*BIO*AUTOMATION, 2009, 12, 1-12

A Genetic Algorithm for Feeding Trajectory Optimisation of Fed-batch Fermentation Processes

Olympia Roeva*, Stoyan Tzonkov

Centre of Biomedical Engineering – Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., Sofia 1113, Bulgaria
E-mail: {olympia, tzonkov}@clbme.bas.bg

**Corresponding author*

Received: December 18, 2008 **Accepted: February 27, 2009**

Published: March 16, 2009

Abstract: *In this work a genetic algorithm is proposed with the purpose of the feeding trajectory optimization during a fed-batch fermentation of E. coli. The feed rate profiles are evaluated based on a number of objective functions. Optimization results obtained for different feeding trajectories demonstrate that the genetic algorithm works well and shows good computational performance. Developed optimal feed profiles meet the defined criteria. The ration of the substrate concentration and the difference between actual cell concentration and theoretical maximum cell concentration is defined as the most appropriate objective function. In this case the final cell concentration of 43 g·l^{-1} and final product concentration of 125 g·l^{-1} are achieved and there is not significant excess of substrate.*

Keywords: *Genetic algorithms, Optimization, Feed rate profile, E. coli.*

Introduction

The problem of determining optimal controls for fed-batch fermentation processes has become an important field of interest in biotechnology that offers a sustainable production of existing and novel products. Today many proteins are produced by genetically modified microorganisms. One of the most used host organisms is the bacterium *E. coli* as it is a well studied and a well-known organism. To achieve a good productivity, high cell concentration and high cell productivity are desired and this is usually obtained from fed-batch cultivations. Fed-batch culture is advantageous in particular when nutrient concentrations strongly affect cell yield or productivity, as both overfeeding and underfeeding would result in growth repression and starvation to cells, respectively [12]. Development of a suitable feeding strategy is critical in fed-batch operation and review on the subject is given in [3].

Currently, the feed rate optimization problem 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 [5, 9, 15, 16]. However, fermentation processes are typically very complex, involving different transport phenomena, microbial components and biochemical reactions. Furthermore, the nonlinear behavior and time-varying 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]. However, besides having a problem of singular control, those methodologies become too complex when the number of state variables increases.

Lately the use of evolutionary algorithms (EA) for optimization has increased [1, 6, 10]. In the work [13], EA are used to achieve optimal feed-forward control in a recombinant bacterial

fed-batch fermentation process that aims at producing a bio-pharmaceutical product. Three different aspects are the target of the optimization procedure: the feeding trajectory, the duration of the fermentation and the initial conditions of the process. The intention of the work [2] is to use the most popular type of EA – genetic algorithms (GA) for identifying the parameters of a seventh-order nonlinear model of fed-batch culture of hybridoma cells, and determining the best feed rate control profiles for glucose and glutamine. Genetic algorithms proved to be a good alternative method for solving such problems. In the work [4] the optimal profile for the substrate feeding rate in a fed-batch culture of *S. baicalensis g.* is determined using a genetic algorithm. The experimental results showed that neurocontrol incorporated with a genetic algorithm improved the flavonoid production compared with a simple fuzzy logic control system.

The main motivation of this paper is to develop a robust and reliable genetic algorithm in order to achieve optimal substrate feeding trajectory. An optimal state of microorganisms' culture for biosynthesis of the desired product can be maintained by using appropriate feed rate profiles. A fed-batch fermentation process of *E. coli* strain *BL21(DE3)pPhyt109* was studied [8]. 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. The specific objective is to obtain the best feed rate profile for considered fed-batch fermentation process based on a number of objective functions.

The fed-batch fermentation process

E. coli strain *BL21(DE3)pPhyt109* is used for fermentation experiments. The experiments are performed in the *Department of Fermentation Engineering, Faculty of Technology, University of Bielefeld*. Plasmid pPhyt109, an expression vector derived from the multi copy plasmid pUC19, contains the gene for *E. coli* phytase under the constitutive promoter of the *bglA* gene of *Bacillus amyloliquefaciens*. In addition, the expression vector contained a secretion cassette of 2.5 kb providing the competence for the secretion of pythase into the culture medium based on the action of the Kil protein expressed under the control of the stationary-phase promoter of the *fic* gene [8].

Fermentation experiments are carried out in a bioreactor with a total volume of 7 l and a working volume of 5 l. The bioreactor is equipped with direct digital control (DDC) from MBR (Multiple Bioreactors and Sterile Plants, Zurich, Switzerland). Glucose mineral salt medium is used as growth medium. The pH is maintained at 6.9 by controlled addition of 4 N NaOH. Antifoam (PE8100, BASF, Germany) is added automatically when required. The temperature is kept at 37ºC. Air flow is kept constant at 10 l·min^{-1}. The stirrer speed is kept constant at 500 rpm.

The rates of cell growth, substrate consumption and phytase production in the *E. coli* fed-batch fermentation are commonly described as follows:

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

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

$$\frac{dPh}{dt} = \frac{1}{Y_{Ph/X}} \mu_{\max} \frac{S}{k_S + S} X - \frac{F}{V} Ph \tag{3}$$

$$\frac{dV}{dt} = F \tag{4}$$

where: X is the biomass concentration, [g·l^{-1}]; S – substrate (glucose) concentration, [g·l^{-1}]; Ph – phytase concentration, [g·l^{-1}]; F – feeding rate, [l·h^{-1}]; V – bioreactor volume, [l]; S_{in} – substrate concentration in the feeding solution, [g·l^{-1}]; $\mu_{\max}$ – maximum growth rate, [h^{-1}]; k_S – saturation constant, [g·l^{-1}]; $Y_{S/X}$ and $Y_{Ph/X}$ – yield coefficients, [g·g^{-1}].

The following assumptions are made in the model development of the fed-batch fermentation of *E. coli BL21(DE3)pPhyt109*:

- The bioreactor is completely mixed.
- Potential mixing effects of the highly concentrated feeds with the fermentation medium are neglected for the sake of the model simplicity.
- The suspension viscosity in the reactor remains constant during the experiment.
- The substrate (glucose) is consumed mainly oxidatively.
- Variations in the growth rate and phytase production, as well as in substrate consumption do not significantly change the elemental composition of biomass, thus balanced growth conditions are only assumed.
- The phytase production is regarded as a one-step enzymatic reaction.
- Parameters, e.g. pH and temperature, are controlled to certain acceptable constant values during the process.

The numerical values of the model parameters used in simulations are presented in Table 1.

Table 1. Model parameters

Parameter	$\mu_{\max}$, [h^{-1}]	k_S, [g·l^{-1}]	$Y_{S/X}$, [g·g^{-1}]	$Y_{Ph/X}$, [g·g^{-1}]
Value	0.74	0.03	1.47	1.54

Genetic algorithms for feeding trajectory optimization

During the fed-batch fermentation of *E. coli BL21(DE3)pPhyt109* 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. An appropriate approach for optimizing the feeding trajectory comes from the use of Evolutionary algorithms.

EA are a very popular class of methods based on the ideas of biological evolution, which is driven by the mechanisms of reproduction, mutation, and the principle of survival of the fittest. EA differ from more traditional optimization techniques in that they involve a search from a "population" of solutions, not from a single point. Each iteration involves a competitive selection that weeds out poor solutions. Similarly to biological evolution, evolutionary computing methods generate better and better solutions by iteratively creating new "generations" by means of those mechanisms in numerical form.

Several different types of evolutionary search methods were developed independently. These include: genetic programming, which evolve programs; evolutionary programming, which

focuses on optimizing continuous functions without recombination; evolutionary strategies, which focuses on optimizing continuous functions with recombination; and genetic algorithms [7], which focuses on optimizing general combinatorial problems.

Genetic algorithms

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 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.

Proposed GA is based on the *Genetic Algorithm Toolbox* for *Matlab* [11]. Outline of the algorithm could be presented as:

1. **[Start]** Generate random population of *n* chromosomes
2. **[Fitness]** Evaluate the fitness *f(x)* of each chromosome *x* in the population
3. **[New population]** Create a new population by repeating following steps until the new population is complete
    1. **[Selection]** Select two parent chromosomes from a population according to their fitness
    2. **[Crossover]** With a crossover probability cross over the parents to form new offspring
    3. **[Mutation]** With a mutation probability mutate new offspring at each locus
    4. **[Accepting]** Place new offspring in the new population
4. **[Replace]** Use new generated population for a further run of the algorithm
5. **[Test]** If the end condition is satisfied, stop, and return the best solution in current population
6. **[Loop]** Go to step **2**

The parameters of a GA significantly affect the speed of convergence to the near optimal solution, and the accuracy of the solution itself. Therefore, there is a need to investigate the effects of the different GA parameters on the outcome of the GA enhanced simulation.

Results and discussion

Configuration of the genetic algorithm

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) – (4). A first set of experiments was carrying out in order to find the best set of genetic operators to tackle the feed rate optimization problem. Each run of the GA is stopped after 100 iterations and the results are given in terms of the mean of 25 runs, with the associated 95% confidence intervals. Moreover, there was performed a lot of tests to choose the appropriate GA parameters for considered here problem. The tests performed held most elements of GA constant while one element was changed. The chosen

GA operators and parameters are summarized in Table 2. All experiments reported were run on a PC with a Pentium IV 3.2 GHz processor.

Table 2. Genetic algorithm elements

Operator	Type	Parameter	Value
encoding	binary	generation gap	0.97
crossover	double point	crossover rate	0.70
mutation	bit inversion	mutation rate	0.05
selection	roulette wheel selection	precision of binary representation	20
fitness function	linear ranking	number of individuals	50
-	-	number of generations	100

A binary 20 bit encoding is considered. Binary representation is the most common one, mainly because of its relative simplicity. The best known selection mechanism, roulette wheel selection, is used in the proposed GA.

The genetic operators used in this GA are namely, reproduction, crossover and mutation. Offspring are normally different from parents due to the genetic information exchange process, e.g. chromosome crossover. However, in GA, the reproduction process is merely a simple coping activity which passes the parent's genetic information to the offspring. The reproduction process usually acts as a complementary process of crossover activity and the offspring are either created by reproduction or crossover.

Crossover is an extremely important component in GA as it is responsible for searching through the solution space. Crossover can be quite complicated and depends (as well as the technique of mutation) mainly on the encoding of chromosomes. Here, double point crossover is employed. After a crossover is performed, mutation takes place. Mutation reintroduces diversity into the population. 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.

Particularly important parameters of GA are the population size (number of individuals) and number of generations. If there is too low number of chromosomes, GA has a few possibilities to perform crossover and only a small part of search space is explored. On the other hand, if there are too many chromosomes, GA slows down. Using the proposed GA, initial genetic parameters are set according Table 2.

Representation of chromosomes is a critical part of GA application. In this work, each chromosome of the population represents a feed rate profile as a sequence of feed rate values. The simplest way to represent it was using a piecewise approximation of the feed rate profile. The profile is divided into equal intervals of 20 minutes 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 20 minutes. In this case, every gene is coded in range $0 – 0.05\ l \cdot h^{-1}$ [8].

An evaluation function plays a role similar to that which the environment pays in natural evolution and it rates chromosome in terms of fitness. The objective functions (*OF*) utilized here, for the simulation tests, are presented as follows:

1. $OF_1 = f(X_{Actual}, X_{Theory})$
2. $OF_2 = f(X_{Actual})$
3. $OF_3 = f(S)$
4. $OF_4 = f(Ph_{Actual}, Ph_{Theory})$
5. $OF_5 = f(X_{Actual}, X_{Theory}, S)$
6. $OF_6 = f(Ph_{Actual})$

The first objective function (OF_1) considers the difference between the actual cell concentration (X_{Actual}) and theoretical maximum cell concentration (X_{Theory}). The second objective function (OF_2) considers only the cell concentration over the fermentation period. The third objective function (OF_3) considers only the substrate concentration (S) over the fermentation period. The fourth objective function (OF_4) considers the difference between the actual phytase concentration (Ph_{Actual}) and theoretical maximum phytase concentration (Ph_{Theory}). The fifth objective function (OF_5) considers the ratio of the substrate concentration and the difference between X_{Actual} and X_{Theory}. The final objective function (OF_6) considers only the phytase concentration over the fermentation period.

Since the evaluation of fitness is a measurement of the individual's suitability to survive in the population, the higher the fitness value, the higher the chance for the individual to survive. However, the dominating effect of some extraordinary individuals in the early generations should be suppressed. In order to maintain the selection pressure throughout the whole evolution process and to help the population to diversify in the early evolution process, dynamic linear scaling is employed. This technique adjusts the fitness value of all the individuals such that only an expected number of offspring will generated from the best individual. Hence, this prevents the dominance of the extraordinary individuals.

Feeding trajectory optimization

All six problems (six *OF*) are running 25 executions with the proposed GA. Average values of best results at a certain evaluation are calculated and presented on the Table 3 and Fig. 1 – Fig. 6. Computational performance of the GA is presented in Table 3.

Table 3. Computational performance

Objective function	OF_1	OF_2	OF_3	OF_4	OF_5	OF_6
CPU time (sec)	84.7190	85.3280	75.1250	75.3750	76.5470	82.1880
floating point operations	45263528	45990118	45078526	45177170	45131382	46097830

The feeding trajectory obtained based on OF_1, as well as the biomass, substrate and phytase concentrations are depicted in Fig. 1. The developed feed profile is acceptable for the whole fermentation period, with an excess substrate in the broth for the first two hours of the fed-batch mode. The cell and the product concentrations have an ideal increase for the complete fermentation period, achieving the values, respectively of 43 $g \cdot l^{-1}$ and 125 $g \cdot l^{-1}$.

The results obtained based on OF_2 are presented in Fig. 2. The developed feed profile is somewhat high for the whole fermentation period and does exhibit a general increase over time. In this instance a consequential excess substrate in the broth is obtained. The cell and the product concentrations have a high increase, achieving the values, respectively of 82 $g \cdot l^{-1}$ and 240 $g \cdot l^{-1}$.

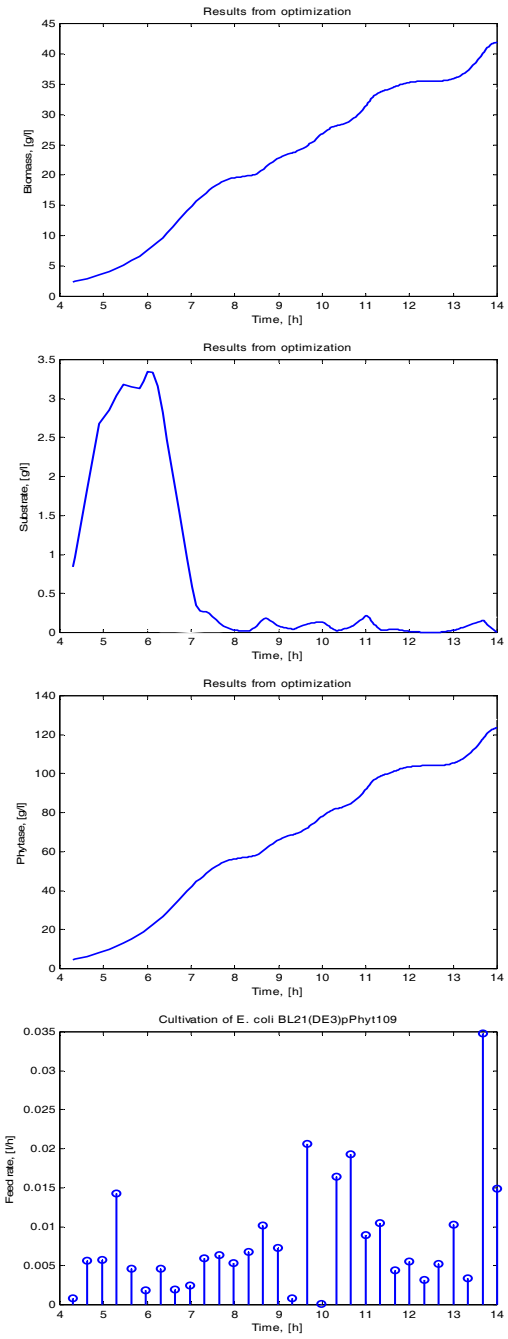

Fig. 1 $f(X_{\text{Actual}}, X_{\text{Theory}})$

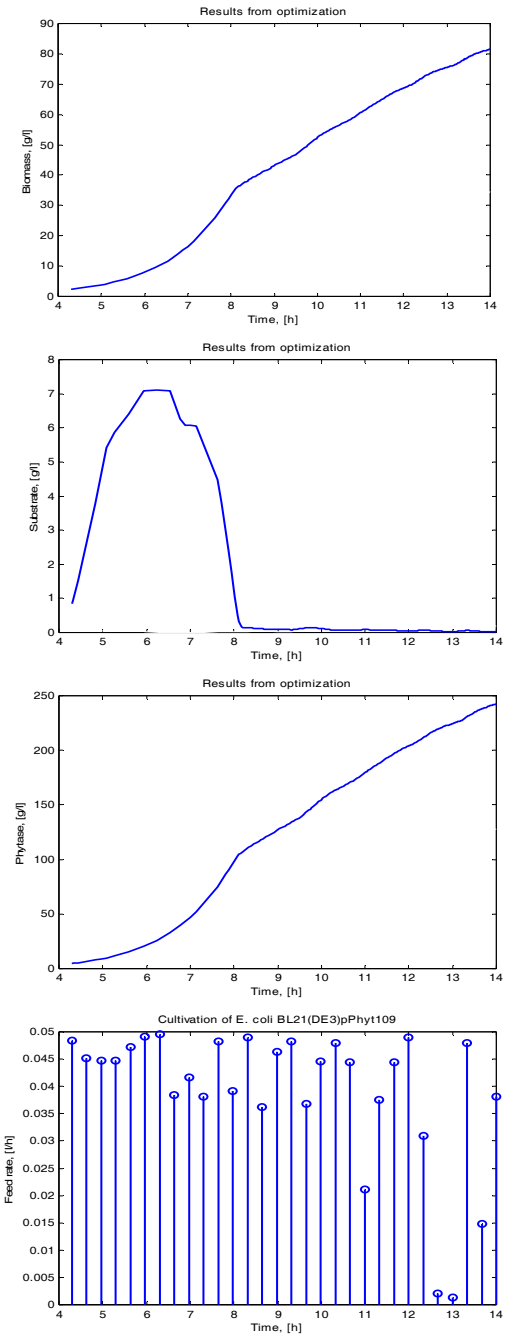

Fig. 2 $f(X_{\text{Actual}})$

In the case of OF_3, the feed profile has a higher rate than the first test (Fig. 3). While there are periods of excess substrate in the broth, basically the substrate is kept to a minimum. However, the final cell concentration is much reduced – 36 $g \cdot l^{-1}$. The final phytase concentration achieves the value of 105 $g \cdot l^{-1}$.

The results obtained based on OF_4 are depicted in Fig. 4. The general level of the feed profile is similar to that for first test. The cell concentration increases over the fermentation period, although its final value is smaller compared to the results obtained based on OF_2. The obtained values are less than these for rest tests – 33 $g \cdot l^{-1}$ for final cell concentration and 90 $g \cdot l^{-1}$ for final phytase concentration.

In case of OF_5 the results are similar to these for first test (Fig. 5). The substrate is kept to a minimum with small periods of excess substrate in the broth. The cell and the product concentrations have increase during the fermentation, achieving the identical to the first test values, respectively of 43 g·l^{-1} and 125 g·l^{-1}.

The feed profile obtained based on OF_6 is the higher than that for the rest tests. This results in a significant amount of excess substrate in the broth, as well as in case of OF_2. The results are presented in Fig. 6. In this instance the maximum of final cell and product concentrations are achieved – 88 g·l^{-1} for final cell concentration and 260 g·l^{-1} for final phytase concentration.

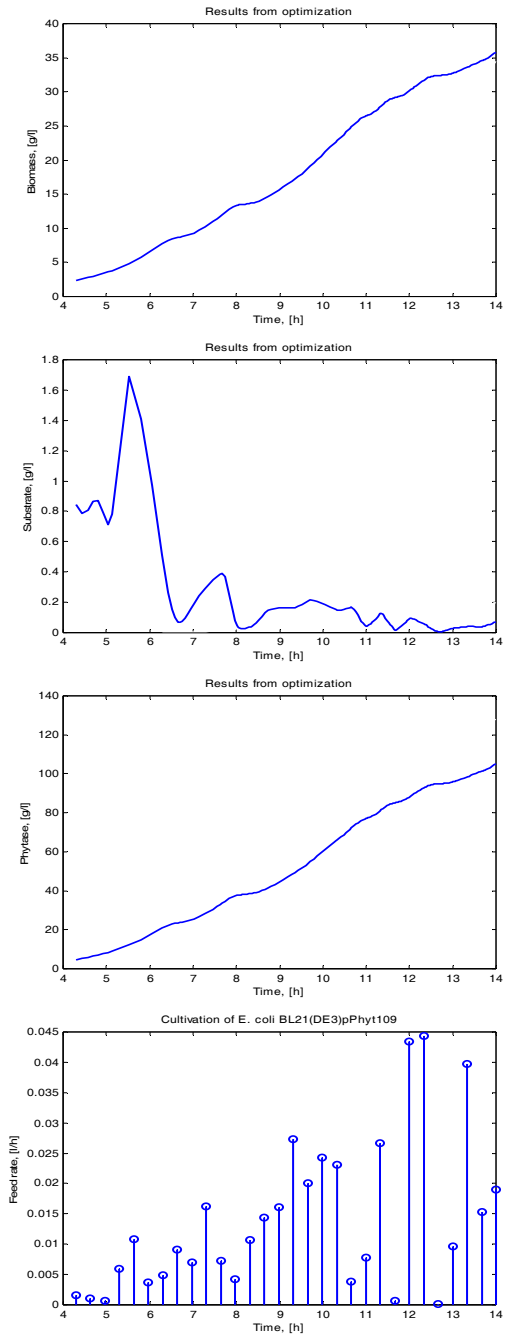

Fig. 3 $f(S)$

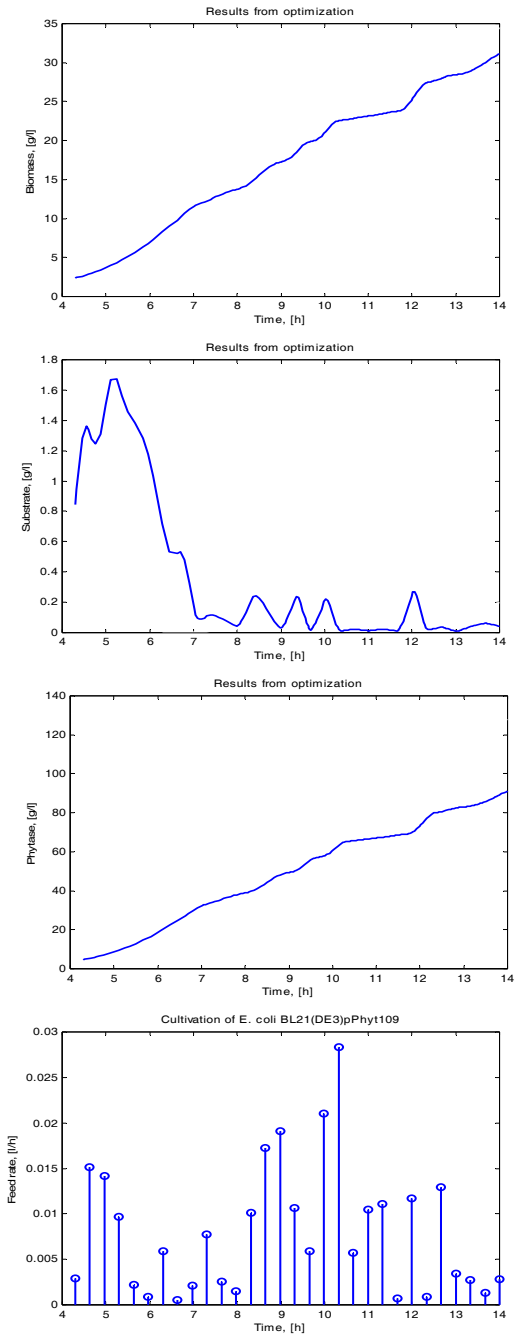

Fig. 4 $f(Ph_{\text{Actual}}, Ph_{\text{Theory}})$

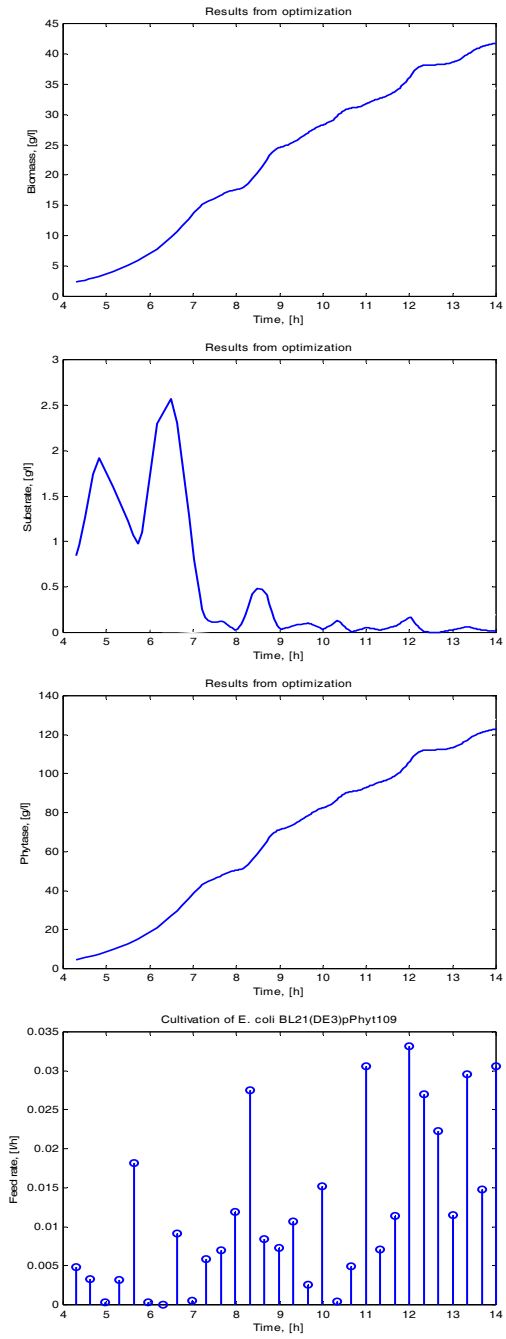

Fig. 5 $f(X_{\text{Actual}}, X_{\text{Theory}}, S)$

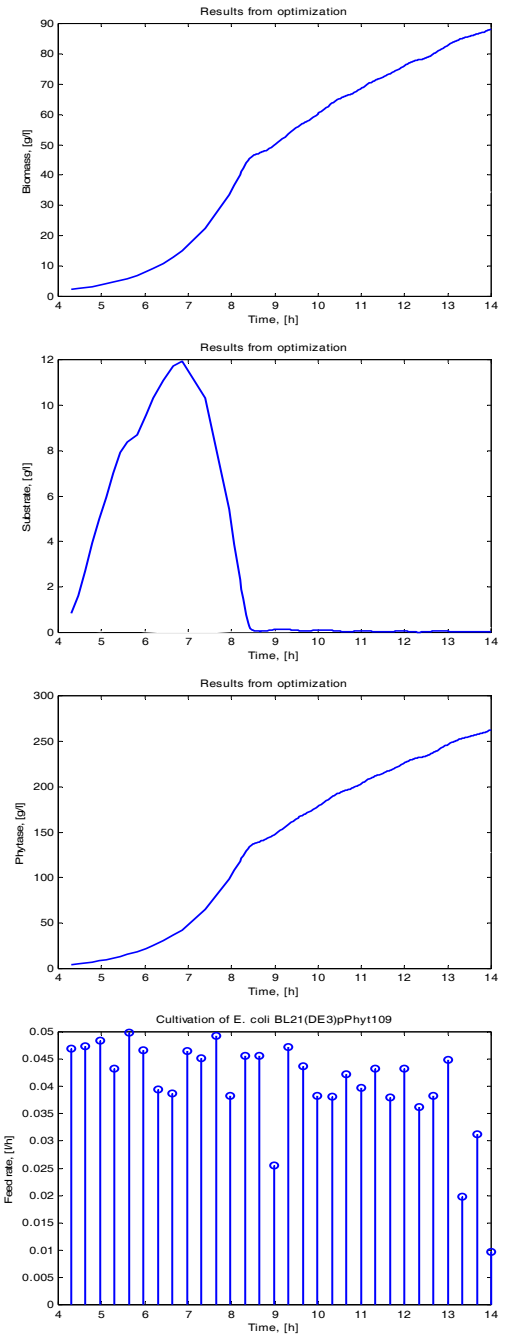

Fig. 6 $f(Ph_{\text{Actual}})$

For genetic algorithm applied to each test, it is clear that the required objective function has been achieved. GA has advantage over other methods in that it does not require any unrealistic assumptions on the objective functions, such as linearity, convexity and differentiability. In addition the problem decision can be reached in a relatively short time running on a PC (Table 3). The proposed approach is found to be an effective and efficient method for solving the optimal feed rate profile problem. However the results seem to indicate that the feed profile formed by the OF_5, considers the ratio of the substrate concentration and the difference between X_{Actual} and X_{Theory}, is superior to rest of the feeding trajectories. The OF_5 gives generally higher final cell and product concentrations and level lower of the excess substrate. This is the fundamental requirement of the fermentation system due to effect economies and process effectiveness.

Conclusion

In this work a genetic algorithm was proposed in order to optimize the feeding trajectory in an *E. coli BL21(DE3)pPhyt109* fermentation process. Technique such as GA is inspired by nature, and has proved themselves to be effective solutions to optimization problems. However, this technique is not a panacea, despite its apparent robustness. There are a lot of parameters involved in the algorithm. In general, some form of trial-and-error tuning is necessary for each particular instance of optimization problem. The appropriate setting of these parameters is a key point for success.

The results, although based on a simulation model, show that the GA is capable of simultaneously optimizing feed rate profile for a given objective function. The main problem in implementation lies in the selection of an appropriate objective function, then once the control parameters have been tuned GA can produce a result.

For all tests the required objective function has been achieved. The results show that the feed profile formed by the objective function considers the ratio of the substrate concentration and the difference between actual cell concentration and theoretical maximum cell concentration is superior to rest of the feed rate profiles. Generally the final cell concentration is higher and the excess substrate level is lower which is the fundamental requirement of the fermentation system. The obtained results of the GA are quite encouraging and its application to these kinds of bioprocesses highly recommended.

Acknowledgements

This work is partially supported from National Science Fund Project № MI – 1505/2005.

References

1. Chen B.-Y., J.-W. You, Y.-T. Hsieh, J.-S. Chang (2008). Feasibility Study of Exponential Feeding Strategy in Fed-batch Cultures for Phenol Degradation using *Cupriavidus taiwanensis*, Biochemical Engineering Journal, 41, 175-180.
2. Chen L. Z., S. K. Nguang, X. D. Chen (2002). 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.
3. Choi J. H., K. C. Keum, S. Y. Lee (2006). Production of Recombinant Proteins by High Cell Density Cultures of *Escherichia coli*, Chemical Engineering Science, 61, 876-885.
4. Choi J.-W., W. Lee, M. C. Jin, Y.-K. Kim, S.-Y. Park, H. L. Won (2002). Control of Feed Rate using Neurocontroller Incorporated with Genetic Algorithm in Fed-batch Cultivation of *Scutellaria baicalensis georgi*, J. of Microbiol. and Biotech., 12(4), 687-691.
5. de Mare L., C. Cimander, A. Elfwing, P. Hagander (2007). Feeding Strategies for *E. coli* Fermentations Demanding an Enriched Environment, Bioprocess Biosyst. Eng., 30, 13-25.
6. Franco-Lara E., D. Weuster-Botz (2005). Estimation of Optimal Feeding Strategies for Fed-Batch Bioprocesses, Bioprocess and Biosystems Engineering, 27, 255-262.
7. Goldberg D. E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley Longman, London.
8. Kleist S., G. Miksch, B. Hitzmann, M. Arndt, K. Freihs, E. Flaschel (2000). Optimization of the Extracellular Production of a Bacterial Phytase with *Escherichia coli* by using Different Fed-batch Fermentation Strategies, Applied Microbiology and Biotechnology, 61, 456-462.

9. Levišauskas D., T. Tekorius (2005). Model-based Optimization of Fed-batch Fermentation Processes using Predetermined Type Feed-rate Time Profiles. A Comparative Study, Information Technology and Control, 34(3), 231-236.
10. Marteijn R., O. Jurrius, J. Dhont, C. de Gooijer, J. Tramper, D. Martens (2003). Optimization of a Feed Medium for Fed-batch Culture of Insect Cells using a Genetic Algorithm, Biotechnol. Bioeng., 81, 269-278.
11. MatWorks Inc. (1999). Genetic Algorithms Toolbox, User's Guide.
12. Miguel A. S. M., M. Vitolo, A. Jr. Pessoa (2007). Fed-batch Culture of Recombinant *Saccharomyces cerevisiae* for Glucose 6-phosphate dehydrogenase Production, Biochem. Eng. J., 33, 248-252.
13. Rocha M., J. Neves, I. Rocha, E. C. Ferreira (2004). Evolutionary Algorithms for Optimal Control in Fed-batch Fermentation Processes, G. R. Raidl et al. (Eds.): EvoWorkshop 2004, LNCS 3005, 84-93.
14. Roeva O., St. Tzonkov (2005). Optimal Feed Rate Control of *Escherichia coli* Fed-batch Fermentation, Bioautomation, 2, 30-36.
15. Samad N. A., A. Ahmad (2006). Comparison Product Optimization Performance of Fed-batch Fermentation Processes for Penicillin G Production, Proc. of the 1st Int. Conf. on Natural Resources Engineering & Technology, 24-25th July 2006, Malaysia, 562-575.
16. Shin H. S., H. C. Lim (2007). Maximization of Metabolite in Fed-batch Cultures Sufficient Conditions for Singular Arc and Optimal Feed Rate Profiles, Biochemical Engineering Journal, 37, 62-74.
17. Vanichsriratana W. (2000). Optimization of a Primary Metabolite Fermentation Process: Effect of Cost Factor on the Optimal Feed Rate Control, Kasetsart J. (Nat. Sci.), 34, 495-499.

Olympia Roeva, Ph.D.
E-mail: olympia@clbme.bas.bg

Olympia Roeva was born in 1974. She received the M. Sc. Degree (1998) and Ph.D. Degree (2007) from the Technical University – Sofia. At present she is a Research Associate at the Centre of Biomedical Engineering – BAS. She has above 50 publications among those 5 books and book chapters. Her current scientific interests are in the fields of Modelling, Optimization and Control of Biotechnological Processes, Genetic Algorithms and Generalized Nets.

Prof. Stoyan Tzonkov, D.Sc., Ph.D.
E-mail: tzonkov@clbme.bas.bg

Prof. Stoyan Tzonkov graduated the Technical University – Sofia in 1966. Since 1984 he is a Doctor of Technical Science and from 1987 he is a Professor. Since 1994 he is the Head of Department of "Modelling and Optimization of Bioprocess Systems", Centre of Biomedical Engineering – BAS. He has more than 300 publications among those 30 books, book chapters and textbooks with more than 258 known citations. Scientific interests: Modeling and Optimization, Control Systems, Complex Control Systems, Variable Structure Systems, Bioprocess Engineering.