Bioresource Technology 101 (2010) 7537–7544

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech

Application of response surface methodology and artificial neural networks for optimization of recombinant *Oryza sativa* non-symbiotic hemoglobin 1 production by *Escherichia coli* in medium containing byproduct glycerol

Pablo C. Giordano [a,b], Hugo D. Martínez [b], Alberto A. Iglesias [b], Alejandro J. Beccaria [b,*], Héctor C. Goicoechea [a,**]

[a] *Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Universidad Nacional del Litoral, Ciudad Universitaria, CC 242 (S3000ZAA), Santa Fe, Argentina*
[b] *Instituto de Agrobiotecnología del Litoral (IAL), UNL-CONICET, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, CC 242 (S3000ZAA), Santa Fe, Argentina*

ARTICLE INFO

Article history:
Received 23 December 2009
Received in revised form 21 April 2010
Accepted 3 May 2010
Available online 23 May 2010

Keywords:
Artificial neural networks
Response surface methodology
Escherichia coli
Byproduct glycerol
Recombinant protein

ABSTRACT

Production of recombinant *Oryza sativa* non-symbiotic hemoglobin 1 (OsHb1) by *Escherichia coli* was maximized in shake-flask cultures in media containing tryptone, yeast extract, sodium chloride and byproduct glycerol from biodiesel production. Response surface methodology (RSM) and artificial neural networks (ANNs), followed by multiple response optimization through a desirability function were applied to evaluate the amount of OsHb1 produced. The results obtained by the application of ANNs were more reliable since better statistical parameters were obtained. The optimal conditions were (g L^{-1}), tryptone, 42.69; yeast extract, 20.11; sodium chloride, 17.77; and byproduct glycerol, 0.33. A maximum recombinant protein concentration of 3.50 g L^{-1} and a minimum biomass concentration of 18.48 g L^{-1} were obtained under these conditions. Although the concentrations of tryptone, yeast extract and sodium chloride are relatively high, the increase in the yield with respect to biomass formed ($Y_{P/X}$) overcomes this disadvantage.

1. Introduction

Escherichia coli is the most common prokaryotic expression system for production of recombinant proteins because physiology and genetics of this bacterium are well-characterized (Cheng et al., 2007; Ma et al., 2006; Tabandeh et al., 2008; Zhao et al., 2008). *E. coli* is frequently grown in LB medium which includes tryptone, yeast extract and sodium chloride (Bertani, 2004). Since this medium has a low carbon-to-nitrogen ratio, inexpensive components that can increase this ratio and the product yield are being sought (Ren et al., 2006). Corn steep liquor (Ren et al., 2006), eucalyptus hemicellulosic hydrolysate (Almeida e Silva et al., 2003) and byproduct glycerol from biodiesel production (Imandi et al., 2007) have been studied as carbon sources.

Byproduct glycerol is becoming increasingly available as a result of the expansion of the biodiesel (Murarka et al., 2008). Since for 10 kg of biodiesel produced, 1 kg of glycerol is obtained (Lamers et al., 2008), the price of glycerol has declined 90% between 2006 and 2009 (Yazdani and Gonzalez, 2007).

When formulating a culture medium, a key step is the optimization of the concentrations of its components since the component ratios directly influence the amount of biomass produced and the yield of the process (Nikerel et al., 2006). The traditional method to achieve optimization consists of varying one factor while keeping others constant. The main problems with this approach are the requirement to carry out a large number of experiments and the uncertainty of actually finding the optimal conditions, because the interactions between factors are not taken into account (Hao et al., 2006; Leardi, 2009; Montgomery, 1991). There are alternative techniques which are more efficient such as response surface methodology (RSM) and artificial neural networks (ANNs) (Havel et al., 1998; Marengo et al., 2004). RSM is a collection of statistical techniques for designing experiments, building models, evaluating the effect of factors and searching for the optimum conditions (Tabandeh et al., 2008). Combining RSM with the so-called Derringer's desirability function (Derringer and Suich, 1980), a very powerful tool is obtained, which can be applied to determine the optimum operative conditions for the studied system (Didier et al., 2007; Tabandeh et al., 2008). ANNs

* Corresponding author.
** Corresponding author. Address: Laboratorio de Desarrollo Analítico y Quimiometría, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Paraje "El Pozo", Ruta Nacional No. 168 S/N, S3000ZAA Santa Fe, Argentina. Tel.: +54 342 4575205; fax: +54 342 4575 221.
E-mail addresses: abeccari@fbcb.unl.edu.ar (A.J. Beccaria), hgoico@fbcb.unl.edu.ar (H.C. Goicoechea).

0960-8524/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biortech.2010.05.001

allow estimating relationships between one or more input or independent variables and one or more output or dependent variables (also called responses). In an ANN, the information is distributed among multiple nodes and connections between them, called weights (Despagne and Massart, 1998). RSM and ANNs have been applied in diverse areas like recombinant protein (Cheng et al., 2007) and bioinsecticides production (Moreira et al., 2007), and in pharmaceutics (Leonardi et al., 2009).

The aim of this work was to define the composition of a culture medium for recombinant *E. coli* through the application of RSM and ANNs. The selected recombinant protein is a non-symbiotic hemoglobin from *Oryza sativa* (rice), because it has a characteristic red color, whose intensity may be spectrophotometrically quantified and related to its concentration. The goal was the maximisation of the $Y_{P/X}$, maximisation of recombinant protein concentration (response 1) and minimisation of biomass concentration (response 2).

2. Methods

2.1. Bacterial strain and media

E. coli BL21 (DE3) harboring plasmid pET28/OsHb1 which allows the expression of *Oryza sativa* non-symbiotic hemoglobin 1 (OsHb1) (Arredondo-Peter et al., 1997) was grown on LB agar and stored at 4 °C. Tryptone, yeast extract, sodium chloride and byproduct glycerol were used as substrates, while lactose was used as an inducer for the production of OsHb1. Stock solutions of 150 g L^{-1} tryptone, 150 g L^{-1} yeast extract, 80 g L^{-1} sodium chloride and 10 g L^{-1} byproduct glycerol, respectively, were prepared in distilled water. Substrates concentrations were analyzed by a central composite design, and the 18 suggested combinations were achieved by combining different volumes of the corresponding stock solutions and adding distilled water to a final volume of 30 mL. The pH was adjusted to 7.00 with 1 M sodium hydroxide or 1 M chlorhydric acid. The media were sterilized at 121 °C for 15 min.

2.2. Inoculum preparation and culture conditions

A loopful of bacteria were inoculated into a 250 mL Erlenmeyer flask containing 30 mL of sterilized LB medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride) and 3.3×10^{-3} g L^{-1} of a kanamycin sulphate and grown for 1 day on a rotary shaker at 200 rpm and 30 ± 1 °C. Aliquots of this culture were pipetted into 250-mL Erlenmeyer flasks containing 30 mL of the different media as suggested by the central composite design to obtain an initial optical density at 600 nm (OD_{600}) of 0.1. The cultures were incubated at 200 rpm at 30 ± 1 °C. Once an OD_{600} of 1 was reached, synthesis of OsHb1 was induced by the addition of 0.3 mL of a 25% lactose solution (Ramchuran et al., 2005). Cultivation was continued for 24 h.

2.3. OsHb1 extraction protocol

Extractions were carried out with an empirically derived (protocol A) and an optimized protocol (protocol B). Protocol A consisted of suspending each gram of the biomass in 4 mL of extraction solution (disodium EDTA, 1×10^{-3} mol L^{-1}; Na_2HPO_4, 5×10^{-2} mol L^{-1}; NaH_2PO_4, 5×10^{-2} mol L^{-1}). Then, each biomass suspension was submitted to four cycles of ultrasound. Each cycle consisted of 80 s (1 s of resting time between 2 pulses of 1 s of ultrasound with an amplitude of 70%).

To arrive at an optimized protocol, a Plackett–Burman design was applied, and concentration of EDTA (E), Na_2HPO_4 (P1) and NaH_2PO_4 (P2); frequency (*F*) (seconds of resting time between two pulses of ultrasound), amplitude (*A*), grams of biomass/mL of extraction solution (XV), extraction time (*T*) and number of cycles (*C*) were analyzed (Table 1). The two levels required for each factor in the Plackett–Burman design were defined taking protocol A as a reference. The design consisted of 12 experiments. The analyzed responses were: yield, referred to biomass formed ($Y_{P/X}$) and the coefficient of variation (CV). $Y_{P/X}$ was calculated according to Eq. (1):

$$Y_{P/X} = \frac{\Delta \text{OsHb1}}{\Delta B} \quad (1)$$

where ΔOsHb1 and ΔB are the increments of OsHb1 and biomass concentrations, respectively, during the culture process. Only final concentrations were taken into account, as the initial concentrations were negligible.

The optimized protocol B consisted of suspending each gram of the biomass in 4 mL of extraction solution (disodium EDTA, 1×10^{-3} mol L^{-1}). Then, each biomass suspension was submitted to five cycles of ultrasound. Each cycle consisted of 90 s (2 s of resting time between 2 pulses of 1 s of ultrasound with an amplitude of 50%). To compare both protocols, the recombinant *E. coli* strain was grown in D1 medium (40.0 g L^{-1} of tryptone (TT), 23.5 g L^{-1} of yeast extract (YE), 16.0 g L^{-1} of sodium chloride (NaCl)) supplemented with 0.70 g L^{-1} of byproduct glycerol (BPG). The harvested biomass was divided into six samples, and OsHb1 was extracted from three samples, each following protocols A and B. Statistical tests were applied: an ANOVA test to identify the significant factors, an *F*-test to compare standard deviations of both protocols, and a *t*-test to compare the average yields of both protocols.

Table 1
Plackett–Burman design built to define the most important factors in the extraction protocol.

Experiment	Factors[a]								Responses[b]	
	E (mol $L^{-1} \times 10^3$)	P1 (mol $L^{-1} \times 10^3$)	P2 (mol $L^{-1} \times 10^3$)	*F* (s)	*A* (%)	XV (g mL^{-1})	*T* (s)	*C*	$Y_{P/X}$	CV (%)
1	1.00	0.00	50.00	1.00	50.00	0.125	50.00	5.00	0.092	4.59
2	0.00	50.00	0.00	1.00	70.00	0.250	30.00	5.00	0.138	8.34
3	1.00	0.00	0.00	2.00	70.00	0.125	30.00	5.00	0.104	6.74
4	1.00	50.00	50.00	1.00	70.00	0.125	30.00	3.00	0.094	1.34
5	1.00	50.00	50.00	2.00	50.00	0.250	30.00	5.00	0.186	7.84
6	0.00	50.00	0.00	2.00	50.00	0.125	50.00	5.00	0.071	2.88
7	1.00	50.00	0.00	1.00	50.00	0.250	50.00	3.00	0.190	9.92
8	0.00	0.00	0.00	1.00	50.00	0.125	30.00	3.00	0.084	18.34
9	0.00	0.00	50.00	2.00	50.00	0.250	30.00	3.00	0.164	24.47
10	0.00	50.00	50.00	2.00	70.00	0.125	50.00	3.00	0.073	14.55
11	1.00	0.00	0.00	2.00	70.00	0.250	50.00	3.00	0.176	8.38
12	0.00	0.00	50.00	1.00	70.00	0.250	50.00	5.00	0.153	9.50

[a] *E*: EDTA, P1: Na_2HPO_4, P2: NaH_2PO_4, *F*: frequency, *A*: amplitude, XV: grams of biomass/mL of extraction solution, *T*: extraction time, *C*: number of cycles.
[b] $Y_{P/X}$: yield referred to biomass formed and CV: coefficient of variation.

2.4. Optimum glycerol concentration

The range of BPG concentrations allowing maximum $Y_{P/X}$ was determined by growing the recombinant *E. coli* strain in D1 medium supplemented with (in g L^{-1}) BPG 0, 0.4, 0.73, 1, 1.5 and 2, respectively.

2.5. Analysis by response surface methodology and artificial neural networks

The optimal combination of the concentrations of TT, YE, NaCl and BPG was determined using RSM through a small central composite design (small CCD) with two center points (Hartley, 1959). This design consisted of 18 experiments with TT, YE, NaCl and BPG at concentration ranges of (in g L^{-1}) 10–50, 5–50, 1–20, and 0.25–2.5, respectively. The concentration of biomass (B) (in g L^{-1}) and concentration of OsHb1 (in g L^{-1}) were measured by gravimetrical and spectrophotometrical procedures, respectively.

The data were analysed by ANNs. The architecture of the ANNs consisted of three layers of nodes, an input layer consisting of four nodes, a hidden layer with a variable number of nodes, and an output layer with one node. Linear transfer functions were used in this work for input and output layers, while the sigmoid transfer function $f(x) = 1/(1 + e^{-x})$ was used for the hidden layer. For each response, different architectures were tested computing errors (RMSEC and RMSEP) for several ANNs obtained varying the number of nodes in the hidden layer. One ANN was selected for modelling each of the responses taking into account the lowest *R*oot *M*ean *S*quare *E*rror for *C*alibration (RMSEC) and *P*rediction (RMSEP). Responses predicted by RSM and ANNs were used to generate partial desirability functions (d_i), and the combination of the four components that predicted the highest value of global desirability function (D) was selected as the best blend of components to be present in the culture medium being developed.

2.6. Analytical

Cells were harvested by centrifugation (5000×*g*, 10 min) and their wet weight was determined. Response B was obtained by dividing this weight by the total medium volume. The cells were resuspended in extraction solution and lysed in a cool bath with a high intensity ultrasonic processor VibraCell™ VCX 130 (Sonics & Materials, Inc., USA). Cell debris was removed by centrifugation at 5000×*g* for 10 min and the supernatants were recovered and stored at −20 °C until OsHb1 determination was carried out.

Owing to the fact that there is no commercial availability for a standard of OsHb1, a reference sample was developed in our laboratory. Briefly, dilutions of a bovine serum albumine (BSA) standard and of supernatant containing the OsHb1 were subjected to electrophoresis under denaturing conditions (PAGE–SDS) (Laemmli, 1970) and staining, and band intensities were compared by densitometry. The concentration of OsHb1 in routine samples was determined spectrophotometrically in a 96 well plate reader at 410 nm. All measures were made in triplicate. Blanks and corresponding dilutions were made with extraction solution.

BPG concentration was monitored by adapting a commercial kit available for the determination of triglycerides (Wiener Lab, Argentina).

The data were transferred to a PC Intel Celeron D for further interpretation. Design Expert™ version 7.0.3 trial (Stat-Ease, Inc., Minneapolis, USA) was used for performing experimental design and data analysis.

3. Results and discussion

3.1. OsHb1 extraction protocol

Results of the application of an ANOVA test to the data presented in Table 1 are tabulated in Table 2. Factors E and XV were significant for response $Y_{P/X}$, while E and C were significant for response CV. The results also show that there was no need for the extraction solution to contain phosphate buffer pH 7.00. To maximize the response $Y_{P/X}$ and minimize CV, E was set at 1 M, XV at 0.25 g mL^{-1}, *C* at 5 cycles, F at 2 s, A at 50% and *T* at 30 s in order to avoid overburdening the equipment. P1 and P2 were excluded from the extraction solution, which only contained 1 M EDTA.

The results obtained with protocols A and B were statistically different when a *F*-test was applied, not different when a *t*-test was used (Table 3). Since protocol B allowed a decrease of 54% in CV (6.2% for protocol B against 13.7% for protocol A), protocol B was selected for extraction of OsHb1.

3.2. Optimum byproduct glycerol concentration

When the concentration of BPG increased up to 0.4 g L^{-1}, $Y_{P/X}$ also increased up to a value of 0.23, at which it stabilized until it

Table 3
Statistical parameters of an *F*-test and a *t*-test applied to compare standard deviations and average yields referred to biomass formed ($Y_{P/X}$) of both protocols, respectively.

Test	Statistical parameter[a]	Protocol	
		A	B
F-test	s	0.018	0.008
	F_c	4.678	
	$F_{t(nA,nB);0.05}$	3.438	
	p-value	0.043 (s_A and s_B are statistically different)[b]	
t-test	$Y_{P/Xaver}$	0.130	0.136
	t_c	0.833	
	$t_{t(11);\ 0.05}$	1.796	
	p-value	0.422 ($Y_{P/XaverA}$ and $Y_{P/XaverB}$ are statistically not different)[c]	

[a] s: standard deviation; F_c: calculated F; F_t: table F; $n_A = n_B = 9$; $Y_{P/Xaver}$: average yield referred to biomass formed ($n = 9$); t_c: calculated t; t_t: table t.
[b] s_A: standard deviation for protocol A; s_B: standard deviation for protocol B.
[c] $Y_{P/XaverA}$: average yield referred to biomass formed for protocol A; $Y_{P/XaverB}$: average yield referred to biomass formed for protocol B.

Table 2
Factors *p*-values for each response obtained by the application of an ANOVA test to the experimental data presented in Table 1.

Response[b]	*p*-Values[a]								
	Model	*E*	P1	P2	*F*	*A*	XV	*T*	*C*
$Y_{P/X}$:	<0.0001	**0.0006**	0.4906	0.9792	0.4298	0.1196	**<0.0001**	0.6505	0.2071
CV	0.0394	**0.0522**	0.1275	0.6604	0.4183	0.2355	0.2323	0.2709	**0.0637**

[a] *E*: EDTA, P1: Na_2HPO_4, P2: NaH_2PO_4, *F*: frequency, *A*: amplitude, XV: grams of biomass/mL of extraction solution, *T*: extraction time and *C*: number of cycles.
[b] $Y_{P/X}$: yield referred to biomass formed and CV: coefficient of variation.

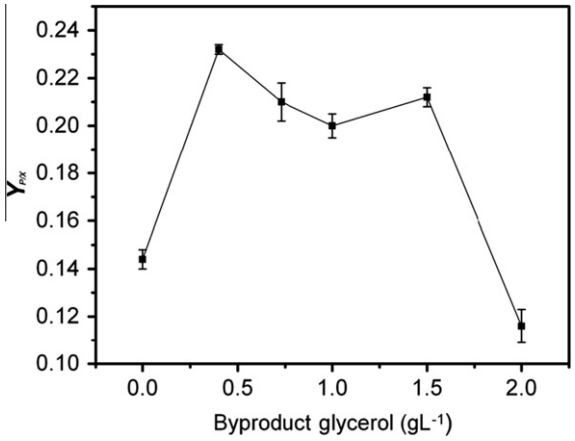

Fig. 1. Yield, referred to biomass formed ($Y_{P/X}$) as a function of the concentration of byproduct glycerol (g L^{-1}).

decreased to 0.11 at 2 g L^{-1} of BPG (Fig. 1). Therefore the optimum concentration of BPG (at which $Y_{P/X}$ reaches its maximum value) is in the range between 0 and 2 g L^{-1} BPG.

3.3. Analysis by response surface methodology

Table 4 shows all the combinations suggested by the small CCD and its respective responses values. Experiment 10 (for response B) and experiment 14 (for response OsHb1) were excluded for further analysis because they were detected as outliers, by applying a difference between fitted values test (DFFITS) (Myers and Montgomery, 1995).

An ANOVA test applied to the data obtained from the design (Table 5) demonstrated that a linear model could fit the response B with YE and BPG as significative factors and that the response OsHb1 could be fitted with a quadratic model with TT (quadratic term), YE (linear and quadratic terms) and NaCl (linear term) as significant factors. These models can be mathematically expressed according to Eqs. (2) and (3):

$$Y_1 = 17.1366 + 0.1988\,X_2 + 1.3379\,X_4 \quad (2)$$

$$Y_2 = -5.0249 + 0.2999X_1 + 0.2657X_2 - 0.0359X_3 - 0.005X_1^2 - 0.004X_2^2 \quad (3)$$

Table 4
Small CCD used for the culture media optimization.

Experiment	Factors[a] (g L^{-1})				Responses[b] (g L^{-1})	
	TT	YE	NaCl	BPG	B	OsHb1
1	41.89	40.88	4.85	0.71	24.21	0.98
2	41.89	40.88	16.15	0.71	27.32	1.16
3	18.11	40.88	16.15	2.04	27.44	0.86
4	30.00	27.50	10.50	1.38	24.85	2.63
5	41.89	14.12	4.85	2.04	22.49	1.17
6	18.11	40.88	4.85	2.04	27.71	2.09
7	18.11	14.12	16.15	0.71	19.90	1.03
8	18.11	14.12	4.85	0.71	18.93	0.64
9	41.89	14.12	16.15	2.04	20.04	1.12
10	30.00	5.00	10.50	1.38	15.96	0.37
11	30.00	50.00	10.50	1.38	27.32	1.39
12	10.00	27.50	10.50	1.38	26.19	1.24
13	50.00	27.50	10.50	1.38	24.01	1.09
14	30.00	27.50	10.50	2.50	27.83	1.06
15	30.00	27.50	1.00	1.38	29.08	4.03
16	30.00	27.50	10.50	0.25	24.94	3.29
17	30.00	27.50	20.00	1.38	22.68	2.80
18	30.00	27.50	10.50	1.38	24.32	2.65

[a] TT: tryptone. YE: yeast extract. NaCl: sodium chloride. BPG: byproduct glycerol.
[b] B: biomass. OsHb1: *Oryza sativa* non-symbiotic hemoglobin 1.

Table 5
Statistical results obtained for the model test.

Statistical parameter	B	OsHb1
Fitted model	Linear	Quadratic
F-Model	12.31	14.16
p-Model[a]	0.001 (significant)	0.0003 (significant)
R^2	0.654	0.876

where Y_1 and Y_2 are responses B and OsHb1, respectively, and X_1, X_2, X_3 and X_4 are TT, YE, NaCl and BPG, respectively. Only the factors that are significant for each response have been included in the above equations.

As examples, Fig. 2a and b represent the response surfaces corresponding to responses B (as a function of BPG and YE) and OsHb1 (as a function of TT and YE, setting NaCl at 1.00 g L^{-1}), respectively. In addition, it can be seen the quadratic contributions of YE and TT.

The maximum concentration of OsHb1 obtained after modelling was 4.03 g L^{-1} (experiment 15, Table 4, with a $Y_{P/X}$ of 0.138), which is 5.5 times greater than the concentration obtained for an LB

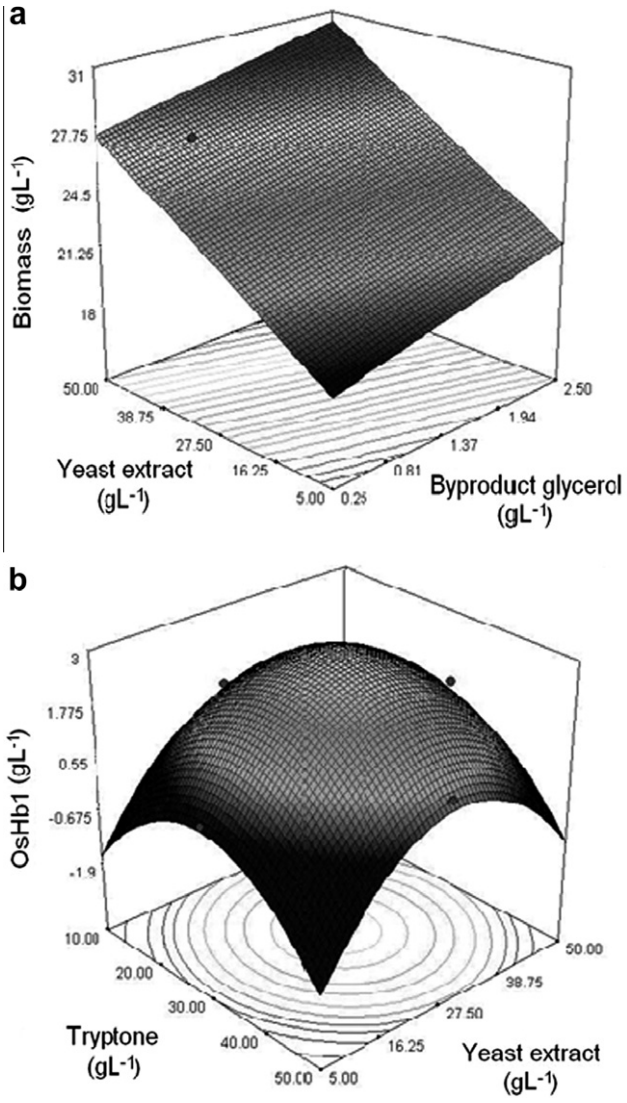

Fig. 2. (a) Concentration of biomass as a function of YE (g L^{-1}) and BPG (g L^{-1}), obtained by the application of RSM techniques. (b) Concentration of OsHb1 as a function of the concentrations of tryptone (g L^{-1}) and yeast extract (g L^{-1}), setting NaCl at 1.00 g L^{-1}, obtained by the application of RSM techniques.

Table 6
Criteria for the optimization of multiple responses.

Factors[a] and responses[b]	Optimization criteria	Lower limit	Upper limit
TT (g L^{-1})	In range	10.00	50.00
YE (g L^{-1})	In range	5.00	50.00
NaCl (g L^{-1})	In range	1.00	20.00
BPG (g L^{-1})	In range	0.25	2.50
B (g L^{-1})	Minimize	18.93	29.08
OsHb1 (g L^{-1})	Maximize	0.37	4.03

[a] TT: tryptone. YE: yeast extract. NaCl: sodium chloride. BPG: byproduct glycerol.
[b] B: biomass. OsHb1: *Oryza sativa* non-symbiotic hemoglobin 1.

medium (10.00 g L^{-1} TT, 5.00 g L^{-1} YE, 10.00 g L^{-1} NaCl, yielding 0.73 g L^{-1} OsHb1 with a $Y_{P/X}$ of 0.057). This justifies the application of RSM in order to improve yields of recombinant proteins, optimizing the formulation of culture media.

Based on these results the media was optimized by the application of the global desirability function (D). Table 6 shows the criteria employed to perform this optimization. The combination predicted by the application of D according to the optimization criteria was, 29.71 g L^{-1} TT, 23.90 g L^{-1} YE, 1.00 g L^{-1} NaCl and 0.25 g L^{-1} BPG (R1 medium). The values predicted for the responses were, 22.22 g L^{-1} B and 3.15 g L^{-1} OsHb1. Fig. 3 represents the response surface for the D, as a function of YE and BPG setting NaCl at 1 g L^{-1} and TT at 29.71 g L^{-1}. Its maximum corresponded to 0.721.

3.4. Analysis by artificial neural networks

Table 7 summarizes the characteristics and parameters of the selected ANNs: number of nodes in the three layers, RMSEC, RMSEP, parameters/data ratio (the number of parameters should not be higher than the number of data), number of epochs at which the lowest errors were reached, learning rate (η) and momentum coefficient (μ). Parameters μ and η were set at 0.5 because at lower values, the learning process would be very slow, and at higher values, oscillations may occur causing the non-convergence of the ANN (del Brío and Sanz Molina, 2002).

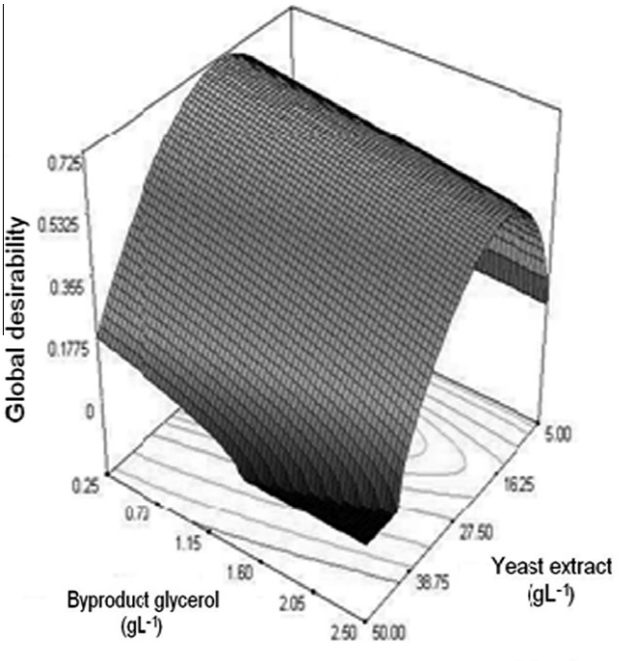

Fig. 3. Global desirability as a function of YE (g L^{-1}) and BPG (g L^{-1}), setting NaCl at 1.00 g L^{-1} and TT at 29.71 g L^{-1}. Its maximum corresponds to 0.721.

Table 7
Architecture and statistical parameters of the selected ANNs.

	ANNB[b]	ANNHb[b]
Net architecture	4/3/1	4/3/1
Parameters/Data	19/56	19/56
Epochs	4332	5241
Learning rate	0.5	0.5
Momentum coefficient	0.5	0.5
RMSEC[a]	0.059	0.025
RMSEP[a]	0.027	0.042
R^2	0.922	0.970

[a] RMSE $= \left[\frac{1}{I}\sum_1^I (c_{act} - c_{pred})^2\right]^{1/2}$ where c_{act} and c_{pred} are the actual and the predicted concentrations in either the training or the monitoring steps, and I is the number of samples used in each procedure.
[b] ANNB: ANN trained with biomass data; ANNHb: ANN trained with OsHb1 data.

The values predicted by the ANNs vs. the actual ones were employed to calculate the determination coefficients (R^2) for both responses. Fig. 4a–d show the correlation between actual and predicted values for both responses, modeled applying RSM and ANNs, respectively; and it can be seen that ANNs improved R^2 values in both cases as compared to RSM. Therefore, ANNs were capable of fitting a reliable model with only 17 experimental points.

Selected ANNs were used for predicting the responses of 600 different simulated combinations of the four factors, which completely covered the experimental space. The $2 \times 600 = 1200$ predicted values obtained were then used to calculate their d_i and D values (Derringer and Suich, 1980). The D was calculated assigning a value of 1 to the parameter r_i, as it was done for RSM. The optimal value found was $D = 0.913$, which corresponds to the combination: TT 42.69 g L^{-1}, YE 20.11 g L^{-1}, NaCl 17.77 g L^{-1} and BPG 0.33 g L^{-1} (A1 medium). The responses values that correspond to this combination were: 19.24 g L^{-1} B and 3.59 g L^{-1} OsHb1. Again, the ANN approach was capable of improving, in this case, the value of D, which was near to 1, indicating that factors and responses have all desirable values, simultaneously. The results given by the RSM implementation are in general agreement with these latter values but the results obtained by ANNs are more reliable due to the better statistical parameters that are more complex as can be seen in Fig. 5a and b, which represent the response surfaces corresponding to responses B (as a function of BPG and YE) and OsHb1 (as a function of TT and YE, setting NaCl at 20.00 g L^{-1}), respectively, obtained by the application of ANNs.

The media predicted by RSM and ANNs have an important content of TT and YE. This observation can be explained taking into account the characteristics of the OsHb1. This protein has heme as prostetic group, and the microorganism must synthesize not only heme for its own proteins but also for the recombinant protein, which increases the requirements for iron and nitrogen. These needs may be supplied by increasing the concentrations of TT (source of nitrogen) and YE (source of various microelements including iron). Since glycerol, rather than peptides, is used as a carbon and energy source, pH changes due to the production of ammonia are minimized (Murarka et al., 2008).

3.5. Experimental verification

The optimal combinations of components predicted by means of RSM and ANNs were verified by additional independent experiments. It was seen that BPG was consumed at the moment of harvesting the biomass (data not shown). This proves that BPG was metabolized by *E. coli*, but does not mean that the culture was carbon limited at that time, since TT and YE include carbon sources in their compositions. Beyond this fact, one possibility for a fed-batch or continuous culture strategy could be the BPG feeding, which should be started when the initial amount of BPG is almost totally

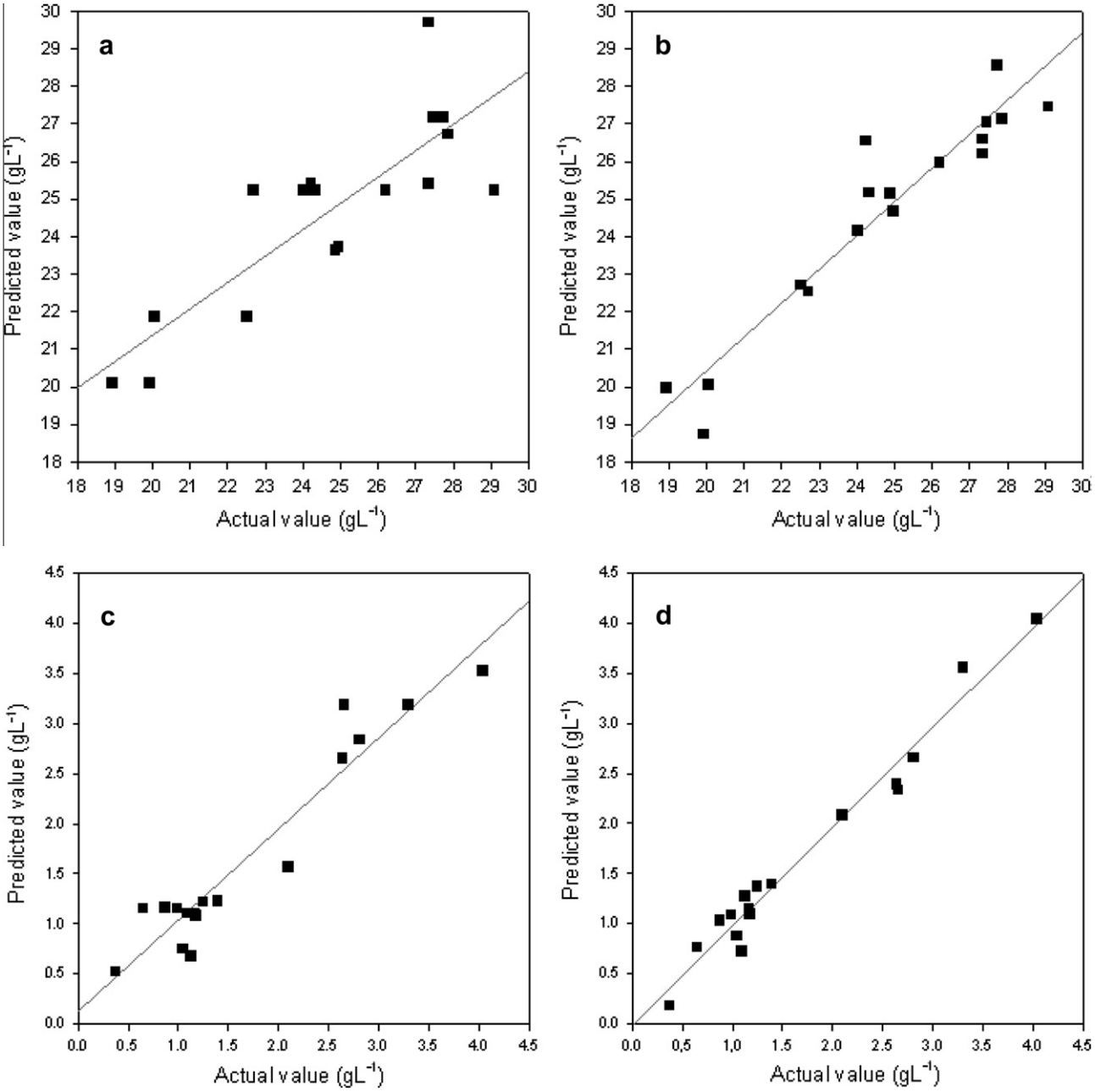

Fig. 4. Correlation between actual and predicted values for response B, fitted applying RSM (a) and ANNs (b); and for response OsHb1, fitted applying RSM (c) and ANNs (d).

depleted. Predicted and empirical responses values are presented in Table 8. The empirical data only showed close agreement with what was predicted by ANNs, thus indicating the high reliability of the fitted models obtained by means of this technique. The following combination, 42.69 g L^{-1} TT, 20.11 g L^{-1} YE, 17.77 g L^{-1} NaCl and 0.33 g L^{-1} BPG, is expected to render optimal yields, maximize the concentration of recombinant protein, and minimize the concentration of biomass.

In a further step, the cost of TT, YE and NaCl were obtained from local suppliers in order to estimate the cost of LB and A1 media, according to Eq. (4):

$$A = C_{TT} * 0.421 + C_{YE} * 0.509 + C_{NaCl} * 0.156 \quad (4)$$

where A is the cost (in U$S/L), and C_{TT}, C_{YE} and C_{NaCl} are the concentrations of TT, YE and NaCl (in g L^{-1}), respectively, in each media. Then, the costs of LB and A1 media are, in U$S/L, 8.3 and 30.9.

Whereas the cost of A1 medium increased 3.7-fold with respect to LB medium, the concentration of OsHb1 increased 4.9-fold. In other words, producing 1 g of recombinant protein in LB medium requires U$S 43.3, while it requires U$S 32.8 in A1 medium. Moreover, the $Y_{P/X}$ was 0.057 and 0.180 for LB and A1 media, respectively. This 3.2-fold increase in yield demonstrates an increase in the volumetric productivities of the bioreactor and should result in effluents reductions (Matsui et al., 2008).

4. Conclusions

ANNs provided more reliable results than RSM in optimization of the yield of recombinant *Oryza sativa* non-symbiotic hemoglobin 1 in medium containing byproduct glycerol. Additional studies should be carried on to evaluate the predicted formulation with fed-batch cultures, which are extensively used in industrial bioprocesses and to include the parameters involved in the feeding strategy.

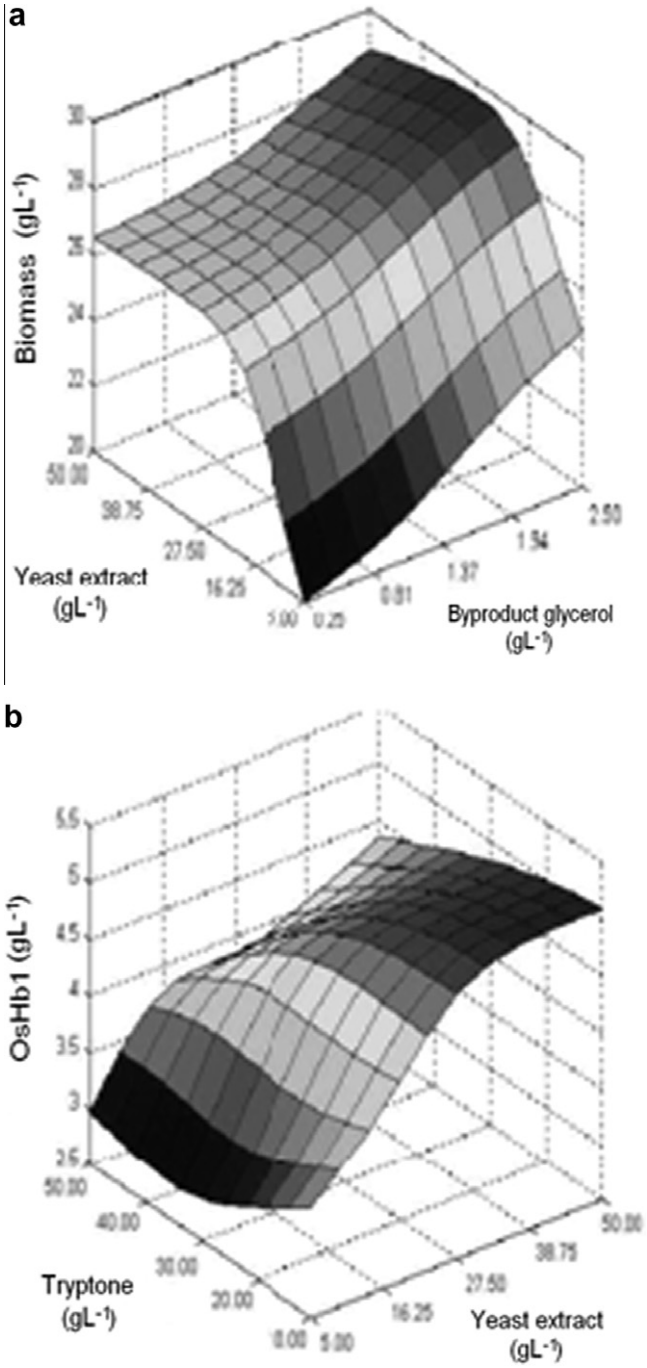

Fig. 5. (a) Concentration of biomass as a function of YE (g L^{-1}) and BPG (g L^{-1}), obtained by the application of ANNs. (b) Concentration of OsHb1 as a function of the concentrations of tryptone (g L^{-1}) and yeast extract (g L^{-1}), setting NaCl at 20.00 g L^{-1}, obtained by the application of ANNs.

Table 8
values of responses obtained by application of RSM and ANNs, and by experimental verification.

		Response[a]		
		B (g L^{-1})	OsHb1 (g L^{-1})	$Y_{P/X}$
RSM	Prediction	22.22	3.15	0.14
	Experimental verification	21.56	2.44	0.11
ANNs	Prediction	19.24	3.59	0.18
	Experimental verification	18.48	3.50	0.19

[a] B: biomass. OsHb1: *Oryza sativa* non-symbiotic hemoglobin 1. $Y_{P/X}$: average yield referred to biomass formed.

Acknowledgements

The authors wish to thank Dr. M.S. Hargrove, Dept. Biochemistry, Biophysics & Molecular Biology, Iowa State University (Ames, IA 50011, USA) for the generous gift of the plasmid expressing rice hemoglobin. This work was supported by grants from CONICET (PIP 112-200801-02988), ANPCyT (PICTO 559 05 N° 35124) and Universidad Nacional del Litoral (CAI+D 2008-O, CAI+D 2009-R2). PCG is a fellow from CONICET, and AAI and HCG are member of the research career from the same Institution.

References

Almeida e Silva, J.B., Lima, U.A., Taqueda, M.E.S., Gauragna, F.G., 2003. Use of response surface methodology for selection of nutrient levels for culturing Paecilomyces variotii in eucalyptus hemicellulosic hydrolysate. Bioresour. Technol. 87, 45–50.

Arredondo-Peter, R., Hargrove, M.S., Sarath, C., Moran, J.F., Lohrman, J., Olson, J.S., Klucas, R., 1997. Rice hemoglobins. Gene cloning, analysis, and O_2-binding kinetics of a recombinant protein synthesized in *Escherichia coli*. Plant Physiol. 115, 1259–1266.

Bertani, G., 2004. Lysogeny at mid-twentieth century: P1, P2 and other experimental systems. J. Bacteriol. 186, 595–600.

Cheng, S., Song, Q., Wei, D., Gao, B., 2007. High-level production penicillin G acylase from *Alcaligenes faecalis* in recombinant *Escherichia coli* with optimization of carbon sources. Enzyme Microbiol. Technol. 41, 326–330.

del Brío, B.M., Sanz Molina, A., 2002. Redes neuronales y sistemas difusos. 2da edicion. Alfaomega Grupo Editor.

Derringer, G., Suich, R., 1980. Simultaneous optimization of several response variables. J. Qual. Technol. 12, 214–219.

Despagne, F., Massart, D.L., 1998. Neural networks in multivariate calibration. Analyst 123, 157R–178R.

Didier, C., Etcheverrigaray, M., Kratje, R., Goicoechea, H.C., 2007. Crossed mixture design and multiple response analysis for developing complex culture media used in recombinant protein production. Chemomet. Intell. Lab. Syst. 86, 1–9.

Hao, D.C., Zhu, P.H., Yang, S.L., Yang, L., 2006. Optimization of recombinant cytochrome P450 2C9 protein production in *Escherichia coli* DH5α by statistically-based experimental design. World J. Microbiol. Biotechnol. 22, 1169–1176.

Hartley, H.O., 1959. Smallest composite designs for quadratic response surfaces. Biometrics 15, 611–624.

Havel, J., Peña, E.M., Rojas-Hernández, A., Doucet, J.-P., Panaye, A., 1998. Neural networks for optimization of high-performance capillary zone electrophoresis methods. A new method using a combination of experimental design and artificial neural networks. J. Chromatogr., A 793, 317–329.

Imandi, S.B., Bandaru, V.V.R., Somalanka, S.R., Garapati, H.R., 2007. Optimization of medium constituents for the production of citric acid from byproduct glycerol using Doehlert experimental design. Enzyme Microbiol. Technol. 40, 1367–1372.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Lamers, P., McCormick, K., Hilbert, J.A., 2008. The emerging liquid biofuel market in Argentina: implications for domestic demand and international trade. Energy Policy 36, 1479–1490.

Leardi, R., 2009. Experimental design in chemistry: a tutorial. Anal. Chim. Acta 652, 161–172.

Leonardi, D., Lamas, M.C., Salomón, C.J., Olivieri, A., 2009. Development of novel formulations for Chagas' disease. Optimization of benznidazol chitosan microparticles based on artificial neural networks. Int. J. Pharm. 367, 140–147.

Ma, X., Zheng, W., Wang, T., Wei, D., Ma, Y., 2006. Optimization and high-level expression of a functional GST-tagged rHLT-B in Escherichia coli and GM1 binding ability of purified rHLT-B. J. Microbiol. 44, 293–300.

Marengo, E., Gianotti, V., Angioni, S., Gennaro, M.C., 2004. Optimization by experimental design and artificial neural networks of the ion-interaction reversed-phase liquid chromatographic separation of twenty cosmetic preservatives. J. Chromatogr., A. 1029, 57–65.

Matsui, T., Sato, H., Yamamuro, H., Misawa, S., Shinzato, N., Matsuda, H., Takahashi, J., Sato, S., 2008. High cell density cultivation of recombinant Escherichia coli for hirudin variant 1 production. J. Biotechnol. 134, 88–92.

Montgomery, D.C., 1991. Diseño y análisis de experimentos. Grupo Editorial Iberoamericana 3.

Moreira, G.A., Micheloud, G.A., Beccaria, A.J., Goicoechea, H.C., 2007. Optimization of the Bacillus thuringiensis var. kurstaki HD-1 δ-endotoxins production by using experimental mixture design and artificial neural networks. Biochem. Eng. J. 35, 48–55.

Murarka, A., Dharmadi, Y., Yazdani, S.S., Gonzalez, R., 2008. Fermentative utilization of glycerol by Escherichia coli and its implications for the production of fuels and chemicals. Appl. Environ. Mibrobiol. 74, 1124–1135.

Myers, R.H., Montgomery, D.C., 1995. Response Surface Methodology. JohnWiley & Sons Inc., New York.

Nikerel, E., Oner, E., Kirdar, B., Yildirim, R., 2006. Optimization of medium composition for biomass production of recombinant *Escherichia coli* cells using response surface methodology. Biochem. Eng. J. 32, 1–6.

Ramchuran, S.O., Holst, O., Nordberg Karlsson, E., 2005. Effect of postinduction nutrient feed composition and use of lactose as inducer during production of thermostable xylanase in *Escherichia coli* glucose-limited fed-batch cultivations. J. Biosci. Bioeng. 99, 477–484.

Ren, X., Yu, D., Han, S., Feng, Y., 2006. Optimization of recombinant hyperthermophilic esterase production from agricultural waste using response surface methodology. Bioresour. Technol. 97, 2345–2349.

Tabandeh, F., Khodabandeh, M., Yakhchali, B., 2008. Response surface methodology for optimizing the induction conditions of recombinant interferon beta during high cell density culture. Chem. Eng. Sci. 63, 2477–2483.

Yazdani, S.S., Gonzalez, R., 2007. Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. Curr. Opin. Biotechnol. 18, 213–219.

Zhao, J., Wang, Y., Chu, J., Zhang, S., Zhuang, Y., 2008. Statistical optimization of medium for the production of pyruvate oxidase by the recombinant *Escherichia coli*. J. Ind. Microbiol. Biotechnol. 35, 257–262.