Model-based and Experimental Optimization Using Genetic Algorithm

Z. Findrik, B. Zelić, S. Bogdan*, and Đ. Vasić-Rački**
Faculty of Chemical Engineering and Technology, University of Zagreb,
Savska c. 16, HR-10000 Zagreb, Croatia
*Pliva d. d., Research and Development,
Prilaz baruna Filipovića 25, HR-10000 Zagreb, Croatia

Original scientific paper
Received: February 11, 2004
Accepted: May 11, 2004

Dedicated to Prof. Egon Bauman

For the purpose of this paper genetic algorithm (GA) was used for a model-based optimization of the enzymatic synthesis of N-acetyl neuraminic acid and experimental optimization of the initial conditions of an aqueous two-phase extraction system.

Model-based optimization of enzymatic synthesis of N-acetyl neuraminic acid was carried out by using an earlier developed and verified mathematical model for this system in a batch reactor.[27] The influence of pyruvate (c_{Pyr}) and N-acetyl glucosamine (c_{GlcNAc}) initial concentrations, epimerase (E_{epi}) and aldolase (E_{ald}) concentrations and reaction time (t) were investigated to optimize volumetric productivity (Q_P) and ratio (R) of the final concentrations of N-acetyl neuraminic acid and pyruvate. In this example parameters were optimized within 40 generations of 8 experiments. Maximal volumetric productivity of Q_P = 47.8 g/(dm^3 d) in the batch reactor and ratio of R = 2.57 were estimated under the following initial operating conditions: c_{Pyr} = 1 mmol/dm^3, c_{GlcNAc} = 460 mmol/dm^3, E_{epi} = 39.9 mg/cm^3, E_{ald} = 0.2 mg/cm^3 and t = 156.8 min. Conversion of N-acetylglucosamine to N-acetylneuraminic acid at these conditions was 0.15 %, and conversion of pyruvate to N-acetylneuraminic acid was 71.99 %.

Experimental optimization using GA was applied to investigate the influence of the mass fraction of PEG-6000 (w_{PEG}), the mass fraction of ammonium sulfate (w_{AS}), pH and BS Albumine (BSA) concentration c_{BSA} on partition coefficient of BSA K_{BSA} in aqueous two-phase system consisting of PEG-6000 and ammonium sulfate. The partition coefficient was optimized within five sets of four experiments. It was found out that the mass fraction of PEG-6000 of 0.1770, the mass fraction of ammonium sulfate of 0.1105, pH of 7.0 and BSA concentration of 2.5 g/dm^3 are the best working conditions to achieve the optimal partition coefficient of K_{BSA} = 0.4793.

Keywords:
Genetic algorithm, model-based optimization, experimental optimization, enzymatic synthesis of N-acetyl neuraminic acid, aqueous two-phase system, BS Albumine

Introduction

One of the basic problems in the bioprocesses development is the optimization of initial experimental conditions. Considering that, there is a great deal of parameters that can influence the outcome of bioprocess, it is of general concern to find a good optimization method. High number of parameters considerably narrows the choice of optimization methods, and it also indicates the system's complexity. The usual methods generally used for this purpose (e.g. simplex,[1] EVOP,[2] response surface methodology[3]) are not good enough when there are too many parameters involved, i.e. they are sometimes not successful in finding the optimum. In these cases it is very useful to use the genetic algorithm (GA) which is not necessary to solve problems of comparatively smaller magnitude, but promising for complicated multidimensional problems.[4]

Optimization itself represents considerable cost, due to a great number of experiments necessary to find optimal conditions. If long-term optimization benefits are considered, it can be stated that its cost is of no importance in comparison to its benefits in large-scale processes.[5]

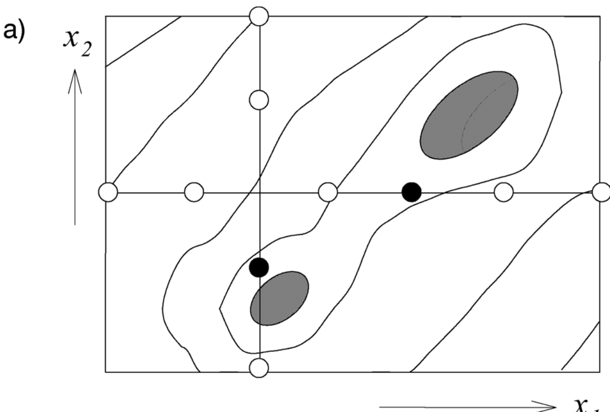
Generally speaking, there are two kinds of optimization methods: statistical and stochastic. Statistical methods include various methods of experimental design. They include four steps: (1) identification of the most important media components,

**Corresponding author: Faculty of Chemical Engineering and Technology, University of Zagreb, Savska c. 16, HR-10000 Zagreb, Croatia, Phone: + 385 1 4597 104, Fax: + 385 1 4597 133,
E-mail: dvracki@marie.fkit.hr

called screening, (2) identification of the optimum variable range or 'narrowing', (3) identification of the optimum or optimum search and (4) experimental verification of identified optimum. The success of these methods depends of a random choice of a variable range, which doesn't have to be always correct. This leads to another limitation of these methods: the target function has to be unimodal, otherwise a local optimum can be identified instead of the global one.[6]

One of the examples of these methods is one-dimensional search with successive variation of variables, the classical procedure for optimizing fermentation media. It has been employed as yet and even seems to be the most frequently used[6]. This is very surprising since many examples show that it is practically impossible for one-dimensional search to achieve the optimum in a finite number of experiments.[7–10] General principle of one-dimensional search is shown in Figure 1a.

All stochastic methods of generations are based on the following procedure: (1) generating a first experimental plan, (2) determining the values of the target function, (3) generating new experimental points with the quality of the previous generation by correlating the number of new experimental points relative to the target function and by fitting the step width around a previous experimental point to a global optimization success.[6]



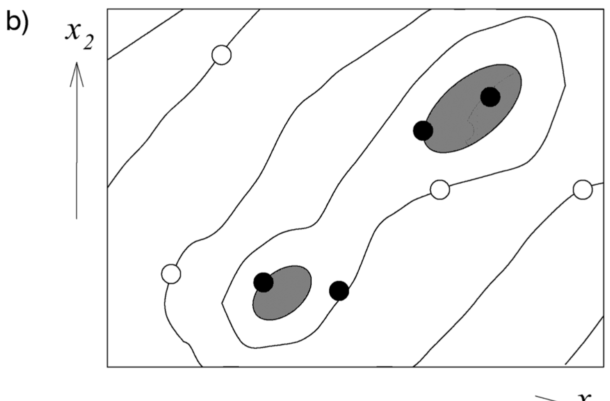


Fig. 1 – *a) Statistical optimization methods: Simplified principle of one-dimensional search; b) Stochastic optimization methods: Simplified principle of genetic algorithm.* ○ *the first generation of experiments* ● *the last generation of experiments*

Genetic algorithm (GA) is a stochastic optimization method based on the principals of evolution. It is quite commonly used for experimental optimization, but is also used for parameter estimation of the nonlinear systems.[11] It can also be used for optimization of initial conditions when mathematical model of the process is available.

GA's were first devised by Holland in 1975. They are non-model-based optimization methods that can search a large parameter space in a highly directed way.[12] Since the early nineties of the last century, greater use of GA has been noticed in various scientific fields.[4,6–24] Using genetic algorithm, it has been experimentally verified, with help of process examples, that process improvements can be achieved for both microbial and enzymatic conversions and for cell cultures, despite the large number of medium components under simultaneous consideration. In exploring a variable space, process improvements of more than 100 % were generally achieved, while in those cases where standard optimization procedures were involved, improvement reached only 20–40 % of the target quantity.[6] In comparison to other methods, GA considerably decreases the number of experiments. There is a good example corroborating this statement. During continuous production of format dehydrogenase, a medium for growth of *Candida boidinii* was optimized in order to improve the volumetric productivity and the activity of the enzyme[14]. The concentrations of 14 medium components were optimized within 125 experiments in shake flasks. It would take 101^{14} experiments for the full experimental plan. The results hereof were improved by 50 %. It can be stated that GA differs from traditional and usual methods of optimization. It offers a number of possible solutions to the problem and it is up to a user to decide which choice will be final. The simplified principal of GA is shown in Figure 1b.

This paper presents in what manner genetic algorithm can be used for model-based (Example 1: Enzymatic synthesis of N-acetyl neuraminic acid) and experimental (Example 2: Aqueous two-phase extraction of BSA) optimization.

Genetic algorithm (GA)

Principals of genetic algorithm

GA uses the rules of evolution on a population of potential solutions to produce better individuals.

The mechanism used for this purpose is called selection. Choosing individuals from each generation considering their qualities, a new set of approximations is being created.

Individuals are encoded as binary strings, i.e. chromosomes, so that the genotypes (chromosome values) are uniquely mapped onto the decision variable, phenotypic domain. For example, a problem with two variables x_1 and x_2 may be mapped onto the chromosome structure as it is shown in Figure 2a. x_1 is coded with 10 bits, and x_2 with 15 bits, possibly reflecting the level of accuracy or range of the individual decision variables. During the reproduction phase, each individual is assigned a fitness value derived from its raw performance measure given by the objective function. This value is used in the selection to bias towards more fit individuals. Highly fit individuals, relative to the whole population, have a high probability of being selected for mating whereas less fit individuals have correspondingly low probability of being selected.[25]

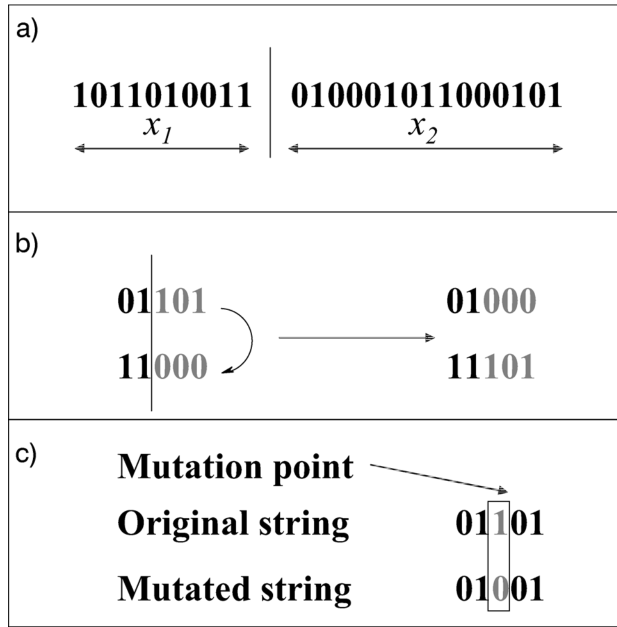


Fig. 2 – *Principals of genetic algorithm. a) The presentation of an individual as a binary string. x_1 is coded with 10 bits, and x_2 with 15 bits; b) Example of crossover. It involves two binary strings, and the place where the crossover takes place is at i = 2; c) Example of mutation. Occasional change of one bit from 1 to 0 takes place.*

Genetic operators manipulate directly with the chromosome genes with the assumption that certain gene codes represent better individuals. The basic operator for the production of new chromosomes is crossover. It creates new individuals that have genetic material from both parents. As the crossover takes place, the place of its occurrence is randomly being chosen. It can be at any position i of the integer, at the interval [1, l–1] where l is the length of the integer. Two new offspring arise hereout.[25] The simplest crossover is shown in Figure 2b. It involves two binary strings, and the place where the crossover takes place is at i = 2. It is important to emphasize that crossover does not occur always and not in all individuals, but with a certain probability p, that is assigned by the algorithm.

Besides crossover, there is another operator that has important role in GA. It is called mutation. The mutation in nature is a random process at which change in genetic code takes place. In genetic algorithms mutation is used as an operator with low occurrence probability. This probability is adjusted in the algorithm, and it is usually between 0.001 and 0.01. Mutation may not occur too often because of the possible loss of good genetic material, which can lead away from optimum values. Mutation enables the appearance of each possible string, i.e. each possible individual or gene combination with a certain possibility.[25] Mutation can also restore a good genetic material that can be lost by selection or crossover. Figure 2c shows the example of mutation where the occasional change of one bit from 1 to 0 takes place.

Use of software GALOP

The software GALOP (Genetic Algorithm for the Optimization of Processes) Version 1.24 developed at the Institute of Biotechnology, Research Centre Jülich, Germany, was used for both model-based and experimental optimization.[26]

Use of software GALOP includes a program that consists of:

– the list of parameters (including lower and upper boundary of each parameter optimized and the step size by which these variables change);

– the list of target functions that need to be optimized;

– the list of target functions that need to be optimized with weighting factors and belonging functions.

After the initiation of the algorithm, the first generation is created. The algorithm can randomly create it, or the author can create it by himself. Having performed model simulations or experiments, values of target functions should be written in GA as return information, which GA then uses for further adjustments of individuals. If objective of optimization is reached, GA will end, otherwise new generation will be proposed. The following diagram shows a typical program run (Figure 3).

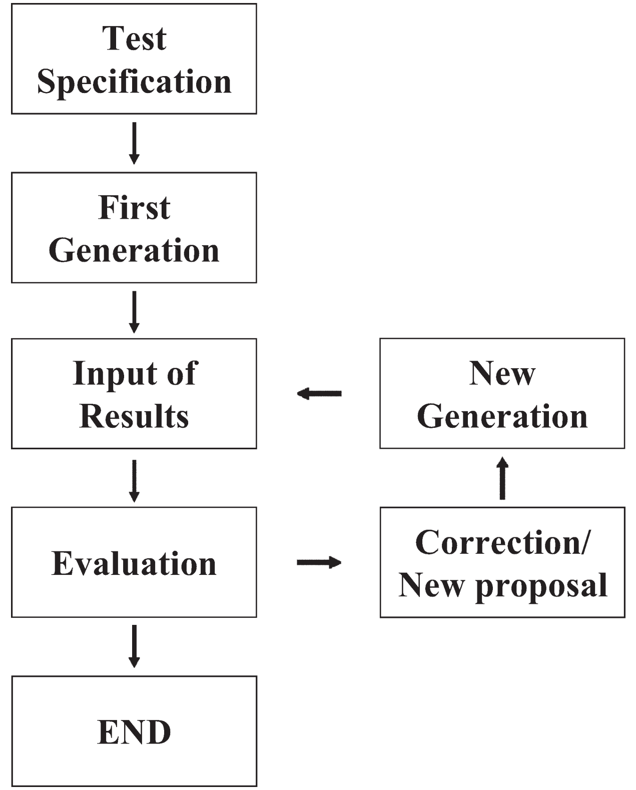


Fig. 3 – *Flowchart of the GALOP software*

Model based optimization for the enzymatic synthesis of N-acetyl neuraminic acid

N-acetyl neuraminic acid (Neu5Ac) is a representative of amino sugars and plays an important role in biological processes. It is a component of drugs used for treating the diseases of ear, nose and throat. It is also used for the production of N-acetyl neuraminic acid derivates known as selective inhibitors of enzyme sialidase and as blockers for its binding sites.[27]

In this paper GA was used for the purpose of optimization of enzyme synthesis of N-acetyl neuraminic acid. It is a two-stage synthesis, which involves the reaction of epimerisation of N-acetyl glucosamine (GlcNAc), which is catalysed by 2-acylamido-2-deoxy D-glucose 2 epimerase (epimerase, EC 5.1.3.8.). This reaction produces N-acetyl mannosamin (ManNAc) necessary for the main reaction. This compound reacts with pyruvic acid (Pyr) in the reaction catalyzed by the enzyme N-acetilneuraminic acid aldolase (aldolase, EC 4.1.3.3.) and produces N-acetyl neuraminic acid (Figure 4). Namely, N-acetyl glucosamine is relatively cheap, and N-acetyl mannosamine is quite expensive, and that is why it is more profitable to carry out this reaction stage wise, especially when it's known that epimerase is an inexpensive enzyme. The experiments were carried out in the enzyme membrane reactor. The model was written and verified by these experiments.[27]

The mathematical model

The mathematical model[27] for this system is given by the equations 1–6. The kinetic model is presented by the equations 1 and 2, and the equations from 3–6 are balance equations for the batch reactor.

Fig. 4 – *The reaction scheme for the enzymatic synthesis of N-acetyl neuraminic acid*

$$\nu_1 = \frac{E_{\text{ald}}\left(\frac{V_{\max}^{\text{F}} c_{\text{Pyr}} c_{\text{ManNAc}}}{K_i^{\text{Pyr}} K_m^{\text{ManNAc}}} - \frac{V_{\max}^{\text{B}} c_{\text{Neu5Ac}}}{K_m^{\text{Neu5Ac}}}\right)\left(\frac{1}{1+(c_{\text{Pyr0}}+c_{\text{ManNAc0}}+c_{\text{Neu5Ac0}})/K_\nu}\right)}{1+\frac{c_{\text{Pyr}}}{K_i^{\text{Pyr}}}+\frac{K_m^{\text{Pyr}} c_{\text{ManNAc}}}{K_i^{\text{Pyr}} K_m^{\text{ManNAc}}}+\frac{c_{\text{Pyr}} c_{\text{ManNAc}}}{K_i^{\text{Pyr}} K_m^{\text{ManNAc}}}+\frac{c_{\text{Neu5Ac}}}{K_m^{\text{Neu5Ac}}}+\frac{c_{\text{Neu5Ac}} c_{\text{ManNAc}}}{K_i^{\text{ManNAc}} K_m^{\text{Neu5Ac}}}} \tag{1}$$

$$\nu_2 = \frac{E_{\text{epi}}\left(\frac{V_{\max}^{\text{GlcNAc}} c_{\text{GlcNAc}}}{K_m^{\text{GlcNAc}}} - \frac{V_{\max}^{\text{ManNAc}} c_{\text{ManNAc}}}{K_m^{\text{ManNAc*}}}\right)}{1+\frac{c_{\text{GlcNAc}}}{K_m^{\text{GlcNAc}}}+\frac{c_{\text{ManNAc}}}{K_m^{\text{ManNAc*}}}} \tag{2}$$

$$\frac{\mathrm{d}c_{\text{GlcNAc}}}{\mathrm{d}t} = -\nu_2 \tag{3}$$

$$\frac{\mathrm{d}c_{\text{ManNAc}}}{\mathrm{d}t} = \nu_2 - \nu_1 \tag{4}$$

$$\frac{\mathrm{d}c_{\text{Pyr}}}{\mathrm{d}t} = -\nu_1 \tag{5}$$

$$\frac{\mathrm{d}c_{\text{Neu5Ac}}}{\mathrm{d}t} = \nu_1 \tag{6}$$

Model parameters were optimized and presented in the literature[27] (Table 1).

Table 1 – *Values of parameters of the mathematical model for the enzymatic synthesis of N-acetyl neuraminic acid*

K_m^{GlcNAc}	[mmol/dm^3]	14.01
$K_m^{\text{ManNAc*}}$	[mmol/dm^3]	35.19
K_m^{Pyr}	[mmol/dm^3]	0.136
K_m^{NeuAc}	[mmol/dm^3]	9.44
K_m^{ManNAc}	[mmol/dm^3]	402.2
K_i^{Pyr}	[mmol/dm^3]	1.301
K_i^{ManNAc}	[mmol/dm^3]	23.76
K_v	[mmol/dm^3]	1556
$K_{\max}^{\text{ManNAc}}$	[μkat/g]	0.829
$K_{\max}^{\text{GlcNAc}}$	[μkat/g]	0.078
$V_{\max}^{\text{F}}$	[μkat/g]	13.8
$V_{\max}^{\text{B}}$	[μkat/g]	8.51

*adapted from ref 27

Model-based optimization

The aim of this work was optimization of initial conditions of the N-acetyl neuraminic acid synthesis. That means finding the best initial concentrations of pyruvic acid and N-acetyl glucosamine, and concentrations of two enzymes: aldolase and epimerase, as well as the reaction time in the batch enzyme membrane reactor. The target function for that purpose was maximum value of volumetric productivity, Q_P, (Equation 7) and maximum value of the ratio, R, of the concentration of N-acetyl neuraminic acid and pyruvic acid at the end of experiment (Equation 8). This ratio is very important from the aspect of downstream processing. Namely, the isolation of N-acetyl neuraminic acid is being carried out by ion-exchange chromatography, and these two compounds have similar pK_a values, so it is not easy to separate them.[27]

$$Q_{\text{P}} = \frac{c_{\text{Neu5Ac}} M_{\text{Neu5Ac}}}{t} \tag{7}$$

$$R = \frac{c_{\text{Neu5Ac}}}{c_{\text{Pyr}}} \tag{8}$$

Start values of investigated parameters, namely, initial concentrations of pyruvic acid and N-acetyl glucosamine, concentrations of aldolase and epimerase, and reaction time, investigated area of parameters (defined by lower and upper border) and optimization step are shown in Table 2. Equal importance was given to both target functions by the choice of the same weighting factors, which was 1 in both cases. Each function was written as linear combination of given parameters. The evaluation functions are generally optimized on the principle that higher number is better than a lower number. Negative coefficients indicate the search for the minimum, and positive the search for the maxi-

Table 2 – *Start values of investigated parameters, investigated area of parameters (defined by lower and upper border) and step of the optimization for model-based optimization*

Parameter	Unit	Step	Lower border	Upper border
c_{Pyr}	mmol/dm^3	1	1	2000
c_{GlcNAc}	mmol/dm^3	1	1	1000
E_{epi}	mg/cm^3	0.1	0.1	40
E_{ald}	mg/cm^3	0.05	0.05	5
t	s	10	10	43200

mum. If there is no coefficient in front of certain parameter that indicates that there is no special requirement concerning it. Coefficient "+1" was written for both target functions (Q_P and R) in the linear combination, and "–1" for the parameter of aldolase concentration. Both target functions had the same linear combination of parameters, i.e. the same coefficients.

Mutation occurrence in the program was set to 0.01, which indicates the probability of 1 % for mutation to appear. The crossover occurrence was set to 0.95 and the number of individuals in a generation was 8.

The principle of optimization was as follows: GA offers the first random population of 8 individuals with given characteristics (initial concentrations of substrates and enzymes and the time of the reaction). These values were assigned to the program written in SCIENTIST software[28] (which consists of kinetic and mass balance equations of the mathematical model) and the reactions at these conditions were then simulated. Values for volumetric productivity, and product ratio estimated by SCIENTIST were written in GA as return information, which GA used for further adjustments of individuals.

The model-based optimization is not usual in cases where model is involved. It is usually used for the cases where it is very difficult to set a model, or there is no way to write it. It can also be used for parameter optimization of mathematical models. One example of parameter optimization in the literature shows the changes of their values through 1500 of generations[11]. This example shows that finding the optimal values of parameters does not necessarily mean uniformity of all parameters through out the generations. If good improvement of the results is achieved, then it doesn't really matter how many generations were made. Besides these facts that are already mentioned, GA can truly achieve the results improvement by 100 % in comparison to other optimization methods.[6]

Graphic analysis of results is given in Figures 5 and 6. Optimization was carried out through 40

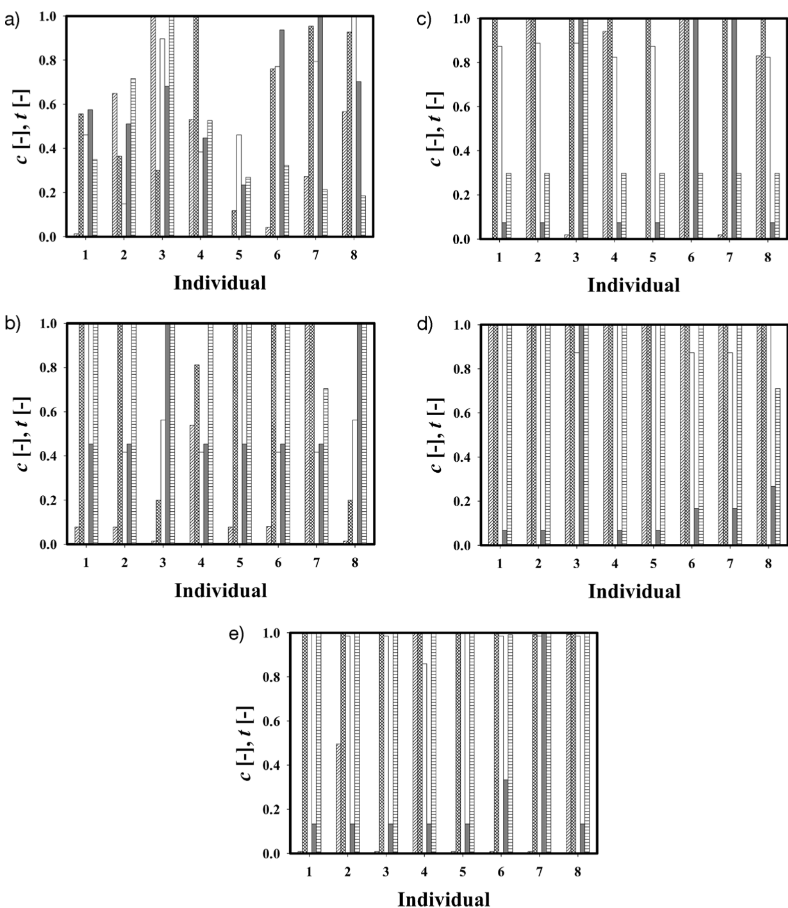
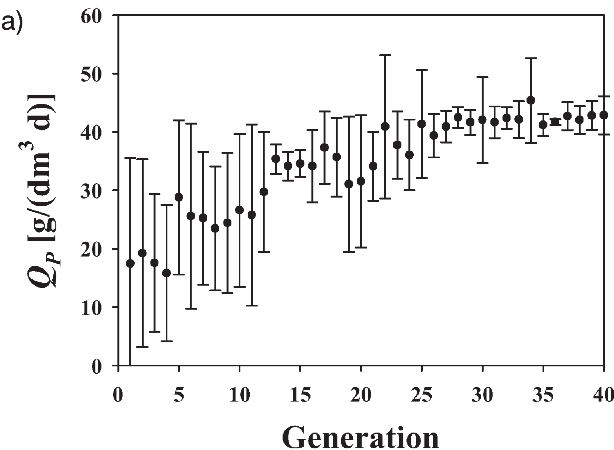


Fig. 5 – *a) Values of optimization parameters for the first generation; b) Values of optimization parameters for the tenth generation; c) Values of optimization parameters for the twentieth generation; d) Values of optimization parameters for the thirtieth generation; e) Values of optimization parameters for the fortieth generation.* ▧ *normalized concentration of pyruvate,* ▩ *normalized concentration of N-acetyl glucosamine,* ▭ *normalized concentration of epimerase,* ■ *normalized concentration of aldolase,* ▥ *normalized reaction time*

a)



b)

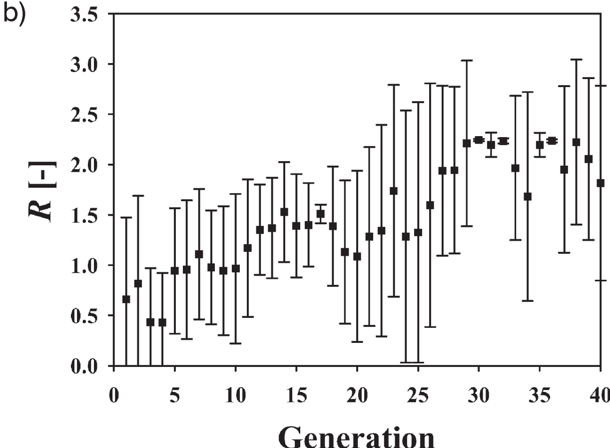


Fig. 6 – *a) Changes in the volumetric productivity (●) through generations; b) Changes in the value of the ratio (■) of N-acetyl neuraminic acid and pyruvic acid through generations.* I *standard deviation of data*

generations. The results in Figure 5 are presented as normalized values (calculated by Equation 9, *y* in the equation is the concentration or reaction time) of all eight individuals of a certain generation: the first, the tenth, the twentieth, the thirtieth and the fortieth. The results in Figure 6 are presented as middle values of target function of a certain generation (middle value of 8 individuals) with standard deviations. It can be seen that uniformity of all parameters did not occur. (The concentration of pyruvate was equal for 5 individuals of the last generation. The concentration of N-acetyl glucosamine was equal in all individuals of the last generation. The concentration of epimerase was practically the same for all individuals of the fortieth generation, only with small deviations seen. The concentration of aldolase was the same for six individuals of the last generation.)

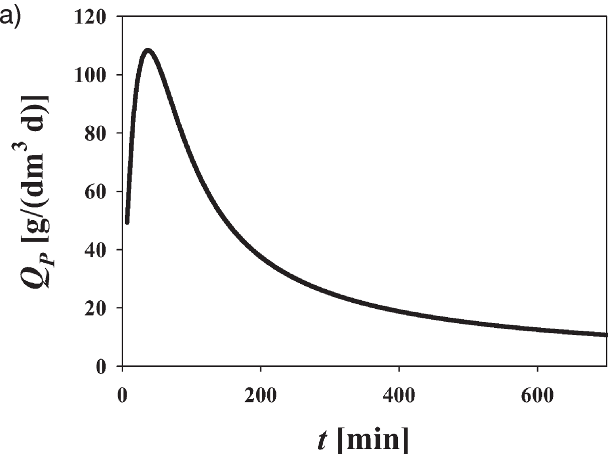
Despite that, great improvements in target functions value through the generations are obvious (Figure 6 a and b). It can be concluded that selected target functions are highly sensitive to small changes of parameters. That especially goes for the sensitivity of the ratio target function, and it is obvious from the standard deviations in Figure 6b. With reference to the target function of volumetric productivity, it can be stated that value of this function was pretty stable through more than last ten generation. That is not the case with the ratio target function. It is obvious that it should take more calculations to reach the optimum for this target function. The target function *R* is highly sensitive to small changes in concentration of aldolase, as is expected, because both concentrations Neu5Ac and Pyr are highly dependant on concentrations of enzymes.

$$y = \frac{y_i}{y_{\max}} \tag{9}$$

Up to this point, the role of reaction time hasn't been mentioned in the discussion. It seems that GA was inert in changing the reaction time, because from the second generation to the end there were only small changes in its value. This is not so obvious from Figure 5 a, b, c and d because these values are normalized, but a great percentage of all individuals had time parameter around 150 minutes. The question is why GA did not alter the reaction time considerably through generations?

The answer is quite simple actually. If one look is taken upon linear combinations of target functions and its requirements, it can be seen that besides maximization of target functions, other condition in finding the optimum was also minimal concentration of aldolase. Considering that requirement, GA almost always offered lower concentration of aldolase, but still within the range where slight changes of its value can highly influence the reaction rate, i.e. in the range where the enzyme concentration is a limiting factor, and not the reaction time itself, for the production of Neu5Ac. This can be proven by the results obtained from the measuring of the dependence of specific activity on the concentration of aldolase.[27] One could conclude that time has no particular influence on the target function. But that is not the case. Figures 7a and b show to what extant volumetric productivity and the ratio depend of the reaction time. Optimal values of substrate concentrations and enzyme concentrations from GA were taken as initial values in simulation. It is obvious that volumetric productivity highly depends on the reaction time, and as for as the ratio is concerned, it also depends on the reaction time, but only at the beginning of the reaction.

In the given example of GA optimization 40 generations were calculated and no further attempt was made. The simulation results include 5 % error. Optimum value of volumetric productivity Q_P = 47.8 g/(dm^3 d) and the ratio of N-acetyl neuraminic acid and pyruvate R = 2.57 were estimated under



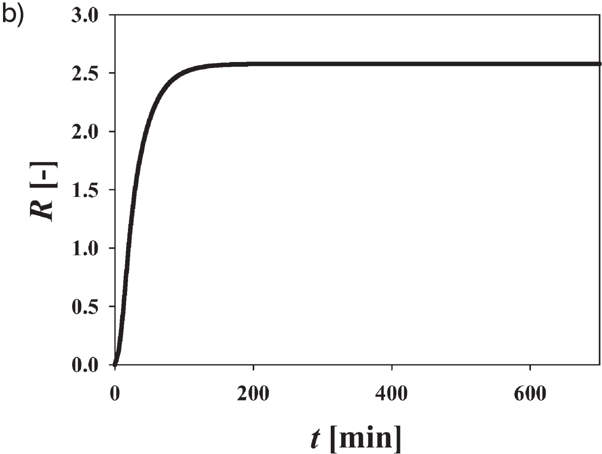


Fig. 7 – *a) Dependence of volumetric productivity on the reaction time at optimal initial concentrations of substrates and enzymes; b) Dependence of ratio of the final concentration of N-acetyl neuraminic acid and pyruvic acid on the reaction time at optimal initial concentrations of substrates and enzymes*

these conditions: c_{Pyr} = 1 mmol/dm^3, c_{GlcNAc} = 460 mmol/dm^3, E_{epi} = 39.9 mg/cm^3 and E_{ald} = 0.2 mg/cm^3. As regards the reaction time GA gave value of 156.8 min. At these optimal conditions conversion of N-acetylglucosamine to N-acetylneuraminic acid was 0.15 %, and conversion of pyruvate to N-acetylneuraminic acid was 71.99 %. It can be concluded from the Figure 7 that another compromise can be made between these two target functions. For example, at reaction time 57.6 min the ratio is 2.21 and volumetric productivity is 99.97 g/(dm^3 d). The choice of optimum here depends on our priority. Considering the earlier mentioned (regarding more calculations being needed for achievement of the stability of both target functions) the obvious about the genetic algorithm could be: GA lands either at global optimum or at a value nearer enough to be considered as global optimum.[24] The more important fact in this problem of finding the optimum is the advantage of GA in finding the global and not local optimum, which is provided by unique mechanisms of evolution.

Weakness of the obtained results lies in the lack of experimental evaluation for which unfortunately there was no opportunity. Considering these results, performing model-based optimization with GA may seem discouraging due to a high number of generations and a long time spent in front of computer. Usually, this is not the case. Namely, in this particular case the situation was more complicated because of two target functions involved and the requirement for low aldolase concentration. If one target function is involved, the optimum could be usually reached up to tenth generation.

Experimental optimization of an aqueous two-phase extraction of BSA

Aqueous two-phase systems have a great potential for the extraction of bioproducts.[1,3,29–39] The aqueous environment provides mild conditions for bio-molecules so that denaturation which often occurs in organic solvents hardly takes place.[40] To form the two aqueous phases, usually polyethylene glycol (PEG) and ammonium sulfate (AS) (or PEG-dextran) are required.

General properties of aqueous two-phase system have been studied by several researchers.[41,42] However, the mechanism governing the partition of biological materials isn't completely understood. The observed partition coefficient is a result of van der Waals, hydrophobic, hydrogen bond, and ionic interactions of bio-molecules with the surrounding phase. Therefore, the partition coefficient is influenced by many factors, including the concentrations and molecular weights of PEG, type and concentration of added salts, temperature and pH.[43]

The aqueous two-phase system, consisting of PEG-6000 and ammonium sulfate, was used to investigate the behavior of batch system for extraction of BS Albumin model solution. This system was used to illustrate how GA can be applied for experimental optimization. GA was proved to be a reliable method for the optimization of process conditions for protein extraction in the PEG-salt system[1]. Since GA is not based on any assumption it can easily cope with irregularities of the aqueous two-phase system. It is neither harmed by very small values nor is limited at the maximum. In comparison to other optimization methods (e.g. steepest ascent, simplex), GA does not need to be further adopted or limited just to begin the optimization process. The batch experiments were used to examine the effects of pH, concentration of PEG-6000, concentration of ammonium sulfate and the enzyme concentration on partitioning in aqueous two-phase system. The partition coefficient (K, ratio of equilibrium concentrations of enzyme in the

extract phase, c^{E*}, and in the rafinate phase, c^{R*}) defined by Eq. 10 was optimized using the genetic algorithm.

$$K = \frac{c^{E*}}{c^{R*}} \tag{10}$$

Materials and methods

PEG-6000 with an average molecular weight of 6000 g/mol, ammonium sulfate, and BSA were obtained from "Merck". Concentration of stock solutions of PEG-6000 and ammonium sulfate were 40 % (mass fraction). The concentration of NaCl of 0.9 % (mass fraction) was used to prepare stock solution of BSA with an accurately known concentration 1.5 g/dm^3. Different stock solutions of K_2HPO_4 and KH_2PO_4 were used to prepare the aqueous two-phase system at different pHs.

Aqueous two phase partitioning experiments were performed at 20 °C by mixing the determined volume of the phase forming polymer solution with solutions of salt and BSA in the graduate cylinder. The buffer solution was then added to obtain the final volume of 5 cm^3. The system was mixed and then left overnight. After 24 hours samples were carefully withdrawn from the top (extract) phase and from the bottom (rafinate) phase. In addition to the BSA concentration determination, the absorbance of each sample was measured using the spectrophotometer (Lamda EZ 210, Perkin Elmer) at 280 nm.

Experimental optimization

Phase diagram and thermodynamic parameters of aqueous two-phase system consisting of PEG-6000 and ammonium sulfate were determined previously and published elsewhere[29,30]. It was assumed that the presence of different concentrations of BSA and pH used in the experimental optimization do not have any influence on thermodynamic equilibrium of system.

To reduce the number of parameters mass fraction of PEG-6000 and ammonium sulfate were replaced with the tie-line length (*TLL*)[3,44]. If two-points of binodal curve, the top and the bottom phase composition, for a particular mixture composition are known distance between them is the tie-line length and may be calculated using equation 11,

$$TLL = \sqrt{\left(w_{AS}^{E*} - w_{AS}^{R*}\right)^2 + \left(w_{PEG-6000}^{R*} - w_{PEG-6000}^{E*}\right)^2} \tag{11}$$

where w^{E*} and w^{R*} are the equilibrium mass fractions (%) of ammonium sulfate (AS) and PEG-6000 in the extract (E) phase and in the rafinate (R) phase of aqueous two-phase system. The total mixture composition was always prepared at the critical point (or plait point) of binodal curve at which volumes of the extract and the rafinate phase theoretically become equal.

List of parameters used in the optimization of partitioning using GA, namely, tie-line length, pH and BSA concentration, investigated area of parameters and optimization steps are shown in Table 3. Weighting factor of target function *K* was set to 1 and it was written as linear combination of investigated parameters. Mutation occurrence in a program was set to 0.01, the crossover occurrence was set to 0.95 and the number of individuals in a generation was 4.

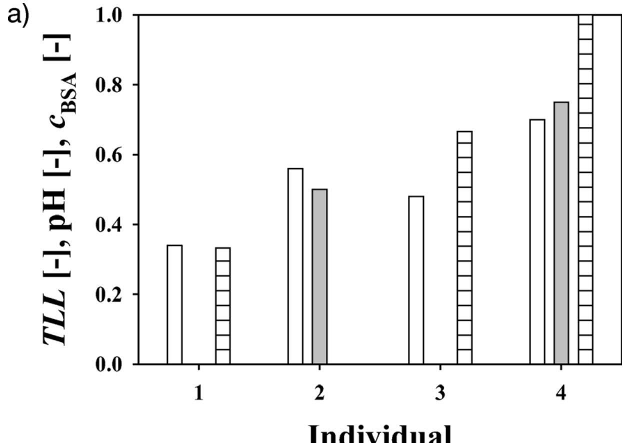
Table 3 – *Start values of investigated parameters, investigated area of parameters (defined by lower and upper border) and step of the optimization for experimental optimization*

Parameter	Unit	Step	Lower border	Upper border
c_{BSA}	g/dm^3	0.5	1.0	2.5
pH	-	0.5	5.5	7.5
TLL	cm	1	1	51

Furthermore, when the tie-line comes close to a plane point (where composition of the extract phase and the rafinate phase are equal) e. g. for small values of *TLL*, the density difference between the phases decreases and hence, time for phase separation increases. That was the reason of setting the distribution function for *TLL* in that way so as to prefer more frequent acceptance of bigger *TLL* value, e. g. to ensure faster phase separation.

Experimental optimization of process parameters was carried out through 5 generations. The principle of experimental optimization was as follows: GA offers the first random population of 4 individuals with given characteristics (tie-line length, pH and concentration of BSA). Experiments were performed under process conditions given by GA. After the equilibrium has been reached, equilibrium concentrations of BSA in the rafinate and in the extract phase were measured and partition coefficient calculated. These experimentally obtained values of partition coefficients were written in GA as return information, which GA used for further adjustments of next generation.

Normalized values of investigated parameters in the first and in the fifth generation are shown in the Figure 8. Normalized values of pH in the first and in the third individual of the first generation are equal to 0 which matches an absolute value of pH of 5.5. The same is for normalized value of the



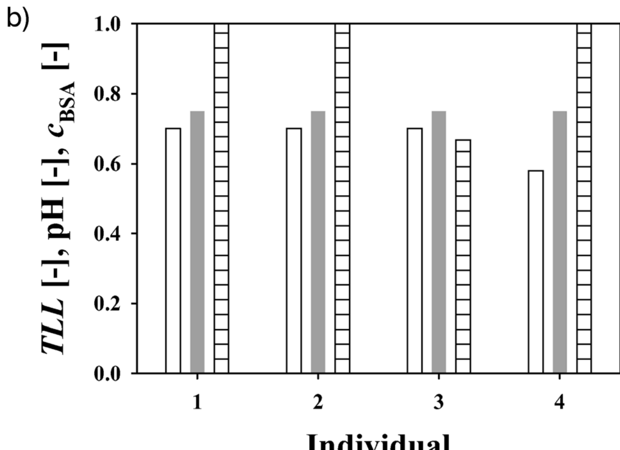


Fig. 8 – *a) Values of the optimized parameters for the first generation; b) Values of the optimized parameters for the fifth generation.* ▭ *normalized tie-line lenght,* ▬ *normalized pH,* ▤ *normalized concentration of BSA*

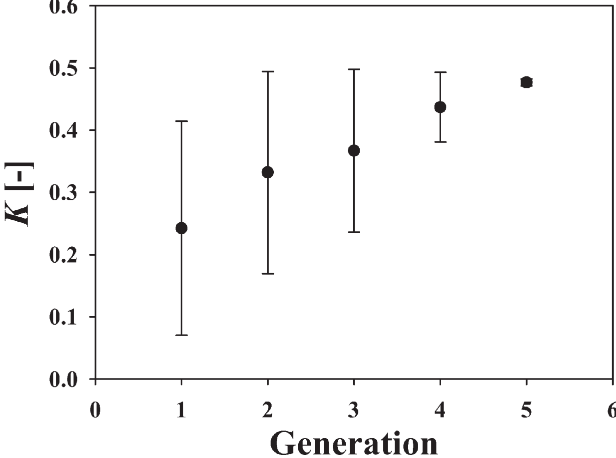


Fig. 9 – *Changes in the partition coefficient (●) through generations.* I *standard deviation of data*

BSA concentration in the second individual of the first generation (equal to 0 which matches an absolute value of c_{BSA} = 1.0 g/dm^3). Average value of K of a certain generation (average value of 4 individuals) with standard deviations is shown in Figure 9.

It can be noticed that uniformity of all parameters did not occur. Experiments 1, 2 and 4 in the fifth generation are the best performing of these four and show striking homology in process conditions. Namely, in the fifth generation BSA concentration differed in one individual. As it can be seen from Figure 9, this change practically did not affect the partition coefficient (small standard deviation). It means that change in BSA concentration has very small effect on K. From the first to the fourth generation a constant and a huge improvement of K is found. Only minor improvements were found going further on to generations four and five.

Data of Figure 8 show that pH and tie-line length have a strong positive influence on the maximum partition coefficient reached. Since a large improvement in the partitioning was not achieved between fourth and fifth generation and process conditions given by GA in fifth generation were quite homolog, the optimization was stopped after fifth generation.

Under optimal process conditions, namely, tie-line length of 36 cm (representing the mass fraction of PEG-6000 of 0.1770 and the mass fraction of ammonium sulfate of 0.1105), pH of 7.0 and BSA concentration of 2.5 g/dm^3, the maximal partition coefficient of K = 0.4793 was achieved. Furthermore, it should be stated that process optimum was achieved in five generations with four experiments, namely in twenty experiments, comparing it to 1530 experiments needed for full experimental plan.

Conclusions

GA is a simple program that makes optimization of highly complex systems possible, i.e. systems with high parameter number. Furthermore, GA allows relatively fast, extensive and effective optimization of process conditions. In this paper model-based and experimental optimization were shown for two different systems, model based optimization of enzyme synthesis of N-acetyl neuraminic acid, and experimental optimization of an aqueous two-phase extraction of BS Albumine respectively.

For model based optimization of enzyme synthesis of N-acetyl neuraminic acid the optimum is given by the value of volumetric productivity, Q_p = 47.8 g/dm^{3}d and R = 2.57 for c_{Pyr} = 1 mmol/dm^3, c_{GlcNAc} = 460 mmol/dm^3, E_{epi} = 39.9 mg/cm^3, E_{ald} = 0.2 mg/cm^3 and t = 156.8 min. At these optimal conditions conversion of N-acetylglucosamine to N-acetylneuraminic acid was 0.15 %, and conversion of pyruvate to N-acetylneuraminic acid was 71.99 %.

For experimental optimization of aqueous two-phase system containing PEG-6000 and ammonium sulfate, the maximal partition coefficient of $K = 0.4793$ was achieved at the tie-line length of 36 cm (representing the mass fraction of PEG-6000 of 0.1770 and the mass fraction of ammonium sulfate of 0.1105), pH of 7.0 and BSA concentration of 2.5 g/dm^3. Comparing the full experimental plan, which would take 1530 experiments, process optimum using GA was achieved only in twenty experiments.

ACKNOWLEDGEMENTS

The authors are greatly indebted to Prof. Dr. Udo Kragl, University of Rostock, Rostock, Germany, for his support during the stay of Mrs. Zvjezdana Findrik at his Department and for providing us with the GALOP software package. The authors would like to thank Mrs. Dubravka Maretić, graduate student at the Faculty of Chemical Engineering and Technology, University of Zagreb, for performing aqueous-two phase extraction experiments. This work was partially supported by Croatian Ministry of Science and Technology (Contract Grant Number 0125 021).

List of symbols

c – concentration of substrate, $mmol/dm^3$
E_{ald} – concentration of aldolase, mg/cm^3
E_{epi} – concentration of epimerase, mg/cm^3
K – partition coefficient, –
K_i^{ManNAc} – Inhibition constant for N-acetyl mannosamine (reaction catalyzed by aldolase), $mmol/dm^3$
K_i^{Pyr} – Inhibition constant for pyruvate (reaction catalyzed by aldolase), $mmol/dm^3$
K_m^{GlcNAc} – Michaelis-Menten constant for N-acetyl glucosamine (reaction catalyzed by aldolase), $mmol/dm^3$
K_m^{ManNAc} – Michaelis-Menten constant for N-acetyl mannosamine (reaction catalyzed by aldolase), $mmol/dm^3$
$K_m^{ManNAc*}$ – Michaelis-Menten constant for N-acetyl mannosamine (reaction catalyzed by epimerase), $mmol/dm^3$
K_m^{Neu5Ac} – Michaelis-Menten constant for N-acetyl neuraminic acid (reaction catalyzed by aldolase), $mmol/dm^3$
K_m^{Pyr} – Michaelis-Menten constant for pyruvate (reaction catalyzed by aldolase), $mmol/dm^3$
K_v – Inhibition constant that describes the influence of initial concentrations of substrates (reaction catalyzed by aldolase), $mmol/dm^3$
l – length of integer, –
M – molar weight, g/mol
Q_P – volumetric productivity, $g/(dm^3\ d)$
p – probability, –
R – ratio of the final concentration of N-acetyl neuraminic acid and pyruvate, –
t – time, s or min
TLL – tie-line length, cm
v_1 – reaction rate of the N-acetyl neuraminic acid synthesis, $mol/(dm^3\ min)$
v_2 – reaction rate of the reaction of epimerisation, $mol/(dm^3\ min)$
V_{max}^{ManNAc} – maximal specific activity of epimerase for the synthesis of N-acetyl mannosamine, μkat/g
V_{max}^{GlcNAc} – maximal specific activity of epimerase for the synthesis of N-acetyl glucosamine, μkat/g
V_{max}^{F} – maximal specific activity of aldolase for the N-acetyl neuraminic acid synthesis, μkat/g
V_{max}^{B} – maximal specific activity of aldolase for the N-acetyl neuraminic acid degradation, μkat/g
w – mass fraction, %
w^* – equilibrium mass fraction, %
x_1, x_2 – parameters, –

Abbreviations

ald – aldolase
AS – ammonium sulfate
BSA – bovine serum albumine
E – extract phase
epi – epimerase
GA – genetic algorithm
GlcNAc – N-acetyl glucosamine
ManNAc – N-acetyl mannosamine
Neu5Ac – N-acetyl neuraminic acid
PEG – polyethylene glycol
Pyr – pyruvic acid
R – rafinate phase

References

1. *Selber, K., Nellen, F., Steffen, B., Thömmes, J., Kula, M.-R.,* J. Chromatogr. B. **743** (2000) 21
2. *Zelić, B., Pavlović, N., Delić, V., Vasić-Rački, Đ.,* Chem. Biochem. Eng. Q. **16** (2002) 7
3. *Ghosh, S., Swaminathan T.,* Chem. Biochem. Eng. Q. **17** (2003) 319
4. *Bates, D. M., Watts, D. G.*, Nonlinear regression analysis and its applications, Wiley, New York, 1988
5. *Brass, J. M., Hoeks, F. W. J. M. M., Rohner, M.,* J. Biotechnol. **59** (1997) 63
6. *Weuster-Botz, D.*, J. Biosci. Bioeng. **90** (2000) 473
7. *Greasham, R. L., Inamine, E.*: Nutritional improvement of processes, In Demain, A. L. And Solomon N. A. (Eds.), pp. 41–48 Manual of industrial microbiology and biotechnology. American Society of Microbiology, Washington, D. C., 1986
8. *Silveira, R. G., Kakizono, T., Takemoto, S., Nishio, N., Nagai, S.*, J. Ferment. Bioeng. **72** (1991) 20
9. *Haltrich, D., Preiss, M, Steiner, W.*, Enzyme Microb. Tech., **15** (1993) 854

10. *Liu, C.-H., Hwang, C.-F., Liao, C.-C.*, Proc. Biochem., **34** (1999) 17
11. *Ranganath, M., Renganathan, S., Gokulnath, C.*, Bioproc. Eng. **21 (**1999) 123
12. *Weuster-Botz, D., Karutz, M., Joksch, B., Schaertges, D., Wandrey, C.*, Biotechnol Prog, **13** (1997) 387
13. *Milavec P., Podgornik A., Štravs R., Koloini T.*, Bioproc. Biosyst Eng, **25** (2002) 69
14. *Weuster-Botz, D., Wandrey, C.*, Process Biochem. **30** (1995) 563
15. *Sörensen, K., Janssens, G. K.*, Eur. J. Oper. Res. **151** (2003) 253
16. *Pendharkar, P. C, Rodger, J. A.*, Comput. Oper. Res. **31** (2004) 481
17. *Shafiei, S., Domenech, S., Koteles, R., Paris, J.*, J. Clean. Prod. **12** (2004) 131
18. *Dimou, C. K., Koumousis, V. K.*, Adv. Eng. Software **34** (2003) 773
19. *Komives, C., Parker, R. S.*, Curr. Opin. Biotech. **14** (2003) 468
20. *Silva, C. M., Biscala, E. C. Jr.*, Computers Chem. Eng. **27** (2003) 1329
21. *Chakraborti, N.*, *Mishra, P., Banerjee, A.*, Mater. Lett. **58** (2003) 136
22. *Pongcharoen, P.*, *Hicks, C., Braiden, P. M.*, Eur. J. Oper. Res. **152** (2002) 215.
23. *Hansen, J. V.*, Comput. Oper. Res. **31** (2004) 445.
24. *Marteijn, R. C. L., Jurrius, O., Dhont, J., de Gooijer, C. D., Tramper, J., Martens, D. E.*, Biotechnol. Bioeng. **81** (2003) 269
25. The MATLAB Genetic Algorithm Toolbox Version 1.2 User's Guide, 1994
26. GALOP, User Manual for Program Version 1.24, 1995
27. *Kragl, U.*, PhD Thesis, University of Bonn, Bonn, 1992.
28. SCIENTIST handbook, Micromath-, Salt Lake City, 1986–1995
29. *Bogdan, S.,* Master Thesis, University of Zagreb, Zagreb, 1997
30. *Bogdan, S., Gosak, D., Vasić-Rački, Đ.,* Computers Chem. Eng. **19** (1995) 791
31. *Fan, W., Bakir, U., Glatz, C. E.,* Biotechnol. Bioeng. **59** (1998) 461
32. *Rämsch, C., Kleinelanghorst, L. B., Knieps, E. A., Thömmes, J., Kula, M.-R.,* Biotechnol. Prog. **15** (1999) 493
33. *Montagna, J. M., Vecchietti, A. R., Iribarren, O. A., Pinto, J. M., Asenjo, J.A.,* Biotechnol. Prog. **16** (2000) 228
34. *Li, C., Bai, J., Li, W., Cai, Z., Ouyang, F.,* Biotechnol. Prog. **17** (2001) 366
35. *Tani, H., Kamidate, T., Watanabe, H.,* Anal. Sci. **14** (1998) 875
36. *Dekker, M., Van 't Riet, K., Bijsterbosch, B. H., Wolbert, R. B. G., Hilhorst, R.,* AIChE J. **35** (1989) 321
37. *Tjerneld, F., Johansson, G., Joelsson, M.,* Biotechnol. Bioeng. **30** (1987) 809
38. *Fitzpatrick, J. J., Engler, C.R.,* Bioprocess Eng. **13** (1995) 149
39. *Larrson, M., Arasaratnam, V., Mattiasson, B.,* Biotechnol. Bioeng. **33** (1989) 758
40. *Hamidi, A., van Berlo, M., Luyben, K. Ch. A. M., van der Wielen, L. A. M.,* J. Chem. Technol. Biotechnol. **74** (1999) 244
41. *Alberttson, P. A.,* Partition of cell particles and macromolecules, John Wiley and Sons, New York, 1986
42. *Walter, H., Brooks, D. E., Fisher, D.,* Partitioning in aqueous two-phase systems, Academic Press, New York, 1985
43. *Gündüz, U., Korkmaz, K.,* J. Chromatogr. B, **743** (2000) 255
44. *Mishima, K., Matsuyama, K., Ezawa, M., Taruta, Y., Takarabe, S., Nagatani, M.,* J. Cromatogr. B. 711 (1998) 313