁻¹</td>
</tr>
<tr>
<td>$k_S^I$</td>
<td>0.12 g/l</td>
<td>$k_S^II$</td>
<td>0.038 g/l</td>
<td>$k_S^IV$</td>
<td>0.01 g/l</td>
</tr>
<tr>
<td>$Y_{S/X}^I$</td>
<td>0.5181 gg⁻¹</td>
<td>$Y_{S/X}^II$</td>
<td>0.0525 gg⁻¹</td>
<td>$Y_{S/X}^IV$</td>
<td>0.4950 gg⁻¹</td>
</tr>
<tr>
<td>$Y_{A/X}^I$</td>
<td>66.67 gg⁻¹</td>
<td>$Y_{A/X}^II$</td>
<td>76.92 gg⁻¹</td>
<td>$Y_{A/X}^IV$</td>
<td>66.67 gg⁻¹</td>
</tr>
<tr>
<td>$Y_{O/X}^I$</td>
<td>2.08 gg⁻¹</td>
<td>$Y_{O/X}^II$</td>
<td>1.66 gg⁻¹</td>
<td>$Y_{O/X}^IV$</td>
<td>1.87 gg⁻¹</td>
</tr>
<tr>
<td>$k_{L_a}^I$</td>
<td>103.06 h⁻¹</td>
<td>$k_{L_a}^II$</td>
<td>125.56 h⁻¹</td>
<td>$k_{L_a}^IV$</td>
<td>98.45 h⁻¹</td>
</tr>
</tbody>
</table>

Wheel Selection:
• XOV_V - Recombination function - Crossover Double Point;
• XOVR - Crossover rate – 0.7;
• MUTR - Mutation rate – 0.1;
• MIGR – Migration rate – 0.2;
• INSR – Insertion rate (how many individuals produced in each generation are reinserted into the population) – 0.9;
• SUBPOP - Number of subpopulations - 10;
• NIND – Number of individuals - 100;
• MAXGEN – Maximum numbers of generations – 200.

Results from the identification of model parameters are presented in Table 3. Both the real fermentation trajectories and the simulated ones are presented in Fig. 1-3. The initial values for simulation in new functional states are the last simulated values in previous functional state so that the trajectories have been continuous. The values of criterion $J$ (6) for each functional state are:

$J_I = 7.5213; \quad J_{II} = 11.1843; \quad J_{IV} = 9.3916.$

Biotechnol. & Biotechnol. Eq. 18/2004/3 212
**Conclusions**

The results indicate that the functional state approach to modelling can help to understand better the behavior of a process and simplify the process modelling. The simulations presented above indicate that the process can be rather well modeled with the functional state approach to modelling. Basic aim of this approach is to overcome the main disadvantage of using global process model, mainly big number of model parameters, which complicate the model simulation and parameter estimation. The main advantage of functional state modelling is that the parameters of each local model can be separately estimated from other local models parameters. To illustrate the concept of functional states...
the simulations of a fed-batch cultivation of *E. coli* are presented. It is shown how this process can be divided into functional states and how the model can be obtained on the basis of genetic algorithms. The effectiveness of the proposed identification scheme is confirmed by the simulation and comparison of the results to experimental data.

Acknowledgements
This work is partially supported from DFG Project №113/135/0-1, NSF-TN-1314/2003 and Project №03-1046/2003 with Thracian University- Stara Zagora.

REFERENCES