International Conference
AUTOMATICS AND INFORMATICS'10
3-7 October 2010, Sofia, Bulgaria

GENETIC ALGORITHMS FOR FEED RATE PROFILES DESIGN

O. Roeva, K. Kosev

Centre of Biomedical Engineering – Bulgarian Academy of Sciences
105 Acad. George Bonchev St., 1113 Sofia, Bulgaria
E-mail: olympia@clbme.bas.bg, kalinkosev@yahoo.com

Abstract. In the paper a genetic algorithm for feed rate profiles design is proposed. An *E. coli MC4110* fed-batch fermentation process is considered. The feed rate profiles based on three different lengths of chromosomes are synthesized. The ration of the substrate concentration and the difference between actual cell concentration and theoretical maximum cell concentration is used as an objective function. As a result the genetic algorithm synthesized optimal feed profiles fulfilling the defined criterion. In the case of 60 genes obtained feed rate profile cell concentration has an ideal increase for the complete fermentation period, achieving final cell concentration of 5.26 $g \cdot l^{-1}$. During the process, 1.38 l feeding solution is used. This is a satisfactory result for the fermentation system due to economical effect and process effectiveness. The obtained results for feed rate profiles based on different chromosome lengths demonstrate good computational performance of the proposed genetic algorithm.

Keywords: Genetic algorithms, Chromosome, Gene, Feed rate profile, *E. coli* fermentation process.

INTRODUCTION

Optimization of fed-batch fermentation processes has been a subject of research since many years. Control opportunities in fed-batch operated fermentations have been reviewed in detail in a number of articles. It is well known that the design of high-performance model-based control algorithms for biotechnological processes is hampered by two major problems which require adequate engineering solutions. First, the process kinetic is too often poorly understood nonlinear functions, while the corresponding parameters are in general time-varying. Second, up to now there is a lack of reliable sensors suited to real-time monitoring of the process variables which are needed in advanced control algorithms.

Currently, the feed rate design is commonly solved by mathematical model based optimization methods. If an accurate model of the system is available optimization procedures can be used to calculate the feeding strategy [4, 6, 8, 15, 16]. However, fermentation processes are typically very complex, involving different transport phenomena, microbial components and biochemical reactions. These properties make processes difficult to control with traditional techniques. For simple mathematical models, the problem can be solved analytically, from the Hamiltonian function, by applying the minimum principle of Pontryagin [14, 17-19]. However, besides having a problem of singular control, those methodologies become too complex when the number of state variables increases. As an alternative to surmount these difficulties the global optimization methods are used.

Global optimization methods can be roughly classified as deterministic and stochastic strategies. Stochastic methods for global optimization ultimately rely on probabilistic approaches and can locate the vicinity of global solutions with good efficiency. There are many different kinds of stochastic methods for global optimization, but the following groups should be highlighted: adaptive stochastic methods; clustering methods; evolutionary algorithms; simulated annealing and other meta-heuristics.

Several different types of evolutionary search methods were developed independently. These include: genetic programming, which evolve programs; evolutionary programming, focused on optimizing continuous functions without recombination; evolutionary strategies, focused on optimizing continuous functions with recombination; and genetic algorithms (GA) [7], focused on optimizing general combinatorial problems.

GA is a global, parallel, stochastic search method, founded on Darwinian evolutionary principles. Since its introduction and subsequent popularization, the GA has been frequently utilized as an alternative optimization tool to conventional methods.

Specific particularities of the fermentation processes lead to estimation of a large-scale problem and as a successful tool for solving this problem are examined GA. The GA effectiveness and robustness have been demonstrated for identification of fed-batch cultivation processes [2, 10-12].

The GA is already used for design of feed rate profiles for glucose and glutamine, based on a seventh-order nonlinear model of fed-batch culture of hybridoma cells [3]. In the work [5] the optimal profile for the substrate feeding rate in a fed-batch culture of *S. baicalensis georgi* is determined using a GA. The experimental results showed that neurocontrol incorporated with a genetic algorithm improved the flavonoid production compared with a simple fuzzy logic control system.

In this paper a genetic algorithm for feed rate profiles design during an *E. coli MC4110* fed-batch fermentation process is proposed. The bacterium *E. coli* is the microorganism of choice for the production of the majority of the valuable biopharmaceuticals. *E. coli* usually grows under fed-batch mode due to the effect of acetic acid, which is produced when glucose is present above certain concentrations. Here an optimal state of microorganisms' culture is maintained by GA synthesized feed rate profiles.

FED-BATCH FERMENTATION PROCESS OF *E. COLI MC4110*

The mathematical model of fed-batch fermentation of *E. coli MC4110* has the form [1, 9, 12, 14]:

$$\frac{dX}{dt} = \mu_{max} \frac{S}{k_S + S} X - \frac{F}{V} X \quad (1)$$

$$\frac{dS}{dt} = -\frac{1}{Y_{XS}} \mu_{max} \frac{S}{k_S + S} X + \frac{F}{V}\left(S_{in} - S\right) \quad (2)$$

$$\frac{dV}{dt} = F \tag{3}$$

where: X is the concentration of biomass, [g·l^{-1}]; S – concentration of substrate (glucose), [g·l^{-1}]; F – feeding rate, [l·h^{-1}]; V – the volume of the content, [l]; S_{in} – substrate concentration of the feeding solution, [g·l^{-1}]; μ_{max} – maximum growth rate, [h^{-1}]; k_S – saturation constant, [g·l^{-}1]; Y_{XS} – yield coefficient, [g·g^{-1}].

The values of the model parameters used in simulations are [13]: $\mu_{max} = 0.52$ h^{-1}; $k_S = 0.023$g·l^{-1}; $Y_{XS} = 0.5$.

Initial conditions of the process variable are [1]:
$X(0) = 1.252$ g·l^{-1}; $S(0) = 0.812$ g·l^{-1};
$V(0) = 1.35$ l; $S_{in} = 100$ g·l^{-1}.

APPLICATION OF GENETIC ALGORITHMS FOR FEED RATE PROFILES DESIGN

Background of the Genetic Algorithm

Genetic algorithms are a class of non-gradient methods. The basic idea of GA is the mechanics of natural selection. Each optimization parameter, (x_n), is coded into a gene as for example a real number or a string of bits. The corresponding genes for all parameters, x_1, …, x_n, form a chromosome, which describes each individual. A chromosome could be an array of real numbers, a binary string, a list of components in a database, all depending on the specific problem. Each individual represents a possible solution, and a set of individuals form a population. In a population, the fittest are selected for mating. Mating is performed by combining genes from different parents to produce a child, called a crossover. Solutions are also "mutated" by making a small change to a single element of the solution. Finally, the children are inserted into the population and the procedure starts over again. The optimization continues until the population has converged or the maximum number of generations has been reached.

A pseudo code of a GA is presented as:

i = 0	set generation number to zero
initpopulation P(0)	initialize a usually random population of individuals
evaluate P(0)	evaluate fitness of all initial individuals
while (not done) do	test for termination criterion (time, fitness, etc.)
begin	
i = i + 1	increase the generation number
select P(i) from P(i – 1)	select a sub-population for offspring reproduction
recombine P(i)	recombine the genes of selected parents
mutate P(i)	perturb the mated population stochastically
evaluate P(i)	evaluate its new fitness
end	

Initial population: A GA starts with a population of strings to be able to generate successive populations of strings afterwards. The initialization is done randomly. A binary 20 bit encoding is considered. Binary representation is the most common one, mainly because of its relative simplicity.

Reproduction: An important aspect is to decide which individuals should be chosen as parents for the purpose of reproduction. With GA, this selection is based on the string fitness: according to the "survival of the fittest" principle. The best known selection mechanism, roulette wheel selection, is used in the proposed GA.

Recombination: Once two parents have been selected, the GA combines them to create two new offspring using crossover operator. The role of the crossover operator is to allow the advantageous traits to be spread throughout the population in order that the population as a whole may benefit from this chance discovery. The crossover is the prime distinguishing factor of a GA from other optimization algorithms. Here, double point crossover is employed.

Mutation: The last operator is the mutation algorithm. The effect of mutation is to reintroduce divergence into a converging population. The biological inspiration behind this operator is the way in which a chance mutation in a natural chromosome can lead to the development of desirable traits which give the individual displaying these characteristics an advantage over its competitors. In accepted encoding here a bit inversion mutation is used. This prevents the solution from converging to some local optimal solutions; thereby the global optimal solution can be obtained.

The GA operators and parameters are summarized in Tables 1 and 2 based on [13].

Representation of chromosomes: Representation of chromosomes is a critical part of GA application. Here, each chromosome of the population represents a feed rate profile as a sequence of feed rate values. The simplest way to represent it was by using a piecewise approximation of the feed rate profile. The profile is divided into equal intervals and the feed rate values at the breakpoints are registered. The sequence of numbers obtained is considered a chromosome and each gene represented the feed rate after definite time. Three chromosomes representations are proposed:

- *1st*: division into equal 30 intervals (30 genes);
- *2nd*: division into equal 60 intervals (60 genes);
- *3rd*: division into equal 100 intervals (100 genes).

Every gene is coded in range $F = 0$ - 0.05 l·h^{-1} [1].

Table 1. Genetic algorithm operators

Operator	Type
encoding	binary
crossover	double point
mutation	bit inversion
selection	roulette wheel selection
fitness function	linear ranking

Table 2. Genetic algorithm parameters

Parameter	Value
generation gap	0.97
crossover rate	0.70
mutation rate	0.05
precision of binary representation	20
number of individuals	100
number of generations	150

Evaluation: After every generated population, the individuals of the population should be evaluated to be able to distinguish between good and bad individuals. The evaluation function plays a role similar to that which the environment plays in natural evolution and it rates chromosome in terms of fitness. This is done by mapping the objective function to a "fitness function": a non-negative figure of merit. Here linear ranking is used.

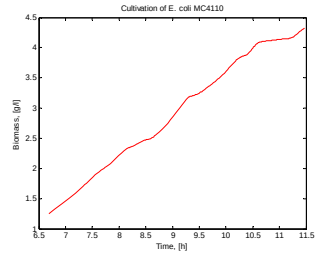
The objective function (J_{OF}) utilized here is presented as:

$$J_{OF} = f(X_{\text{Actual}}, X_{\text{Theory}}, S) \rightarrow \min \quad (4)$$

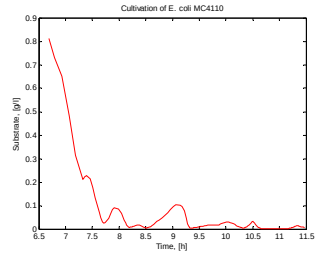
The genetic algorithm syntheses feed rate profile based on minimization of the ration of the substrate concentration (S) and the difference between actual cell concentration (X_{Actual}) and theoretical maximum cell concentration (X_{Theory}). The results in [13] show that the feed profile formed by the objective function (4) is superior to the feed rate profiles formed by the rest five objective functions. Therefore, here the objective function (4) is considered.

Feed Rate Profiles Design

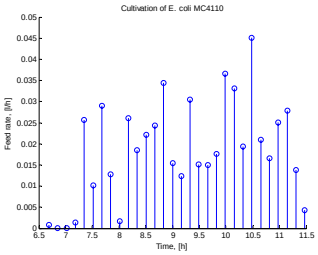
Since GA are stochastic, their performance usually varies from generation to generation. Extensive simulation tests have been conducted on the GA to test the effectiveness of the algorithm, using the model (1) – (3) and objective function (4). Three problems (30, 60 and 100 genes in chromosome) are running 50 executions with the GA. The resulting feed rate profiles, as well as the biomass and substrate dynamics are depicted on Figs. 1 – 3.



a) biomass concentration



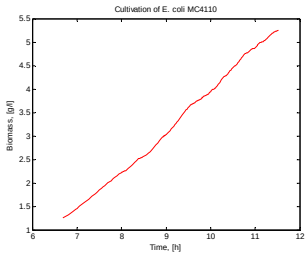
b) substrate concentration



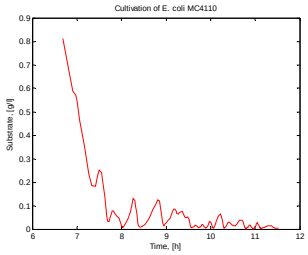
c) feed rate profile

Fig. 1. Resulting dynamics of biomass and substrate and feed rate profile in case of 30 genes in chromosome

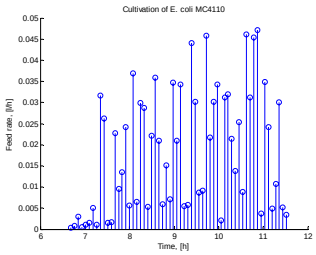
All experiments reported were run on a PC with a Pentium IV 3.2 GHz processor in Matlab environment. Average values of the best results at a certain evaluation are calculated. In Table 2 are presented the average values of the objective function (J_{OF}), the biomass concentration in the end of the fermentation process (X_{end}) and the total amount of substrate used for process feeding (F_{Total}). The genetic algorithm produce the same results with more than 85% coincidence.



a) biomass concentration

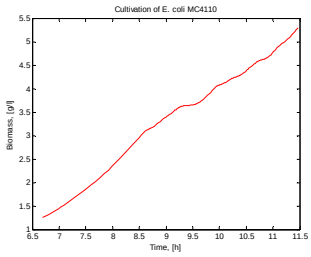


b) substrate concentration

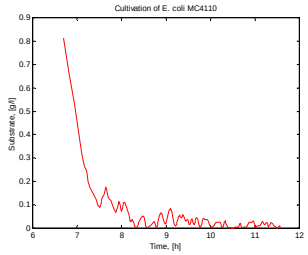


c) feed rate profile

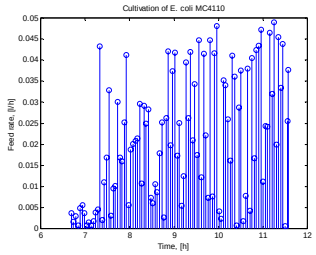
Fig. 2. Resulting dynamics of biomass and substrate and feed rate profile in case of 60 genes in chromosome



a) biomass concentration



b) substrate concentration



c) feed rate profile

Fig. 3. Resulting dynamics of biomass and substrate and feed rate profile in case of 100 genes in chromosome

Table 2. Results from the feed rate design.

Gene	J_{OF}	X_{end}, $g \cdot l^{-1}$	F_{Total}, l
30	0.0308	4.32	0.51
60	0.0295	5.26	1.38
100	0.0295	5.29	1.96

The results show that for all tests the required objective function has been achieved. In the case of a chromosome with 30 genes, the resulting feed rate profile is assumed as not satisfactory result. In this case, considered piecewise approximation of the feed rate profile is not appropriate. Better results are achieved in the cases of chromosome with 60 and 100 genes. The genetic algorithm synthesized feed rate profile resulting in generally higher final cell concentration (5.29 $g \cdot l^{-1}$) using chromosome with 100 genes. Nevertheless, the feed rate profile achieved by chromosome with 60 genes is defined as the superior to the rest of profiles. The objective function is the same as in the case of 100 genes and the biomass concentration in the end of the fermentation process is close to biomass concentration achieved in the case of 60 genes. In the same time the higher final cell concentration is achieved using only 1.38 l feeding solution. In the case of chromosome with 100 genes almost 2 l solution is used, which is worse result from the economical point of view and process effectiveness.

CONCLUSIONS

In this work genetic algorithms are used for design of feed rate profiles during an *E. coli MC4110* fermentation process. Development of a suitable feeding strategy is critical in fed-batch operation modes. During the fed-batch fermentation of *E. coli* the system states change considerably, from a low initial to a very high biomass and product concentration. This dynamic behavior motivates the development of optimization methods to find the optimal input feeding trajectories in order to improve the process. A genetic algorithm using different chromosome lengths is proposed in order to optimize the feeding trajectory of the fermentation process. The ration of the substrate concentration and the difference between actual cell concentration and theoretical maximum cell concentration is used as an objective function.

The proposed genetic algorithm is found to be an effective and efficient method for solving the optimal feed rate profile problem. The GA is capable of simultaneously optimizing feed rate profile for a given objective function. However, the results seem to indicate that the feed profile formed using chromosome with 60 genes is superior to the rest feeding trajectories. Based on obtained feed rate profile cell concentration has an ideal increase for the complete fermentation period, achieving final cell concentration of 5.26 $g \cdot l^{-1}$ using 1.38 l feeding solution. This is a satisfactory result for the fermentation system due to the economical effect and process effectiveness.

ACKNOWLEDGEMENT

This work is partially supported by National Science Fund Grants DMU 02/4 “High quality control of biotechnological processes with application of modified conventional and metaheuristics methods” and DID-02-29 “Modelling Processes with Fixed Development Rules”.

REFERENCES

1. Arndt M., B. Hitzmann, Feed Forward/Feedback Control of Glucose Concentration during Cultivation of *Escherichia coli*, In: *8th IFAC Int. Conf. on Comp. Appl. in Biotechn.*, Quebec City, Canada, 425-429, 2001.
2. Carrillo-Ureta G. E., P. D. Roberts, V. M. Becerra, Genetic Algorithms for Optimal Control of Beer Fermentation, In: *2001 IEEE Int. Symp. on Intell. Cont.*, 391-396, 2001.
3. Chen L., S. Nguang, X. Chen, On-line Identification and Optimization of Feed Rate Profiles for High Productivity Fed-batch Culture of Hybridoma Cells using Genetic Algorithms, *ISA Trans.*, 41(4), 409-419, 2002.
4. Choi J. H., K. C. Keum, S. Y. Lee, Production of Recombinant Proteins by High Cell Density Cultures of *Escherichia coli*, *Chem. Eng. Sci.*, 61, 876-885, 2006.
5. Choi J.-W., W. Lee, M. C. Jin, Y.-K. Kim, S.-Y. Park, H. L. Won, Control of Feed Rate using Neurocontroller Incorporated with Genetic Algorithm in Fed-batch Cultivation of *Scutellaria baicalensis georgi*, *J. Microbiol. Biotech.*, 12(4), 687-691, 2002.
6. de Mare L., C. Cimander, A. Elfwing, P. Hagander, Feeding Strategies for *E. coli* Fermentations Demanding an Enriched Environment, *Bioprocess Biosyst Eng.*, 30, 13-25, 2007.
7. Goldberg D. E., Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley Longman, London, 1989.
8. Levišauskas D., T. Tekorius, Model-based Optimization of Fed-batch Fermentation Processes using Predetermined Type Feed-rate Time Profiles. A Comparative Study, *Information Technology&Control*, 34(3), 231-236, 2005.
9. Miksch G., S. Kleist, K. Friehs, E. Flaschel, Overexpression of the Phytase from *Escherichia coli* and its Extracellular Production in Bioreactors, *Applied Microbiology and Biotechnology*, 59, 685-694, 2002.
10. Rocha I., E. Ferreira, Optimization Methods for Improving Fed-batch Cultivation of *E. coli* Producing Recombinant Proteins, In: *10th Mediterranean Conference on Control and Automation – MED2002*, Lisabon, Portugal, 2002.
11. Roeva O., A Modified Genetic Algorithm for a Parameter Identification of Fermentation Processes, *Biotechnology and Biotechnological Equipment*, 20(1), 202-209, 2006.
12. Roeva O., Parameter Estimation of a Monod-type Model based on Genetic Algorithms and Sensitivity Analysis, *LNCS*, Springer, 4818, 601-608, 2008.
13. Roeva O., St. Tzonkov, A Genetic Algorithm for Feeding Trajectory Optimization of Fed-batch Fermentation Processes, *Int. J. Bioautomation*, 12, 1-12, 2009.
14. Roeva O., St. Tzonkov, Optimal Feed Rate Control of *Escherichia coli* Fed-batch Fermentation, *Int. J. Bioautomation*, 2, 30-36, 2005.
15. Samad N., A. Ahmad, Comparison Product Optimization Performance of Fed-batch Fermentation Processes for Penicillin G Production, In: *1st Int. Conf. on Natural Resources Engineering&Technology*, 562-575, 2006.
16. Shin H., H. Lim, Maximization of Metabolite in Fed-batch Cultures Sufficient Conditions for Singular Arc and Optimal Feed Rate Profiles, *Biochemical Eng. J.*, 37, 62-74, 2007.
17. Smets I. Y., G. P. Bastin, J. F. Van Impe, Feedback Stabilization of Fed-Batch Bioreactors: Non-Monotonic Growth Kinetics, *Biotechnol. Prog.*, 18, 1116-1125, 2002.
18. Smets I., J. Claes, E. November, G. Bastin, J. Van Impe, Optimal Adaptive Control of (Bio)chemical Reactors: Past, Present and Future, *J. of Proc. Contr.*, 1-43, 2003.
19. Vanichsriratana W., Optimization of a Primary Metabolite Fermentation Process: Effect of Cost Factor on the Optimal Feed Rate Control, *Kasetsart J. (Nat. Sci.)*, 34, 495-499, 2000.