SIMULATION AND PREDICTION OF PROTEIN PRODUCTION IN FED-BATCH *E. coli* CULTURES: AN ENGINEERING APPROACH

Daniel Calleja[1], John Kavanagh[2], Carles de Mas[1], Josep López-Santín[1*].

[1]Departament d'Enginyeria Química. Escola d'Enginyeria. Edifici Q.

Universitat Autònoma de Barcelona, 08193 Bellaterra (Catalunya), Spain.

[2]School of Chemical and Biomolecular Engineering. Chemical Engineering Building.

The University of Sydney. NSW, 2006 Australia.

* Corresponding author

phone: (+34) 93 581 1806

fax: (+34) 93 581 2013

e-mail: josep.lopez@uab.cat

Running title: Simulation of *E. coli* protein production

ABSTRACT

An overall model describing the dynamic behavior of fed-batch *E. coli* processes for protein production has been built, calibrated and validated. Using a macroscopic approach, the model consists of three interconnected blocks allowing simulation of biomass, inducer and protein concentration profiles with time. The model incorporates calculation of the extra and intracellular inducer concentration, as well as repressor –inducer dynamics leading to a successful prediction of the product concentration. The parameters of the model were estimated using experimental data of a rhamnulose-1-phosphate aldolase-producer strain, grown under a wide range of experimental conditions. After validation, the model has successfully predicted the behavior of different strains producing two different proteins: fructose-6-phosphate aldolase and ω-transaminase. In summary, the presented approach represents a powerful tool for *E. coli* production process simulation and control.

INTRODUCTION

Escherichia coli still remains as one of the preferred expression systems in industrial biotechnology for relatively simple proteins not requiring post-translational modifications. Efficient production processes have been developed using fed-batch operation to reach high cell concentration cultures and, for *lac*-operon derived expression systems, pulse induction of an inducer (typically isopropyl-β-D-thiolgalactopyranoside, IPTG) is a well established procedure (Tripathi et al., 2009; Ruiz et al., 2011 ; Huang et al., 2012).

From an engineering point of view, modeling of such processes is an essential tool to simulate, predict, optimize and, eventually, design model-based control systems. Several modeling approaches have been published at different levels: i) fed-batch operation models, describing biomass and substrate concentration evolution and using different optimization algorithms (Modak and Lim, 1987; Modak and Lim, 1992; Sarkar and Modak, 2003; Chen et al., 2004); ii) *lac* operon dynamics models, including IPTG transport (Vilar et al.,2004;

2

Santillán and Mackey, 2004; Noel et al., 2009); iii) non-segregated (Lee and Ramírez, 1992; Kavanagh and Barton, 2008; Ruiz et al., 2011) or segregated models (Zheng et al., 2005; Fan et al., 2007) describing the effect of protein production on microbial growth.

Nevertheless, until now, there is not a model of the whole process able to predict protein productivity as well as simulate the time behavior of all macroscopic process variables. The main limitations of the previously mentioned approaches come from the difficulty in quantifying two main aspects. First, the determination of intracellular IPTG concentrations has been mostly modeled without direct inducer measurements, and the transport mechanisms have been derived employing indirect measurements. Second, although the bistability behavior of the *lac* operon has been realized (Laurent et al., 2005) , the relationship between inducer concentration and protein expression rate is still contentious affecting proper protein expression prediction as well as quantification of the deleterious effect of induction and protein production on growth rates.

Most of the published works on *lac* operon are focused on the different elements (operator, repressor, *lac*Z, *lac*Y, *lac*A and inducers) separately or proposing theoretical models with little direct experimental data. In addition, few works are concerned with *lac* operon derived systems with plasmids potentially carrying additional copies of its constituents- i.e. *lac*I (repressor) gene or operator site sequences. On the other hand, the relationship between IPTG concentration (extracellular or intracellular) and protein production is quite complex (Nadri et al., 2006) and IPTG dosage is a determinant factor in the overexpression of recombinant proteins (Donovan et al., 1996; Durany et al., 2004), as intracellular IPTG concentration contributes to the probability of IPTG to bind the repressor molecules (Vilar et al., 2003).

With regards to inducer transport, our research group has recently developed a model to describe and simulate intracellular IPTG concentrations with time (Calleja et al., 2014), this

3

model was calibrated with previously published experimental IPTG measurements (Fernández et al., 2010).

On the other hand, it is worth to mention a model in literature (Ceroni et al., 2010) describing the transcription of the *lac*I gene, producing repressor molecules. It is an example of unsegregated and unstructured model: cellular population differences are not considered, and it is supposed that all the reactions, production and degradation of the different components take place in the same block. Ceroni's approach includes equations for the gene transcription, as well as expressions for the production of free repressor (not bound to IPTG) and repressor-IPTG molecules, as a function of the initial IPTG concentration added to the culture. This model allows the calculation of a time evolution of the repressor-IPTG molecules that are going to be used in the protein production model. Moreover, the production of a green fluorescence protein is described, making a relationship between the IPTG-repressor binding and the production of recombinant proteins. However, Ceroni's model employs the initial IPTG concentration as variable, which will likely limit the utility of the model, instead of using the intracellular IPTG concentration evolution along time.

The aim of this paper is to report an overall model able to describe the whole behavior of a fed-batch IPTG-induced protein production process by simulating the variation of the major macroscopic properties: biomass and substrate concentration, volume, medium and intracellular inducer concentration, and recombinant protein concentration (and activity in case of enzymes), with time. Such a model was built by coupling IPTG uptake model of Calleja to a biomass growth model (Ruiz et al., 2011) and to a newly proposed protein production model. The protein production model was built by modifying Ceroni et al's approach to include the calculated intracellular IPTG concentration.

The experimental system for calibration and validation of the overall model is the exponential fed-batch production of rhamnulose-1-phosphate aldolase (RhuA) by a K-12 derived *E. coli*

4

strain, under different induction and growth conditions. The validity of the proposed protein production model will be extended to simulate the behavior of *E. coli* BL21 strains (one of the most employed hosts for protein expression) producing ω-transaminases and fructose-6-phosphate aldolase (FSA).

MATERIALS AND METHODS

Strains

E. coli M15 Δ*glyA,* derived from K-12, harboring the plasmids [pQEαβrham] (Vidal et al., 2003) and [pREP4] was used for RhuA expression as a fusion protein to a 6xhistidine tag at its N-terminal end, under the control of the strong promoter T5 (Vidal et al., 2003; Vidal et al., 2008). *E. coli* BL21(DE3) was used for production of fructose-6-phosphate aldolase (FSA) (from now BL21 (DE3) FSA. The FSA encoding *mipB* gene was inserted into a modified expression vector pET22b(+) (Kreimeyer et al., 2007; Sánchez-Moreno et al., 2012). This strain was supplied by Prof. Dr. Pere Clapés from IQAC-CSIC, Barcelona. ω-transaminase (ATA) was produced using *E. coli* BL21 (DE3) strain (from now BL21 (DE3) ATA), harboring pLE1-A10-AcATA-D4 plasmid (supplied by c-LEcta, Leipzig) (Casablancas et al., 2013).

Culture media and fermentation conditions

Lysogeny Broth (LB), containing 10 g·L^{-1} of peptone, 5 g·L^{-1} of yeast extract and 10 g·L^{-1} of NaCl, was used for all pre-inoculum growth. It was sterilized by autoclaving (120°C, 30 min). When necessary, it was supplemented with antibiotic: 0.1 g·L^{-1} ampicillin for M15 and BL21 (DE3) FSA and 0.1 and 0.05 g·L^{-1} kanamycin for M15 and BL21 (DE3) ATA, respectively. Stock solution of kanamycin was prepared with a concentration of 100 g·L^{-1} and stored at -20°C. Ampicillin at 100 g·L^{-1} concentration in 50% (v/v) ethanol-water stock was prepared and stored at -20°C. Kanamycin and ampicillin stocks were sterilized by filtration (0.20 μm, Minisart® NY25. Sartorius Stedim).

A Defined Minimal medium (DM) using glucose as a sole carbon source (Durany et al., 2004) was employed for inoculum and fermentor batch growths. Table I shows the DM composition used for the different strains. Concentrations used for inoculum growth were one quarter of that shown in Table I for batch growths. All the components in Table I were prepared separately (600 $g \cdot L^{-1}$ glucose; 500 $g \cdot L^{-1}$ $MgSO_4 \cdot 7H_2O$; 5 $g \cdot L^{-1}$ $FeCl_3$; 100 $g \cdot L^{-1}$ $CaCl_2 \cdot H_2O$; 100 $g \cdot L^{-1}$ thiamine), save K_2HPO_4, KH_2PO_4, NaCl and $(NH_4)_2SO_4$, that were prepared as a unique solution (17.91 $g \cdot L^{-1}$ K_2HPO_4; 3.59 $g \cdot L^{-1}$ KH_2PO_4; 4.52 $g \cdot L^{-1}$ $(NH_4)_2SO_4$; 2.76 $g \cdot L^{-1}$ NaCl), called macroelements. Sterilization of glucose and macroelements were made by autoclaving (120°C, 30 min), while all the other components were sterilized by filtration (0.20 µm, Minisart® NY25. Sartorius Stedim). Trace elements shown in Table I were prepared as a unique solution, and sterilized by filtration (0.20 µm, Minisart® NY25. Sartorius Stedim). Antifoam (Antifoam 204, Sigma) was added to bioreactor whenever foam formation was observed. Antifoam was autoclaved (120°C, 30 min).

Feeding medium composition for fed-batch growth is also shown in Table I. In order to avoid co-precipitation with magnesium salts, 5 mL of phosphates solution (500 gL^{-1} K_2HPO_4 and 100 gL^{-1} KH_2PO_4) were added manually to the culture every 30 OD_{600} increment. The solution was autoclaved (120°C, 30 min) for its sterilization. IPTG stock was prepared at 100 mM, sterilized by filtration (0.20 µm, Minisart® NY25. Sartorius Stedim) and stored at -20°C.

From cryostock (-80 °C) in commercial Cryobilles (AES Chemunex), strains were grown in Falcon tubes containing 10 mL of LB medium and the supplementary antibiotic. Growths were performed at 37 °C and 150 rpm, reaching 1.8 to 2.1 units of optical density (OD_{600}) after around 16 hours of incubation. Then, 5 mL of pre-inoculum were transferred to shake flasks, containing 100 mL of DM and with same cultivation conditions as pre-inoculum cultures, until reaching 1.2 OD_{600}, that ensures the culture to be in the exponential growth phase.

6

For the batch phase, 80 mL of inoculum were added to 720 mL of DM. Bacterial growth was carried out in a Biostat® B bioreactor (at UAB) with 2L jar for *E. coli* M15 strain and in a New Brunswick BioFlo® fermenter (at the University of Sydney) with 3L vessel for *E. coli* DE3 strains. The bioreactors were equipped with stirring, temperature, dissolved oxygen and pH controllers. Temperature was maintained at 37 °C and pH was kept at 7.00±0.05 by adding 15% (v/v) NH_4OH solution. Dissolved oxygen value was maintained at 60% of saturation by adapting the stirring speed between 350 and 1120 rpm and supplying 1.5 $L·h^{-1}$ of air (enriched with pure oxygen when necessary, only at UAB). A reduction in oxygen consumption and an increase in pH can be used as identification for the batch phase ending. For the glucose limited feeding, a microburette at UAB, and a peristaltic pump at the U. of Sydney, was used. Feeding rate was performed according to a predefined exponential feeding profile based on mass balances and substrate uptake (Pinsach et al., 2006), keeping the specific growth rate at a fixed value, μ_{fix}. When culture reached the desired biomass concentration, a pulse of IPTG was added, in order to have the desired inducer concentration into the reactor.

Analytical methods

Biomass. Cellular concentration was determined by optical density OD_{600}, using a spectrophotometer (Uvicon 941 Plus, Kontrol at UAB and a Cary 50 at the University of Sydney). OD_{600} values were converted to biomass concentration expressed as Dry Cell Weigh (DCW), with 1 OD_{600} equivalent to 0.3 gDCW L^{-1} (Pinsach et al., 2008a).

Glucose. Biomass was removed from a 1.5 mL sample of broth culture by centrifugation (13400 rpm, 5 min) and filtration (0.45 µm). Glucose concentration in culture supernatant was determined enzymatically using a YSI 2700 (Yellow Spring Instruments)

Protein. Samples from bioreactor were diluted with milliQ water to an OD_{600} of 4. 1 mL of diluted sample was centrifugated (13400 rpm, 5 min) and the pellet resuspended with 1mL of lysis buffer (100 mM Tris-HCl; pH=7.5; OD_{600}=4). Cell suspension was kept in ice, and

sonicated in order to disrupt the cells. Sonication consisted in 4 cycles of 15 s pulses at 50 W (with 2 min between pulses), using a VC50 (Vibracell®, Sonics & Material) with microtip probe. Cellular debris was separated from supernatant by centrifugation (14000 rpm, 10 min, 4°C).

Total protein content in cells was determined by Bradford method, using Coomasie® Protein Assay Reagent Kit (Pierce, Thermo Fisher Scientific, Rockford, IL, USA)), in a microplate system (Microtiter Plate Flat, SUDELAB 900011). The percentage of RhuA amongst the rest of intracellular soluble proteins was determined by using NuPAGE® 12% Bis-Tris gels, following the manufacturer's manual (Invitrogen) and quantified by using Image Lab© (Bio-Rad Laboratories). RhuA, FSA and ATA activity assays were carried out as previously described (Vidal et al., 2003; Sánchez-Moreno et al., 2012; Casablancas et al., 2013).

Inducer. IPTG analysis was performed by MS-HPLC, using a Shimadzu Prominence liquid chromatograph SHPLC (Shimadzu), with an UV/Vis detector operating at 210 nm wavelength, coupled to a mass Spectrometer, equipped with ESI (Electro Spray Ionization) (Shimadzu) interface and a single quadrupole. Sample preparation, chromatographic and Mass Spectrometer conditions are described elsewhere (Fernández et al., 2010).

Modeling and parameter estimation

Model parameters' estimation and simulation were carried out using PSE gPROMS® ModelBuilder. MathWorks™Matlab®2012b was used to adjust experimental data by splines when needed.

RESULTS AND DISCUSSION

Model development

The proposed model is an unsteady, unsegregated, unstructured and based on first principles model composed by three blocks as depicted in the scheme shown in Figure 1. In the last block, specific protein production profile - expressed either in mass units (P) or activity units

8

(U) per gram of cells- is calculated as a function of the intracellular inducer concentration ($[IPTG]_i$) and the repressor-inducer dynamics. The inducer uptake model block calculates these dynamics from biomass concentration and total volume that are outputs of the growth model block. Differential equations of the model are presented in Table II and the values of the parameters in Table III. These parameter values have been sequentially determined in each block and kept constant in the next blocks.

Biomass Growth Model (Ruiz et al., 2011) For non-induced growth phase, the equations in this block need, as input, the initial batch biomass concentration (X_0), the initial volume (V_0), the initial substrate concentration (S_0), the fixed specific growth rate for the fed-batch (μ_{fix}) and the glucose concentration in the feeding solution (S_f). Equations (1), (4-a) and (5-a) correspond to volume (V), biomass (X) and substrate (S) balances and equation (6) to substrate-inhibited growth kinetics for specific growth rate (μ). The model includes equation (2-a) to determine the exponential fed-batch addition flow rate of feeding solution (F_S) and equation (3) to calculate the alkali addition flow rate (F_B) for pH control. Using these equations, a prediction of biomass, volume, and substrate dynamics is obtained for batch and fed-batch phases of the non-induced culture. Non-induced fed-batch phase starts when the initial glucose of the batch phase has been completely consumed.

Induced phase starts when the culture reaches the desired biomass concentration (X_{ind}) and IPTG pulse addition is done. For induced growth, equations (2-a) and (5-a) are substituted with (2-b) and (5-b) and equation (4-a) is replaced by (4-b) that needs (4-b.1) as complementary equation to take into account the decrease of specific growth rate due to the metabolic burden caused by overexpression of recombinant protein using a shock function, g_s (Lee & Ramirez, 1992). Initial conditions for induced growth are the values obtained in the non-induced growth phase when biomass reaches the desired concentration (X_{ind}).

9

IPTG Uptake Model. The IPTG uptake model (Calleja et al., 2014) is employed here with some improvements, which slightly modify the values of the parameters. It requires as inputs the biomass and IPTG concentration at induction time and uses the volume and biomass concentration predicted by the growth equations.

Equations (7) and (8) predict the time evolution of total volume of cells, V_{cel} (L_{cell}) and medium volume, V_m (L_m) , using a specific cell volume of 0.0023 $L_{cell} \cdot g^{-1}$DCW (Bennett et al., 2008). Equations (9) and (10) are the material balances for IPTG (e:extracellular; i:intracellular, μM) as a function of IPTG transport rate, r(μM·h^{-1}) and a dilution term. Finally, equation (11) represents the IPTG transport rate described by a diffusion term plus a non-specific active transport and a specific active transport.

Using equations (7) to (11) it is possible to calculate the intracellular IPTG concentration evolution along time, which is a key variable in protein production. Intracellular IPTG molecules can bind the repressor, making RNA polymerase able to transcribe the target gene and producing the target protein, RhuA in this case.

Protein Production Model Using a modification of Ceroni's model (Ceroni et al., 2010) it is possible to estimate the concentration of free and IPTG-bound repressor molecules with time and, from this information, the specific protein amount, either in mass (P) or activity (U) units, can be predicted.

Equation (12) calculates the number of LacI mRNA molecules (M, molecules·cell^{-1}) into the cell along time. For its calculation, the transcription rate (α_{mRNA}, h^{-1}), number of *lac*I gene copies -pREP4 plasmid copy number (N, plasmid·cell^{-1}) in this case- and the degradation rate of LacI mRNA molecules (λ_{mRNA}, h^{-1}) are needed. N has been assumed almost constant along growth according to experimental determinations (Pinsach et al., 2008b).

The amount of free repressor (R_{free}, molecule·cell^{-1}) and repressor bound to IPTG (R_{IPTG}, molecule·cell^{-1}) are calculated using equations (13) and (14), respectively. The original

equations used by Ceroni et al have been modified to introduce model-calculated dynamics of intracellular concentration of IPTG from equation (10). In these equations, the values found in literature (Ceroni et al., 2010) for the IPTG-LacI equilibrium binding constant ($K_{R\text{-}IPTG}$, mM), the protein degradation rate (λ_R, h^{-1}), the protein transcription rate (α_R, h^{-1}) and the cooperativity of the binding L*ac*I-IPTG (n) have been used (Table III), because they depend on the interaction between IPTG and repressor. Otherwise, the value for the time constant of LacI binding to the operator ($\tau_{R\text{-}O}$, h^{-1}) will be a model-fitting parameter because it depends on the operator used in the particular expression system and has to be fitted to experimental data. Equation (15) is a mass balance for the specific protein amount in mass units (P, mgRhuA·g^{-1}DCW), which is a function of the specific protein production rate (q_p, mgRhuA·g^{-1}DCW·h^{-1}) and a dilution term. The observed exponential decay for the specific production rate is described by equation (16) taking as initial condition q_{p0} that is calculated by equation (17) that takes into account the experimentally observed dependencies on μ_{fix} and $[IPTG]_{e,0}/X_{ind}$ (Ruiz et al., 2011).

The production of protein in activity units -equation (18)- is related to equation (15) through the protein quality (K_q, U·mg^{-1}RhuA). Experimental values of K_q ranged from 4.5 to 7.5 U·mg^{-1}RhuA and did not show a clear dependency on any other variable, instead they show a Gauss-like distribution around a value of 6. Because of that K_q was considered as a constant parameter with a value to be found in the model fitting step.

Model fitting

A set of 15 *E. coli* fed-batch experiments has been employed to get experimental data for the recombinant rhamnulose-1-phosphate aldolase production process, monitoring their performance by determining specific protein in mass and in activity along time. The target aldolase was intracellularly overexpressed, representing up to 30% of the total protein content. To give the model the capability to predict the dynamic behavior of a specific protein,

11

a wide range of experimental conditions must be used in the fitting process, expecting any experiment located into the defined space to follow the prediction of the model. The selected experimental variables were inducer and biomass concentration at induction time and the specific growth rate in the fed-batch phase of the culture because they are key in protein production (Ruiz et al., 2011). The inducer range was from 8 μM to 1 mM, 4 μM being the minimum concentration for RhuA overexpression (Pinsach et al., 2008a). Biomass concentration at induction time ranged from 8 to 47 gDCW·L^{-1}, being the later the maximum allowing a final biomass concentration compatible with the aeration capacity of the fermentor. Finally, specific growth rate ranged from 0.1 to 0.22 h^{-1}, avoiding too slow processes and excessive acetic acid production (Ruiz et al., 2009).

The numerical values for the parameters corresponding to the two first blocks of the model have been determined previously and are listed in Table III. The protein production block - equations (12) to (18)- has only four unknown parameters to be fitted (N is known). K_{P1} and K_{P2} are related to specific protein production rate in mass units, K_q is the protein quality, and τ_{R-O} is the time constant for the linkage between the repressor and the operator, which depends directly on the expression system, or, more specifically the operator-repressor bond and so is operator specific. The values for all other parameters in the protein block are also listed in Table III.

The mathematical model was fitted to the experimental data of aldolase concentration and activity evolution with time for the mentioned 15 fed-batch experiments. The values found for the four unknown parameters are those shown in Table III. The estimated value for specific activity per unit mass of aldolase, K_q, was 6.10 U·mg^{-1} in agreement with the experimental values.

12

The predictions of the model reproduced properly the experimental results. Figure 2 shows the parity plots for specific protein in mass and in activity units in order to give an overall idea of the goodness of the fitting.

Model validation

Once the model was calibrated, a validation of the whole model was conducted from the beginning of the batch to the end of induction phase. For that purpose, a new experiment -using different conditions at the beginning of induction phase than the ones of the fitting experiments- was performed. The experimental conditions were selected to be within the experimental space used and are depicted in Table IV, which also shows the initial values for the different input variables. In the induction phase, the initial values of M and R_{free} were calculated from equations (12) and (13) at equilibrium, when degradation rate equals formation rate and time derivative is zero. On the other hand, the initial values for P_{ind} and U_{ind} are the mean experimental values of the basal specific protein in mass or activity units, respectively.

Figures 3-A to 3-E show the model prediction and experimental data of the different process variables. The model is able to predict the experimental data well, confirming the usefulness of the developed simulation model.

Model extension to *E. coli* BL21 (DE3) strains producing ATA and FSA

The prediction capability of the built model has been tested by extending the protein production simulation to two different *E. coli* BL21 strains expressing ω-transaminases (ATA) and fructose-6-phosphate aldolase (FSA). In these cases, experimental biomass concentration and total volume data have been used as inputs of the IPTG uptake and protein production models.

E. coli BL21 (DE3) strain carries LacI repression system in its genome (Jeong H et al., 2009; Lebedeva M I et al., 1994; Xu J et al., 2012). Moreover, in the target strains, the inserted

13

plasmids also contain additional copies of the *lac*I gene. As both BL21 based expression systems use the same repressor as the M15 strain producing RhuA, the parameters describing the interaction between IPTG and repressor molecules ($K_{R\text{-}IPTG}$, n) and the ones describing LacI production and degradation (α_{mRNA}, λ_{mRNA}, α_R, λ_R) were be kept at their former values.

For each case, 5 parameters of the model had to be estimated because they are depending on the specific expression system or protein: i) N, the number of copies of *lac*I gene for both BL21 (DE3) strains; ii) $\tau_{R\text{-}O}$, the time constant for the linkage between the repressor and the operator; iii) K_{P1} and K_{P2}, the parameters related to specific protein production, and iv) K_q, the specific activity of the protein produced.

For both strains, three fed-batch production experiments at different operational conditions were conducted, detailed in Table V. It should be noted that all values of variables ($[IPTG]_{e,0}$, X_{ind} and μ_{fix}) are inside the experimental space employed for model calibration.

The model estimation of the different parameters is also shown in Table V. As can be seen, the obtained N value for both strains is much higher than was the case for the M15, meaning that a large number of copies of *lac*I gene is present into cells (it can be related to a large number of copies of the corresponding plasmid). This fact, taking into account that the number of copies of the gene present in the genome is the same for both BL21 (DE3) strains, means that the number of copies of plasmid in both strains is similar. The values of the remaining parameters are harder to explain, because they are depending on the particular expression system. Figures 4 and 5 present the results of the fitting using the value of the parameters presented in Table V. It can be seen that the model is again able to properly predict specific protein production both in mass and activity for the two *E. coli* BL21 (DE3) strains.

CONCLUSIONS

A macroscopic model for *E. coli* production of protein in fed-batch cultures has been developed. The model is built with three blocks of equations describing growth, IPTG and

14

repressor behavior with time. The parameters of the model have been fitted and validated for RhuA production in *E. coli* M15 Δ*glyA*.

The model has subsequently been tested for the production of FSA and ATA in *E. coli* BL21(DE3). In these cases only the five parameters that are strain dependent have been tuned. In all cases the agreement between experimental and predicted values was good and the mathematical model can be considered a powerful tool for simulation of such processes with potential exploitation in process control.

ACKNOWLEDGEMENTS

NOMENCLATURE

	units	description
F_B	$L \cdot h^{-1}$	alkali flow rate
F_s	$L \cdot h^{-1}$	feeding flow rate
g_s	--	function describing metabolic burden
$[IPTG]_m^e$	$\mu mol\ IPTG \cdot L^{-1}_{mediu}$	extracellular IPTG concentration
$[IPTG]$	$\mu mol\ IPTG \cdot L^{-1}_{mediu}$	extracellular IPTG concentration

15

Symbol	Units	Description
$[T_mG]_{e,0}$		at induction
$[IPTG]_i$	µmol IPTG·L$^{-1}_{cell}$	intracellular IPTG concentration
K_{R-IPTG}	mM	equilibrium constant for the repressor-IPTG bond
M	molecules·cell^{-1}	mRNA molecules of lacI
N	plasmid·cell^{-1}	plasmid copy number
n	--	cooperativity of the binding lacI-IPTG
P	mgRhuA·g^{-1}DCW	specific protein in mass units
P_{ind}	mgRhuA·g^{-1}DCW	basal specific protein in mass units
q_p	mgRhuA·g^{-1}DCW·h^{-1}	specific protein production rate
q	mgRh	initial

16

Symbol	Units	Description
p_0	$uA \cdot g^{-1}DCW \cdot h^{-1}$	specific protein production rate
r	$\mu mol\ IPTG \cdot h^{-1} \cdot L^{-1}_{medium}$	net transport rate
R_{free}	$molecule \cdot cell^{-1}$	free repressor
$R_{IPT G}$	$molecule \cdot cell^{-1}$	repressor bond to IPTG
S	$g \cdot L^{-1}$	glucose concentration
S_0	$g \cdot L^{-1}$	initial batch glucose concentration
S_f	$g \cdot L^{-1}$	glucose concentration in the feeding solution
t	h	time
t_{ind}	h	time since induction
U_d	$U \cdot g^{-1}DCW$	specific protein in activity units
U_{ind}	$U \cdot mg^{-1}RhuA$	basal specific protein in activity units
V	L	total volume
V_0	L	initial batch total volume
V	L_{cell}	volume of

17

Symbol	Units	Description
cell		cells
V_m	L_{medium}	culture medium volume
X	gDCW·L⁻¹	biomass concentration
X_0	gDCW·L⁻¹	initial batch biomass concentration
X_{ind}	gDCW·L⁻¹	biomass concentration at induction
Y_{XS}	gDCW·g⁻¹ glucose	biomass-substrate yield
$(Y_{XS})_{app}$	gDCW·g⁻¹ glucose	apparent biomass-substrate yield
α_{mRNA}	h⁻¹	rate of formation of M
α_R	h⁻¹	repressor protein synthesis rate
λ_{mRNA}	h⁻¹	mRNA degradation rate

18

Symbol	Units	Description
λ_R	h^{-1}	repressor protein degradation rate
μ	h^{-1}	specific growth rate
μ_{fix}	h^{-1}	fixed specific growth rate during the fed-batch
μ_{max}	h^{-1}	maximum specific growth rate for M15 strain
$\tau_{R\text{-}O}$	h^{-1}	time constant for the repressor-IPTG bond

REFERENCES

Bennett B D, Yuan J, Kimball E H, Rabinowitz J D (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nature Protocols, 3(8):1299–311.

Calleja D, Fernández-Castañé A, Pasini M, de Mas C, López-Santín J (2014). Quantitative modeling of inducer transport in fed-batch cultures of *Escherichia coli*. Biochemical Engineering Journal, 91:210–219

Casablancas A, Cárdenas-Fernández M, Álvaro G, Benaiges M D, Caminal G, De Mas C,González G, López C, López-Santín J (2013). New ammonia lyases and amine transaminases: Standardization of production process and preparation of immobilized biocatalysts. Electronic Journal of Biotechnology, 16(3). doi:10.2225/vol16-issue3-fulltext-4

Ceroni F., Furini S, Giordano E, Cavalcanti S (2010). Rational design of modular circuits for gene transcription: A test of the bottom-up approach. Journal of Biological Engineering, 4(1), 14. doi:10.1186/1754-1611-4-14

Chen L, Nguang SK, Chen XD, Li XM (2004). Modelling and optimization of fed-batch fermentation processes using dynamic neural networks and genetic algorithms. Biochemical Engineering Journal 22:51-61

Donovan R S, Robinson C W, Glick B R (1996). Optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter. Journal of Industrial Microbiology, 16(3):145–54.

20

Durany O, Caminal G, de Mas C, López-Santín J (2004). Studies on the expression of recombinant fuculose-1-phosphate aldolase in *E. coli*. Process Biochemistry, 39(11):1677–1684.

Fan J, Tuncay K, Ortoleva PJ (2007) Chromosome segregation in *Escherichia coli* division: a free energy-driven string model. Computational Biology and Chemistry 31:257-264.

Fernández A, Ruiz J, Caminal G, López-Santín J (2010) "Development and validation of a Liquid Chromatography-Mass Spectrometry assay for the quantitation of IPTG in *E. coli* fed-batch cultures". Analytical Chemistry, 82, 5728-5734

Huang C-J, Lin H, Yang X (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. Journal of Industrial Microbiology Biotechnology 39:383-399

Kavanagh JM, Barton GW (2008) Productivity improvement of recombinant *Escherichia coli* fermentation via robust optimization. Bioprocess and Biosystems Engineering 31:137-143

Jeong H, Barbe V, Lee C H, Vallenet D, Yu D S, Choi S-H, Kim J F (2009). Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). Journal of Molecular Biology, 394(4):644–52.

Kreimeyer A, Perret A, Lechaplais C, Vallenet D, Médigue C , Salanoubat M, Weissenbach J (2007). Identification of the last unknown genes in the fermentation pathway of lysine. The Journal of Biological Chemistry 282:7191-7197.

Lebedeva M I, Rogozhkina E V, Tsyba N A, Mashko S V (1994). A new T7 RNA polymerase-driven expression system induced via thermoamplification of a recombinant plasmid carrying a T7 promoter-*Escherichia coli lac* operator. Gene, 142(1):61–66.

Lee J, Ramirez W F (1992). Mathematical modeling of induced foreign protein production by recombinant bacteria. Biotechnology and Bioengineering, 39(6):635–46.

Modak JM, Lim HC (1987) Feedback optimization of fed-batch fermentation. Biotechnology and Bioengineering 30:528-540

Modak JM, Lim HC (1992). Optimal-mode of operation of bioreactor for fermentation processes.Chemical Engineering Science, 47: 3869-3884

Nadri M, Trezzani I, Hammouri H, Dhurjati P, Longin R, Lieto J (2006). Modeling and observer design for recombinant *Escherichia coli* strain. Bioprocess and Biosystems Engineering, 28(4):217–25.

Noel JT, Pilyugin SS, Narang A (2009) The diffusive influx and carrier eflux have a strong effect on the bistability of the *lac* operon in *Escherichia coli*. Journal of Theoretical Biology, 256:14-28

Pinsach J, de Mas C, López-Santín J (2006). A simple feedback control of *Escherichia coli* growth for recombinant aldolase production in fed-batch mode. Biochemical Engineering Journal, 29:235–242.

Pinsach J, de Mas C, López-Santín J (2008a). Induction strategies in fed-batch cultures for recombinant protein production in *Escherichia coli*: Application to rhamnulose 1-phosphate aldolase. Biochemical Engineering Journal, 41(2):181–187.

Pinsach J, de Mas C, López-Santín J (2008b). Influence of process temperature on recombinant enzyme activity in *Escherichia coli* fed-barch cultures. Enzyme and Microbial Technology, 43:507–512.

Ruiz J, González G, de Mas C, López-Santín J (2011). A semiempirical model to control the production of a recombinant aldolase in high cell density cultures of *Escherichia coli*. Biochemical Engineering Journal, 55(2):82–91.

Ruiz J, Pinsach J, Álvaro G, González G, de Mas C, Resina D, López-Santín J (2009). Alternative production process strategies in *E. coli* improving protein quality and downstream yields. Process Biochemistry, 44(9):1039–1045.

Sánchez-Moreno I, Nauton L, Théry V, Pinet A, Petit J-L, de Berardinis V, Lemaire M.(2012). FSAB: A new fructose-6-phosphate aldolase from *Escherichia coli*. Cloning, over-expression and comparative kinetic characterization with FSAA. Journal of Molecular Catalysis B: Enzymatic, 84:9–14.

Santillán M, Mackey D (2004). Influence of catabolite repression and inducer exclusion on the bistable behavior of the *lac* operon.Biophysical Journal 86:1282-1292.

Sarkar D, Modak JM (2003). Optimisation of fed-batch bioreactors using genetic algorithms. Chemical Engineering Science 58: 2283-2296.

Tripathi N K, Sathyaseelan K, Jana A M, Rao P V L (2009) High Yield Production of Heterologous Proteins with *Escherichia coli*. Def Sci J 59: 137-146

Vidal L, Durany O, Suau T, Ferrer P, Benaiges M, Caminal G (2003). High-level production of recombinant His-tagged rhamnulose 1-phosphate aldolase in *Escherichia coli*. Journal of Chemical Technology & Biotechnology, 78(11):1171–1179.

23

Vidal L, Ferrer P, Alvaro G, Benaiges M D, Caminal G (2005). Influence of induction and operation mode on recombinant rhamnulose 1-phosphate aldolase production by *Escherichia coli* using the T5 promoter. Journal of Biotechnology, 118(1):75–87.

Vidal L, Pinsach J, Striedner G, Caminal G, Ferrer P (2008). Development of an antibiotic-free plasmid selection system based on glycine auxotrophy for recombinant protein overproduction in *Escherichia coli* . Journal of Biotechnology, 134(1-2):127–36.

Vilar J M G, Guet C C, Leibler S (2003). Modeling network dynamics: the *lac* operon, a case study. The Journal of Cell Biology, 161(3):471–476.

Xu J, Banerjee A, Pan S-H, Li Z J (2012). Galactose can be an inducer for production of therapeutic proteins by auto-induction using *E. coli* BL21 strains. Protein Expression and Purification, 83(1):30–6.

Zheng Z, Yao S, Lin D (2005). Using a kinetic model that considers cell segregation to optimize hEGF expression in fed-batch cultures of recombinant *Escherichia coli*.Bioprocess Biosystems Engineering 27:143-152.