Vaccine 29 (2011) 7136–7143

Contents lists available at ScienceDirect

Vaccine

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

Cloning and expression of protease ClpP from *Streptococcus pneumoniae* in *Escherichia coli*: Study of the influence of kanamycin and IPTG concentration on cell growth, recombinant protein production and plasmid stability

Karen Einsfeldt[a,*], João Baptista Severo Júnior[a], Ana Paula Corrêa Argondizzo[b], Marco Alberto Medeiros[b], Tito Lívio Moitinho Alves[a], Rodrigo Volcan Almeida[c], Ariane Leites Larentis[b,*]

[a] *Programa de Engenharia Química - COPPE - Universidade Federal do Rio de Janeiro (UFRJ) - Av. Horácio Macedo 2030, Bloco G, Sala 115 - Centro de Tecnologia (CT) - Cidade Universitária, Ilha do Fundão, Caixa Postal 68502 - 21941-972, Rio de Janeiro, RJ, Brazil*

[b] *Bio-Manguinhos (Instituto de Tecnologia em Imunobiológicos) - Fundação Oswaldo Cruz (FIOCRUZ) - Laboratório de Tecnologia Recombinante (LATER) - VDTEC (Vice-Diretoria de Desenvolvimento Tecnológico) - Av. Brasil 4365, Pavilhão Rockfeller Sala 202 - 21040-360, Rio de Janeiro, RJ, Brazil*

[c] *Departamento de Bioquímica - Instituto de Química - Universidade Federal do Rio de Janeiro (UFRJ) - LaMMP (Laboratório de Microbiologia Molecular e Proteínas) - Av. Athos da Silveira Ramos 149, Bloco A, Sala 541 - Centro de Tecnologia (CT) - Cidade Universitária, Ilha do Fundão - 21941-909, Rio de Janeiro, RJ, Brazil*

ARTICLE INFO

Article history:
Available online 7 June 2011

Keywords:
Plasmid stability
Experimental design
ClpP
Recombinant *E. coli*

ABSTRACT

Infections caused by *Streptococcus pneumoniae* are one of the main causes of death around the world. In order to address this problem, investigations are being made into the development of a protein-based vaccine. The aims of this study were to clone and express ClpP, a protein from *S. pneumoniae* serotype 14 in *Escherichia coli*, to optimize protein expression by using experimental design and to study plasmid segregation in the system. ClpP was cloned into the pET28b vector and expressed in *E. coli* BL21 Star (DE3). Protein expression was optimized by using central composite design, varying the inducer (IPTG) and kanamycin concentration, with a subsequent analysis being made of the concentration of heterologous protein, cell growth and the fraction of plasmid-bearing cells. In all the experiments, approximately the same concentration of ClpP was expressed in its soluble form, with a mean of 240.4 mg/L at the center point. Neither the IPTG concentration nor the kanamycin concentration was found to have any statistically significant influence on protein expression. Also, higher IPTG concentrations were found to have a negative effect on cell growth and plasmid stability. Plasmid segregation was identified in the system under all the concentrations studied. Using statistical analysis, it was possible to ascertain that the procedures for determining plasmid stability (serial dilution and colony counting) were reproducible. It was concluded that the inducer concentration could be reduced tenfold and the antibiotic eliminated from the system without significantly affecting expression levels and with the positive effect of reducing costs.

1. Introduction

Infections caused by *Streptococcus pneumoniae* are among the main causes of death in the world. This gram-positive, encapsulated bacterium is the main cause of bacterial pneumonia and is often implicated in other diseases such as otitis, sinusitis and meningitis, as well as septicemia and bacteremia [1–4]. In recent years, an increasing number of strains have shown resistance to different kinds of antibiotics, making the treatment of pneumococcal pneumonia infections a major public health issue [1,3]. At the present time there are two kinds of vaccines available on the world market against *S. pneumoniae*: polysaccharide vaccines and conjugate vaccines. Polysaccharide vaccines are made from the capsular polysaccharides of *S. pneumoniae*, but their protection period is limited and they are not very effective on children under 2 years of age or with a compromised immune system [2,4,5]. Conjugate vaccines, made from capsular polysaccharides conjugated to proteins [6], have proved effective with children and adults. However, they are limited to certain serotypes and are very expensive [5,7]. In recent years several groups have been investigating different proteins associated with this bacteria for their potential use in a protein-based vaccine with broader effectiveness at a lower cost [1,2,5,7–10]. One such protein with the potential to be used as an antigen for a vaccine is the protease ClpP, which appears well con-

* Corresponding authors. Tel.: +55 21 25628343; fax: +55 21 25628300.
E-mail addresses: karen@peq.coppe.ufrj.br (K. Einsfeldt), jsevero@peq.coppe.ufrj.br (J.B. Severo Júnior), acorrea@bio.fiocruz.br (A.P. Corrêa Argondizzo), medeiros@bio.fiocruz.br (M.A. Medeiros), tito@peq.coppe.ufrj.br (T.L.M. Alves), volcan@iq.ufrj.br (R.V. Almeida), ariane@bio.fiocruz.br, ariane@peq.coppe.ufrj.br (A.L. Larentis).

0264-410X © 2011 Elsevier Ltd. Open access under the Elsevier OA license.
doi:10.1016/j.vaccine.2011.05.073

served and prevalent among the different *S. pneumoniae* serotypes. ClpP is known as a heat shock protein, which protects bacteria against adverse effects caused by elevated temperatures (for example), raising their survival levels. Mutant strains of *S. pneumoniae* in this protein become less virulent, and immunizations of mice using ClpP show that it can provide protection against pneumococcal infections [11,12].

In order to obtain high concentrations of recombinant proteins in soluble form and also reduce production costs, the many variables that influence the expression of these proteins must be understood and controlled. The strategy of assessing one factor at a time while keeping the others constant may not be efficient, as it fails to take account of the interaction between the process variables and more experiments have to be done to obtain the information required. The best approach is to use experimental design, which can be used to assess the effect and interaction of the variables involved, yielding the maximum amount of information from a minimum of experiments, while also allowing experimental errors to be assessed in order to enhance process effectiveness [13].

In recombinant bioprocesses, antibiotics like kanamycin are widely used on a bench scale to put selective pressure on the culture medium, preventing plasmid segregation, since most of the plasmids used have an antibiotic resistance marker gene. Plasmid segregation may have an impact on the recombinant protein yield, especially on an industrial scale. However, the use of these antibiotics is unfeasible on an industrial scale because they are costly and also contaminate the product and have to be completely removed in the food or drug purification process [14]. This is why studying the antibiotic concentration used in recombinant processes is so important, even though the variation of the antibiotic in the culture may affect plasmid stability. Another important variable in the process, especially on a large scale, is the inducer used in the expression system, since some inducers, like IPTG, are expensive and may be toxic to the host cell [15,16].

In view of these considerations, the aim of this study was to clone and express ClpP using *Escherichia coli* as a host, optimize protein production using experimental design and study the plasmid stability of the system. As such, central composite design was used for two variables: concentration of the inducer of the recombinant system (IPTG) and the concentration of the antibiotic (kanamycin) in the culture medium.

2. Materials and methods

2.1. Bacterial strains

E. coli TOP 10 (Invitrogen) was used as the host for the cloning procedures. *E. coli* BL21 Star (DE3)^{TM} (Invitrogen) was used as the bacteria for expressing the recombinant protein ClpP.

2.2. Chemicals

Bacto^{TM} yeast extract and tryptone were purchased from BD (Becton, Dickinson and Company), the glucose and NaCl were from Merck, the glycerol was from Invitrogen, the kanamycin was from Sigma and the IPTG (isopropyl β-D-1-thiogalactopyranoside) was purchased from Promega.

2.3. PCR and cloning

The gene that codifies protein ClpP was amplified by PCR using genomic DNA from *S. pneumoniae* serotype 14 (strain 113/95 deposited at Instituto Adolfo Lutz) as a template. The primers used were: 5′-CCCATGGTTCCTGTAGTTATTGAACAAAC-3′ and 5′-CACTCGAGGTTCAATGAATTGTTGGC-3′. The *Nco*I and *Xho*I restriction sites are underlined in the forward and reverse primers, respectively. The PCR product was digested by enzymes *Nco*I and *Xho*I and inserted into the vector pET28b (Novagen). The *E. coli* TOP10 strain was transformed by electroporation with the constructed plasmid (pET28b/*clpP*). The constructed plasmid pET28b/*clpP* was confirmed by digestion and sequenced with fluorescent terminators (Big Dye, Applied Biosystems) using the ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems). Once analyzed, the plasmid was transformed into *E. coli* BL21 Star (DE3)^{TM}.

2.4. Cell viability test

The cell viability of the stock of recombinant *E. coli* BL21 Star (DE3)^{TM}/pET28b/*clpP* in LB (5 g/L yeast extract, 10 g/L tryptone, 5 g/L NaCl, pH 7) with 25% glycerol, stored at −70 °C, was assessed by counting the colony forming units (CFUs) for all the experimental design experiments. Serial dilutions were made in PBS pH 7.4 and transferred to Petri plates containing LB Agar and 50 μg/mL kanamycin (concentration of stocks around 10^{10} CFU/mL).

2.5. Expression and solubility test

Recombinant *E. coli* BL21 Star (DE3)^{TM}/pET28b/ClpP was pre-inoculated (10 μL) in 10 mL of the LB medium enriched with 1% glucose, 0.4% glycerol and 50 μg/mL kanamycin. The pre-inoculum was incubated for 16 h at 37 °C and 200 rpm in 50 mL flasks under agitation.

The inoculum was prepared in 500 mL flasks with 2 mL pre-inoculum and 100 mL of the LB medium enriched with 1% glucose, 0.4% glycerol and different kanamycin concentrations according to the experimental design (as described in the next section). The culture was incubated at 37 °C and 200 rpm until it reached the exponential growth phase (Abs600 nm between 0.65 and 0.75). At this point, expression was induced with IPTG for 4 h under different induction concentrations according to the experimental design. *E. coli* BL21 (DE3) Star/pET28a was used as a negative control.

1 mL samples were taken from each experiment before and after the 4 h expression period to assess cell growth, ClpP expression (by SDS-PAGE) and solubility.

The cells were harvested by centrifugation at 20,817 × *g* for 5 min to separate the culture medium. In order to assess the solubility of the expressed protein the cells were resuspended in a lysis buffer (20 mM Tris, 1 mM EDTA, pH 8.0) at a ratio of 25 μL buffer to each 0.1 of Abs600 nm (normalizing to Abs600 nm), to obtain the total protein extract. The total extract was put through five 10 s ultrasound cycles at 30% amplitude in an ultrasonic cell disruptor (Sonics & Materials, Inc.). The soluble and insoluble fractions of the total protein were separated from the cultures by centrifugation (20,817 × *g* for 10 min at 10 °C). The samples were added to 12% SDS-PAGE [17], stained with Coomassie Blue R-250.

2.6. Experimental design

The influence of kanamycin and IPTG concentration on cell growth, the concentration of expressed protein and plasmid stability was assessed by using a central composite design for two variables. Eight experiments were performed, four of which were replications at the center point (CP), as described in the previous section. The expression cultures were prepared using different concentrations of kanamycin and were then induced with different concentrations of IPTG as shown in the details of the central composite design experiments in Table 1.

The statistical analyses were performed using STATISTICA 9.1 software (Statsoft), using the normalized variables. The effect of each variable was estimated, as was standard error, and was

Table 1
Central composite design and its response variables.

Run	IPTG (mM)	Kanamycin (μg/mL)	Dry mass of cells (mg/L)	ClpP (mg/L)	Φ (Fraction of plasmid-bearing cells)
1	0.1	0	905.4	247.3	0.37
2	0.1	50	867.2	251.1	0.56
3	1	0	706.4	271.6	0.05
4	1	50	761.8	225.9	0.11
5 (CP)	0.55	25	744.6	266.3	0.08
6 (CP)	0.55	25	723.6	244.2	0.18
7 (CP)	0.55	25	766.7	229.7	0.07
8 (CP)	0.55	25	759.8	221.2	0.11

assessed by the *t*-test, with all results giving $p < 0.05$ being considered statistically significant.

2.7. Analysis of expression by SDS-PAGE using densitometry and assessment of cell growth

Cell growth was measured by absorbance at 600 nm. This was converted to dry mass of cells using a standard calibration curve. Samples of cells from 1 mL culture were resuspended in a sample buffer (60 mM Tris–HCl, pH 6.8, 10% glycerol, 5% β-mercaptoethanol, 2% SDS, 0.5% Bromophenol Blue) to obtain the total protein extract, at a ratio of 25 μL buffer to each 0.1 Abs600 nm. These samples were added to 12.5% SDS-PAGE [17], stained with Coomassie Blue R-250. The same gel also had 2 μL low molecular weight marker (LMW, Amersham Bioscience) added, with 97 kDa, 66 kDa, 45 kDa, 30 kDa, 20.1 kDa and 14.4 kDa bands and 1340 ng, 1660 ng, 2940 ng, 1660 ng, 1600 ng and 2320 ng protein weight in each band, respectively, for the purpose of comparing with the bands corresponding to ClpP. The amount of protein expressed under each condition was analyzed by densitometry using a Bio-Rad GS-800 calibrated densitometer and Quantity-One 4.4.1 software. The concentration of expressed protein was obtained using the ratio (mg/L) = (Abs600 nm × band in densitometry)/4, where 4 was the concentration factor used in the preparation of the total protein extract samples.

2.8. Plasmid stability analysis

In order to analyze plasmid segregation, 100 μL samples were taken from each experiment at the end of the 4 h expression period, with analysis done on two aliquots from each experiment. Each aliquot was serially diluted in sterile PBS to 10^{-6} (Fig. 1). 10 μL samples of each dilution with at least three replications were added to LB Agar plates with kanamycin (50 μg/mL) and without it. Plasmid stability was measured as the fraction of plasmid-bearing cells (Φ) by calculating the ratio between the number of colony forming units (CFU/mL) on the plate with the antibiotic and on the plate without the antibiotic.

A statistical evaluation was made with the aim of checking the reproducibility and variability of the procedures for assessing plasmid stability (serial dilution and colony count). Student's *t*-test was used to find out whether the mean values from the colony count were equivalent, while the *F*-test (Fisher) was used to find out whether the errors made at each stage of the count were equivalent. These tests were done using the values obtained from CFU/mL in the experiments at the center point of the experimental design, comparing different aliquots diluted to the same degree from the same culture, and the same aliquots diluted to different degrees from the same culture, as shown in the diagram in Fig. 1.

In order to do the *F*-test, *F* was calculated using Eq. (1):

$$F_{xy} = \frac{\sigma_x^2}{\sigma_y^2} \quad (1)$$

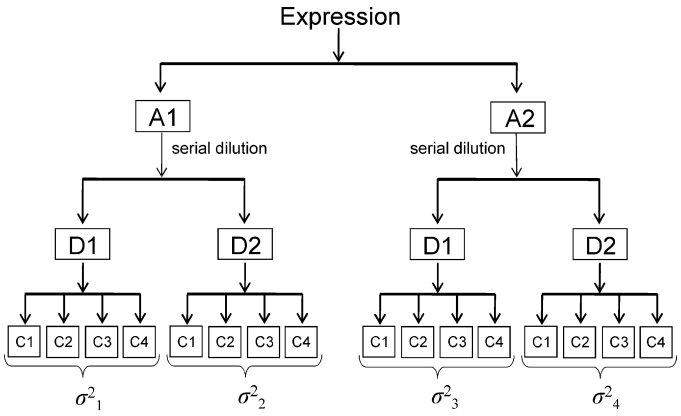

Fig. 1. Diagram used for statistical evaluation of serial dilution and colony counting procedures. This evaluation was performed on the experiments at the center point (5, 6, 7 and 8) in LB agar plates with and without the antibiotic. A1 and A2 represent aliquots 1 and 2, respectively, taken from the culture after 4 h expression to perform serial dilution. D2 represents the level of dilution obtained between 0 and 10 colonies. D1 represents the dilution obtained between 10 and 100 colonies. C1, C2, C3 and C4 are replications of CFU/mL. σ^2_1, σ^2_2, σ^2_3 and σ^2_4, represent the variances of CFU/mL obtained in A1D1, A1D2, A2D1 and A2D2, respectively.

where σ_x^2 and σ_y^2 are the variances obtained for the different CFU/mL counts, as shown in Fig. 1. The variances were considered to be statistically equivalent when F_{xy} was between the confidence limits set (95% confidence level) as described by Fisher's *F*-distribution [18].

The confidence intervals for the mean were obtained using the *t*-test as shown by Eq. (2):

$$\mathrm{CI}[\mu]_{95\%} = \bar{x} \pm t\frac{s}{\sqrt{n}} \quad (2)$$

where μ is the estimated mean population (95% confidence), $\bar{x}$ is the sample mean, *t* is the value described by the Student's *t* distribution, *s* is standard deviation, and *n* is the sample size. The means were regarded as statistically equivalent if the confidence intervals crossed.

2.9. Validation of experimental design

Having conducted the analyses of the experimental design, replications were performed of the optimal cultivation condition to validate the results obtained from the experimental design. Once the cultures were induced, samples were taken every hour to assess the ClpP protein production rate, cell growth and plasmid segregation.

3. Results and discussion

3.1. Expression and solubility

ClpP was expressed in *E. coli* BL21 Star (DE3)™ by induction with IPTG. At the end of the expression period samples were taken for the preparation of protein extracts, and the soluble and insoluble fractions of the total protein were also separated out. These samples were analyzed using SDS-PAGE, as shown in Fig. 2. The ClpP protein was not expressed in the negative control using *E. coli* BL21 (DE3) Star/pET28a. The results show that the size of ClpP expressed was as expected (22.4 kDa), as can be seen from the gel between bands 18.4 kDa and 24 kDa of the molecular weight marker. Also, the band that corresponds to ClpP cannot be seen before expression was induced (non-induced sample), as the RNA polymerase of bacteriophage T7 was used in the system, which is highly regulated and repressed by the glucose added to the culture medium, only allowing the recombinant protein to be expressed when the inducer was

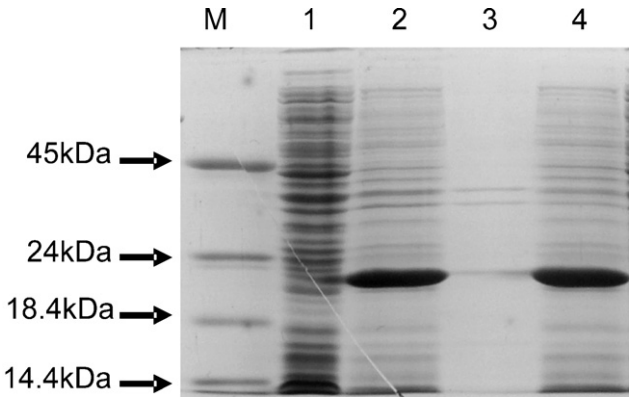

Fig. 2. SDS-PAGE analysis of ClpP expression in *E. coli* BL21 Star (DE3). M, molecular weight marker; 1, sample of uninduced culture; 2, total protein extract after 4 h induction; 3, insoluble fraction of total protein after 4 h induction; 4, soluble fraction of total protein after 4 h induction. The samples added to the gel were normalized to Abs600 nm.

added. The solubility analysis (Fig. 2) shows that the protein was expressed in a soluble form in high concentrations and that no inclusion bodies were formed. It is known that one of the problems associated with overexpressing heterologous proteins in this bacterial cytoplasm is the formation of insoluble protein aggregates (inclusion bodies) caused by the mal-conformation of the protein [19,20]. This problem was not identified in the study in question.

3.2. Experimental design: influence of the concentration of IPTG and kanamycin on cell growth, ClpP expression and plasmid segregation

Experimental design was used to assess the influence of the concentration of IPTG and kanamycin on cell growth, protein production and plasmid segregation. The conditions for each of the central composite design experiments are shown in Table 1, as are the responses of the dependent variables under analysis.

3.2.1. Effect of variables on cell growth

The effects of IPTG and kanamycin on cell growth are shown in Table 2. By analyzing these effects it was possible to infer, within the 95% confidence interval, that the IPTG concentration had a significant negative influence on cell growth. Some authors have already reported on the fact that IPTG is toxic and that in high concentrations it can drastically reduce cell growth in some cultures of recombinant *E. coli* [16,21–24]. Meanwhile, the kanamycin concentration and the interaction of IPTG with kanamycin was not found to have any appreciable effect on cell growth within the ranges under study.

3.2.2. Effect of variables on ClpP Production

The protein concentration stayed at similar levels under all the conditions tested in the factorial design (Table 1). This means that when the effects of the variables on protein concentration were analyzed, it could be concluded that neither the IPTG concentration nor the kanamycin concentration nor the interaction of IPTG with kanamycin had any appreciable influence on protein expression in the ranges tested, since the *p*-value was higher than 0.05 (Table 3). This suggests that under the conditions tested the inducer concentration could be reduced up to tenfold (also advantageous because of its negative effect on cell growth) and kanamycin could even be completely eliminated from the system without this having any major impact on the protein concentration. Costs would be reduced and expression levels would remain about the same.

Also, high concentrations of protein in a soluble form were expressed in all the concentrations at 37 °C and 200 rpm, which is positive, since *E. coli* normally expresses insoluble proteins under such conditions. The concentration of ClpP obtained in these experiments fell within the concentration range obtained in other studies in the literature, which report on experiments to optimize the expression of other proteins in *E. coli* using agitated flasks and batch cultivation [22,25–28]. Data from the literature show that higher protein concentrations from *E. coli* are achieved when the heterologous protein is expressed and optimized in bioreactors, where conditions can be controlled [26].

Table 3
Effects of variables on ClpP concentration (mg/L).

	Effects	Standard error	$t(4)$	p-value
Mean	244.69	6.41	38.14	<0.0001
IPTG	−0.49	18.14	−0.03	0.98
Kanamycin	−20.98	18.14	−1.17	0.31
IPTG × Kanamycin	−24.79	18.14	−1.37	0.24

3.2.3. Effect of variables on plasmid segregation

Independent of the kanamycin concentration, the experiments carried out using lower concentrations of IPTG yielded higher cell growth than the others and greater plasmid stability, since the values for Φ (fraction of plasmid-bearing cells) in experiments 1 and 2 were higher than in the others. In the experiments with lower IPTG concentrations, the fraction of plasmid-bearing cells was found to be between 37% and 56%, while in the other experiments induced with higher IPTG concentrations, Φ was found to be lower than this. In other words, in the experiments with lower IPTG concentrations there was less plasmid segregation (Table 1).

The effects of the variables on Φ are shown in Table 4. It can be concluded from these analyses, within a 95% confidence interval, that the IPTG concentration in the range tested had a negative effect on plasmid stability, while the kanamycin concentration in the system and the IPTG/kanamycin interaction had no appreciable impact on Φ in the range under study. This means that the closer the IPTG concentration was to 0.1 mM, the more plasmid stability was achieved for any kanamycin concentration within the range under study.

The negative effect of induction with IPTG on plasmid segregation identified in this study was already mentioned in the literature [14,29,30]. Marí et al. [29] found that when they used vectors pYMK5 and pYMK7, which contain brain-derived neurotrophic factor (BDNF) and nerve growth factor (NGF) genes, respectively, plasmid stability declined in the presence of the inducer (1 mM IPTG) in *E. coli*, with or without the antibiotics ampicillin and kanamycin. Data on the stability of plasmid pED-GnRH3 (obtained from vector pET28a), transformed in *E. coli*, indicate that plasmid segregation is far more dependent on induction than the pres-

Table 2
Effects of variables on cell growth (mg/L of dry mass of cells).

	Effect	Standard error	$t(4)$	p-value
Mean	779.43	16.45	47.37	<0.0001
IPTG	−152.19	46.54	−3.27	0.03
Kanamycin	8.61	46.54	0.18	0.86
IPTG × Kanamycin	46.81	46.54	1.01	0.37

Table 4
Effects of variables on Φ value (fraction of plasmid-bearing cells).

	Effects	Standard error	$t(4)$	p-value
Mean	0.19	0.04	4.34	0.01
IPTG	−0.38	0.12	−3.10	0.03
Kanamycin	0.12	0.12	0.97	0.38
IPTG × Kanamycin	−0.07	0.12	−0.56	0.61

ence or absence of kanamycin, and that after 10 h cultivation in non-induced cultures, plasmid stability was as high as 95% with antibiotics and 90% without them. However, stability levels in induced cultures were far lower after 10 h induction, dropping as low as 15% with antibiotics and 10% without them [30].

If one looks at the values for Φ obtained in the experiments at the center point (Table 1), one might think that the value obtained in experiment 6 (CP) is an outlier since it differs from the trend seen for all the other Φ values from the replications performed at the center point. An outlier is defined as an experimental point that would seem not to fit into a particular distribution pattern of probabilities defined by the vast majority of the other experimental points [18]. However, the identification of outliers is a controversial issue and the elimination of a putative outlier could result in a misinterpretation of the data. For this reason, the effects of the variables on the plasmid-bearing cells (Φ) were analyzed both taking account of and discarding the Φ value obtained from experiment 6 (CP), resulting in the same conclusions about the effects.

Also, it can be perceived from the Φ values (fraction of plasmid-bearing cells) (Table 1) that the behavior of the Φ values was not linear, which was confirmed by the low value of the linear adjustment coefficient (R^2). As it is only possible to assess linear regression coefficients for each variable when analyzing central composite design, the low R^2 indicates that the linear model does not adjust well to the data. According to the studied ranges, in order to obtain lower plasmid segregation levels, 0.1 mM IPTG should be used. These data do not rule out the possibility of there being an optimal point lower than 0.1 mM IPTG that would still assure minimum plasmid segregation and good protein expression levels.

3.2.4. Statistical analysis of the procedure (colony count and serial dilution) for determining plasmid stability

The results of the statistical analysis showed that according to the Student's *t*-test, the mean CFU/mL values obtained from the experiments were equivalent, meaning that for most of the data they were statistically equivalent (within a 95% confidence level), as can be seen from Fig. 3. The only statistically different values were the mean CFU/mL values obtained from aliquots 1 and 2 at dilution D1 (A1D1 and A2D1) from experiment 8, on the plates with and without kanamycin, and the mean values for aliquot 1 at dilution 1 and aliquot 2 at dilution 2 (A1D1 and A2D2) from experiment 8, on the plates with kanamycin (Fig. 3). It can also be seen from Fig. 3 that the confidence intervals of the means for the D2 dilutions were always higher than those for the D1 dilutions, independent of the aliquots, showing that the variability of the mean for dilution D2 was higher than for dilution D1, which means that the errors made in dilution D2 were greater than in D1.

When the variance in the data on CFU/mL was assessed using the *F*-test, when different aliquots with the same dilution were compared (Fig. 4A) the calculated *F* values were within the *F* value limits for 95% confidence, except when aliquots 1 and 2 at dilution D1 were compared (A1D1 and A2D1) from experiment 8 with no antibiotic. This means that the errors incurred during the dilution and colony count procedures were the same when compared between the same dilutions. However, when different dilutions of the same aliquot were compared, the data showed different variance levels in most cases (Fig. 4B). The calculated *F* values were outside the pre-established *F* interval at 95% confidence level. As already reported and shown in Fig. 3, the errors in the CFU/mL data were greater at dilution D2 than they were at dilution D1, with standard deviation about ten times higher in the data for dilution D2 than for dilution D1 (data not shown). This variability is owing to the fact that at the higher dilution (D2), between 0 and 10 colonies were counted, while at D1, between 10 and 100 colonies were counted. This being the case, only the data on CFU/mL obtained from dilution D1 were used for calculating Φ values in the experimental design experiments.

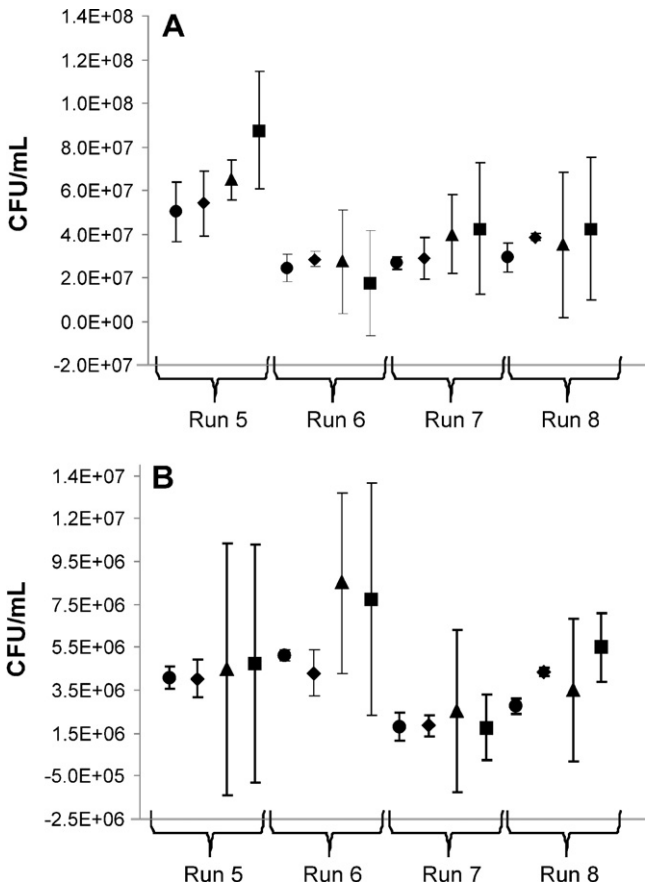

Fig. 3. Mean CFU/mL obtained from the experiments at the center point. (A) CFU/mL in LB agar plates without kanamycin. (B) CFU/mL in LB agar plates with kanamycin. Vertical bars represented the confidence intervals, with 95% confidence level. The symbols represent the different aliquots (A1 and A2) and their dilutions (D1 and D2) for each experiment: A1D1 (●), A2D1 (♦), A1D2 (▲), A2D2 (■).

This statistical analysis shows that when the data from dilution D1 were used, the procedures for determining plasmid stability (serial dilutions and colony count) were reproducible, meaning that the CFU/mL data obtained had statistically equivalents means and variances, within a 95% confidence interval.

3.3. Validation of experimental design

The optimal condition as identified by the experimental design was the condition used in experiment 1 (0.1 mM IPTG and 0 μg/mL kanamycin). This condition permitted a tenfold reduction in the inducer concentration and the elimination of kanamycin from the system, keeping the protein concentration and cell growth at similar levels while also keeping plasmid stability at levels that would not harm recombinant protein production over the 4 h expression period.

In order to validate the optimal condition as identified by experimental design, replications of the culture were produced under this condition (0.1 mM IPTG and 0 μg/mL kanamycin). The cultures were allowed to grow until they reached exponential growth (Abs600 nm approximately 0.7), at which point they were induced with 0.1 mM IPTG. From this moment on, samples were taken every hour for analysis until the end of the 4 h induction period.

The mean cell growth (expressed as dry mass of cells – mg/L) obtained for these replications was 912 mg cells/L at the end of 4 h induction, with 13.7% relative standard deviation, which is in agree-

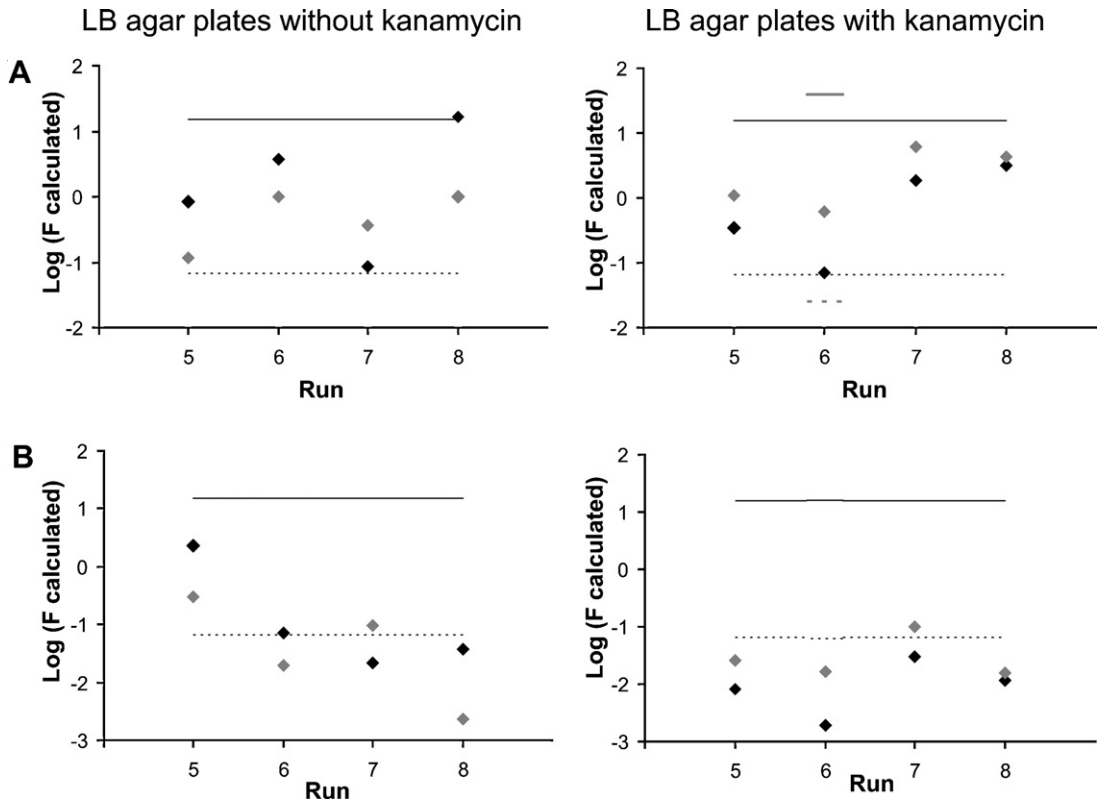

Fig. 4. *F*-test values for each experiment at the center point. Values for confidence limits for the *F*-test at confidence 95% (low limit – Log($F_{0,025}$) (- - -) and high limit – Log($F_{0,975}$) (—)). (A) Comparison between different aliquots with the same dilutions; the symbols representing the values of the *F*-test obtained when comparing the data on the different aliquots at same dilutions for each experiment are as follows: A1D1 and A2D1(♦), A1D2 and A2D2 (◆); in this case the gray lines, (▪ ▪ ▪ ▪) and (▬), represents the confidence limits for comparison between A1D1 and A2D1(♦). (B) Comparison between different dilutions of one aliquot; the symbols representing the values of the *F*-test obtained when comparing the data on the different dilutions of one aliquot for each experiment are as follows: A1D1 and A1D2(♦), A2D1 and A2D2 (◆).

ment with the final value obtained for experiment 1 of the initial experimental design. Cell growth was also monitored throughout the experiment and the graph of the cell growth rate is shown in Fig. 5A.

The analysis of cell growth (Fig. 5A) shows that after 2 h induction (242 min of culture), the cells started to reach the stationary growth phase. Some authors argue that when systems with strong promoters are used, as is the case of T7 promoters, when the system is induced the growth rate drops because the host cell's metabolism is overburdened [31]. The specific growth rate obtained in this study was 0.72 h^{-1} while the generation time was 0.96 h. Similar values to these have been obtained in other studies during the expression of heterologous proteins in *E. coli* [32].

The mean protein production over 4 h expression can be seen in Fig. 5A, with this value reaching around 294 mg/L ClpP at the end of this period. This is slightly higher than the value obtained in experiment 1 from the experimental design. However, taking into account the errors associated with the densitometry measurements, which varied from 10% to 13% in these experiments, and the estimated 8% error in experiment 1 from the experimental design, it can be stated that the values obtained in the validation experiment were similar to those obtained from the original experimental design experiment. It can be seen (Fig. 5A) that after the second hour of induction (242 min of culture) the protein production rate and cell growth rate both started to fall, coming close to the stationary phase during the fourth hour of induction. It can therefore be concluded that there would be nothing to be gained by extending the expression time further, since the protein concentration would remain constant and the overall productivity of the process would fall.

By calculating the ratio of protein concentration to dry mass of cells, the yield factor $Y_{P/X}$ was obtained (production of product per cell) throughout the induction time. The plasmid segregation in the cultures was also studied over time, starting from the moment protein expression was induced. Fig. 5B shows the graph of variable Φ (fraction of plasmid-bearing cells) and yield factor $Y_{P/X}$ as a function of culture time after induction. Fig. 5B shows that over 4 h expression the fraction of plasmid-bearing cells reached around 45%. The great variability of the values calculated for Φ over the 242 min of culture time could be associated with the physiological state of the cells, since it was at this point that the cell growth rate fell most sharply (Fig. 5A). The system also presented plasmid segregation in the negative control using *E. coli* BL21 (DE3) Star/pET28a.

The Φ values obtained in these experiments were similar to those obtained in a study that also used a pET vector in the expression of recombinant phytase [33]. The authors reported that stability levels had fallen to 10% by 4 h of induction. They added that before induction the plasmid was stable for over 96 h, but that after induction it started to show signs of segregation. The greater level of instability after induction could be attributed to the fact that recombinant protein expression imposes a metabolic burden on the host cells, resulting in higher segregation levels. Other authors have also shown that vector pET101 is more stable in non-induced cultures [34], showing that when the system is induced, plasmid stability reaches around 30% when the pH is not controlled and around 60% when the pH is kept at 7.0 after 4 h expression. These results imply that the pH may have been behind the low stability levels seen in our study, since this factor was not kept constant. In the experiments to validate the optimal condition obtained from factorial planning, the initial pH of the cultures was 7.0, but by the

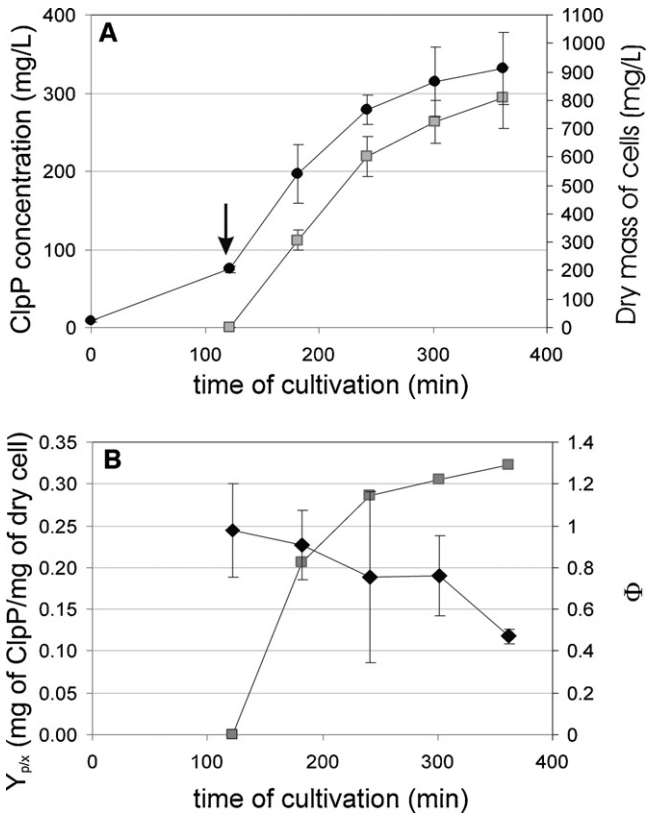

Fig. 5. (A) ClpP concentration and dry mass of cells as a function of culture time. The arrow indicates the time when expression started to be induced; (●) dry mass of cells (mg/L), (■) ClpP concentration (mg/L). (B) Yield factor of ClpP to cell mass ($Y_{P/X}$) and fraction of plasmid-bearing cells (Φ) as function of culture time; (♦) Φ (fraction of plasmid-bearing cells), (◆) $Y_{P/X}$ (mg of ClpP/mg of dry cells).

end of the 4 h expression period it had dropped to 5.1. There may be other factors associated with the low plasmid stability found in our experiments, such as the drop in dissolved oxygen in the cultures, which some authors suggest could have an impact on plasmid stability [14]. As the experiments were conducted in agitated flasks and this does not allow dissolved oxygen in the culture medium to be controlled, this could have been one of the causes behind the high segregation levels encountered throughout the culture period. In order to control aeration, pH and monitor other process variables, bioreactors should be employed, as should experimental design tools to define the optimal operation conditions. Aside from the factors presented here, there are many others that may have an impact on plasmid stability. Some authors claim that more complex culture mediums may result in lower plasmid stability [35]. The other factors that might affect stability are the growth rate, number of plasmid copies, the insert size and the recombinant protein expression level [35].

The yield factor ($Y_{P/X}$), obtained throughout the culture time can be seen in Fig. 5B. It can be seen that after the second hour of induction (242 min of culture), the yield factor no longer increased at the same rate, again indicating that longer expression times would bring no particular benefit. As expected, as segregation increased, the product formation rate per dry mass of cells dropped and the yield factor ($Y_{P/X}$) came close to constant levels (Fig. 5B). The yield factor still increased even during the third and fourth hours of expression, albeit at a slower rate. This may have been because of the increased protein production by the remaining plasmid-bearing cells. In studies of phytase expression in *E. coli* [33] the authors found that in the first 2 h of induction, phytase production increased from 0 to 800 U/L while plasmid stability fell to 60%, i.e. even with 40% segregation, phytase production continued to rise. After two and a half hours' induction, phytase production rose again to 1000 U/L, while segregation increased to 80%. It was only after this point that phytase activity started to drop [33]. The data presented in Fig. 5 show that after 4 h induction the fraction of plasmid-bearing cells stood at around 45%, while the yield factor was still rising. However, as shown by other authors [33], if segregation were to rise even higher, the yield factor could start to fall.

4. Conclusions

High levels of a soluble form of ClpP were expressed in all the experiments from the experimental design used. Plasmid segregation was identified in the system throughout the kanamycin concentration range tested. The lowest concentration of IPTG (0.1 mM) tested in this study resulted in greater plasmid stability. The statistical analyses made of the procedures used to determine plasmid segregation confirmed that they are reproducible. By using experimental design it was possible to conclude that the optimal point of the system was with 0.1 mM IPTG and 0 μg/mL kanamycin, which yielded 247.3 mg/L ClpP; this optimal condition was validated with success. It should therefore be possible to reduce the inducer concentration tenfold and eliminate the antibiotic from the system while still keeping protein expression at similar levels and reducing overall process costs. It is also important to highlight the importance of the study of plasmid segregation in recombinant systems, since plasmid stability is one of the lynchpins of recombinant protein production.

Experimental design proved to be a powerful tool for determining the optimal conditions for expressing recombinant protein in *E. coli* using a minimum number of experiments, enabling an assessment to be made of the effect of each of the variables, their interactions and experimental errors. It is still common practice in molecular biology for each variable to be evaluated separately, which may result in misinterpretations of the data obtained, because it fails to take account of their interactions. Experimental design enables the selection of the best test conditions for detecting the interactions between the variables, which is not possible empirically by adopting the methods usually used in the area that treat variables independently. These techniques have universal application in the production of recombinant proteins.

Acknowledgments

This work received financial support from Bio-Manguinhos and PAPES V (Programa Estratégico de Apoio à Pesquisa em Saúde) from Fundação Oswaldo Cruz (FIOCRUZ). Karen Einsfeldt and João B. Severo Júnior received scholarships from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), respectively. We also wish to thank the Setor de Imunobiológicos/Departamento de Química/INCQS/Fiocruz for making available the Bio-Rad Densitometer and the QuantiOne software; Laboratório de AIDS & Imunologia Molecular/IOC/Fiocruz and Alfredo Jabor MSc for the DNA sequencing; Laboratório de Tecnologia Bacteriana (LATEB)/Bio-Manguinhos/Fiocruz for providing *S. pneumoniae* serotype 14 growth; Dr. Maria Isabel Rodrigues (PROTIMIZA) for her assistance with the statistics.

References

[1] Jedrzejas MJ. Pneumococcal virulence factors: structure and function. Microbiol Mol Biol Rev 2001;65(2):187–207.
[2] García-Suárez MM, Vázquez F, Méndez FJ. *Streptococcus pneumoniae* virulence factors and their clinical impact: an update. Enferm Infecc Microbiol Clin 2006;24(8):512–7.

[3] Braido F, Bellotti M, Maria AD, Cazzola M, Canonica GW. The role of Pneumococcal vaccine. Pulm Pharmacol Ther 2008;21:608–15.
[4] Pletz MW, Maus U, Krug N, Welte T, Lode H. Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species. Int J Antimicrob Agents 2008;32(3):199–206.
[5] Bogaert D, Hermans PWM, Adrian PV, Rümke HC, Groot R. Pneumococcal vaccines: an update on current strategies. Vaccine 2004;22:2209–20.
[6] Poland GA. The burden of pneumococcal disease: the role of conjugate vaccines. Vaccine 1999;17(13–14):1674–9.
[7] Barocchi MA, Censini S, Rappuoli R. Vaccines in the era of genomics: the pneumococcal challenge. Vaccine 2007;25(16):2963–73.
[8] Paton JC. Novel pneumococcal surface proteins: role in virulence and vaccine potential. Trends Microbiol 1998;6:85–7.
[9] Di Guilmi AM, Dessen A. New approaches towards the identification of antibiotic and vaccine targets in *Streptococcus pneumoniae*. EMBO rep 2002;3:728–34.
[10] Ogunniyi AD, Grabowicz M, Briles DE, Cook J, Paton JC. Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. Infect Immun 2007;75(1):350–7.
[11] Kwon HY, Ogunniyi AD, Choi M, Pyo SN, Rhee DK, Paton JC. The ClpP protease of *Streptococcus pneumoniae* modulates virulence gene expression and protects against fatal Pneumococcal challenge. Infect Immun 2004;72(10):5646–53.
[12] Cao J, Chen D, Xu W, Chen T, Xu S, Luo J, et al. Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC and ClpP. Vaccine 2007;25:4996–5005.
[13] Rodrigues MI, Iemma AF. Planejamento de Experimentos e Otimização de Processos – Uma estratégia seqüencial de planejamentos. 1st ed. Campinas SP: Casa do Pão Editora; 2005.
[14] Friehs K. Plasmid copy number and plasmid stability. Adv Biochem Eng Biotechnol 2004;86:47–82.
[15] Volontè F, Marinelli F, Gastaldo L, Sacchi S, Pilone MS, Pollegioni L. Optimization of glutaryl-7-aminocephalosporanic acid acylase expression in *E. coli*. Protein Expr Purif 2008;61(2):131–7.
[16] Pan H, Xie Z, Bao W, Zhang J. Optimization of culture conditions to enhance cis-epoxysuccinate hydrolase production in *Escherichia coli* by response surface methodology. Biochem Eng J 2008;42:133–8.
[17] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680–5.
[18] Schwaab M, Pinto JC. Análise de Dados Experimentais. I. Fundamentos de estatística e estimação de parâmetros. 1st ed. Rio de Janeiro: e-papers editora; 2007.
[19] Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 2006;72:211–22.
[20] Peti W, Page R. Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. Protein Expr Purif 2007;51:1–10.
[21] Islam RS, Tisi D, Levy MS, Lye GJ. Framework for the rapid optimization of soluble protein expression in *Escherichia coli* combining microscale experiments and statistical experimental design. Biotechnol Prog 2007;23:785–93.
[22] De León A, Jiménez-Islas H, González-Cuevas M. Analysis of the expression of the *Trichoderma harzianum ech42* gene in two isogenic clones of *Escherichia coli* by surface response methodology. Process Biochem 2004;39:2173–8.
[23] Andersson L, Yang S, Neubauer P. Impact of plasmid presence and induction on cellular responses in fed batch cultures of *Escherichia coli*. J Biotechnol 1996;46:255–63.
[24] Bentley WE, Davis RH, Kompala DS. Dynamics of CAT expression in *E. coli*. Biotechnol Bioeng 1991;38:749–60.
[25] Kotik M, Kocanová M, Maresová H, Kyslík P. High-level expression of a fungal pyranose oxidase in high cell-density fed-batch cultivations of *Escherichia coli* using lactose as inducer. Protein Expr Purif 2004;36:61–9.
[26] Choi JH, Keum KC, Lee SY. Production of recombinant proteins by high cell density culture of *Escherichia coli*. Chem Eng Sci 2006;61:876–85.
[27] Han GH, Shin H, Kim SW. Optimization of bio-indigo production by recombinant *E. coli* harboring *fmo* gene. Enzym Microb Technol 2008;42:617–23.
[28] Pistorino M, Pfeifer BA. Efficient experimental design and micro-scale medium enhancement of 6-deoxyerythronolide B production through *Escherichia coli*. Biotechnol Prog 2009;25(5):1364–71.
[29] Marí YM, Espinosa AE, Ubieta R, Batista OF, Fernández ML. Effect of the selection marker on the viability and plasmid stability of two human proteins with neurotrophic action expressed in *Escherichia coli*. Biochem Biophys Res Commun 1999;258:29–31.
[30] Xu J, Li W, Wu J, Zhang Y, Zhu Z, Liu J, et al. Stability of plasmid and expression of a recombinant gonadotropin-releasing hormone (GnRH) vaccine in *Escherichia coli*. Appl Microbiol Biotechnol 2006;73(4):780–8.
[31] Toksoy E, Önsan ZÍ, Kirdar B. High-level production of TaqI restriction endonuclease by three different expression systems in *Escherichia coli* cells using the T7 phage promoter. Appl Microbiol Biotechnol 2002;59:239–45.
[32] Micheletti M, Barrett T, Doig SD, Baganz F, Levy MS, Woodley JM, et al. Fluid mixing in shaken bioreactors: implications for scale-up predictions from microlitre-scale microbial and mammalian cell cultures. Chem Eng Sci 2006;61:2939–49.
[33] Sunitha K, Kim Y-O, Lee J-K, Oh T-K. Statistical optimization of seed and induction conditions to enhance phytase production by recombinant *Escherichia coli*. Biochem Eng J 2000;5:51–6.
[34] Tomazetto G, Mulinari F, Stanisçuaski F, Settembrinib B, Carlinic CR, Ayub MAZ. Expression kinetics and plasmid stability of recombinant *E. coli* encoding urease-derived peptide with bioinsecticide activity. Enzym Microb Technol 2007;41:821–7.
[35] Gupta R, Sharma P, Vyas VV. Effect of growth environment on the stability of a recombinant shuttle plasmid, pCPPS-31, in *Escherichia coli*. J Biotechnol 1995;41:29–37.